



Published in final edited form as:

J Speech Lang Hear Res. 2007 February ; 50(1): 58–73.

Children's Weighting Strategies for Word-Final Stop Voicing Are Not Explained by Auditory Sensitivities

Susan Nittrouer and Joanna H. Lowenstein

Utah State University, Logan

Abstract

Purpose: It has been reported that children and adults weight differently the various acoustic properties of the speech signal that support phonetic decisions. This finding is generally attributed to the fact that the amount of weight assigned to various acoustic properties by adults varies across languages, and that children have not yet discovered the mature weighting strategies of their own native languages. But an alternative explanation exists: Perhaps children's auditory sensitivities for some acoustic properties of speech are poorer than those of adults, and children cannot categorize stimuli based on properties to which they are not keenly sensitive. The purpose of the current study was to test that hypothesis.

Method: Edited-natural, synthetic-formant, and sine wave stimuli were all used, and all were modeled after words with voiced and voiceless final stops. Adults and children (5 and 7 years of age) listened to pairs of stimuli in 5 conditions: 2 involving a temporal property (1 with speech and 1 with nonspeech stimuli) and 3 involving a spectral property (1 with speech and 2 with nonspeech stimuli). An AX discrimination task was used in which a standard stimulus (A) was compared with all other stimuli (X) equal numbers of times (method of constant stimuli).

Results: Adults and children had similar difference thresholds (i.e., 50% point on the discrimination function) for 2 of the 3 sets of nonspeech stimuli (1 temporal and 1 spectral), but children's thresholds were greater for both sets of speech stimuli.

Conclusion: Results are interpreted as evidence that children's auditory sensitivities are adequate to support weighting strategies similar to those of adults, and so observed differences between children and adults in speech perception cannot be explained by differences in auditory perception. Furthermore, it is concluded that listeners bring expectations to the listening task about the nature of the signals they are hearing based on their experiences with those signals.

Keywords

children; weighting strategies; auditory; speech perception

When two words differ by just a single phonetic segment, numerous differences can arise in articulatory organization across the entire lengths of the words. For example, English permits a phonetic distinction in the voicing of final stops (e.g., *cap/cab*, *wait/wade*, and *duck/dug*). Superficially, this distinction would seem trivial: The vocal folds abduct sooner in the production of words with voiceless rather than voiced final stops. In truth, the organization of the whole word differs as a function of the voicing of the final stop. The jaw lowers faster and moves to more open positions in syllables with voiceless, rather than voiced, final stops (Gracco, 1994; Summers, 1987). In words with voiceless final stops the jaw is also quicker to close (Summers, 1987), and the tongue is quicker to move away from its vowel-related posture

(Raphael, 1975). For words with voiceless final stops, laryngeal vibration ceases before vocal-tract closure is achieved, whereas laryngeal vibration continues into closure for words with voiced final stops. All these articulatory differences create numerous acoustic differences between words with voiced and voiceless final stops: Words with voiceless final stops have shorter vocalic portions than words with voiced final stops. All of the formant transitions at word endings differ depending on the voicing of the final stop. In particular, the first formant (F1) is higher at voicing offset when the final stop is voiceless rather than voiced. Other formants also differ in frequency at voicing offset depending on the voicing of the final stop, but relations between voicing and offset frequency are not consistent for these formants; rather, differences between voicing conditions in offset frequencies depend on place of closure. The frequency of F1 at the center of the word is higher for words with voiceless final stops than for words with voiced final stops. All of these acoustic differences can be recruited into the perceptual process of recovering the phonetic structure intended by the speaker. In fact, most, if not all, of the signal components are commonly used in speech perception, although the amount of attention (i.e., weight) any one component receives can vary depending on numerous factors. In particular, native language background strongly affects the weighting of the various acoustic properties in the speech signal (e.g., Gordon, Keyes, & Yung, 2001; MacKain, Best, & Strange, 1981). For example, some languages do not permit syllable-final obstruents. As a result, when native speakers of those languages learn a second language, they commonly fail to attend to some of the correlates of syllable-final consonant voicing. In particular, these listeners fail to attend to the vowel-length distinction that accompanies syllable-final voicing in many languages (Crowther & Mann, 1992, 1994; Flege, 1989).

These findings across listeners with different first languages have led to the hypothesis that weighting strategies for speech perception are modified during first language acquisition, and there is strong evidence from developmental studies to support that hypothesis. Specifically, children have been found to weight formant transitions heavily, while not assigning as much weight to other signal properties. Then as they gain experience with their first language, children learn to focus perceptual attention on (i.e., weight) those acoustic properties that facilitate the recovery of phonetic structure from the signal in that language. As early as 1978, Parnell and Amerman showed that 4-year-olds were unable to recognize place of articulation for word-initial stops much above chance when presented with only the burst and aspiration noise: Mean correct recognition was 47%. Adults' recognition with these cues, on the other hand, was highly accurate, at 90% mean correct recognition. When formant transitions were added, 4-year-olds' recognition scores improved to 76%. Numerous studies have also shown that young children (i.e., 3–8 years of age) do not weight fricative noise as much as adults in decisions of place for word-initial fricatives, but instead weight formant transitions at voicing onset greatly (e.g., Mayo, Scobbie, Hewlett, & Waters, 2003; Nittrouer, 1992; Nittrouer, Miller, Crowther, & Manhart, 2000; Siren & Wilcox, 1995).

When it comes to temporal properties, words differing in the voicing of final stops have been used by several investigators to study differences between adults and children in weighting strategies. Although several acoustic properties distinguish between words with voiceless and voiced final stops, the two most prominent perceptual properties seem to be vocalic duration and formant transitions going into vocal tract closure. Again, vocalic duration is shorter before voiceless than before voiced stops. Formant transitions just before voicing offset also differ depending on the voicing of the final stop. F1 is always higher at voicing offset before voiceless stops than before voiced stops because F1 is largely a consequence of how open the vocal tract is. The vocal tract has not achieved complete closure at the time that voicing ends in words with final voiceless stops, and so F1 is high. For words with voiced final stops, on the other hand, complete closure has been reached before voicing ends, and so F1 is at its lowest possible value. Several studies have reported that young children (generally 3–8 years of age) weight vocalic duration less than adults in making decisions about the voicing of word-final stops

(Greenlee, 1980; Krause, 1982; Lehman & Sharf, 1989; Nittrouer, 2004; Wardrip-Fruin & Peach, 1984). With the exception of Lehman and Sharf, these studies also demonstrated that children weight formant offsets more than adults in these decisions. Both sorts of results are illustrated in Figure 1, displaying labeling functions for adults and 6-year-olds from the Nittrouer (2004) study. The stimuli used in that study were largely modeled after those of Crowther and Mann (1992, 1994): Vocalic duration varied from 100 ms (most appropriate for voiceless *buck*) to 260 ms (most appropriate for voiced *bug*), and F1 was the only formant to differ in transition extent at word's end. It was either 650 Hz or 450 Hz at offset. Steady-state F1 was 650 Hz, and so stimuli with the first of these values had no transition and clearly signaled the voiceless *buck*; stimuli with the 450-Hz F1 offset clearly signaled the voiced *bug*. Figure 1 shows that adults based their voicing decisions entirely on vocalic duration: Labeling functions were steep, but not separated at all based on F1 offset frequency. Children's labeling functions were shallower than those of adults, indicating that less weight was assigned to the acoustic property represented on the x-axis, which in this case was vocalic duration. This figure also shows that children's functions for the two different F1 offsets were more separated than those of adults. This finding indicates that children weighted these offset transitions more than adults.

Collectively, studies examining developmental changes in speech perception have led to the general conclusion that children modify the relative amounts of attention they pay to acoustic signal properties as they gain experience with their native language. The fact that the weighting strategies eventually acquired by children vary depending on the language they are learning suggests that developmental changes are likely consequences of experience with that language. Nonetheless, an alternative explanation can be offered: Perhaps the age-related differences in speech perception results reported in the literature are due to age-related differences in auditory sensitivities. That is, perhaps children are poorer than adults at resolving some acoustic attributes, such as noise spectra, gaps in signals, or duration of periodic signals. Accordingly, if children are poor at resolving some properties, it follows that they would not attend to those properties in making phonetic decisions. In fact, there is evidence that children require greater differences in temporal and spectral properties to detect those differences in nonspeech acoustic signals than adults require (e.g., Allen, Wightman, Kistler, & Dolan, 1989; Jensen & Neff, 1993; Wightman & Allen, 1992; Wightman, Allen, Dolan, Kistler, & Jamieson, 1989). But demonstrations that children are not as skilled as adults in resolving selected properties may not automatically warrant attributing age-related differences in speech perception to those observed differences in auditory sensitivities. A primary basis for this suggestion is that it is not clear how sensitive a listener needs to be to any given acoustic property in order to be able to rely on that property in phonetic decisions. In essence, the question can be asked "how sensitive is sensitive enough?", and that question is extremely difficult, if not impossible, to answer. Nonetheless, it seems reasonable to think that a listener needs to have some level of sensitivity to an acoustic property in order to use that property in the course of speech perception. So it remains appropriate to ask if children's weighting strategies might somehow reflect their developing auditory sensitivities. Exploring that relation between nonspeech sensitivities and speech perception strategies is a terribly tricky business, however, because of myriad procedural difficulties.

First, task variables generally differ between experiments with speech and nonspeech stimuli. Studies of speech perception commonly use labeling tasks in which listeners hear one stimulus at a time. As the term suggests, listeners need to assign a label (usually out of two possibilities) to that stimulus. In discrimination tasks, on the other hand, listeners are asked to make decisions regarding similarity and/or differences among two or more stimuli. We do not have, at the moment, a thorough understanding of the differences in nonsensory influences on performance in the two kinds of task, nor do we know what the differences in those influences might be for children and adults. Strictly regarding discrimination of non-speech signals, Wightman, Allen,

and colleagues, who have themselves reported on differences between children and adults in auditory perception (e.g., Allen et al., 1989; Wightman et al., 1989), concluded that “the apparent differences between adults and children in auditory skills may reflect nothing more than the influence of nonsensory factors such as memory and attention” (Wightman & Allen, 1992, p. 133).

Besides the fact that the procedures used to measure auditory sensitivities and speech perception proclivities differ, there are also obvious differences between the two kinds of stimuli used. Speech signals have inherent ecological validity; most nonspeech signals used in perception studies do not. These stimuli are generally composed of signals that can only be generated by a computer. It is possible, for example, that children have difficulty processing signals that do not correspond to something in the environment. We do not know how nonsensory influences might interact with the kinds of signals being heard (speech or nonspeech), and whether those interactions might differ for children and adults.

Another reason why it is hard to make connections between results of nonspeech and speech studies is that upon finding a difference in auditory sensitivities, few investigators go the next step of asking if that difference can predict specific differences in speech perception. One example of a study that did try to make that sort of connection is Rosen and Manganari (2001). Although this study investigated auditory sensitivities and speech perception abilities in children (11–14 years of age) with and without dyslexia, it illustrates the approach. Rosen and Manganari explicitly asked if previously described increases in backward masking for children with dyslexia, compared with children without dyslexia, could explain the difficulty encountered by children with dyslexia in consonant recognition. First, Rosen and Manganari replicated the increased backward masking effect for children with dyslexia (compared with children without dyslexia) that Wright et al. (1997) had reported. Next, Rosen and Manganari reasoned that such a difference should only affect the abilities of children with dyslexia to label syllable-initial consonants, rather than syllable-final consonants, for words presented in isolation. Accordingly, they looked at recognition of consonants at the beginnings and ends of syllables. Results showed that children with dyslexia were poorer at recognition than their peers with normal reading abilities, but there was no difference depending on syllable place. Consequently, the reported differences in auditory sensitivities could not explain the differences in speech perception for children with and without dyslexia.

A study that tested the possibility that there is a connection between speech and nonspeech perception by children and adults was a study by Nittrouer and Crowther (1998). Specifically, that study looked at adults' and children's (5- and 7-year-olds') sensitivities to static and dynamic spectral properties. Following the descriptions of Kewley-Port, Pisoni, and Studdert-Kennedy (1983), the term *static* was used to refer to those properties that remain stable for several tens or hundreds of milliseconds, such as release bursts or fricative noises. The term *dynamic* referred to formant transitions, which are constantly changing in spectral composition. Two specific hypotheses were tested: (1) Weighting strategies across age groups could reflect absolute differences between children and adults in auditory sensitivities, or (2) weighting strategies across age groups could reflect differences in relative sensitivities for listeners in the two age groups. The kind of evidence that would have supported the first hypothesis would have been evidence that showed that children are less sensitive to changes in static spectral properties than adults, but more sensitive than adults to changes in dynamic spectral properties (i.e., frequency glides). That finding would have supported the absolute-differences hypothesis because children have been found to weight static spectral properties less and dynamic spectral properties more than adults (e.g., Nittrouer, 1992; Nittrouer et al., 2000; Parnell & Amerman, 1978; Siren & Wilcox, 1995). Supporting the second hypothesis would have been evidence that adults are more sensitive to static than they are to dynamic spectral properties, and that children are more sensitive to dynamic than they are to static spectral properties. How children and

adults compare to each other in sensitivities to these properties is irrelevant under the second hypothesis. Findings of Nittrouer and Crowther showed (a) that children were less sensitive than adults to both static and dynamic spectral properties, thus failing to support the first hypothesis, and (b) that children were less sensitive to dynamic properties than they were to static ones, thus failing to support the second hypothesis. Consequently, that study failed to demonstrate any kind of a relation between auditory sensitivities and speech perception across age groups. However, that study used a discrimination task, which is very different from the task of speech perception, and the perception of non-speech and speech stimuli was not compared in the same groups of listeners.

The purpose of the current study was to test the general hypothesis that age-related differences in the weighting of some acoustic properties in speech perception could be explained by age-related differences in auditory sensitivities. The specific hypothesis tested in this study was that previously reported differences between children and adults in the weighting of vocalic duration and formant-offset transitions for decisions about word-final stop voicing could predict differences in sensitivities between children and adults for duration of periodic signals and glides at signal offset. Children weight vocalic duration less than adults (e.g., Nittrouer, 2004), and so it would be predicted, under the hypothesis being tested, that children would be less sensitive than adults to variation in duration of periodic signals. At the same time, children weight syllable-final formant transitions more than adults. Therefore, under the hypothesis tested, it would be predicted that children would be more sensitive than adults to variation in frequency glides. This second prediction runs counter to both intuition and the collective results of studies showing that children generally perform more poorly than adults on tests of auditory perception. Of particular relevance, children need more extensive frequency glides to discriminate between two glides (e.g., Nittrouer & Crowther, 1998; Sussman & Carney, 1989). Nonetheless, if age-related differences in auditory sensitivities are to be used to explain age-related differences in the weighting of various acoustic properties, then this prediction (that children are more sensitive than adults to the acoustic properties that they weight more strongly) would need to be supported.¹

In the current study, perception of nonspeech and speech stimuli was compared for the same groups of listeners. Furthermore, efforts were made to keep the listening tasks as much alike as possible with speech and nonspeech stimuli. Stimuli were modeled after those of Nittrouer (2004), which investigated the labeling of CVCs differing in the voicing of word-final stops. That study used six word pairs with different vowels and places of stop closure. In the study reported here, the abilities of adults and children to discriminate speech (or speech-like) stimuli and comparable nonspeech stimuli varying in duration and in glides/transitions at stimulus offsets were examined. A discrimination task was used for both speech and nonspeech stimuli, using the method of constant stimuli. That is, all stimuli from along the continua were presented equal numbers of times in random order. Because all stimuli are presented equal numbers of times in this method, it is more similar to the labeling task used with speech stimuli than the adaptive tasks generally used to investigate auditory sensitivities. This method should also be easier for children because they are not listening to stimuli close to their discrimination thresholds for as much of the test session. Finally, parameters used to create speech and nonspeech stimuli for this experiment were kept as similar as possible. Thus, task and stimuli were both similar across speech and nonspeech conditions.

In summary, this study was undertaken to test the hypothesis that children do not weight vocalic duration in decisions of syllable-final stop voicing as greatly as adults do because children are less sensitive to durational properties of periodic signals. This hypothesis would be supported

¹The relative-sensitivity hypothesis proposed by Nittrouer and Crowther (1998) could not be tested in this study because there is no obvious way to compare sensitivity for temporal and spectral properties.

if children required greater duration differences in nonspeech stimuli than adults to judge two stimuli as different, and if those results were comparable to results for speech-like stimuli. Sensitivities for spectral properties mimicking formant transitions at syllable offsets for words with voiced and voiceless final stops were also measured. Strictly speaking, an auditory-based hypothesis for labeling results would predict that children would have greater sensitivity than adults to spectral properties because children weight formant-offset transitions more in decisions of final stop voicing. That is, if weighting strategies for speech signals are constrained by auditory sensitivities, then children should have keener sensitivities to these properties than adults.

Method

Participants

Adults and 5- and 7-year-old children participated in this experiment. Participants needed to be native speakers of American English and 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 to each ear separately. Children needed to be within -1 and $+5$ months of their birthdays: For example, all 5-year-olds were between 4 years, 11 months and 5 years, 5 months. Children needed to score at or above the 30th percentile on the Sounds-in-Words subtest of the Goldman-Fristoe Test of Articulation 2 (Goldman & 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 were between 18 and 31 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 & Wilkinson, 1984). Meeting these criteria were 25 adults (11 males, 14 females), 36 seven-year-olds (16 males, 20 females), and 45 five-year-olds (22 males, 23 females). As listener age decreased, the number of participants in the groups increased because it was expected that younger participants would be less likely to be able to do the task.

Equipment

Testing took place in a sound-treated booth with the computer that controlled the experiments 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. This system provided a good, distortion-free signal through 11.025 kHz, the Nyquist filtering rate. The experimenter recorded responses using a keyboard.

A cardboard response card 4 in. \times 14 in. with a line dividing it into two 7-in. halves was used with all participants. On one half of the card was two black squares, and on the other half was one black square and one red circle. Participants pointed to the two black squares when the two stimuli were heard as the same, and pointed to the square and the circle when they were heard as different. Participants also said “same” or “different” as part of their response. Ten other cardboard cards (4 in. \times 14 in., not divided in half) were used for training with children. On 6 cards were two simple drawings each of common objects (e.g., hat, flower, ball). On 3 of these cards the same object was drawn twice (identical in size and color) and on the other 3 cards two different objects were drawn (different in size and color). On 4 cards were two drawings each of simple geometric shapes: two with the same shape in the same color and two with different shapes in different colors. These cards were used to ensure that all children knew the concepts of same and different. A gameboard with 10 steps was also used with children. They moved a marker to the next number on the board after each block of stimuli (10 blocks in each condition). 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 so served as additional reinforcement for responding.

Stimuli

Five sets of stimuli were used: two sets of stimuli that varied in duration and three sets that varied in offset glides (or formants). For both kinds of stimuli (duration and spectral), one set was created to be speech-like and one set (in the case of the duration stimuli) or two sets (in the case of the spectral stimuli) were created to be nonspeech analogs of real speech stimuli. All stimuli were modeled to some extent after the speech stimuli used by Nittrouer (2004) in a study investigating the labeling of words with final voiced and voiceless stops. For all five sets of stimuli in this study mid-central vowels were created or replicated. Because these vowels have high F1s, they result in the most extensive F1 transitions going into a closure of any vowel. Nonspeech stimuli were composed of sine waves, which can be designed so that listeners are predisposed to hear them as speech. In this study, however, none of the sine waves were designed in such a way that they evoked a speech-like percept for any of the listeners.

Nonspeech duration stimuli.—Eleven stimuli were created varying in length from 110 ms to 310 ms in 20-ms steps. This range represents the range of durations of vocalic syllable portions found in monosyllabic words ending with voiced or voiceless stops. In particular, it covers the range of vocalic durations of all six words used in Nittrouer (2004). These stimuli consisted of three steady-state sinusoids of the frequencies 650 Hz, 1130 Hz, and 2600 Hz. These stimuli will be termed the *dur* stimuli in this report.

Speech duration stimuli.—Two natural words from Nittrouer (2004) were selected for use: one *cob* and one *cop*. These words are displayed in Figure 2. The first three formants fell just before voicing offset in both tokens, as would be expected given that the vocal tract moves to a bilabial closure at the end of the word. Both tokens had fundamental frequencies close to 133 Hz, and so each pitch period was approximately 7.5 ms. To create stimuli, final release bursts and any voicing during closure were removed. Pitch periods in the steady-state regions of the words were manipulated as necessary to make 13 stimuli from each token with the vocalic portions varying between 120 ms to 300 ms in 15-ms steps.² This range is similar to that of the nonspeech *dur* stimuli. To shorten stimuli, pitch periods were deleted. To lengthen stimuli, a single pitch period was reiterated. Pilot testing with adults and 6-year-olds showed that extremely similar results were obtained with stimuli created from *cob* and from *cop*, and so only stimuli created from *cob* were used in the experiment.

Nonspeech spectral stimuli, all tones falling.—These stimuli were all 150 ms long and consisted of the same steady-state sinusoids as the nonspeech *dur* stimuli, but the last 50 ms of these stimuli had falling tones. In this way they replicated the falling formants of words with mid-central vowels followed by bilabial stops. In particular, these glides replicated the formant patterns of the *cob* stimuli in Nittrouer (2004). The falling formant pattern of *cob* was replicated, rather than those of a word with either an alveolar or velar stop (i.e., those with rising second and/or third formants), because it was thought that having all three tones fall would create the most salient condition: that is, there would be overall shifts in the spectral centers of gravity. Thirteen stimuli were created. One of these had flat tones throughout the stimulus, and so was identical to the *dur* stimuli. The other stimuli all had tones that fell near offset in 12 equal-sized steps, with each step representing a change of 20 Hz in ending frequency. Consequently, the lowest tone varied between 650 Hz and 410 Hz at offset across stimuli. The stimulus with the

²It was not possible to have the step size for the speech stimuli be exactly the same as for the nonspeech *dur* stimuli because sections to be deleted or reiterated needed to be marked at the zero-crossings of pitch periods. This speaker had 7.5-ms pitch periods.

most change in each sinusoid is shown in the left panel of Figure 3. These stimuli will be referred to as the *Tall* stimuli.

Nonspeech spectral stimuli, T1 falling only.—Nonspeech stimuli were also created in which only the first tone (T1) fell over the last 50 ms. These stimuli were included in this experiment because many experiments investigating the perception of voicing for word-final stops use synthetic speech stimuli with the same second and third formant transitions across all stimuli; that is, only the first formant differs across stimuli. That was the case for the synthetic stimuli of Nittrouer (2004). There were 11 of these stimuli in this experiment. The frequency of the steady-state portion of T1 was 650 Hz, as in the *dur* stimuli, and T2 and T3 were the same as T2 and T3 in the *dur* stimuli. T1 at offset varied between 650 Hz (a flat tone) and 400 Hz in 25-Hz steps. The stimulus with the most change in T1 at offset is shown in the center panel of Figure 3. These stimuli will be referred to as the *Tl* stimuli. Step size was slightly larger for these stimuli than for the *Tall* stimuli in order to replicate the steps in the speech-like spectral stimuli. (Step size was actually abbreviated to 20 Hz in the *Tall* stimuli from what it was for the *Tl* and speech-like spectral stimuli because it was thought that listeners might place thresholds at much briefer transitions when all three tones fell.)

Speech-like spectral stimuli.—These stimuli were synthetic, formant stimuli derived from the synthetic *buck/bug* stimuli used by Nittrouer (2004). Using these stimuli meant that F2 was rising near voicing offset, rather than falling, as was the case for the second sinusoid of the *Tall* stimuli, but the decision to use them was nonetheless made, for several reasons. First, stimuli with the same synthesis parameters were used before, and so labeling results are available for comparison. In addition, any concern that not having all formants transition in the same direction would diminish listeners' abilities to discriminate the most different tokens was dismissed based on the results of Nittrouer (2004). In that study, four sets of natural stimuli were used, and of them, the stimuli with a velar place of closure (*pick/pig*) resulted in the fewest 5-year-olds failing to label the best exemplars reliably. Four out of 18 five-year-olds failed to label the best exemplars (i.e., the stimuli that were most different from each other) of *pick/pig* reliably, whereas 7 out of 17 five-year-olds failed to label the best exemplars of *cop/cob* reliably. Consequently, there was no reason to suspect that children would show more sensitive discrimination for speech stimuli with all three formants falling than for the pattern used here. In particular, it seemed that the stimuli to which young listeners would be most sensitive were selected for use in the cases of both the nonspeech and speech-like stimuli. This manipulation bolstered our confidence in our abilities to attribute any differences in discrimination between these two conditions to the speech-like nature of the stimuli.

All these speech-like stimuli were created using the Sensyn Laboratory Speech Synthesizer, and all were 160 ms long. The fundamental frequency began at 130 Hz and fell throughout to 100 Hz. F2 was constant at 1000 Hz until 50 ms before offset, at which time it rose linearly to 1400 Hz. F3 was constant at 2700 Hz until 50 ms before offset, at which time it fell to its ending frequency of 2400 Hz. F1 started at 400 Hz, rose linearly to 650 Hz over the first 50 ms, where it remained for the next 60 ms. Then F1 fell over the last 50 ms. Final F1 frequency varied in 25-Hz steps between 650 Hz and 400 Hz, just as T1 did in the *Tl* stimuli. Consequently, there were 11 of these speech-like stimuli. The one with the most change in F1 at voicing offset is shown in the right panel of Figure 3.

Procedure

Participants attended two sessions. In the first session the screening measures were administered, followed by two test conditions: one duration and one spectral. In the second session the remaining three test conditions were administered, with the duration condition

presented between the remaining two spectral conditions. Within these constraints, the order of presentation of the test conditions varied across participants.

An AX procedure was used. In this procedure, participants judge one stimulus, which varies across trials (X), to a constant standard (A). Therefore, although it is a discrimination task, as is needed for nonspeech stimuli, which have no natural labels, it is as close to a labeling task as can be had: The constant “A” stimulus serves as the category exemplar. For the duration stimuli, the “A” stimulus was the shortest member of the continuum. For the spectral stimuli, the “A” stimulus was the one without formant or glide transitions at offset. The onset-to-onset interval was 600 ms for all trials in all conditions. The participant responded by pointing to the picture of the two black squares and saying “same” if the stimuli were judged as being the same, and by pointing to the picture of the black square and the red circle and saying “different” if the stimuli were judged as being different. Both pointing and verbal responses are used because each serves as a check on the reliability of the other. Although infrequent in occurrence, when there was a discrepancy between the pointing and the verbal response the stimulus was repeated and the listener was asked to respond again. The left–right placements of the “same” and “different” sides of the response card were randomly varied across participants such that for half of the participants the “same” side was on the left and for half it was on the right. The experimenter controlled the presentation of stimuli and entered responses using the keyboard. During pilot testing, 20 blocks of stimuli were presented. However, no differences in response patterns were observed for the first 10 and the last 10 blocks, and so only 10 blocks of stimuli were presented during the experiment (i.e., each stimulus was presented 10 times each).

Several kinds of training were provided. Before any testing with the acoustic stimuli was done with children, they were shown the drawings of the six same and different objects and asked to report if the two objects on each card were the same or different. Then they were shown the cards with drawings of same and different geometric shapes and asked to report if the two shapes were the same or different. Finally, children were shown the card with the two squares on one side and a circle and a square on the other side and asked to point to “same” and to “different.” Adults were simply shown the card with the two squares on one side and a circle and a square on the other side and asked to point to “same” and to “different.”

Before testing with stimuli in each condition, all participants were presented with five pairs of stimuli that were identical and five pairs of stimuli that were maximally different, in random order. Participants were asked to report whether the stimuli were the same or different and were given feedback. Next, these same training stimuli were presented, and participants were asked to report if they were the same or different, only without feedback. Participants needed to respond correctly to 9 of the 10 training trials without feedback in order to proceed to testing. During testing participants needed to respond correctly to at least 16 of these physically same and maximally different stimuli (80%) to have their data included in the final analysis.

The discrimination functions of each participant formed cumulative normal distributions, and probit functions were fit to these distributions (Finney, 1971). From these fitted functions distribution means were calculated and termed *difference thresholds*. These thresholds were the 50% points on the discrimination functions and were used in statistical analyses to examine potential between-groups and within-groups differences.

Results

All children demonstrated that they understood the concepts of same and different by labeling all the pictures of common objects and of geometric patterns correctly. In addition, they were able to label the side of the response card with two squares as “same” and the side with a square and a circle as “different” on the first try.

Seventeen five-year-olds were unable to meet the training and testing criteria for having their data included in the study for any of the five conditions. Consequently data were included from only 28 five-year-olds (15 males). All of the seven-year-olds and adults were able to meet the training and testing criteria for at least one condition. Table 1 shows how many participants of each age group were able to participate in each condition, and shows that the *buck/bug* condition resulted in the lowest numbers of participants meeting criteria. Although the proportions of listeners who could not discriminate the most different stimuli were high, these proportions were comparable to those of earlier studies looking at the labeling of synthetic speech stimuli in which F1 is the only formant with an offset transition. In particular, Nittrouer (2004) reported that 14 out of 29 six-year-olds were unable to label the most *buck*-like and most *bug*-like synthetic stimuli consistently. Although older children and adults fared better in that study, it was still the case that 6 out of 29 adults were unable to label the most *buck*-like and most *bug*-like stimuli consistently. These numbers for adults are very similar to the proportion of adults who were unable to discriminate the most different *buck/bug* stimuli in the current study. Although it is tempting to blame this difficulty on the part of listeners to the synthetic nature of the stimuli, we believe that the problem stems from the fact that formants higher than F1 do not have transitions typical of natural *buck* and *bug*. Children may be more affected by this deficit in dynamic properties (i.e., formant transitions) for speech-like stimuli (i.e., those consisting of broad-band formants) precisely because they rely on dynamic signal properties more than adults. We are currently testing this hypothesis.

Duration Stimuli

Figure 4 shows discrimination functions for all three age groups for the two duration conditions, and shows that listeners in all three groups placed functions in roughly the same location for the *dur* stimuli. For the *cob* stimuli, children's labeling functions appear farther to the right than those of adults, indicating that children needed more of a difference in duration to judge stimuli as different.

Figure 5 shows discrimination functions for the two duration conditions for each age group separately. From this figure it appears that adults' mean *dur* and *cob* functions were in the same locations, but that children's mean *cob* function was farther to the right than the mean *dur* function. In other words, children required more of a duration difference to judge two stimuli as different when listening to speech rather than nonspeech. The first two rows of Table 2 show mean thresholds for each age group for these duration conditions, and indicates the same trends as observed in Figures 4 and 5: Mean thresholds are similar for all three age groups for the *dur* stimuli, adults' mean thresholds remain constant across the two duration conditions, and children's mean thresholds are at longer durations for the *cob* than for the *dur* stimuli.

One-way analyses of variance (ANOVAs) with age as the main effect and post hoc comparisons were done on thresholds for each condition separately. Results for the *dur* and the *cob* stimuli are shown at the top of Table 3. No age effect was found for thresholds for the *dur* stimuli, but a significant age effect was observed for the *cob* stimuli. The post hoc *t* tests show that both children's groups placed these *cob* thresholds in different locations than adults did, but that the two children's groups performed similarly to each other.

In addition to the ANOVAs performed to test for possible age effects, we wanted a way to index the magnitude of the difference in thresholds for the *dur* and *cob* conditions for each age group separately. We decided that a simple ratio of the difference between condition means (M) divided by the pooled standard deviation (SD) could be used: $M_1 - M_2 / [\sqrt{(SD_1 + SD_2) / 2}]$. This metric is often referred to as Cohen's d , which is commonly used to index the size of a treatment effect (Cohen, 1988). However, it is more broadly a way of providing a normalized (based on SD) index of the magnitude of the difference between any two means. The top line of Table 4 shows Cohen's d for the difference between *dur* and *cob* thresholds, for each age

group separately. These *ds* indicate that the magnitude of the difference decreases with increasing age. Although *p* values can not be computed for *d*, anything larger than .2 is generally considered a sizeable difference that would result in statistical significance if a *t* test were used.³

Spectral Stimuli

A striking result regarding these spectral stimuli is the large number of listeners who responded randomly to the physically identical and maximally different speech-like *buck/bug* stimuli. Almost all of the 5-year-olds and roughly half of the 7-year-olds showed this random pattern of responding. It is tempting to attribute the difficulty encountered by these children solely to the fact that only the first formant varied across tokens, but the same thing can be said of the *Tl* stimuli, and many more children were able to respond correctly to the endpoints in that condition. Clearly, then, the factors influencing response behavior differed for the speech and nonspeech stimuli. In this case, it seems that children were predisposed to confusion when all formants did not vary as they do in natural speech, but this confusion arose only when the stimuli were speech-like. It is likely that this confusion was due to listeners expecting all formants to be dynamic in nature. This expectation did not exist for the nonspeech tones.

Figure 6 shows results for the spectral conditions, and the bottom portion of Table 2 lists mean thresholds for each group for these conditions. From both this table and figure it appears that listeners in all three age groups had similar functions for the *Tall* stimuli. For the *Tl* stimuli, it appears that children in both age groups required a more extensive T1 offset transition than adults to judge the stimuli as different. Similarly, children appear to have required a more extensive F1 transition than adults did to judge the *buck/bug* stimuli as different.

Figure 7 shows functions for each age group separately. It appears that adults needed more extensive transitions for the *buck/bug* stimuli than for either of the nonspeech conditions to hear the stimuli as different. For both children's groups, there seems to be a progression: More extensive transitions were needed to hear a difference for *Tl* than for *Tall*, and even more extensive transitions still were needed (for the children who could do the task at all) to hear the *buck/bug* stimuli as different.

The lower section of Table 3 shows results for the oneway ANOVAs done on data from each of the spectral conditions. No significant effect of age was found for the *Tall* stimuli, indicating that listeners of all ages placed thresholds in the same location for these stimuli. For the *Tl* stimuli, a significant age effect was found. The difference in results for the two conditions appears traceable to the fact that children placed their thresholds at more extensive transitions for the *Tl* stimuli than they had for the *Tall* stimuli, whereas adults' thresholds for the two conditions are similar. Of the two children's groups, 7-year-olds showed a slightly greater mean difference in placement of thresholds for the two conditions, leading to a significant post hoc comparison between adults and 7-year-olds for the *Tl* condition. Although 5-year-olds' thresholds were at more extensive transitions for *Tl* than for *Tall*, the adults versus 5-year-olds' post hoc comparison for *Tl* did not reach statistical significance. The post hoc comparison of placement for *Tl* thresholds for 7- and 5-year-olds did not reach statistical significance, and so it must be concluded that both groups of children placed these thresholds in roughly the same location.

³One way to demonstrate the relation between values for Cohen's *d* and statistical significance is to look at results of matched *t* tests for two comparisons for adults: *dur* versus *cob* and *Tl* versus *buck/bug*. Results for adults are used because the largest numbers of participants in the adults' group were able to meet criteria for participation in each condition. For the *dur* versus *cob* comparison, Cohen's *d* was .13. A matched *t* test done for these conditions resulted in a $t(24) = 0.93$, which was clearly not significant. For the *Tl* versus *buck/bug* comparison, Cohen's *d* was .33. A matched *t* test done for these conditions resulted in a $t(20) = 2.99$, which has an associated *p* value of .007. These results demonstrate that a Cohen's *d* roughly halfway between .13 and .33 would be the boundary for statistical significance.

The middle row of Table 4 shows Cohen's *ds* for *Tall* versus *Tl* means for each age group separately. Clearly no difference existed between adults' thresholds for *Tall* and *Tl*, but differences did exist between thresholds for these conditions for both groups of children. Children required more of a difference between stimuli to recognize them as different when only the first tone, rather than all three, transitioned at offset.

For the *buck/bug* condition, no ANOVA results are reported because only four 5-year-olds were able to achieve the criteria needed to have their data included in the analysis. Instead, a *t* test comparing thresholds for adults and 7-year-olds was done. This result is shown at the bottom of Table 3 and indicates that there was a significant difference in thresholds between the two groups for the *buck/bug* condition.

The critical condition against which *buck/bug* should be compared is *Tl* because these two conditions involve differences in the extent of transition for the first resonance only. The last row of Table 4 displays *ds* for this comparison. All three groups of listeners showed differences between the two conditions, with the magnitude of the effect increasing with decreasing age. Although many children were unable to perform the discrimination task with the speech stimuli in this experiment, it was important to include results for these stimuli because the comparison of children's results for the nonspeech and speech spectral stimuli so clearly illustrates the independence of responding for the two types of stimuli.⁴ This independence is apparent both for children who discriminated the *buck/bug* sufficiently to have their data included in this analysis and for those who did not.

Discussion

The purpose of the current study was to test the hypothesis that previously reported differences between children and adults in the labeling of speech stimuli stem from age-related differences in auditory sensitivities to the properties manipulated in those experiments. Specifically, the primary hypothesis being explored was that children are less sensitive than adults to changes in duration, and so do not base decisions about the voicing of syllable-final stop consonants on that property as strongly as adults do. At the same time, the possibility was tested that children are more sensitive to frequency glides than adults are, thus explaining children's enhanced weighting of those properties (compared with adults) in phonetic decisions, including those of syllable-final stop voicing. The major outcome of this study was that adults and children performed similarly for nonspeech stimuli that closely replicated the acoustic attributes of speech stimuli being examined in this study: *dur* and *Tall*. The *dur* stimuli replicated the vowel-length differences of words ending in voiced or voiceless stops, and the *Tall* stimuli replicated the spectral change across all formants that is typical of natural speech. Therefore, the conclusion is reached that any genuine age-related differences in sensitivities to temporal properties or spectral glides that exist cannot explain differences in the perception of words ending in voiced and voiceless final stops.

The finding that children and adults showed similar sensitivities to acoustic properties in nonspeech stimuli in this study contradicts other studies in which age-related differences have been observed. Possible reasons for this difference across studies include the facts that the task here may have been easier for children than tasks used in earlier studies (because children were not listening to as many stimuli near threshold) and that measured thresholds, particularly for

⁴Under some circumstances it can be inappropriate to include results when a majority of participants are unable to do the task, because it is impossible to know if that failure is due to task or stimulus variables. In this case, however, it is clear that the children unable to make discriminations among the *buck/bug* stimuli were largely able to do the task: They demonstrated those abilities with other sets of stimuli. Consequently we may attribute their failure to discriminate these particular stimuli precisely to a failure to hear a difference among them. For that reason, it is also supportable to conclude that these children performed differently from the majority of adults for the *buck/bug* stimuli and differently from how they performed themselves for the nonspeech spectral stimuli.

adults, were likely not as close to their absolute thresholds as in earlier studies (because the method used in this study was not particularly sensitive). However, the method used in this study was closer to the method of labeling generally used in studies with speech stimuli. For that reason it is appropriate to conclude that previously observed age-related differences in labeling of syllable-final stop voicing cannot be explained by auditory sensitivities; children's sensitivities are at least adequate to support the same labeling responses as those observed for adults in those experiments. In general then, the proposal that it is the weighting of various acoustic properties in the speech signal that differs for adults and children remains viable.

This study also served to emphasize other points worth remembering. In particular, perception is influenced by factors other than how sensitive listeners are to sensory inputs. Of course, the procedures used in the current study were not sensitive themselves; that is, we were presumably not testing near the limits of listeners' true sensitivities. But even when extremely sensitive procedures are used, the influences of nonsensory factors can be seen. For example, Nelson, Nittrouer, and Norton (1995) measured thresholds for gaps in composite signals of noise (simulating frication) followed by three sine waves (simulating the first three formants of the voiced portion of "stay") presented to adults with normal and impaired hearing. In that study an adaptive three-interval forced-choice procedure was used to obtain thresholds for those nonspeech stimuli. The same listeners were asked to label comparable speech stimuli as *say* or *stay*. In spite of these procedures, listeners with sloping hearing losses placed *say/stay* phoneme boundaries at much shorter gap durations than their discrimination of nonspeech signals indicated they could hear. These listeners had mean gap thresholds of 30.9 ms and *say/stay* phoneme boundaries of 18.5 ms for stimuli with formant transitions appropriate for "stay." Thus the use of nonspeech stimuli with very "sensitive" procedures does not ensure that the absolute limits of the auditory system are being measured; non-sensory factors still influence perception.

There was one possibly confounding difference between the nonspeech and speech stimuli in this experiment: the nonspeech stimuli were made of sine waves, whereas the speech stimuli all consisted of formants (either natural or synthetic). Remez, Pardo, Piorkowski, and Rubin (2001) showed that adults are able to segregate individual components of tonal complexes when listening to sine wave stimuli (regardless of whether they are heard as speech-like or not) so that auditory judgments can be made about individual components, but are unable to do so when listening to formant stimuli. The results of this study are consistent with that explanation. Adults had similar thresholds for both sets of nonspeech spectral stimuli (*Tall* and *TI*), even though only the lowest tone had a transition at stimulus offset in *TI* whereas all three tones had transitions in *Tall*. These same adults required more extensive transitions to hear signals as different when listening to formant stimuli. Thus sensitivity to spectral stimuli used in this study was consistent as long as adult listeners could segregate at least one component of the stimulus complex, which they apparently did when listening to the sine waves. A greater difference between members of the pair was required for adults to hear formant stimuli as different from each other. This finding may reflect the inability to segregate a single component from the complex in formant stimuli, and so might indicate that the nature of the stimuli is a factor affecting performance—at least for adults.

On the other hand, this decreased sensitivity for offset transitions in the speech-like stimuli could reflect the limited role that offset transitions play in adults' voicing decisions for syllable-final stops. As Figure 1 shows, adult listeners do not attend to F1-offset transition much at all when labeling synthetic *buck/bug* stimuli. Thus, the non-sensory factor of whether or not stimuli are heard as speech may have influenced responding for these adults. Similar results have been reported by others. For example, a study by Jaramillo et al. (2001) with native Finnish speakers showed that duration changes in vowels elicited larger mismatched negativity potentials than did comparable duration changes in nonspeech stimuli. Because vowel length

is phonemically relevant in Finnish, Jaramillo et al. concluded that their Finnish participants responded differently to changes in duration, depending on whether or not the stimuli were speech. Results for adults in this study support that conclusion and highlight the inherent flaw in attempting to use results from experiments with nonspeech stimuli to explain observations made with speech stimuli. Other studies have also shown differences in perception results for physically identical stimuli, depending on whether listeners treat those stimuli as speech or nonspeech. For example, Best, Studdert-Kennedy, Manuel, and Rubin-Spitz (1989) presented adult, native-English listeners with three-tone sine wave patterns in which the third tone varied in onset frequency as would be appropriate for a /la/-/ra/ continuum. Listeners were divided into two groups and biased to hear the tonal complexes as either speech or music. The two groups categorized the sine wave stimuli differently, even though all listeners labeled stimuli from a synthetic, formant /la/-/ra/ continuum similarly. Best et al. concluded that listeners recognize acoustic patterns that arise from articulatory gestures of the vocal tract, and so respond to speech-like stimuli differently from how they respond to nonspeech stimuli. Results of the current study are consistent with that conclusion.

This study is also notable for what it tells us about age-related differences in perception, both to speech and nonspeech stimuli. For children listening to the spectral stimuli, thresholds were at more extensive transitions for *Tl* than for *Tall*. Perhaps this indicates that children are not as skilled as adults at segregating individual components of tonal complexes, even when sine wave stimuli are used. That is, the overall spectra remained more similar for *Tl* than for *Tall* stimuli, and children may be obligated to respond based on those overall spectra.

One particularly striking result from the spectral conditions is the large number of children who were unable to meet training and testing criteria for the *buck/bug* stimuli. This finding seems inconsistent with results such as those in Figure 1 showing that children weight formant transitions more strongly than adults do. However, in that study (Nittrouer, 2004), 48% of the 6-year-olds and 43% of the 8-year-olds participating were unable to meet the criteria for participation with the synthetic stimuli that had transitions for F1 only. In other words, they were unable to label the best exemplars of the *buck* and *bug* categories reliably. It is tempting to attribute this high failure rate to the fact that the stimuli were synthetic; perhaps young children have difficulty with synthetic stimuli. Currently we are examining that possibility by testing children of these same ages with synthetic *buck/bug* stimuli in which the first three formants all have similar offsets to those found in natural speech. So far, only 3 six-year-olds out of 23 and 1 eight-year-old out of 24 have failed to meet test criteria. This failure rate is similar to those found for stimuli created by editing natural speech samples (see Nittrouer, 2004). Consequently a more plausible explanation for the failure rates observed for stimuli with transitions for F1 only may be that children have difficulty discriminating or labeling speech-like stimuli when only one spectral component has a transition. There is now ample evidence that children rely on formant transitions (i.e., the dynamic components of the signal) for making phonetic decisions (see Nittrouer, 2005, for a lengthy discussion). Speech-like stimuli that lack these dynamic attributes may be hard for children to deal with.

In summary, this study was undertaken to test the specific hypothesis that reported age-related differences in the weighting of some acoustic properties of speech could be explained by age-related differences in sensitivities to those properties. However, no evidence was found to support that hypothesis. Through use of a discrimination task that was as similar as could be to that of the labeling task used in speech perception studies, it was found that children's thresholds were similar to those of adults for nonspeech acoustic properties emulating speech properties. Consequently, the theoretical position that the way children weight the various acoustic properties of speech signals in their native language changes as they gain experience with those signals remains supportable. Furthermore, it was observed that sensitivity to sensory inputs is difficult to compare across stimuli differing in their fundamental nature. It seems that

listeners bring expectations about the nature of certain stimuli to the listening task, whatever it may be. These expectations likely arise from the simple truth that listeners generally listen to ecologically valid stimuli. We learn about the general properties of sounds created by the objects and events in our environments, and we come to expect those properties from those objects and events. This observation could influence how we view experiments seeking to examine potential age-related differences in sensitivities to acoustic properties overall. Customarily, these experiments involve nonspeech signals and highly sensitive procedures meant to help focus listeners' attention on the acoustic property being examined. Perhaps adults and children differ in their abilities to discover structure in these signals that lack ecological validity.

Acknowledgment

This work was supported by Grant R01 DC00633 from the National Institute on Deafness and Other Communication Disorders.

References

- Allen P, Wightman F, Kistler D, Dolan T. Frequency resolution in children. *Journal of Speech and Hearing Research* 1989;32:317–322. [PubMed: 2739383]
- Best CT, Studdert-Kennedy M, Manuel S, Rubin-Spitz J. Discovering phonetic coherence in acoustic patterns. *Perception & Psychophysics* 1989;45:237–250. [PubMed: 2710622]
- Cohen, J. *Statistical power analysis for the behavioral sciences*. 2nd ed.. Erlbaum; Hillsdale, NJ: 1988.
- Crowther CS, Mann V. Native language factors affecting use of vocalic cues to final consonant voicing in English. *Journal of the Acoustical Society of America* 1992;92:711–722. [PubMed: 1506526]
- Crowther CS, Mann V. Use of vocalic cues to consonant voicing and native language background: The influence of experimental design. *Perception & Psychophysics* 1994;55:513–525. [PubMed: 8008553]
- Finney, DJ. *Probit analysis*. Cambridge University Press; Cambridge, United Kingdom: 1971.
- Flege JE. Chinese subjects' perception of the word-final English /t-/d/ contrast: Performance before and after training. *Journal of the Acoustical Society of America* 1989;86:1684–1697. [PubMed: 2808918]
- Goldman, R.; Fristoe, M. *Goldman-Fristoe Test of Articulation*. 2nd ed.. American Guidance Service; Circle Pines, MN: 2000.
- Gordon PC, Keyes L, Yung YF. Ability in perceiving nonnative contrasts: Performance on natural and synthetic speech stimuli. *Perception & Psychophysics* 2001;63:746–758. [PubMed: 11436743]
- Gracco VL. Some organizational characteristics of speech movement control. *Journal of Speech and Hearing Research* 1994;37:4–27. [PubMed: 8170129]
- Greenlee M. Learning the phonetic cues to the voiced-voiceless distinction: A comparison of child and adult speech perception. *Journal of Child Language* 1980;7:459–468. [PubMed: 7440672]
- Jaramillo M, Ilvonen T, Kujala T, Alku P, Tervaniemi M, Alho K. Are different kinds of acoustic features processed differently for speech and non-speech sounds? *Cognitive Brain Research* 2001;12:459–466. [PubMed: 11689306]
- Jastak, S.; Wilkinson, GS. *The Wide Range Achievement Test–Revised*. Jastak Associates; Wilmington, DE: 1984.
- Jensen JK, Neff DL. Development of basic auditory discrimination in preschool children. *Psychological Science* 1993;4:104–107.
- Kewley-Port D, Pisoni DB, Studdert-Kennedy M. Perception of static and dynamic acoustic cues to place of articulation in initial stop consonants. *Journal of the Acoustical Society of America* 1983;73:1779–1793. [PubMed: 6223060]
- Krause SE. Vowel duration as a perceptual cue to postvocalic consonant voicing in young children and adults. *Journal of the Acoustical Society of America* 1982;71:990–995. [PubMed: 7085987]
- Lehman ME, Sharf DJ. Perception /production relationships in the development of the vowel duration cue to final consonant voicing. *Journal of Speech and Hearing Research* 1989;32:803–815. [PubMed: 2601311]

- MacKain KS, Best CT, Strange W. Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics* 1981;2:369–390.
- Mayo C, Scobbie JM, Hewlett N, Waters D. The influence of phonemic awareness development on acoustic cue weighting strategies in children's speech perception. *Journal of Speech, Language, and Hearing Research* 2003;46:1184–1196.
- Nelson PB, Nittrouer S, Norton SJ. “Say-stay” identification and psychoacoustic performance of hearing-impaired listeners. *Journal of the Acoustical Society of America* 1995;97:1830–1838. [PubMed: 7699164]
- Nittrouer S. Age-related differences in perceptual effects of formant transitions within syllables and across syllable boundaries. *Journal of Phonetics* 1992;20:351–382.
- Nittrouer S. The role of temporal and dynamic signal components in the perception of syllable-final stop voicing by children and adults. *Journal of the Acoustical Society of America* 2004;115:1777–1790. [PubMed: 15101656]
- Nittrouer S. Age-related differences in weighting and masking of two cues to word-final stop voicing in noise. *Journal of the Acoustical Society of America* 2005;118:1072–1088. [PubMed: 16158662]
- Nittrouer S, Crowther CS. Examining the role of auditory sensitivity in the developmental weighting shift. *Journal of Speech, Language, and Hearing Research* 1998;41:809–818.
- Nittrouer S, Miller ME, Crowther CS, Manhart MJ. The effect of segmental order on fricative labeling by children and adults. *Perception & Psychophysics* 2000;62:266–284. [PubMed: 10723207]
- Parnell MM, Amerman JD. Maturation influences on perception of coarticulatory effects. *Journal of Speech and Hearing Research* 1978;21:682–701. [PubMed: 745369]
- Raphael LJ. The physiological control of durational differences between vowels preceding voiced and voiceless consonants in English. *Journal of Phonetics* 1975;3:25–33.
- Remez RE, Pardo JS, Piorkowski RL, Rubin PE. On the bistability of sine wave analogues of speech. *Psychological Science* 2001;12:24–29. [PubMed: 11294224]
- Rosen S, Manganari E. Is there a relationship between speech and nonspeech auditory processing in children with dyslexia? *Journal of Speech, Language, and Hearing Research* 2001;44:720–736.
- Siren KA, Wilcox KA. Effects of lexical meaning and practiced productions on coarticulation in children's and adults' speech. *Journal of Speech and Hearing Research* 1995;38:351–359. [PubMed: 7596100]
- Summers WV. Effects of stress and final-consonant voicing on vowel production: Articulatory and acoustic analyses. *Journal of the Acoustical Society of America* 1987;82:847–863. [PubMed: 3655120]
- Sussman JE, Carney AE. Effects of transition length on the perception of stop consonants by children and adults. *Journal of Speech and Hearing Research* 1989;32:151–160. [PubMed: 2704190]
- Wardrip-Fruin C, Peach S. Developmental aspects of the perception of acoustic cues in determining the voicing feature of final stop consonants. *Language and Speech* 1984;27:367–379. [PubMed: 6536845]
- Wightman, F.; Allen, P. Individual differences in auditory capability among preschool children. In: Werner, LA.; Rubel, EW., editors. *Developmental psychoacoustics*. American Psychological Association; Washington, DC: 1992. p. 113-157.
- Wightman F, Allen P, Dolan T, Kistler D, Jamieson D. Temporal resolution in children. *Child Development* 1989;60:611–624. [PubMed: 2737011]
- Wright BA, Lombardino LJ, King WM, Puranik CS, Leonard CM, Merzenich MM. Deficits in auditory temporal and spectral resolution in language-impaired children. *Nature* 1997;387:176–178. [PubMed: 9144287]

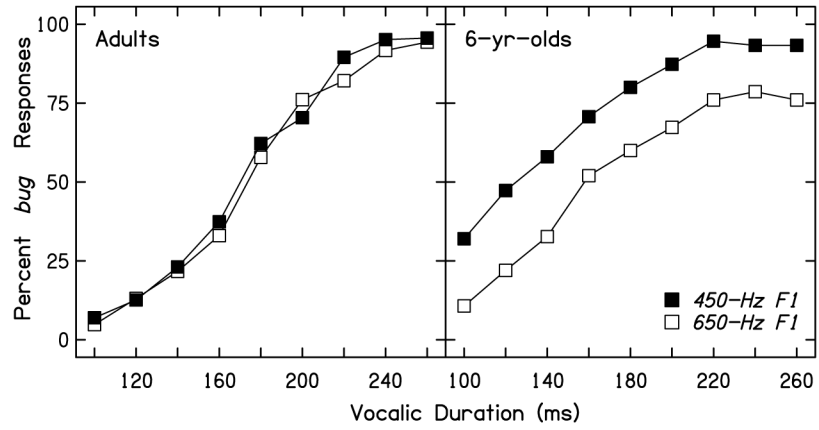


Figure 1. Results of the labeling experiment for synthetic *buck/bug* stimuli (from Nittrouer, 2004).

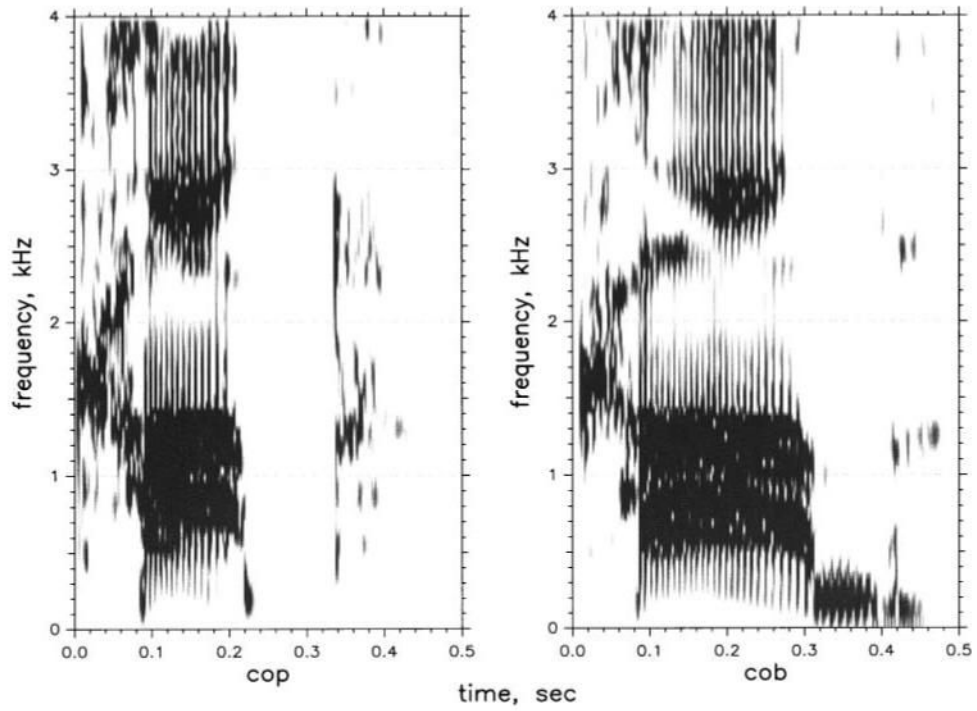


Figure 2.
Cop and *cob* stimuli used to create speech stimuli with varying durations used in the pilot test.

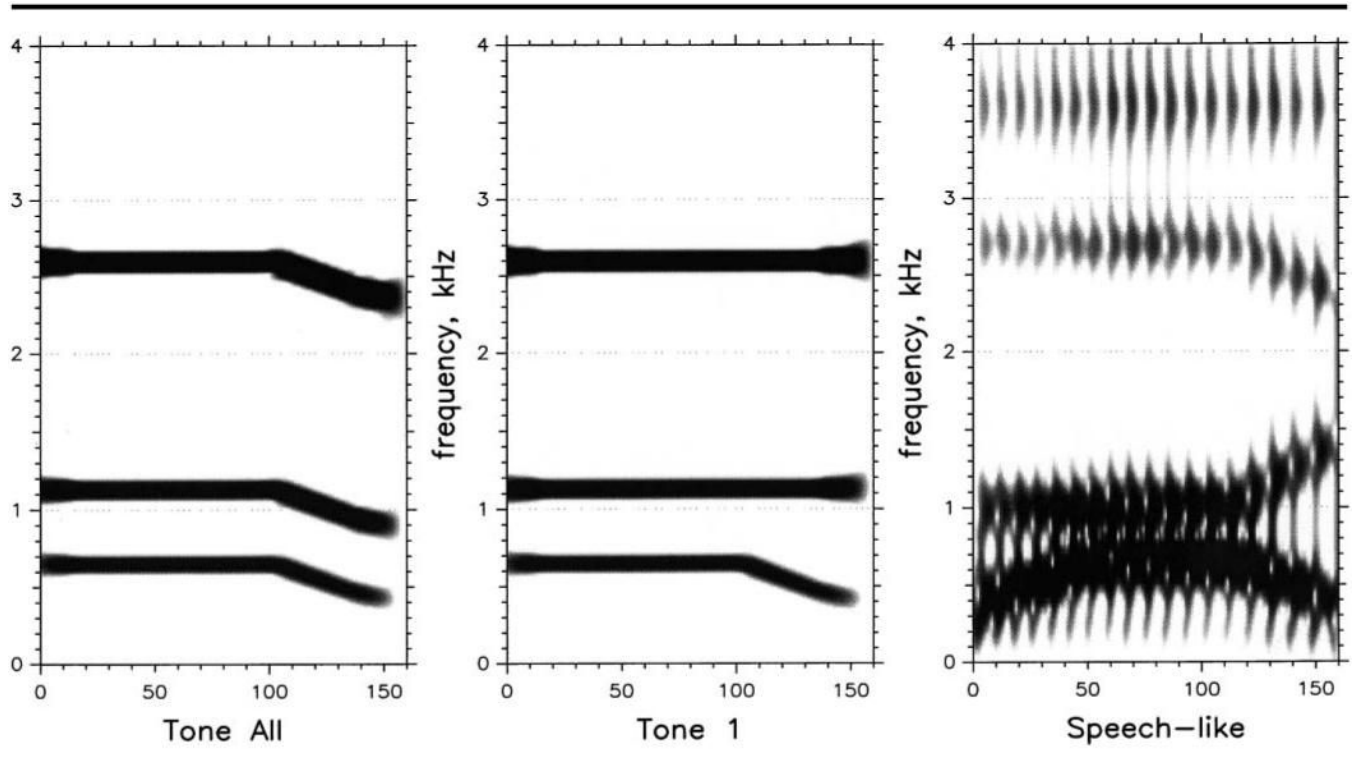


Figure 3. Stimuli used in the three spectral conditions. In each case the stimulus with the most extreme transitions is shown.

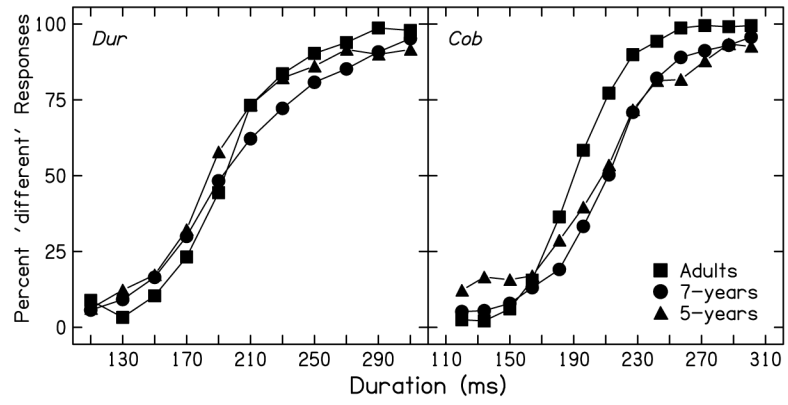


Figure 4.
Discrimination functions for the duration conditions, plotted by condition.

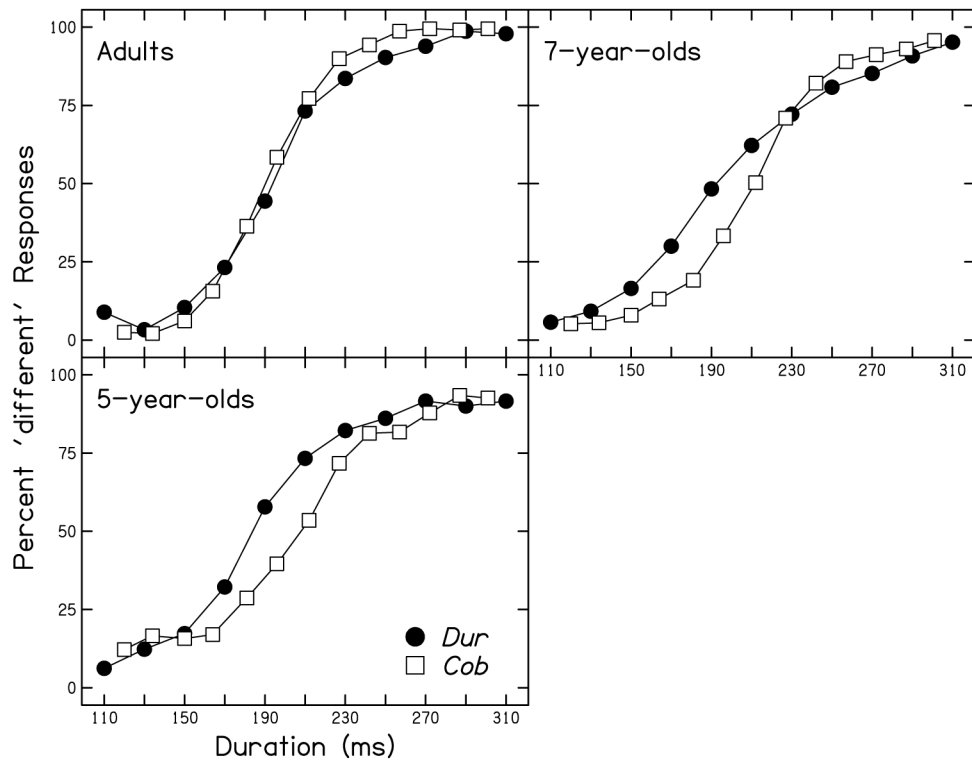


Figure 5. Discrimination functions for the duration conditions, plotted by listener age.

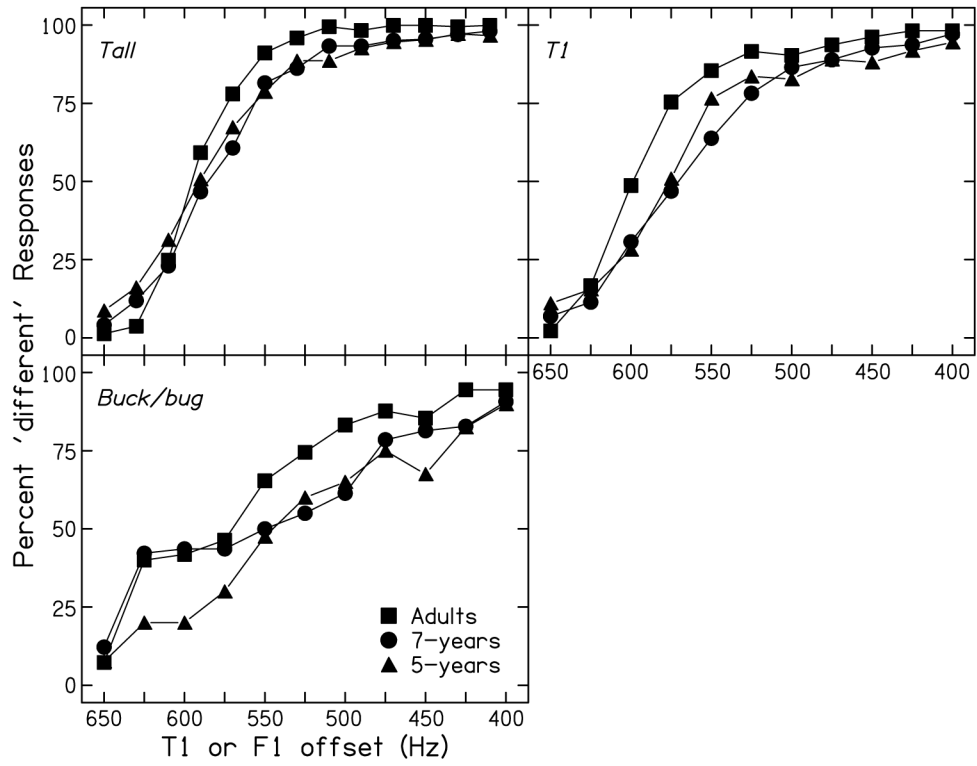


Figure 6. Discrimination functions for the spectral conditions, plotted by condition.

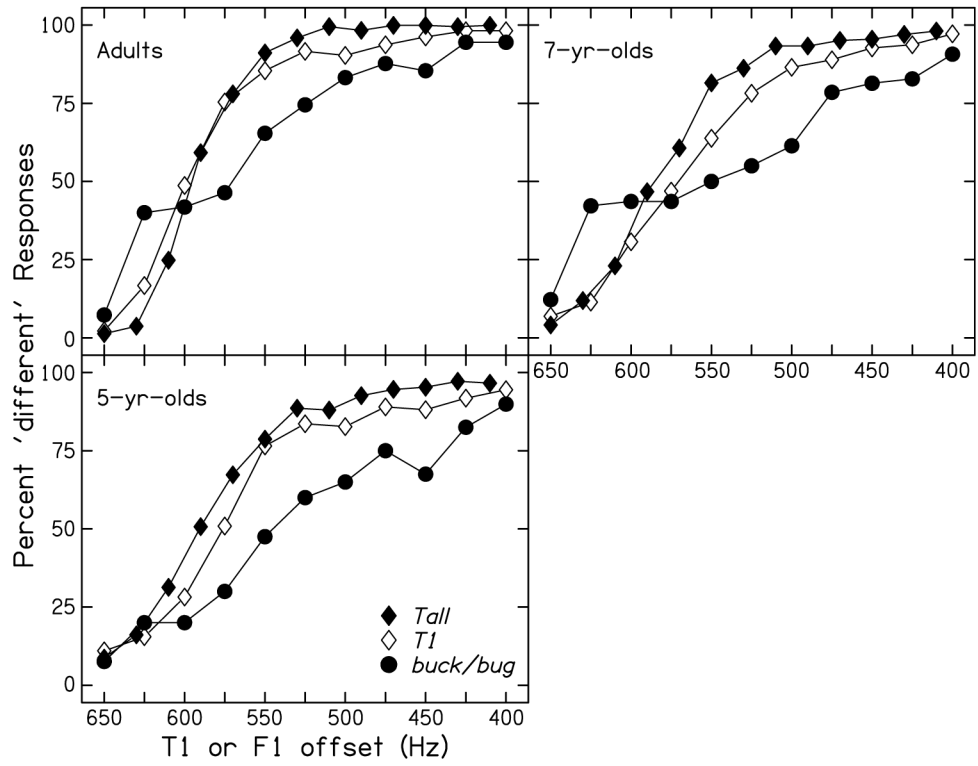


Figure 7. Discrimination functions for the spectral conditions, plotted by listener age.

Table 1

Number of participants in each age group who were able to pass training and testing criteria in each condition. These numbers are out of 25 adults, 36 seven-year-olds, and 28 five-year-olds.

	Adults	7-year-olds	5-year-olds	Total
<i>Dur</i>				
Training	25	31	21	
Testing	25	23	18	66
<i>Cob</i>				
Training	25	33	26	
Testing	25	33	23	81
<i>Tall</i>				
Training	25	34	21	
Testing	25	27	15	67
<i>Tl</i>				
Training	24	31	17	
Testing	24	29	11	64
<i>Buck/bug</i>				
Training	22	20	6	
Testing	22	14	4	40

Table 2

Mean thresholds for each age group, with standard deviations in parentheses.

	Adults	7-year-olds	5-year-olds
<i>Dur</i>	194 (25)	199 (28)	190 (30)
<i>Cob</i>	191 (20)	209 (17)	203 (17)
<i>Tall</i>	589 (16)	580 (27)	584 (28)
<i>Tl</i>	588 (31)	563 (39)	572 (28)
<i>Buck/bug</i>	575 (46)	540 (54)	522 (77)

Note. Thresholds for *dur* and *cob* stimuli are given in milliseconds; those for *Tall*, *Tl*, and *buck/bug* are given in Hertz.

Table 3
Results of ANOVAs and post hoc *t* tests done on thresholds in each condition.

	<i>df</i>	<i>F</i> or <i>t</i>	<i>p</i>	<i>Bonferroni</i> <i>significance</i>
<i>Dur</i>				
Age effect	2, 63	0.47	<i>ns</i>	
<i>Cob</i>				
Age effect	2, 78	7.36	.001	
Adults vs. 7-year-olds	78	-3.81	<.001	<.001
Adults vs. 5-year-olds	78	-2.35	.021	.10
7- vs. 5-year-olds	78	1.22	<i>ns</i>	
<i>Tall</i>				
Age effect	2, 64	1.02	<i>ns</i>	
<i>Tl</i>				
Age effect	2, 61	3.55	.035	
Adults vs. 7-year-olds	61	2.65	.010	.05
Adults vs. 5-year-olds	61	1.32	<i>ns</i>	
7- vs. 5-year-olds	61	-0.71	<i>ns</i>	
<i>Buck/bug</i>				
Adults vs. 7-year-olds ^a	34	-2.10	.043	

^aThis comparison was done as a simple *t* test rather than as a post hoc comparison.

Table 4

Cohen's d for each comparison, for each age group.

	Adults	7-year-olds	5-year-olds
<i>Dur vs. Cob</i>	.13	.43	.53
<i>Tall vs. Tl</i>	.04	.51	.43
<i>Tl vs. Buck/bug</i>	.33	.49	.86