Age-related differences in weighting and masking of two cues to word-final stop voicing in noise[a)]

Susan Nittrouer[b)]
Utah State University, UMC 6840, Logan, Utah 84322-6840

(Received 30 August 2004; revised 2 May 2005; accepted 4 May 2005)

Because laboratory studies are conducted in optimal listening conditions, often with highly stylized stimuli that attenuate or eliminate some naturally occurring cues, results may have constrained applicability to the "real world." Such studies show that English-speaking adults weight vocalic duration greatly and formant offsets slightly in voicing decisions for word-final obstruents. Using more natural stimuli, Nittrouer [J. Acoust. Soc. Am. **115**, 1777–1790 (2004)] found different results, raising questions about what would happen if experimental conditions were even more like the real world. In this study noise was used to simulate the real world. Edited natural words with voiced and voiceless final stops were presented in quiet and noise to adults and children (4 to 8 years) for labeling. Hypotheses tested were (1) Adults (and perhaps older children) would weight vocalic duration more in noise than in quiet; (2) Previously reported age-related differences in cue weighting might not be found in this real-world simulation; and (3) Children would experience greater masking than adults. Results showed: (1) no increase for any age listeners in the weighting of vocalic duration in noise; (2) age-related differences in the weighting of cues in both quiet and noise; and (3) masking effects for all listeners, but more so for children than adults. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1940508]

PACS number(s): 43.71.Ft, 43.71.An [ALF] Pages: 1072–1088

I. INTRODUCTION

The way speech is perceived depends upon the native language of the listener. For example, native Japanese speakers fail to use the third formant ($F3$) transition when deciding whether a word initial liquid is /ɹ/ or /l/ (MacKain, Best, and Strange, 1981; Miyawaki *et al.*, 1975). Native English speakers, on the other hand, show /ɹ/–/l/ labeling responses that reveal a strong dependence on the $F3$ transition. This language-related difference in attention paid (or weight assigned) to the $F3$ transition is observed even though the same sets of English- and Japanese-speaking listeners (who provided the labeling results) have been found to discriminate nonspeech spectral glides in the region of $F3$ equally well (Miyawaki *et al.*, 1975).

Cross-linguistic differences have also been described for decisions regarding the voicing of word-final stops. When a speaker produces words that differ phonetically only in the voicing of a final stop (e.g., cap/cab, wait/wade, and duck/dug), differences in articulation occur throughout the syllable: The jaw lowers faster and farther in syllables with voiceless, rather than voiced, final stops (Gracco, 1994; Summers, 1987). The jaw remains open (Summers, 1987) and the tongue retains its vowel-related posture longer in syllables with voiced final stops (Raphael, 1975). And, the relative timing of the offset of laryngeal vibration and of vocal-tract closure differs across words depending on voicing of final stops. For words with voiceless final stops, laryngeal vibration is halted before vocal-tract closure is achieved. For words with voiced final stops, laryngeal vibration continues into the closure until sub- and supraglottal pressures are equalized. All these articulatory differences create numerous acoustic differences between words with voiced and voiceless final stops: Words with voiced final stops have longer vocalic segments than words with voiceless final stops. Formant transitions at the ends of the vocalic portions differ depending on the voicing of the final stop; in particular, the first formant ($F1$) is generally higher at voicing offset when the final stop is voiceless, rather than voiced. Words with voiced final stops have voicing present during the closure; words with voiceless final stops do not. The frequency of ($F1$) at syllable center tends to be higher for words with voiceless final stops than for words with voiced final stops. But, of all these acoustic differences between words with voiced and voiceless final stops, the perceptual influence on adults' voicing decisions of the duration of the vocalic syllable portion and of syllable-final formant transitions (particularly $F1$) have been most studied (e.g., Crowther and Mann, 1992; 1994; Denes, 1955; Fischer and Ohde, 1990; Flege and Wang, 1989; Hillenbrand *et al.*, 1984; Raphael, 1972; Raphael, Dorman, and Liberman, 1980; Wardrip-Fruin, 1982). Collectively these studies have shown that adult speakers of all languages examined weight $F1$ transitions at voicing offset similarly. However, listeners differ in the extent to which they weight the duration of the vocalic portion depending on native language background. Speakers of languages that permit obstruents in syllable-final position and that demonstrate a vocalic-length distinction based on the voicing of those final obstruents, such as English, weight vocalic length more strongly than speakers of languages that either do not permit syllable-final obstruents, such as Mandarin (Flege and Wang, 1989), or that do not demonstrate a vocalic-length distinction based on the voicing

[a)]Portions of this work presented at the 145th meeting of the Acoustical Society of America, Nashville, April–May, 2003.
[b)]Electronic mail: nittrouer@cpd2.usu.edu

 0001-4966/2005/118(2)/1072/17/$22.50

of those final obstruents, such as Arabic (Crowther and Mann, 1992; 1994; Flege and Port, 1981). Such cross-linguistic results suggest that some learning must be involved for speakers of specific languages to know which properties of the signal demand our perceptual attention, and which may largely be ignored.

Studies comparing labeling results for children and adults support that suggestion. Young children do not weight properties of the speech signal as adults do who share their native language. A characterization of these age-related differences is that children tend to prefer the dynamic resonances arising from the continuously changing cavities of the vocal tract over other acoustic properties, such as static, aperiodic noises, and durational differences.[1] Adults, on the other hand, seem to know when a nondynamic property can come in handy in making a phonetic decision in their native language. Several lines of investigation bolster this characterization of developmental changes in perceptual weighting strategies for speech. Studies investigating the labeling of fricatives have shown that children rely more on the voiced formant transitions in the vicinity of those fricatives than adults do, but rely less on the fricative noises themselves (Mayo *et al.*, 2003; Nittrouer, 1992; Nittrouer *et al.* 2000; Siren and Wilcox, 1995). Studies investigating voicing decisions for syllable-final obstruents have shown that children (3 to 6 years of age) rely more on voiced formant transitions preceding vocal-tract closure and less on the length of the vocalic portion than adults (Greenlee, 1980; Krause, 1982; Nittrouer, 2004; Wardrip-Fruin and Peach, 1984), although these strategies may begin to take on the characteristics of adults in the native language community by 5 years of age (Jones, 2003).

The idea that children initially focus on dynamic signal components in perception parallels suggestions that children first master global vocal-tract movements in their productions. Dynamic signal components arise from global movements of upper vocal-tract articulators. There is evidence from investigators such as MacNeilage and Davis (e.g., 1991) to suggest that articulators operate as a common coordinative structure in children's early speech production: that is, articulators work in synchrony, largely following jaw action. This suggestion is perfectly consistent with more general ideas concerning the development of movement control. For example, the work of Thelen and colleagues on the development of leg movements shows that initially these movements are "global and inflexible," but gradually "…limb segments become both disassociated from these global synergies and reintegrated into more complex coalitions." (from abstract of Thelen, 1985.) Similarly, children gradually acquire the ability to organize movements of isolated regions of the vocal tract in order to make refined constriction shapes, with precise timing patterns. However, in the case of speech, these precise patterns are language specific. Consequently, their acquisition is likely shaped by the child's emerging attention to details of the speech signals produced by others. It appears that burgeoning abilities both to produce more precise articulatory gestures and to attend to details of the speech signal develop in lock step, with each facilitating the other.

But, not all experiments examining developmental changes in perceptual strategies for speech have found differences between children and adults in their perceptual weighting of dynamic and nondynamic signal components. In some cases, this is as expected. For example, adults and children alike weight formant transitions greatly and fricative noises hardly at all in place decisions for /f/ versus /θ/ (Harris, 1958; Nittrouer, 2002). This result is not surprising because /f/ and /θ/ noises are spectrally indistinguishable (Nittrouer, 2002). In other experiments, results are difficult to interpret. For example, Mayo and Turk (2004) reported that children weighted formant transitions less and acoustic voice onset time more than adults in voicing decisions for syllable-initial stops. However, it is not clear why these investigators would ever have expected formant transitions to influence voicing decisions for syllable-initial stops much, if at all, given that voiced and voiceless initial stops that share the same place of constriction share the same formant trajectories. The only difference is that larger portions of those trajectories are excited by aspiration noise rather than by a voiced source in words with voiceless initial stops. In Mayo and Turk, the stimuli did not replicate natural tokens in that they contained no portion of aspirated formant transitions. Instead, vocalic segments of identical length were constructed with different onset frequencies for the formants and placed at different temporal distances from a preceding burst noise. This design meant that there were significant silent gaps separating the two syllable portions. Results from Murphy, Shea, and Aslin (1989) showed that children are incapable of integrating acoustic segments in speech stimuli that are separated by silent gaps of several tens of milliseconds. Consequently, the children in Mayo and Turk's study were likely unable to integrate the initial release burst with the following vocalic segment for stimuli with silent gaps longer than their integration thresholds, and so may have been basing decisions on whether they heard one segment or two. Mayo and Turk's stimulus design also meant that syllable duration was perfectly confounded with voice onset time, making it impossible to know whether responses of listeners of any age were due to changes in voice onset time or to changes in overall syllable duration.

In another study, Sussman (2001) investigated vowel perception by adults and 4-to-5-year-olds developing language normally. Stimuli were synthetic /bib/ and /bæb/, with 40-ms transitions on either side of 280-ms steady-state formants. In one condition, 220-ms sections of the steady-state vocalic portions were inserted between transitions for the incongruent vowel. When asked to label the vowel in this condition, all listeners responded with the label associated with the 220-ms steady-state section. From this result, Sussman concluded that listeners of all ages weight steady-state formants most strongly in vowel recognition. However, there is another obvious explanation: The steady-state stimulus sections so overwhelmed the dynamic regions of the syllables that it is little wonder that listeners based their decisions on those steady-state sections. In sum, there remains no strong evidence contradicting the suggestion that as children gain experience with a native language they modify the relative amounts of perceptual attention paid to various signal

properties. Besides, it must be the case *a fortiori* that children's perceptual strategies for speech change through childhood because adults have different perceptual weighting strategies depending on their native language.

In particular, young listeners seem to prefer dynamic resonances of the speech signal, and then learn what additional properties of their native language can help in phonetic decisions. This suggestion makes sense in light of experiments showing that dynamic components play a central role in speech perception, even for adults. In laboratory experiments, investigators have traditionally crafted single-syllable stimuli with great attention to nondynamic signal components, such as aperiodic noises and length distinctions. These stimuli tend to have long steady-state vocalic segments. Rarely do experimental stimuli have more than one region of spectral change (i.e., formant transitions). But, natural speech is intrinsically dynamic. Vocal-tract resonances are constantly changing, rarely, if ever, exhibiting regions of stable formants as long as 220 ms. Relatively recently in the history of speech research, stimulus generation techniques, such as sine wave speech, have been developed to capture the continuously changing nature of these resonances while eliminating other signal attributes. Results of experiments using these stimuli demonstrate that the dynamic resonant patterns by themselves can support accurate speech recognition for adults listening to their native language (e.g., Remez *et al.*, 1981). In turn, this kind of finding suggests that dynamic resonances may be viewed as the "backbone" of the speech signal, so to speak, providing the listener with necessary and almost sufficient information for speech perception. From this perspective it makes sense that dynamic signal components would be what children focus on first.

Of course, there are challenges to the suggestion that speech perception can be accomplished with only dynamic resonances. For one, some experience is generally required for listeners to be able to interpret time-varying sinusoids as phonetically relevant. Even with experience, the perceptual organization required to hear these signals as phonetically coherent forms remains remarkably fragile. Listeners can easily be provoked into abandoning the perceptual posture needed to hear the signals as indivisible structures, instead segregating an individual component from the spectral whole (Remez *et al.*, 2001). In addition, the fact is that speakers do not produce signals that provide information only about global changes in vocal-tract shapes. Instead, speakers go to the trouble of fashioning precise constrictions and carefully timed syllables. Presumably these behaviors serve a purpose, or else they would not have been selected through evolution. In fact, more than 50 years of traditional research into speech perception has shown us that nondynamic components of the speech signal, such as release bursts, fricative noises, and length differences, can affect phonetic perception for adult listeners. Finally, there is no evidence that listeners can understand impoverished signals such as time-varying sinusoids in the noisy listening conditions that generally exist in the real world. Thus, it may be that properties of the signal other than dynamic resonances provide phonetic information more immune to degradation by natural listening environments. This notion is generally referred to under the general heading of "speech redundancy," and has been the prevailing account of why there are several different "cues" to any one phonetic decision. For example, Edwards (1981) wrote of redundancy, "By integrating information from many acoustic cues… the perceptual mechanism is able to accommodate a large variety of source and channel variations." (p. 535). Assman and Summerfield (2004) wrote, "Speech is a highly efficient and robust medium for conveying information because it combines strategic forms of redundancy to minimize loss of information." (p. 231). In particular, it has been suggested that these other properties might help the listener when listening to speech in noisy backgrounds (e.g., Coker and Umeda, 1976).

In this study the roles of vocalic duration and formant offsets in voicing decisions for syllable-final stops were examined when words were presented in noise. Noise was used as a contextual variable that might influence the relative amounts of attention given to various acoustic properties (or cues) precisely because it is the most commonly offered natural condition in which speech redundancy is thought to provide an advantage: if one property is masked, listeners can use a different one. Stimuli varying in the voicing of syllable-final stops provided a particularly appropriate way of exploring the possibility that children might gradually increase the weight given to signal properties other than dynamic resonances because other properties might be more immune to degradation in natural conditions, such as in noisy backgrounds. Several studies have reported that children (3 to 6 years of age) from American English backgrounds fail to weight vocalic duration as strongly as adults from the same language background when making voicing decisions about final stops (Greenlee, 1980; Krause, 1982; Lehman and Sharf, 1989; Wardrip-Fruin and Peach, 1984). Three of these studies also reported that children weight syllable-final formant transitions more than adults (Greenlee, 1980; Krause, 1982; Wardrip-Fruin and Peach, 1984). An earlier study from this laboratory extended those results, showing that children (ages 6 and 8 years) weighted formant transitions more and vocalic duration less than adults when synthetic stimuli were used (Nittrouer, 2004). However, that study also found that adults weighted syllable-final formant transitions more than reported by earlier investigations and more similarly to children when stimuli were created by editing natural tokens; that is, by reiterating or deleting pitch periods in the steady-state vocalic portion of words ending in voiceless and voiced stops, respectively. Figure 1 illustrates findings for adults and 6-year-olds for synthetic and edited-natural *buck/bug* stimuli. In this figure, vocalic duration changes in a continuous fashion. Steps along this continuum are represented on the *x* axis from shortest to longest. Separate functions are plotted for stimuli with offset transitions appropriate for either voiced (filled symbols) or voiceless (open symbols) final stops. In this case the separation between functions is an index of the weight assigned to offset transitions.[2] Slope (i.e., change in units on the *y* axis per unit of change on the *x* axis) is an index of the weight assigned to vocalic duration. The functions on the left, for synthetic stimuli, are fairly close together and steep, although they are more separated and shallower for 6-year-olds than for adults.

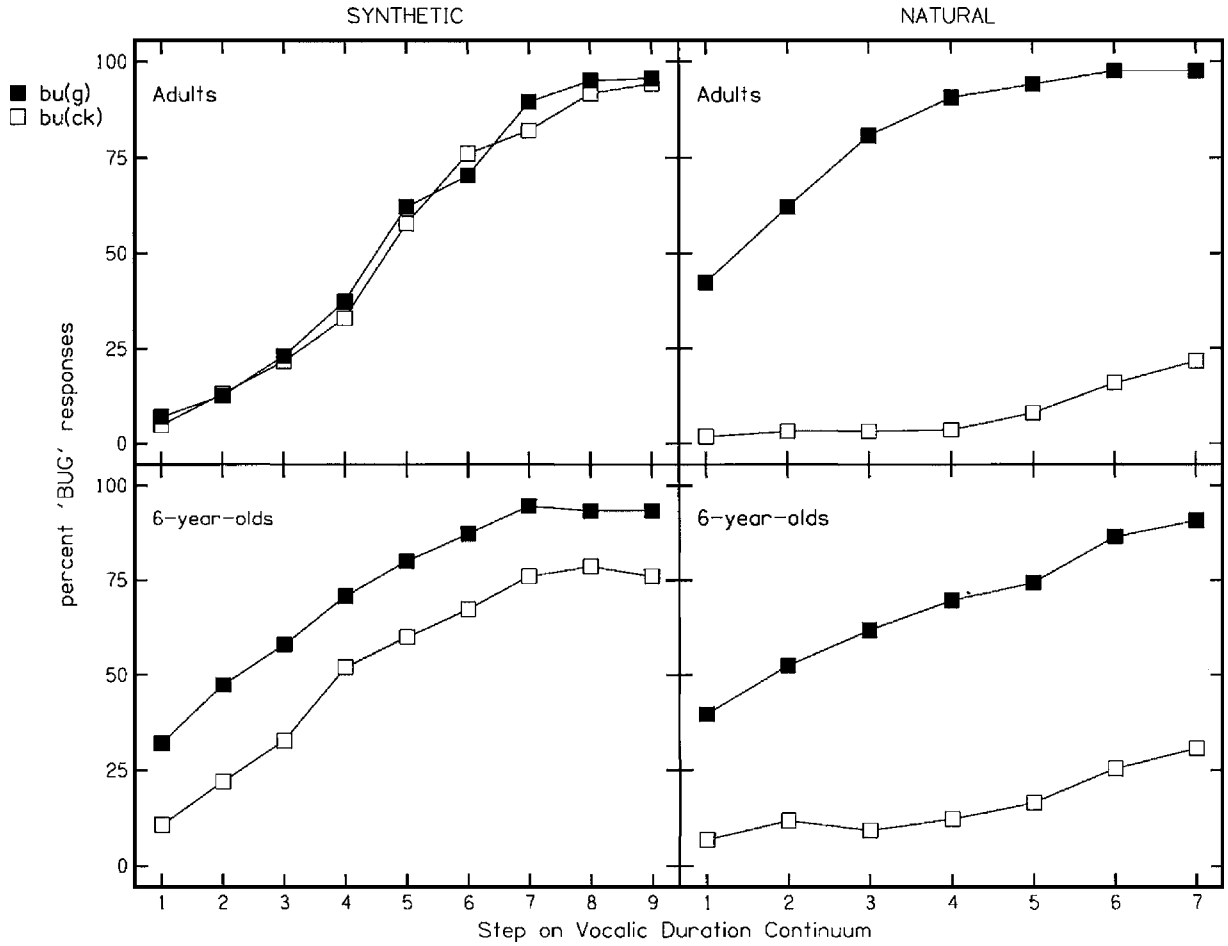

FIG. 1. Labeling functions for *buck/bug* stimuli, synthetic and edited-natural, for adults and 6-year-olds, adapted from Nittrouer (2004).

This pattern indicates that listeners weighted vocalic duration greatly and offset transitions much less so. The functions on the right, for edited-natural stimuli, are more widely separated and shallower, particularly for functions obtained from originally voiceless final stops. The greater separation between functions indicates that both adults and children increased the weight assigned to formant offset transitions when natural stimuli were used instead of synthetic stimuli: Mean separation between functions changed (from synthetic to natural stimuli) from 0.24 to 7.89 steps on the vocalic duration continuum for adults and from 2.80 to 7.27 for 6-year-olds. The age-related difference was statistically significant for synthetic stimuli (0.24 versus 2.80), but not for natural stimuli (7.89 versus 7.27). The shallower functions for natural stimuli indicate that listeners decreased the weight they assigned to vocalic duration when listening to these stimuli instead of the synthetic ones: Mean slopes (across both functions) changed from 0.53 to 0.48 for adults (from synthetic to natural stimuli) and from 0.37 to 0.28 for 6-year-olds. In summary, listeners of all ages (but more so adults than children) increased the weight assigned to offset transitions and decreased the weight they assigned to vocalic duration when edited-natural stimuli were used instead of synthetic stimuli. Working backwards, these results suggest that traditional stimulus synthesis might have caused adults in earlier studies to decrease the weight they normally assign to offset transitions from what they normally do, likely because only $F1$ at syllable offset differed across the ostensibly voiced and voiceless stimuli. Although this effect may have led to erroneous conclusions about what listeners do in the real world, it might also be a demonstration of just how well adults are able to use signal redundancy.

A review of the literature on the perception of voicing for syllable-final stops shows that most studies commonly cited to support the notion that adults base voicing judgments largely on vocalic length have used either synthetic stimuli (e.g., Crowther and Mann, 1992; 1994; Denes, 1955; Fischer and Ohde, 1990; Raphael, 1972) or edited-natural stimuli created only from originally voiced final stops (e.g., Hillenbrand *et al.*, 1984; Hogan and Rozsypal, 1980). In fact, Hogan and Rozsypal explicitly stated that pilot work failed to evoke voiced judgments from adult listeners when stimuli were created by lengthening the steady-state portion of syllables with originally voiceless final stops, so they did not use those syllables in stimulus generation. Findings from Nittrouer (2004) reflect their observation: Figure 1 shows that adults' labeling function for stimuli created from natural, originally voiceless final stops (open symbols in the right-hand plot) is almost completely flat with few "voiced" judgments, even at longer vocalic durations. This pattern contrasts with what is seen for responses to synthetic stimuli with offset transitions appropriate for voiceless final stops (open symbols in the left-hand plot): that function is much steeper, with voiced judgments at longer vocalic durations.

The change in adults' functions across stimulus sets for stimuli with voiced offset transitions (filled symbols) was not as remarkable. In particular, functions were similarly steep for the edited-natural and the synthetic stimuli. Although mean slopes of 0.53 and 0.48 were obtained for synthetic and natural stimuli, respectively, by averaging adults' responses across stimuli with voiced and voiceless offset transitions, a different picture is obtained by looking at functions for voiced and voiceless offsets separately. Mean slopes for adults were the following: synthetic, voiced offset transitions=0.55; natural, voiced offset transitions=0.58; synthetic, voiceless offset transitions=0.51; and natural, voiceless offset transitions=0.37. Thus, all of the change in mean slopes from the synthetic to the natural stimulus conditions was due to functions for stimuli with voiceless offset transitions being much shallower when edited-natural, rather than synthetic, stimuli were used. Nittrouer concluded that this finding (along with the dramatic increase in weight assigned to offset transitions in the natural, compared to the synthetic, condition) mandates modification in our commonly and collectively held assumptions about the roles of vocalic duration and offset transitions in voicing decisions for final stops: Even adult native speakers of languages with a vocalic-length distinction weight formant transitions strongly, at least when words are naturally produced and heard in quiet.

At the same time, the finding gleaned by looking across studies, that adults greatly attenuate the weight they assign to offset transitions when synthetic stimuli are used, served as the impetus for the current study. Perhaps, the thinking went, there are natural conditions in which formant-offset cues are attenuated, as they have been in the synthetic stimuli of earlier experiments. Under these circumstances it would benefit listeners to shift their perceptual attention to vocalic duration. Accordingly, the current study was designed to examine whether adults (and perhaps older children) would show increased weighting of vocalic duration in voicing decisions for word-final stops when listening in a naturalistic condition that might mask formant offsets.

This study also permitted the examination of other possibilities. For one, it is possible that the whole idea that age-related differences in weighting strategies for speech exist might be an artifact, so to speak, of laboratory methods. Perhaps it is only because adults are so skilled at adjusting their listening strategies to fit the conditions that differences between children and adults have ever been found in laboratory studies. When some cues are artificially constrained, perhaps adults are able to turn their attention to other cues, but children cannot. In the real world, where the availability of cues might differ from that of the laboratory, perhaps adults and children use the same strategies. By replicating one condition in the real world this possibility could be tested. And so, another purpose of the current study was to see if age-related differences in weighting strategies exist in natural listening conditions. Although the hypothesis that adults and children would weight acoustic properties similarly in natural conditions (which in this case meant noisy conditions) conflicts with the hypothesis that adults, but not children, would use cue redundancy to make voicing judgments in noise, both hypotheses could be tested by the experimental design.

There is a solid basis for suggesting that noise might mask offset transitions more than it masks vocalic duration. Transitions tend to be lower in amplitude than syllable nuclei because they occur at syllable margins: in this case, when the vocal tract is closing, and so amplitude is falling. Differences in the durations of syllable nuclei primarily account for differences in vocalic duration between voicing conditions. Consequently, voicing-related differences in vocalic duration should remain salient even in noisy conditions. Mature listeners might benefit from paying particular attention to this cue in noise. Accordingly, children would need to learn to attend perceptually to this cue, a skill that native listeners of languages that either do not have syllable-final obstruents or that do not differentiate vocalic duration based on the voicing of those obstruents apparently never acquire.

To test this idea, stimuli differing in vocalic duration created by editing natural words with voiced or voiceless final stops were presented in quiet and in noise to adults and children for labeling. For the quiet condition, listeners of all ages were expected to show labeling functions similar to those for edited-natural stimuli in Fig. 1: that is, functions were expected to be widely separated and shallow (at least those for stimuli created from words with voiceless final stops). If adults (and older children) showed a perceptual weighting shift when words were presented in noise, labeling functions would be expected to resemble those for synthetic stimuli in Fig. 1. That is, they would be less separated and both would be steep. This pattern would indicate that perceptual attention shifted away from offset transitions, and towards vocalic duration.

Two word pairs were used that differed in the frequency of *F*1 at syllable center: *boot/booed* and *cop/cob*. The reason for this was that the extent of the *F*1 transition near voicing offset might affect how robust the *F*1-transition cue is to masking. The frequency of *F*1 at voicing offset is lower for voiced than for voiceless final stops, but this voicing-related difference is greater for words with high medial *F*1 frequencies. Words with low medial *F*1 frequencies fail to show much of a difference in final *F*1 frequency because *F*1 is low throughout the syllable. Consequently, it may be that words with lower medial *F*1 frequencies (such as *boot/booed*) might be more affected by noise masking than words with higher medial *F*1 frequencies (such as *cop/cob*). For these words with low medial *F*1 frequencies listeners, especially adults, might show more of a weighting shift for vocalic duration.

Care was given to the decision of what kind of noise to use as a masker. In general, studies of speech perception in noise have used either speech babble or speech-shaped noise. The reason is that often the speech of others masks the speech signal of interest in natural environments. However, environmental noises (e.g., air-handling equipment, computers, fax machines, traffic, wind, etc.) can mask speech, as well, and these environmental noises have flatter spectra. Therefore, the decision was made to use flat-spectrum noise, which should replicate the combined effects of speech and other environmental maskers.

The decision regarding which signal-to-noise ratio(s) (SNRs) to use when presenting words in noise was also carefully made. Children generally recognize speech less accurately than adults when speech is presented at the same SNRs to both groups (Nittrouer and Boothroyd, 1990). However, this result was found for speech-shaped noise maskers. There was no way of knowing how children would perform with a flat-spectrum masker before this experiment was undertaken, but one report suggested that adults could be expected to perform more poorly with a flat-spectrum than with a speech-shaped masker (Kuzniarz, 1968). Ideally, the SNR selected for each listener would provide the same amount of masking across listeners, indicated by similar overall speech recognition scores across age groups. Initially, the belief was that SNR would likely need to be adjusted among age groups to provide the same amount of masking. Pilot testing, however, showed that recognition scores for consonant–vowel–consonant words were similar for listeners of different ages at a variety of SNRs. Consequently, the decision was made to present words ending with voiced and voiceless stops at one SNR (in addition to quiet) to all listeners: 0 dB, which resulted in roughly 45%–50% correct recognition for listeners of all ages. In addition, adults heard the stimuli at one poorer SNR: −3 dB. General speech recognition scores for CVC words were obtained from all listeners participating in the labeling experiment, even though pilot testing showed similar results for listeners of all ages, just to document that these specific listeners showed similar recognition scores at each SNR. Finally, recognition scores in quiet were also obtained to ensure that recognition scores in noise actually reflected masking effects, rather than merely indexing how well listeners can recognize the particular words used.

In summary, the hypothesis was tested that listeners (especially adults) would decrease the weight they assigned to offset transitions and increase the weight they assigned to vocalic duration when conditions changed from quiet to noise. This hypothesis would be supported by steeper functions that were closer together. At the same time, the hypothesis was tested that adults and children might perform similarly when real-world conditions were simulated. The hypothesis was also tested that children might experience more masking of formant transitions than adults experienced. Nittrouer and Boothroyd (1990) found that children's recognition scores were generally poorer at every SNR than those of adults, although the effects of linguistic context were similar, leading to the conclusion of greater masking for children. Unfortunately, Nittrouer and Boothroyd had no way of determining which part(s) of the speech signal was particularly masked for children. In the current experiment, the signal properties that could be used for phonetic decisions were restricted to formant offsets and vocalic duration. As already proposed, there was good reason to suspect that formant offsets would be vulnerable to masking, but vocalic duration would not be.

II. METHOD

A. Listeners

Adults and children of the ages 8, 6, and 4 years participated in this experiment. To participate, listeners needed to be native speakers of American English. They had to pass a hearing screening of the pure tones 0.5, 1.0, 2.0, 4.0, and 6.0 kHz presented at 25 dB HL. Children needed to be within −1 and +5 months of their birthdays: for example, all 4-year-olds were between 3 years, 11 months and 4 years, 5 months. Children needed to score at or above the 30th percentile on the Goldman-Fristoe 2 Test of Articulation, Sounds-in-Words subtest (Goldman and Fristoe, 2000). Children had to be free from significant, early histories of otitis media with effusion, defined as six or more episodes during the first 2 years of life. Adults needed to be between 18 and 40 years of age. Adults needed to demonstrate at least an 11th-grade reading level on the reading subtest of the Wide Range Achievement Test-Revised. (Jastak and Wilkinson, 1984). Meeting these criteria were 22 adults (mean age =26 years), 20 8-year-olds, 22 6-year-olds, and 24 4-year-olds. However, four of the 4-year-olds were unable to reach the minimum criteria for participation in two of the three tasks they were asked to do (word recognition in quiet and noise, *boot/booed* labeling in quiet and noise, and *cop/cob* labeling in quiet and noise), and so their data were not included for any of the tasks.

B. Equipment and materials

Testing took place in a sound-proof booth with the computer that controlled the various tasks in an adjacent control room. The hearing screening was done with a Welch Allen TM262 audiometer and TDH-39 earphones. All stimuli were stored on a computer and presented through a Creative Labs Soundblaster card, a Samson headphone amplifier, and AKG-K141 headphones at a 22.05-kHz sampling rate. The experimenter recorded responses using a keyboard.

For the labeling tasks, two hand-drawn pictures (8 ×8 in) were used to represent each response label: for example, a picture of a police officer was used for *cop* and a picture of a corncob was used for *cob*. Game boards with ten steps were also used with children. They moved a marker to the next number on the board after each block of stimuli. Cartoon pictures were used as reinforcement and were presented on a color monitor after completion of each block of stimuli. A bell sounded while the pictures were being shown and served as additional reinforcement.

C. Stimuli

1. General speech recognition in noise

For evaluating speech recognition at various SNRs, 20 lists were used, each with ten phonetically balanced consonant–vowel–consonant (CVC) words. These word lists were taken from Mackersie, Boothroyd, and Minniear (2001), and were similar to ones used by Boothroyd and Nittrouer (1988) and Nittrouer and Boothroyd (1990). Each word was recorded three times by a male adult speaker, and the token of each word with the flattest intonation but without any vocal glitches was selected for use in this study. Level was equalized for all words, and then words were mixed with randomly generated white noise (i.e., flat spectrum) low-pass filtered with a cutoff frequency of 11.03 kHz (the upper cutoff of the speech stimuli). The level of the

noise relative to the speech stimuli varied in five equal steps between −6 and +6 dB, a range that results of Boothroyd and Nittrouer (1988) and Nittrouer and Boothroyd (1990) suggested should provide recognition scores between 25% and 75% correct for all listeners in this experiment. Four word lists were presented at each of the five SNRs used. Speech stimuli were mixed with the noise for each listener separately such that different lists were presented at each of the five SNRs across listeners. Furthermore, order of presentation of the lists varied across listeners so that the order of presentation of SNR was randomized. Word level was held constant at 68 dB SPL during testing.

2. Labeling words with voiced and voiceless final stops in noise

Stimuli used for the labeling tasks were taken from the second experiment of Nittrouer (2004). These were natural tokens of a male adult speaker saying *cop*, *cob*, *boot*, and *booed*. Three tokens of each word were used so that there was natural variation in properties such as fundamental frequency and intonation. Although efforts are always made to select tokens with similar fundamental frequencies and flat intonation contours, some variation inevitably exists. When listeners are hearing tokens from only two word categories, these slight variations could influence phonetic decisions if only one token of each word is used. Having several tokens of each word, with all the natural variability that entails, controls for this possible confound.

For each word, the release burst and any voicing during closure was deleted. Vocalic length was manipulated either by reiterating a single pitch period from the most stable region of the vocalic portion or by deleting pitch periods from that stable region. Thus, formant offset transitions were left intact. Care was taken to ensure that the points in the waveform where pitch periods were either reiterated or deleted subsequently lined up at zero crossings to avoid any clicks in the signal. Seven stimuli were created for each token in this way, varying in length from the mean length of the three tokens of the word ending in a voiceless stop to the mean length of the three tokens ending in a voiced stop. For *cop/cob*, the continua varied from 82 to 265 ms. For *boot/booed*, the continua varied from 97 to 258 ms. Steps were kept as equal in size as possible across the continua. Mean *F*1 frequency at voicing offset was 300 Hz across the three tokens of *boot*, 268 Hz across the three tokens of *booed*, 801 Hz across the three tokens of *cop*, and 625 Hz across the three tokens of *cob*. Clearly there was a greater difference in *F*1-offset frequency between *cop* and *cob* than between *boot* and *booed*. In summary, four continua were generated: one each with *boot*, *booed*, *cop*, and *cob* formant offsets. Each continuum had seven stimuli of different lengths, and three tokens of each of those stimuli.

For listening conditions that required that these stimuli be presented in noise, noise was generated in the same way as for the speech recognition task. As with those stimuli, all labeling stimuli were equalized in amplitude before being combined with the noise. And again, the level of the words was held constant at 68 dB SPL. All listeners heard the labeling stimuli presented in noise at 0-dB SNR; adults also heard the stimuli presented at −3-dB SNR.

D. Procedures

Testing took place over two test sessions on different days at least 3 days apart, but not more than 2 weeks apart. This separation between sessions was used to diminish the possibility that there would be learning effects for the word lists, which were presented at both sessions.

The screening tasks were the first things done on the first day, followed by the 20 word lists, either in quiet or in noise. These word lists were the first thing presented on the second day. Half the listeners heard them in quiet on the first day and in noise on the second day, and half heard them in the opposite order. After hearing the 20 word lists, listeners were presented with the labeling stimuli. Children were presented with four sets of stimuli for labeling: *cop/cob* and *boot/booed*, both in quiet and at 0-dB SNR. Adults were presented with six sets of stimuli for labeling: *cop/cob* and *boot/booed*, in quiet and at both 0- and −3-dB SNR. The order of presentation of these stimulus sets was randomized across listeners, with certain restrictions. Children had to hear one set of each word pair (*cop/cob* or *boot/booed*) at each session, and one of these sets had to be presented in quiet and the other at 0-dB SNR. Adults were restricted from hearing the same word pair consecutively, and they had to hear stimuli in each listening condition at each of the two sessions. So, for example, at the first session an adult listener might hear *cop/cob*, *boot/booed*, and then *cop/cob* again in the conditions of 0-dB SNR, quiet, and −3-dB SNR, respectively. At the next session this listener would hear *boot/booed*, *cop/cob*, and *boot/booed*, in the listening conditions of −3-dB SNR, quiet, and 0-dB SNR, respectively.

The task when listening to the 20 word lists at varying SNRs was to repeat the word. The experimenter recorded onto the computer whether the response was correct or not. The word had to be completely correct to be counted as such. Listeners had to recognize correctly at least 180 words on the 20 word lists (90%) when presented in quiet to have their data included in this analysis. This requirement served as a check that all listeners could understand the words used and perform the repetition task.

During the labeling tasks, listeners responded by saying the label and pointing to the picture that represented their selection. Listeners had to pass preliminary tasks with two sets of stimuli in order to proceed to testing. First, unedited versions of the words (i.e., with the release bursts and voicing during closures intact) were presented. Each of the six words (e.g., three tokens of *boot* and three tokens of *booed*) was presented twice. The listener had to respond correctly to at least 11 of the 12 (92%) without feedback to proceed to the next preliminary task. This requirement ensured that all listeners could perform the task and that they all recognized the voicing distinction for word-final stops presented in quiet. This first preliminary task was administered only prior to the first presentation of either set of words (*boot/booed* or *cop/cob*). The second preliminary task was administered

prior to the presentation of each set of words, in each condition. In this task the best exemplars of the six stimuli in the listening condition about to be tested were presented twice each. The term "best exemplar" is used here to refer to the stimulus in which formant transitions and vocalic duration most clearly signaled a specific voicing decision. These stimuli had the release bursts and any voicing during closure removed. So, for example, the best exemplars of *cop* were the three tokens taken from the speaker saying *cop* (so that formant transitions were appropriate for the voiceless stop), with the shortest vocalic portions. The listener needed to respond correctly to at least 11 of the 12 presentations of best exemplars (92%) to proceed to testing. This requirement ensured that all listeners were able to make voicing judgments based on one or the other of the available cues, or a combination of those cues. If listeners do not base the phonetic decision they are being asked to make on the cues available to them, it is pointless to ask questions about the relative weighting of those cues. This preliminary task also serves as a general check on the quality of the stimuli created: If a large number of listeners, particularly adults, cannot hear the presumed best exemplars of each category with near-perfect accuracy, it suggests that the stimuli do not validly replicate natural tokens.

During testing, ten blocks of the 14 stimuli were presented (i.e., stimuli with formant transitions appropriate for a voiced or voiceless final stop, at each of the seven vocalic durations). Because there were actually three tokens with each kind of offset transitions (voiced or voiceless), the program was designed to select randomly one of the three to present during the first block, and then repeat this random selection during the next block without replacement. After three blocks the process was repeated until ten blocks had been presented. For children, cartoon pictures were displayed on the monitor and a bell sounded at the end of each block. They moved a marker to the next space on a game board after each block as a way of keeping track of how much more time they had left in the test. Listeners had to respond correctly to at least 80% of the best exemplars during testing to have their data included in the final analysis. This requirement is commonly viewed as providing a check that the listener paid attention to the task. The reasoning is that, if listeners could respond with better than 90% accuracy to these best exemplars during the preliminary task, then they should be able to respond with at least 80% accuracy during testing, if general attention is maintained during testing. It might also be argued that a listener who labels the best exemplars of each phonetic category with better than 90% accuracy during the preliminary task and then fails to label accurately those same tokens during testing was operating on the perceptual edge, so to speak, during the preliminary task. That is, the listener may have just barely been able to label the best exemplars using the available cues during the preliminary task when no additional demands were present. The increased demands of having to listen to many ambiguous tokens might be enough to disrupt his/her abilities to integrate those cues into a phonetic percept. Regardless, however, of whether the cause of uncertain responding to best exemplars is a general lack of attention or disrupted perceptual processing, there is little to be learned from labeling functions that hover around the 50% line for the length of the function. That sort of responding only means that the listener could not perform the labeling task with the available cues.

Each listener's labeling responses were used to construct cumulative distributions of the proportion of one response (the voiced response in this experiment) across levels of the acoustic property manipulated in a continuous fashion (vocalic duration in this experiment) for each level of the acoustic property manipulated in a dichotomous fashion (formant offsets in this experiment). Best-fit lines were then obtained using probit analysis (Finney, 1964). From these probit functions slopes and distribution means (i.e., phoneme boundaries) were computed. Generally, phoneme boundaries are given in physical units for the property manipulated in a continuous fashion, such as Hz or ms. However, in this experiment step size differed slightly for the two sets of stimuli. Consequently, phoneme boundaries are given using steps as the units of description. Similarly, slope is generally given as the change in probit units per unit change on the physical continuum. In this experiment, slope is given as change in probit units per step. Probit analysis can extrapolate so that phoneme boundaries outside of the range tested can be obtained. For this work, the values that extrapolated phoneme boundaries could take were limited to 3.5 steps beyond the lowest and highest values tested.[3] Mean slope of the function is taken as an indication of the weight assigned to the continuously varied property: the steeper the function, the more weight that was assigned to that property. The separation between functions at the phoneme boundaries (for each level of the dichotomously set property) is taken as an indication of the weight assigned to that dichotomously set property, as long as settings of the property clearly signal each phonetic category involved: the greater the separation, the greater the weight that was assigned. Because these stimuli were created from natural tokens, offset transitions clearly signaled either voiced or voiceless final stops.

Some investigators (e.g., Turner *et al.*, 1998) have computed partial correlation coefficients between each acoustic property and the proportion of one response as a way to describe the weighting of acoustic properties. Nittrouer (2004) compared results for partial correlation coefficients and slopes/phoneme boundaries and found that conclusions reached by the two kinds of metrics were largely the same. However, slopes and separations in phoneme boundaries were found to provide slightly more sensitive estimates of weighting strategies. Furthermore, slopes and separations in phoneme boundaries correspond more directly to visual impressions gleaned from graphed labeling functions. For both these reasons the decision was made to analyze slopes and phoneme boundaries in this study.

III. RESULTS

A. SNR

One 6-year-old and one 4-year-old failed to meet the requirement that they recognize correctly 90% of the words presented in quiet, and so their data were not included. Figure 2 shows mean percent-correct recognition scores for each

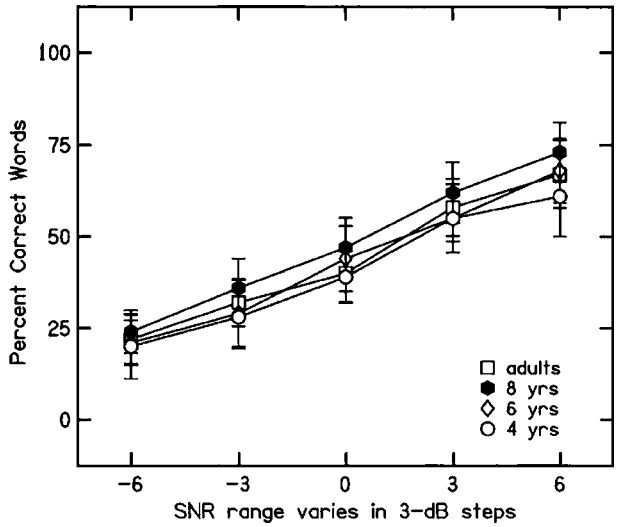

FIG. 2. Percent-correct word recognition for CVC words heard at five SNRs.

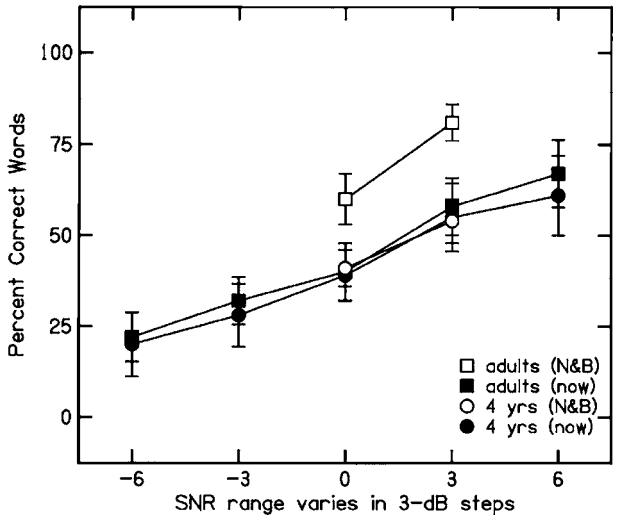

FIG. 3. Percent-correct word recognition for CVC words for adults and 4-year-olds from this experiment (now) and from Nittrouer and Boothroyd (1990) (N&B).

age group at each SNR. Recognition scores are similar across age groups at all SNRs, with the minor exception that 8-year-olds' scores are roughly 5 percentage points better than the other groups at all SNRs. Across SNRs, mean recognition scores (and standard deviations) were: 43.7 (1.7) for adults; 48.6 (1.9) for 8-year-olds; 43.6 (1.9) for 6-year-olds; and 40.6 (1.8) for 4-year-olds. A two-way analysis of variance (ANOVA) was performed on these recognition scores, with age as the between-subjects factor and SNR as the within-subjects factor. The main effect of age was significant, $F(3,78)=10.80$, $p<0.001$, as was the main effect of SNR, $F(4,312)=498.64$, $p<0.001$.[4] Most likely, the significant age effect was largely due to the better recognition scores exhibited by 8-year-olds, rather than to a linear developmental trend. The age $\times$ SNR interaction was not significant.

Although not the focus of this study, it is interesting to compare these results for speech recognition in flat-spectrum noise with those from Nittrouer and Boothroyd (1990) for speech recognition in speech-shaped noise. Figure 3 shows results for adults and 4-year-olds from the current study, and for adults and 4-year-olds from Nittrouer and Boothroyd. Results from Nittrouer and Boothroyd are shown for only 0- and 3-dB SNR because these are the only SNRs that study used. Four-year-olds from the two studies performed identically, but adults from Nittrouer and Boothroyd showed roughly a 20% advantage over adults from this study. It seems that adults benefit from the high-frequency signal portions that are readily available when speech is embedded in speech-shaped noise (as in Nittrouer and Boothroyd) rather than in flat-spectrum noise (as in this experiment). Children, on the other hand, apparently do not achieve this benefit.

B. Labeling results for all age groups, quiet and 0-dB SNR

1. Boot/booed

One 6-year-old and one 4-year-old failed to meet the requirement that they recognize correctly 80% of the best exemplars during testing, and so their data were not included. These were different children from those who failed to meet the criterion for participation with the word lists presented in varying SNRs.

a. Adults versus children. Figure 4 shows labeling functions for all age groups for *boot* and *booed* presented in quiet and at 0-dB SNR. This figure indicates that functions were similarly placed for children and adults when stimuli were heard in quiet, particularly for stimuli with *boot* offset tran-

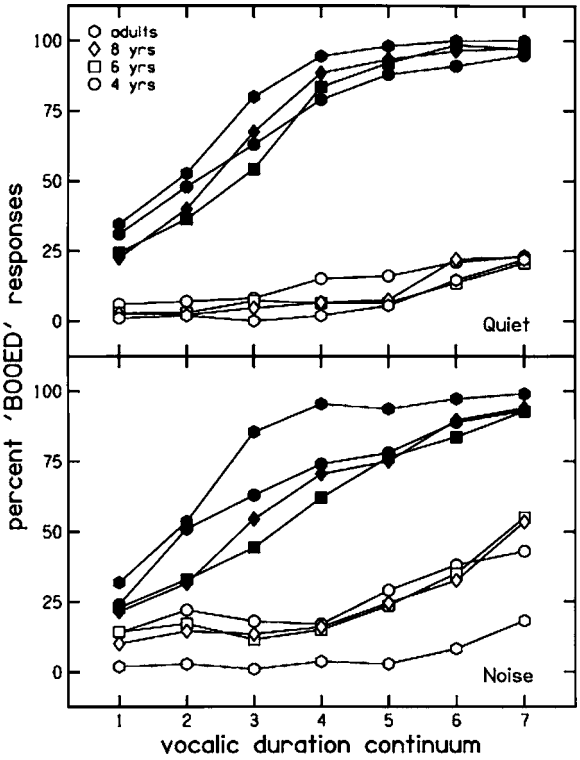

FIG. 4. Labeling functions for *boot/booed* stimuli presented in quiet and at 0-dB SNR, plotted with all age groups together. Filled symbols indicate responses to stimuli with booed formant offsets; open symbols indicate responses to stimuli with *boot* formant offsets.

sitions. Children's labeling functions for stimuli with *booed* offset transitions are slightly more to the right (i.e., towards longer vocalic durations) than those of adults, indicating that children did not weight these offset transitions quite as strongly as adults did. When these stimuli were heard in noise, children's labeling functions for stimuli with *booed* and *boot* offset transitions show less separation than those of adults. Graphically, children's functions are closer to the center of the plot than are those of adults. That is, functions for stimuli with *booed* offset transitions are more to the right (towards longer vocalic durations) and stimuli with *boot* offset transitions are more to the left (towards shorter vocalic durations). This pattern indicates that children did not weight offset transitions as strongly as adults. It is difficult to tell how slopes of the functions may have changed, if at all, across noise conditions, but it does appear as if adults' function for the originally voiced stimuli might be steeper than those of children in the noise condition.

To investigate the apparent age-related effects observed in Fig. 4, simple effects analyses were done on phoneme boundaries for each listening condition separately, with age as a between-subjects' factor and formant offsets as a within-subjects' factor. Simple effects analysis is often a reasonable selection of statistical test for experiments with several independent factors because it permits the examination of effects for one or more of those factors at each level of another factor, while using the overall estimate of error variance.

For stimuli presented in quiet, only the main effect of formant offsets was significant for phoneme boundaries, $F(1,78)=808.30$, $p<0.001$. The age × formant offsets interaction was close to significant, $F(3,78)=2.47$, $p=0.068$, probably reflecting the slight difference in placement of adults' and children's functions for stimuli with *booed* offset transitions. For stimuli presented in noise, the main effect of formant offsets was again significant, $F(1,78)=731.70$, $p<0.001$, and this time the age × formant offsets interaction was significant, $F(3,78)=12.40$, $p<0.001$. Therefore, it seems fair to conclude that listeners of all ages placed labeling functions in roughly the same locations when stimuli were heard in quiet, indicating that formant offset transitions were weighted similarly. However, when stimuli were heard in noise, children's labeling functions were actually less separated, indicating that they decreased the weight they assigned to those formant offset transitions from the quiet condition.

Simple effects analysis was done on slopes for each of the four functions separately with age as the between-subjects factor. Only the function for stimuli with *booed* offsets presented in noise showed a significant age effect, $F(3,78)=10.35$, $p<0.001$, although it was close to significant for stimuli with *booed* offsets presented in quiet, $F(3,78)=2.37$, $p=0.077$. Clearly adults weighted vocalic duration more than children for stimuli presented in noise (at least for those stimuli with *booed* offsets): Mean slopes for stimuli with *booed* offsets presented in noise for individual age groups were 0.79 (0.39) for adults; 0.46 (0.21) for 8-year-olds; 0.40 (0.11) for 6-year-olds; and 0.43 (0.25) for 4-year-olds. For stimuli presented in quiet, mean slopes for stimuli with *booed* offsets presented in quiet were 0.74 (0.36) for adults; 0.70 (0.31) for 8-year-olds; 0.55 (0.31) for 6-year-olds; and 0.52 (0.32) for 4-year-olds. Thus, it appears that children actually assigned slightly more weight to vocalic duration for stimuli presented in quiet than they did for stimuli presented in noise.

b. Effects for each age group. The information above compared results for children and adults, which was necessary to do in order to address two of the three hypotheses posed. However, the third hypothesis to be tested, that adults (and perhaps older children) would weight vocalic duration more in noise than in quiet, could only be addressed by comparing results in quiet and noise for each age group separately.

Figure 5 shows labeling functions for each age group separately for *boot/booed* when stimuli were presented in quiet and at 0-dB SNR. Regarding the weight assigned to offset transitions, evidence can be gathered from the separation between labeling functions. Adults' labeling functions appear similar for stimuli presented in quiet and at 0-dB SNR, but children's labeling functions appear less separated for stimuli presented in noise, rather than in quiet. To see whether these noise-related changes were significant, simple effects analyses were performed on phoneme boundaries for each age group separately. The term of most interest was the noise × formant offset interaction because a significant interaction would indicate that indeed the direction of change in phoneme boundaries across listening conditions was different for stimuli with *boot* and *booed* offset transitions (i.e., functions were "moving towards the center"). Results are shown in Table I and show that all three children's groups had significant noise × formant offset interactions. This pattern of the functions moving closer to one another when the stimuli were presented in noise, rather than in quiet, indicates that children weighted those offset transitions less when stimuli were heard in noise than when they were heard in quiet. Because adults' labeling functions did not differ in location for stimuli heard in noise and in quiet, it can be concluded that they weighted offset transitions equally in both conditions.

Of course, the focus of this particular experiment was on the possibility that listeners (in particular, adults) might increase the weight assigned to vocalic duration in decisions of word-final voicing when speech is heard in noisy backgrounds. If the perceptual weight for vocalic duration increased when listening in noise, then the slopes of labeling functions for stimuli heard in noise would be steeper than those of stimuli heard in quiet. To evaluate this possibility, a simple effects analysis was performed on slopes for each age group separately. The effect of noise was examined separately for the *booed* and *boot* functions. Only results from 8-year-olds showed a significant noise effect, and only for stimuli with *booed* offset transitions, $F(1,78)=8.48$, $p=0.004$, although 6-year-olds' results for stimuli with *booed* offset transitions were close to significant, $F(1,78)=3.42$, $p=0.068$. However, instead of functions being steeper when stimuli were heard in noise, rather than in quiet, they were shallower: for 8-year-olds, mean slope=0.46 probit units in noise versus 0.70 in quiet; for 6-year-olds, mean slope =0.40 probit units in noise versus 0.55 in quiet. These results

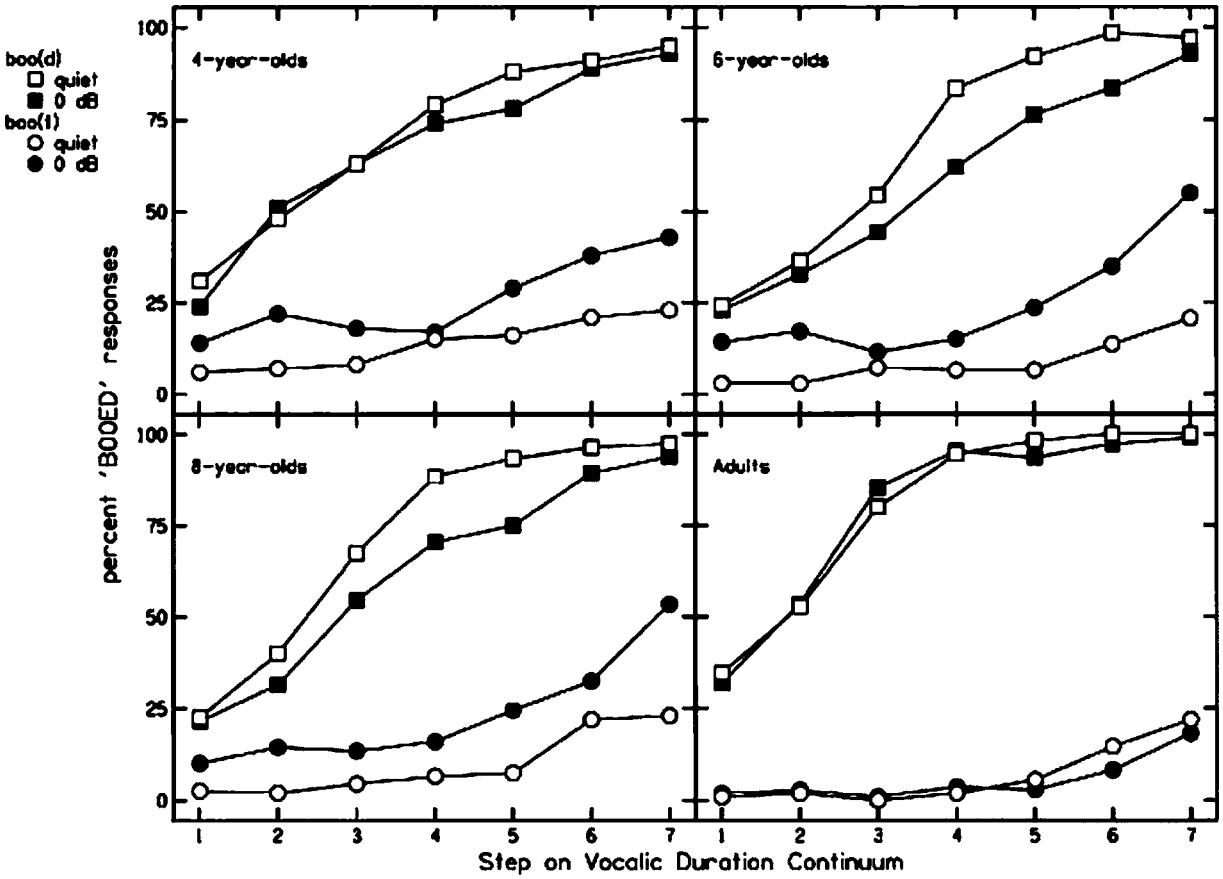

FIG. 5. Labeling functions for *boot/booed* stimuli presented in quiet and at 0-dB SNR, plotted with each age group separately.

indicate that 8-year-olds (and probably 6-year-olds) weighted vocalic duration less when stimuli were in noise, rather than in quiet. These changes are opposite to predictions.

2. Cop/cob

All listeners were able to complete this task.

a. Adults versus children. Figure 6 shows labeling functions for all age groups for *cop* and *cob* presented in quiet and at 0-dB SNR. As with *boot/booed* stimuli presented in quiet, it appears that listeners of all ages had similarly placed labeling functions for these *cop/cob* stimuli when they were presented in quiet. The results of the simple effects analysis done on phoneme boundaries for each listening condition separately confirmed this impression: for the quiet condition, only the main effect of formant offsets was significant, $F(1,80)=680.22$, $p<0.001$. In particular, the age $\times$ formant offsets interaction was not significant, nor close to significant. Although it appears from the lower half of Fig. 6 that children's labeling functions for *cop/cob* stimuli presented in noise may be less separated than those of adults, the simple effects analysis done on phoneme boundaries does not confirm this impression: As with stimuli presented in quiet, only the main effect of formant offsets was significant, $F(1,80)=692.13$, $p<0.001$. Consequently, the conclusion may be drawn that adults and children weighted formant offsets similarly for these stimuli, in both noise and quiet.

As with results for *boot/booed* stimuli, simple effects

TABLE I. Results of simple effects analysis (for each age group separately) for phoneme boundaries, *boot/booed* stimuli presented in quiet and in noise at 0-dB SNR. The main effect of noise refers to whether stimuli were heard in quiet or in noise. The main effect of formant offsets refers to whether formant offset transitions were consistent with a final voiced or voiceless stop. Degrees of freedom were 1, 78 for all effects.

	Noise		Formant offsets		Noise $\times$ formant offsets	
	F	*p*	*F*	*p*	*F*	*p*
Adults	0.12	NS	437.35	<0.001	1.03	NS
8-year-olds	0.46	NS	190.34	<0.001	11.26	=0.001
6-year-olds	0.04	NS	226.86	<0.001	26.29	<0.001
4-year-olds	2.24	NS	237.36	<0.001	6.86	=0.011

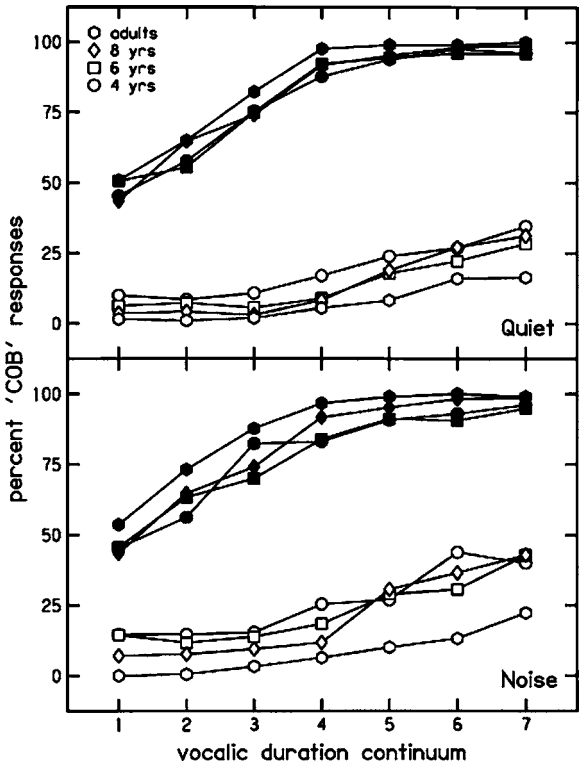

FIG. 6. Labeling functions for *cop/cob* stimuli presented in quiet and at 0-dB SNR, plotted with all age groups together. Filled symbols indicate responses to stimuli with *cob* formant offsets; open symbols indicate responses to stimuli with *cop* formant offsets.

analyses were conducted on slopes for each function separately, with age as the between-subjects factor. The main effect of age was significant for both functions from stimuli presented in noise: for stimuli with *cop* offsets, $F(3,80)=5.00$, $p=0.003$; for stimuli with *cob* offsets, $F(3,80)=7.02$, $p<0.001$. For stimuli with *cop* offsets presented in noise, mean slopes were 0.34 (0.22) for adults; 0.31 (0.22) for 8-year-olds; 0.19 (0.11) for 6-year-olds; and 0.17 (0.11) for 4-year-olds. For stimuli with *cob* offsets presented in noise, mean slopes were 0.66 (0.34) for adults; 0.50 (0.24) for 8-year-olds; 0.33 (0.14) for 6-year-olds; and 0.41 (0.23) for 4-year-olds. Looking at stimuli presented in quiet, the main effect of age was close to significant for stimuli with *cob* offsets only, $F(3,80)=2.30$, $p=0.084$. For these stimuli, mean slopes were 0.74 (0.37) for adults; 0.61 (0.28) for 8-year-olds; 0.46 (0.35) for 6-year-olds; and 0.56 (0.41) for 4-year-olds. Consequently, the conclusion may be drawn that adults weighted vocalic duration more than children for stimuli presented in noise, and possibly for stimuli presented in quiet when formant offsets were appropriate for a voiced final stop.

b. Effects for each age group. Figure 7 shows labeling functions for *cop/cob* for each age group separately. Unlike Fig. 5 showing functions for *boot/booed*, it appears that listeners in all groups performed similarly for stimuli presented in quiet and noise. Statistical analyses support that conclusion. Table II shows results of the simple effects analysis done on phoneme boundaries for *cop/cob*. No age group showed a significant noise × formant offsets interaction. As with results for *boot/booed*, simple effects analyses were

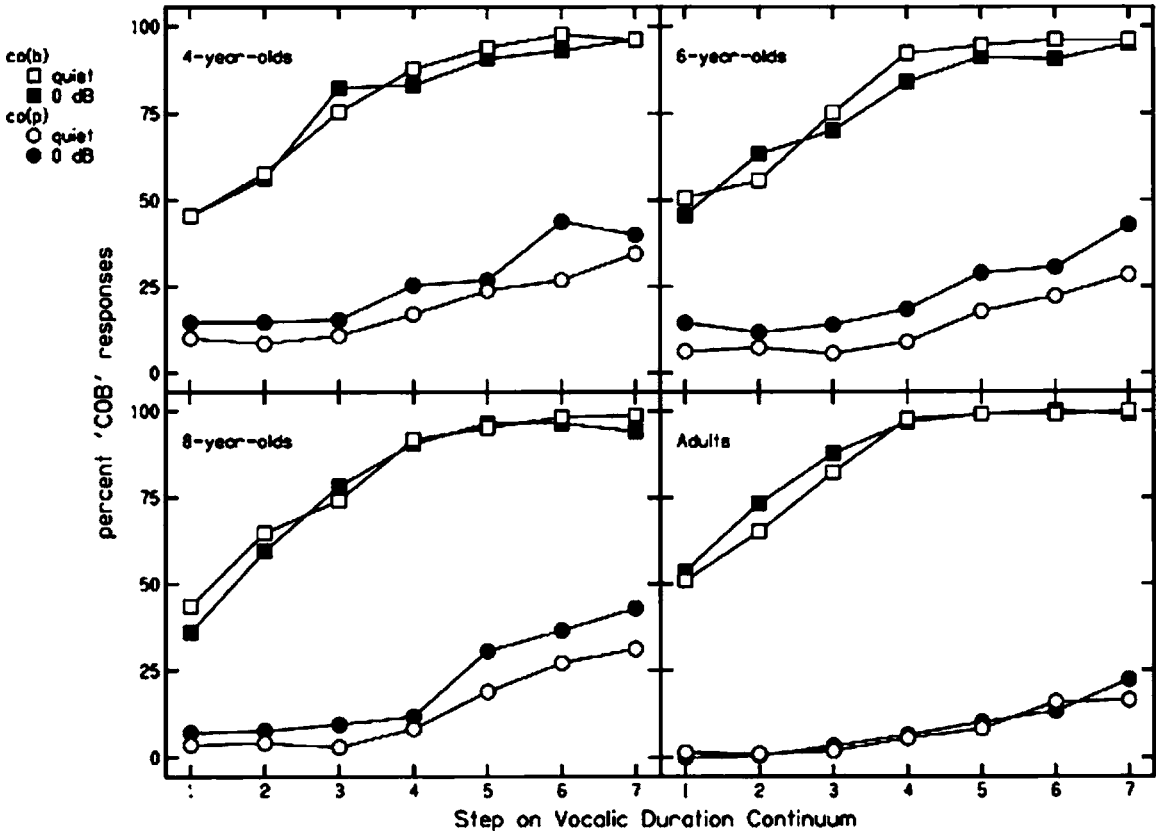

FIG. 7. Labeling functions for *cop/cob* stimuli presented in quiet and at 0-dB signal-to-noise ratio, plotted with each age group separately.

TABLE II. Results of simple effects analysis (for each age group separately) for phoneme boundaries, *cop/cob* stimuli presented in quiet and in noise at 0-dB SNR. Degrees of freedom were 1,80 for all effects.

	Noise		Formant offsets		Noise × formant offsets	
	F	*p*	*F*	*p*	*F*	*p*
Adults	2.41	NS	252.03	<0.001	0.06	NS
8-year-olds	3.27	=0.074	169.81	<0.001	0.76	NS
6-year-olds	1.80	NS	245.16	<0.001	0.31	NS
4-year-olds	1.88	NS	212.30	<0.001	0.04	NS

conducted on slopes for *cop* and *cob* functions separately, for each age group. No significant or marginally significant noise effects were found. Thus, listeners of all ages weighted vocalic duration similarly in noise and quiet.

The finding of significant age effects for slopes when stimuli were presented in noise, but not in quiet (other than the marginally significant age effect for stimuli with *cob* offsets presented in quiet) suggests that listeners modified their weighting of vocalic duration based on whether stimuli were presented in noise or in quiet, and children showed greater shifts than adults. These shifts were enough to create significant age effects for stimuli presented in noise that were not seen for stimuli presented in quiet, but not enough for any one listener group to show a significant noise effect. Of importance to the current study, the slight, nonsignificant shifts in weighting of vocalic duration from the quiet to the noise condition were in the direction of less weight being assigned to vocalic duration in noise than in quiet for all groups. This shift is opposite to the prediction.

C. Labeling results for adults, quiet, 0-dB SNR, and −3-dB SNR

Results of the analyses done on labeling functions for stimuli presented in quiet and at a 0-dB SNR (described above) revealed no significant differences for adults in placement or steepness of functions for *boot/booed* or *cop/cob* presented in these two conditions. However, adults also heard stimuli at an even poorer SNR (−3 dB) to see if this decrement in SNR would affect their performance.

Figure 8 shows labeling functions for adults for stimuli presented in the three listening conditions (quiet, 0-dB SNR, and −3-dB SNR), for *boot/booed* and *cop/cob*. From this figure it appears that functions are similar across listening conditions for the *cop/cob* stimuli. For the *boot/booed* stimuli, functions for the quiet and 0-dB SNR conditions are similar, as demonstrated by the simple effects analysis for adults described in the previous section, but functions for the −3-dB SNR condition appear to be closer to the middle of the figure. This trend is similar to that observed for children at the 0-dB SNR. To evaluate these impressions, two-way ANOVAs were performed on phoneme boundaries for the *boot/booed* and *cop/cob* stimuli separately, with noise and formant offsets as within-subjects' factors. Results of these analyses are shown in Table III. Of particular importance, the noise × formant offsets interaction was significant for *boot/booed* phoneme boundaries, as it had been for children for stimuli presented in quiet and at 0-dB SNR. This result supports the observation that functions were less separated (i.e., closer to the middle of the plot) when stimuli were presented at −3-dB SNR. This trend was not found for phoneme boundaries for *cop/cob* stimuli.

Simple effects analyses were done on slopes for each function separately. Only the *booed* slopes showed a significant effect of noise, $F(2,42)=6.24$, $p=0.004$. Slopes for the *booed* functions were 0.74 (0.36), 0.79 (0.39), and 0.50 (0.28) for the quiet, 0-dB, and −3-dB conditions, respectively. The noise effect was close to significant for *cob* slopes, $F(2,42)=3.08$, $p=0.057$. Slopes for the *cob* functions were 0.74 (0.37), 0.66 (0.34), and 0.50 (0.20) for the quiet, 0-dB, and −3-dB conditions, respectively. Clearly there is evidence that adults decreased the weight assigned to vocalic duration at the poorest SNR. As with the shift in weighting for vocalic duration observed for children at the 0-dB SNR,

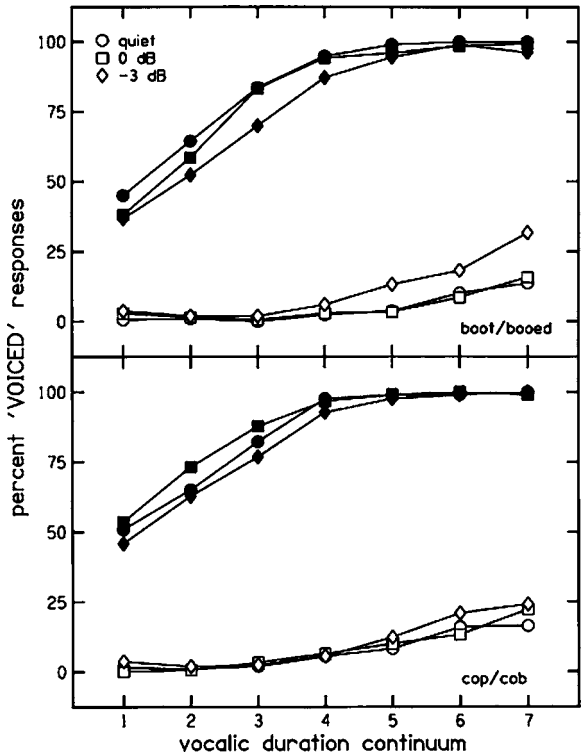

FIG. 8. Labeling functions for *boot/booed* and *cop/cob* stimuli presented in quiet and in noise at 0-dB and −3-dB SNR, adults only. Filled symbols indicate responses to stimuli with voiced formant offsets; open symbols indicate responses to stimuli with voiceless formant offsets.

TABLE III. Results of ANOVAS for phoneme boundaries for *boot/booed* and *cop/cob* stimuli heard in quiet and at 0-dB and −3-dB SNR, adult listeners only. Degrees of freedom were 2, 42 for the main effect of noise, and the noise × formant offsets interaction, and 1,21 for the main effect of formant offsets.

	Noise		Formant offsets		Noise × formant offsets	
	F	*p*	*F*	*p*	*F*	*p*
Boot/booed	2.11	NS	268.20	<0.001	6.36	=0.004
Cop/cob	1.41	NS	363.19	<0.001	0.61	NS

this shift for adults is in the opposite direction to the prediction.

IV. DISCUSSION

The purpose of this experiment was to test the hypothesis that developmental changes in perceptual strategies for speech occur because some acoustic properties in the speech signal are more resistant to noise than others, and mature speech perception takes advantage of this signal redundancy by shifting weight away from the more vulnerable properties and towards the more resistant properties as listening conditions dictate. According to this account, the ability to adjust perceptual weighting strategies depending on listening conditions would be a skill that children would need to develop. However, no evidence was found to support the suggestion that adults make use of redundant cues in this way, at least not for voicing decisions of word-final stops. It does appear, however, that noise masked the signal property weighted most heavily by all listeners in decisions of syllable-final stop voicing: formant offset transitions. That is, the weight assigned to this acoustic property decreased when stimuli were heard in noise, and presumably this decrease was due to those transitions being less available perceptually (i.e., they were masked). This masking could degrade speech recognition in naturally noisy settings. But, even when faced with this degradation, listeners did not shift the focus of their perceptual attention to vocalic duration.

Differences were observed between the two sets of stimuli in how resistant offset transitions were to masking by the noise. *Cop/cob* stimuli had high *F*1 frequencies at syllable center, and so *F*1 fell substantially for cob. As a result, the difference in the frequency of *F*1 at voicing offset between *cop* and *cob* was considerable (176 Hz). *Boot/booed* stimuli, on the other hand, had low *F*1 frequencies at syllable center, and so *F*1 did not fall very much going into closure for the voiced cognate. As a result, there was little difference in the frequency of *F*1 at voicing offset between *boot* and *booed* (32 Hz). These latter stimuli showed greater masking effects for formant offset transitions than did the former stimuli. Whether this is because the offset transitions did not distinguish as strongly between the voiced and voiceless cognates for *boot/booed* as for *cop/cob* or because *F*1 offset transitions are not as extensive for *boot/booed* as for *cop/cob* cannot be determined by this study because the two factors covaried.

Eight-year-olds demonstrated two results that seemed to run counter to developmental trends. First, their recognition scores for words presented in noise were 5 percentage points better than those of adults, at every SNR. No obvious explanation for this result can be offered, but it is probably not important. In spite of having better overall recognition scores than adults, 8-year-olds performed similarly to younger children on the labeling task in noise. Second, 8-year-olds were the only group to show a significant effect of noise on the slope of labeling functions (although the effect was marginally significant for 6-year-olds), and this effect was restricted to stimuli with *booed* offset transitions. For these stimuli, mean slopes in quiet and noise, respectively, were 0.74 vs 0.79 for adults, 0.70 vs 0.46 for 8-year-olds, 0.55 vs 0.40 for 6-year-olds, and 0.52 vs 0.43 for 4-year-olds. Thus, 8-year-olds were able to attend to vocalic duration when stimuli were presented in quiet, but this attention was disrupted when stimuli were presented in noise. A similar trend is observed for 6-year-olds, and to even a lesser extent for 4-year-olds, although effects for these listeners did not reach statistical significance.

A secondary question addressed by this work was whether age-related differences in perceptual weighting strategies would be found when stimuli were presented in more naturalistic conditions than most laboratory studies offer. The most revealing finding in these data is that adults had steeper labeling functions than children (with *p* values for age effects of <0.10) for five of the eight functions obtained. Three of these functions were obtained in the noise condition. In fact, only one labeling function obtained in noise failed to show an age effect on slope, supporting the proposal that there is a genuine age-related difference in the weighting of vocalic duration, even in real-world conditions. This finding fits with the more general notion that children initially attend primarily to the slowly changing, global resonances of the vocal tract, and gradually incorporate information from other sources (such as vocalic duration) as they acquire experience with their native language.

There was also an age-related difference in the amount of masking produced by noise, particularly for *boot/booed* stimuli. This result matches findings of Nittrouer and Boothroyd (1990), who concluded that peripheral masking likely accounts for this difference in masking effect between children and adults. But, greater central masking for children could also explain this difference. Since Nittrouer and Boothroyd was published, two studies have shown that children experience greater masking effects than adults for multitonal maskers (Oh, Wightman, and Lutfi, 2001; Wightman *et al.*, 2003). This sort of masking is thought to be central in nature.

However, these experiments with multitonal maskers used nonspeech stimuli presented simultaneously with the maskers, and so the implications for speech stimuli are not clear. In addition, Wright *et al.* (1997) showed that children with specific language deficits experienced greater backward masking than children developing language normally. Backwards masking is also considered a central effect (Plack, Carlyon, and Viemeister, 1995). Perhaps the difference reported by Wright *et al.* between two groups of children based on the presence or absence of a disorder could reflect a more general, developmental trend. Perhaps children experience more central masking than adults, accounting for the greater reduction in weighting of offset transitions when stimuli are presented in noise. However, Wright *et al.*'s finding was also obtained with nonspeech stimuli, possibly limiting its relevance to speech.

Still one other possible explanation should be considered for the apparent age-related difference in noise masking. Brady, Shankweiler, and Mann (1983) examined recognition in noise of words and environmental sounds by 8-year-olds with reading disorders, and by 8-year-olds learning to read normally. Both groups of listeners were able to recognize words and environmental sounds in quiet with near-perfect accuracy. When the environmental sounds were presented in flat-spectrum noise at a 0-dB SNR, both groups showed similar masking effects. When the words were presented in noise, again at 0-dB SNR, the children who were learning to read normally showed better recognition, compared to their results for environmental sounds. The children with reading disorders showed no improvement in recognition for words in noise over environmental sounds in noise. Brady *et al.*'s conclusion was that the ability to recognize phonetic structure in the acoustic speech signal actually provides some "release from masking," so to speak, for skilled listeners. According to this explanation, the increased masking for speech signals experienced by younger listeners is a consequence of poorer (less-mature) language abilities, rather than a source of those poorer abilities. Unfortunately, this study can shed no light on whether the age-related differences found in the reduction of weighting of formant offset transitions when stimuli were presented in noise can best be explained by central masking effects or by differences in linguistic processing (specifically, in abilities to recover phonetic structure).

In the end, this study leaves open the question of why children gradually modify their perceptual weighting strategies for speech. The reason that has been suggested previously arises from evidence indicating that developmental changes in speech perception strategies and in the abilities to recover and use phonetic structure co-occur. Evidence from different investigators shows that children gradually modify their perceptual weighting strategies for speech (e.g., Greenlee, 1980; Krause, 1982; Nittrouer, 1992; Parnell and Amerman, 1978; Siren and Wilcox, 1995; Wardrip-Fruin and Peach, 1984) and that they gradually acquire skills such as counting word-internal phonetic units (Liberman *et al.* 1974), judging similarity of phonetic structure between different words (Walley, Smith, and Juszcyk, 1986), and using phonetic structure for storing words in working memory (Nittrouer and Miller, 1999). One study showed that these developmental changes in speech perception and abilities to access and use phonetic structure co-occur in the same group of children (Mayo *et al.*, 2003). Finally, several studies have found that when one developmental change is delayed, the other is as well (Nittrouer, 1999; Nittrouer and Burton, 2001; 2005). In light of such evidence, the suggestion has been made that the acoustic properties that gradually, through childhood, come to be weighted more are ones that facilitate the recovering of phonetic structure in the child's native language. Certainly none of the data reported here contradict that suggestion.

In conclusion, this experiment was designed largely to test the hypothesis that children's perceptual weighting strategies for speech change through childhood to allow them to take advantage of signal redundancy in natural listening environments where some acoustic properties may be masked. This hypothesis would have been supported if adults (and perhaps older children, as well) shifted perceptual attention away from formant offset transitions and toward vocalic duration when listening to signals in noise. However, this result was not observed, and so the suggestion that the need to take advantage of redundancy in speech signals motivates developmental shifts in perceptual weighting strategies for speech is not supported. In fact, no evidence was found to support the generally held perspective that skilled perceivers of speech shift the focus of their attention from one acoustic cue to another as listening conditions dictate. The data reported here were consistent, however, with the notion that there is a developmental shift in perceptual weighting strategies for speech in which phonetically relevant signal properties, other than global resonance patterns, come to be weighted more strongly. Finally, a developmental decrease in the masking effects of environmental noise was observed.

ACKNOWLEDGMENTS

The author wishes to thank the following people for their help on this project: Tom Creutz at Boys Town National Research Hospital for writing the software to present stimuli in noise; Melanie Wilhelmsen, Kathi Bodily, and Jennifer Smith for help with data collection; and Carol A. Fowler, John Kingston, Joanna H. Lowenstein, and Donal G. Sinex for their comments on an earlier draft of this manuscript. This work was supported by Research Grant No. R01 DC00633 from the National Institute on Deafness and Other Communication Disorders, the National Institutes of Health.

[1]In keeping with investigators such as Kewley-Port, Pisoni, and Studdert-Kennedy (1983), the term "dynamic" is used here to refer to signal properties involving formant movement. These authors took this term directly from distinctive feature theory. According to this theory, "static" properties contrast with dynamic properties, and refer to broadband spectral patterns that remain stable over at least a few milliseconds, such as steady-state vowel formants, fricative noises, and release bursts. In this work, a temporal property (vocalic length) is considered, and along with static properties is subsumed under the general term of "nondynamic."

[2]In order for the separation between labeling functions to be a valid indicator of the weight assigned to the dichotomously set property (i.e., the property that defines each continuum), the two settings of that property must unambiguously signal each of the two phonetic labels that listeners are

being asked to use. The dichotomously set property in Nittrouer (2004) was formant offsets. Because the stimuli were constructed from natural stimuli, this property clearly signaled voiced and voiceless final stops unambiguously.

[3]Although the value of these limits is somewhat arbitrary, they essentially establish numerical markers for functions that never cross the 50% line. Importantly, the use of these markers only serves to constrain the probability of obtaining statistically significant results, and so cannot bias procedures to show effects where there are none.

[4]F and p values are reported for any results with $p < 0.10$. Results with $p > 0.10$ are described as "not significant" (NS).

Assmann, P., and Summerfield, Q. (**2004**). "The perception of speech under adverse acoustic conditions," in *Speech Processing in the Auditory System, Springer Handbook of Auditory Research*, edited by S. Greenberg and W. Ainsworth (Springer, New York), pp. 231–308.

Boothroyd, A., and Nittrouer, S. (**1988**). "Mathematical treatment of context effects in phoneme and word recognition," J. Acoust. Soc. Am. **84**, 101–114.

Brady, S., Shankweiler, D., and Mann, V. (**1983**). "Speech perception and memory coding in relation to reading ability," J. Exp. Child Psychol. **35**, 345–367.

Coker, C. H., and Umeda, N. (**1976**). "Speech as an error-correcting process," in *Auditory Analysis and Perception of Speech*, edited by G. Fant and M. A. A. Tatham (Academic, New York), pp. 349–364.

Crowther, C. S., and Mann, V. (**1992**). "Native language factors affecting use of vocalic cues to final consonant voicing in English," J. Acoust. Soc. Am. **92**, 711–722.

Crowther, C. S., and Mann, V. (**1994**). "Use of vocalic cues to consonant voicing and native language background: The influence of experimental design," Percept. Psychophys. **55**, 513–525.

Denes, P. (**1955**). "Effect of duration on the perception of voicing," J. Acoust. Soc. Am. **27**, 761–764.

Edwards, T. J. (**1981**). "Multiple features analysis of intervocalic English plosives," J. Acoust. Soc. Am. **69**, 535–547.

Finney, D. J. (**1964**). *Probit Analysis* (Cambridge University, Cambridge, England).

Fischer, R. M., and Ohde, R. N. (**1990**). "Spectral and duration properties of front vowels as cues to final stop-consonant voicing," J. Acoust. Soc. Am. **88**, 1250–1259.

Flege, J. E., and Port, R. (**1981**). "Cross-language phonetic interference: Arabic to English," Lang Speech **24**, 125–146.

Flege, J. E., and Wang, C., (**1989**). "Native-language phonotactic constraints affect how well Chinese subjects perceive the word-final English /t/–/d/ contrast," J. Phonetics **17**, 299–315.

Goldman, R., and Fristoe, M. (**2000**). *Goldman Fristoe 2: Test of Articulation* (American Guidance Service, Inc., Circle Pines, MN).

Gracco, V. L. (**1994**). "Some organizational characteristics of speech movement control," J. Speech Hear. Res. **37**, 4–27.

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.

Harris, K. S. (**1958**). "Cues for the discrimination of American English fricatives in spoken syllables," Lang Speech **1**, 1–7.

Hillenbrand, J., Ingrisano, D. R., Smith, B. L., and Flege, J. E. (**1984**). "Perception of the voiced-voiceless contrast in syllable-final stops," J. Acoust. Soc. Am. **76**, 18–26.

Hogan, J. T., and Rozsypal, A. J. (**1980**). "Evaluation of vowel duration as a cue for the voicing distinction in the following word-final consonant," J. Acoust. Soc. Am. **67**, 1764–1771.

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

Jones, C. (**2003**). "Development of phonological categories in children's perception of final voicing," Unpublished doctoral dissertation, University of Massachusetts, Amherst.

Kewley-Port, D., Pisoni, D. B., and Studdert-Kennedy, M. (**1983**). "Perception of static and dynamic acoustic cues to place of articulation in initial stop consonants," J. Acoust. Soc. Am. **73**, 1779–1793.

Krause, S. E. (**1982**). "Developmental use of vowel duration as a cue to postvocalic stop consonant voicing," J. Speech Hear. Res. **25**, 388–393.

Kuzniarz, J. (**1968**). "Masking of speech by continuous noise," Pol. Med. J. **7**, 1001–1008.

Lehman, M. E., and Sharf, D. J. (**1989**). "Perception/production relationships in the development of the vowel duration cue to final consonant voicing," J. Speech Hear. Res. **32**, 803–815.

Liberman, I. Y., Shankweiler, D., Fischer, F. W., and Carter, B. (**1974**). "Explicit syllable and phoneme segmentation in the young child," J. Exp. Child Psychol. **18**, 201–212.

MacKain, K. S., Best, C. T., and Strange, W. (**1981**). "Categorical perception of English /r/ and /l/ by Japanese bilinguals," Appl. Psycholinguist. **2**, 369–390.

Mackersie, C. L., Boothroyd, A., and Minniear, D. (**2001**). "Evaluation of the Computer-Assisted Speech Perception Assessment Test (CASPA)," J. Am. Acad. Audiol **12**, 390–396.

MacNeilage, P. F., and Davis, B. (**1991**). "Acquisition of speech production: Frames, then content," in *Attention & Performance XIII*, edited by M. Jeanerod (Erlbaum, New York), pp. 453–476.

Mayo, C., and Turk, A. (**2004**). "Adult–child differences in acoustic cue weighting are influenced by segmental context: Children are not always perceptually biased toward transitions," J. Acoust. Soc. Am. **115**, 3184–3194.

Mayo, C., Scobbie, J. M., Hewlett, N., and Waters, D. (**2003**). "The influence of phonemic awareness development on acoustic cue weighting strategies in children″s speech perception," J. Speech Lang. Hear. Res. **46**, 1184–1196.

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A. M., Jenkins, J. J., and Fujimura, O. (**1975**). "An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English," Percept. Psychophys. **18**, 331–340.

Murphy, W. D., Shea, S. L., and Aslin, R. N. (**1989**). "Identification of vowels in 'vowel-less' syllables by 3-year-olds," Percept. Psychophys. **46**, 375–383.

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. (**1999**). "Do temporal processing deficits cause phonological processing problems?" J. Speech Lang. Hear. Res. **42**, 925–942.

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. (**2004**). "The role of temporal and dynamic signal components in the perception of syllable-final stop voicing by children and adults," J. Acoust. Soc. Am. **115**, 1777–1790.

Nittrouer, S., and Boothroyd, A. (**1990**). "Context effects in phoneme and word recognition by young children and older adults," J. Acoust. Soc. Am. **87**, 2705–2715.

Nittrouer, S, and Burton, L. (**2001**). "The role of early language experience in the development of speech perception and language processing abilities in children with hearing loss," Volta Review **103**, 5–37.

Nittrouer, S., and Burton, L. (**2005**). "The role of early language experience in the development of speech perception and phonological processing abilities: Evidence from 5-year-olds with histories of otitis media with effusion and low socioeconomic status," J. Commun. Disord. **38**, 29–63.

Nittrouer, S., and Miller, M. E. (**1999**). "The development of phonemic coding strategies for serial recall," Appl. Psycholinguist. **20**, 563–588.

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.

Oh, E. L., Wightman, F., and Lutfi, R. A. (**2001**). "Children's detection of pure-tone signals with random multitone maskers," J. Acoust. Soc. Am. **109**, 2888–2895.

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

Plack, C. J., Carlyon, R. P., and Viemeister, N. F. (**1995**). "Intensity discrimination under forward and backward masking: Role of referential coding," J. Acoust. Soc. Am. **97**, 1141–1149.

Raphael, L. J. (**1972**). "Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English," J. Acoust. Soc. Am. **51**, 1296–1303.

Raphael, L. J. (**1975**). "The physiological control of durational differences between vowels preceding voiced and voiceless consonants in English," J. Phonetics **3**, 25–33.

Raphael, L. J., Dorman, M. F., and Liberman, A. M. (**1980**). "On defining the vowel duration that cues voicing in final position," Lang Speech **23**, 297–307.

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., Pisoni, D. B., and Carrell, T. D. (**1981**). "Speech perception without traditional speech cues," Science **212**, 947–949.

Siren, K. A., and Wilcox, K. A. (**1995**). "Effects of lexical meaning and practiced productions on coarticulation in children's and adults' speech," J. Speech Hear. Res. **38**, 351–359.
Summers, W. V. (**1987**). "Effects of stress and final-consonant voicing on vowel production: Articulatory and acoustic analyses," J. Acoust. Soc. Am. **82**, 847–863.
Sussman, J. E. (**2001**). "Vowel perception by adults and children with normal language and specific language impairment: Based on steady states or transitions?" J. Acoust. Soc. Am. **109**, 1173–1180.
Thelen, E. (**1985**). "Developmental origins of motor coordination: Leg movements in human infants," Dev. Psychobiol. **18**, 1–22.
Turner, C. W., Kwon, B. J., Tanaka, C., Knapp, J., Hubbartt, J. L., and Doherty, K. A. (**1998**). "Frequency-weighting functions for broadband speech as estimated by a correlational method," J. Acoust. Soc. Am. **104**, 1580–1585.
Walley, A. C., Smith, L. B., and Jusczyk, P. W. (**1986**). "The role of phonemes and syllables in the perceived similarity of speech sounds for children," Mem. Cognit. **14**, 220–229.
Wardrip-Fruin, C. (**1982**). "On the status of temporal cues to phonetic categories: Preceding vowel duration as a cue to voicing in final stop consonants," J. Acoust. Soc. Am. **71**, 187–195.
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.
Wightman, F. L., Callahan, M. R., Lutfi, R. A., Kistler, D. J., and Oh, E. (**2003**). "Children's detection of pure-tone signals: Informational masking with contralateral maskers," J. Acoust. Soc. Am. **113**, 3297–3305.
Wright, B. A., Lombardino, L. J., King, W. M., Puranik, C. S., Leonard, C. M., and Merzenich, M. M. (**1997**). "Deficits in auditory temporal and spectral resolution in language-impaired children," Nature (London) **387**, 176–178.