

## Research Article

# All Cues Are Not Created Equal: The Case for Facilitating the Acquisition of Typical Weighting Strategies in Children With Hearing Loss

Joanna H. Lowenstein<sup>a</sup> and Susan Nittrouer<sup>a</sup>

**Purpose:** One task of childhood involves learning to optimally weight acoustic cues in the speech signal in order to recover phonemic categories. This study examined the extent to which spectral degradation, as associated with cochlear implants, might interfere. The 3 goals were to measure, for adults and children, (a) cue weighting with spectrally degraded signals, (b) sensitivity to degraded cues, and (c) word recognition for degraded signals.

**Method:** Twenty-three adults and 36 children (10 and 8 years old) labeled spectrally degraded stimuli from /ba/-to-/wa/ continua varying in formant and amplitude rise time (FRT and ART). They also discriminated degraded stimuli from FRT and ART continua, and recognized words.

**Results:** A developmental increase in the weight assigned to FRT in labeling was clearly observed, with a slight decrease in weight assigned to ART. Sensitivity to these degraded cues measured by the discrimination task could not explain variability in cue weighting. FRT cue weighting explained significant variability in word recognition; ART cue weighting did not.

**Conclusion:** Spectral degradation affects children more than adults, but that degradation cannot explain the greater diminishment in children's weighting of FRT. It is suggested that auditory training could strengthen the weighting of spectral cues for implant recipients.

Speech perception is a type of skilled behavior. Listeners must not only be accurate in their abilities to recover linguistic structure from an acoustic speech signal, they must also be able to perform this action (a) without expending significant cognitive resources and (b) in the presence of interfering sounds. The first of these requirements—that speech perception proceed without great cognitive effort—frees listeners to devote cognitive resources to other mental operations, such as assessing the environment or solving problems. The second requirement—that speech perception function efficiently in the face of interfering sounds—arises due to the fact that the listening landscape typically consists of sound sources other than the target speech signal. These other sources can involve multiple talkers as well as nonspeech signals.

Because of the need for efficient as well as accurate speech perception, language users typically develop highly

specific listening strategies for at least one language, the one they learn first. In particular, language users attend strongly to components of the acoustic speech signal that are linguistically informative for their first language, and largely ignore other kinds of signal structure. In fact, upon learning a second language, most individuals continue to apply the listening strategies acquired as part of learning their first language, and fail to ever develop strong perceptual attention to the acoustic components most important in the second language. The common paradigm for examining this phenomenon is categorical perception, often involving multiple cues. In this paradigm, syllable-sized stimuli are created by varying the acoustic settings of linguistically relevant cues. For example, Crowther and Mann (1992) examined the extent to which native talkers of Japanese and Mandarin Chinese who had learned English as a second language were able to use two cues to the categorization of syllable-final voicing: vowel length and first-formant frequency at voicing offset. Neither Japanese nor Mandarin Chinese allows stop consonants at the ends of syllables; however, Japanese does have a vowel-length distinction. Consequently, talkers of neither language would have had opportunity to learn about the cues to syllable-final stop voicing, but

<sup>a</sup>The Ohio State University, Columbus

Correspondence to Joanna H. Lowenstein: lowenstein.6@osu.edu

Editor: Nancy Tye-Murray

Associate Editor: Richard Dowell

Received September 10, 2014

Revision received November 12, 2014

Accepted November 21, 2014

DOI: 10.1044/2015\_JSLHR-H-14-0254

**Disclosure:** The authors have declared that no competing interests existed at the time of publication.

Japanese talkers would have had reason to acquire sensitivity to vowel-length distinctions. In the Crowther and Mann study, stimuli spanning a *pot*-to-*pod* continuum were generated. Nine values were used to represent vowel duration, and three values were used to represent first-formant offset frequency. Outcomes showed that native Japanese and Mandarin talkers alike weighted both cues less than native English talkers, but that the Japanese talkers were slightly closer to the English talkers in their patterns of outcome. Thus it was concluded that there is a strong effect of the first language on how perceptual attention, or weight, is distributed in speech perception. Similar outcomes have been observed by others, for different phonemic contrasts, across a variety of first and second languages (e.g., Beddor & Strange, 1982; Cho & Ladefoged, 1999; Gottfried & Beddor, 1988).

These findings have led to suggestions that listeners develop selective attention for the language-specific and phonologically relevant cues of their first language. That suggestion, in turn, means that these strategies must be learned during childhood, and evidence for that proposal is strong. Starting shortly after birth, infants' attention to acoustic structure in the speech signal begins to shift from being language general to being specific to the ambient language (e.g., Aslin & Pisoni, 1980; Eimas, Siqueland, Jusczyk, & Vigorito, 1971; Kuhl, 1987; Maurer & Werker, 2014; Werker & Tees, 1984). An example of this phenomenon comes from Hoonhorst et al. (2009), who showed that 4-month-old infants in a French-speaking environment discriminated nonnative voice-onset-time boundaries better than boundaries typically found in French. By 8 months of age, infants were showing sensitivity to language-specific (French) boundaries. Shifts in perceptual strategies for speech continue to be observed through roughly the first decade of life, during which time children learn how to adjust the relative amount of attention given to the various cues to phonemic decisions (Bourland Hicks & Ohde, 2005; Greenlee, 1980; Mayo, Scobbie, Hewlett, & Waters, 2003; Nittrouer & Studdert-Kennedy, 1987; Ohde & German, 2011; Wardrip-Fruin & Peach, 1984). For example, 3-year-old children base their decisions concerning place of constriction for syllable-initial fricatives (*/s/* vs. */ʃ/*) more strongly than adults on consonant-to-vowel formant transitions, and less on the spectral shape of those fricative noises. Over the next few years, however, that attentional strategy shifts to match that of adults, who attend more to the shape of the fricative noise (for a review of these findings, see Nittrouer, 2002).

But in spite of this robust evidence for developmental shifts in cue weighting, age-related differences are not observed for all phonemic contrasts. There are some phonemic decisions for which mature language users rely strongly on formant transitions, and in those cases, children show similar patterns to adults. When it comes to syllable-initial fricatives, for example, adults weight formant transitions greatly in their place decisions for weak fricatives (*/f/* vs. */θ/*), as demonstrated first by Harris (1958). In that case, adults and children show similar weighting strategies (Nittrouer, 2002). Another example of similar weighting strategies involves

the labeling of syllable-medial vowels. In this case, adults typically weight the formant transitions on either side of the vowel target more than the brief, static spectral section termed the *steady-state region* (e.g., Jenkins, Strange, & Edman, 1983; Strange, Jenkins, & Johnson, 1983). Consequently, when children and adults are asked to label static vowel targets and syllables with those targets excised so all that remains are the formant transitions, listeners in both age groups show similar perceptual strategies; that is, adults and children attend most to the formant transitions flanking the (missing) vowel target (Nittrouer, 2007).

One more instance where adults and children have been observed to show similar perceptual strategies involves consonant decisions with covarying formant transitions and amplitude structure. These phonemic contrasts involve the manner distinction of stops versus glides. Although several contrasts have been used to study manner contrasts, */b/*-vowel versus */w/*-vowel is the most common, with */a/* the most frequently used vowel. Because this contrast is one of manner rather than one of place, the starting and ending frequencies of formant transitions are the same for both initial consonants. What differs across syllables depending on whether they start with */b/* or */w/* is the rate of change, for both formant transitions and amplitude. In the case of formant frequencies, the first two formants have low starting values because the vocal tract is tightly constricted near the front of the mouth. As the vocal tract opens, both formant frequencies rise, and by a fair amount when the following vowel is */a/*. However, that rise is more rapid for syllable-initial */b/* than for syllable-initial */w/*, because the vocal tract is opened more quickly for */b/*. For the same reason, signal amplitude rises to its maximum more rapidly for */b/* than for */w/*. These acoustic properties have been termed formant rise time (FRT) and amplitude rise time (ART; e.g., Nittrouer, Lowenstein, & Tarr, 2013).

Early investigations into the roles played by the acoustic cues underlying the */b/*-*/w/* labeling decision showed that adults primarily attended to FRT, with little attention paid to ART (Nittrouer & Studdert-Kennedy, 1986; Walsh & Diehl, 1991). That strategy makes perceptual sense because patterns of formant change are likely more resistant to adverse listening conditions than patterns of amplitude change, thus satisfying one of the requirements of what constitutes an efficient listening strategy. When children have been tested with this contrast, it has been found that they weight both FRT and ART similarly to adults (Bourland Hicks & Ohde, 2005; Nittrouer et al., 2013). To illustrate this outcome, logistic regression coefficients were computed on the proportion of */wa/* responses given to each stimulus in the Nittrouer et al. (2013) study. These coefficients can be used as weighting factors, as shown in the top two rows of Table 1. From these values, it appears as if 4- and 5-year-olds weighted both cues slightly less than adults, but there was only a significant age effect for FRT,  $t(45) = 3.27$ ,  $p = .002$ . When these regression coefficients are given as ratios of the total amount of perceptual weight distributed across the two acoustic cues, however, patterns are found to be very similar for both listener groups, as shown in the

**Table 1.** Mean weighting factors and cue ratios (and standard deviations) for formant rise time (FRT) and amplitude rise time (ART) computed for listeners in Nittrouer et al. (2013).

Variable	FRT		ART	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
<b>Weighting factors</b>				
Adults	9.62	2.51	1.09	1.28
4/5-year-olds	7.11	2.67	0.73	1.14
<b>Cue ratios</b>				
Adults	0.89	0.09	0.11	0.09
4/5-year-olds	0.86	0.10	0.14	0.10

*Note.* Twenty adults and 27 4- and 5-year-olds met the criteria to have their data included in analysis in that study. Weighting factors are the logistic regression coefficients derived from the analysis. Cue ratios are computed as follows: FRT cue ratio = (FRT weighting factor)/(FRT weighting factor + ART weighting factor); ART cue ratio = (ART weighting factor)/(FRT weighting factor + ART weighting factor).

bottom two rows of Table 1. These ratio values do not differ for adults and children. This last finding indicates that the proportion of total perceptual weight given to each cue is similar across age groups.

The experiment described in this article used a subset of the stimuli from Nittrouer et al. (2013), but the focus was expanded from that earlier study. The specific question examined in this experiment focused on what happens to those weighting strategies when cue saliency is affected by signal degradation, as occurs when an individual must listen through a simulated cochlear implant (CI). Acoustic cues to phonemic categories vary in the extent to which they are degraded by conditions of hearing loss and subsequent amplification, especially CIs. The stimuli from Nittrouer et al. (2013) were selected for the current experiment because the cue that listeners attended to strongly in that earlier study (FRT) would be expected to be severely degraded by CI processing. At the same time, the cue that listeners barely attended to (ART) should be rather well preserved by CI processing. These predictions arise from the nature of CI processing: These devices divide the spectrum into a relatively small set of channels, recover time-varying amplitude structure from each of those channels, and use the recovered amplitude values to drive the individual electrodes. As a result, spectral structure is only crudely preserved at the output of the CI, with time-varying amplitude structure across the electrode array (i.e., the gross temporal envelope) more accurately preserved.

The central focus of the current study was on how spectral degradation might affect listeners' weighting strategies, and how any effect on those strategies would affect word recognition. There is reason to suspect that children would be more negatively affected than adults by degradation in spectral structure for these stimuli. In both the Nittrouer et al. (2013) study and one other (Carpenter & Shahin, 2013), children were less accurate in their labeling of /b/ versus /w/ when original (and natural) /ba/ syllables had the slower ART of /wa/ superimposed on them, but the opposite was not found: When original /wa/ syllables were

given the more rapid ART of /ba/, adults and children alike responded consistently with /wa/. Nittrouer et al. speculated that the lower amplitude at onset of /wa/ made the formant structure less accessible perceptually. Thus all listeners, but especially children, were less certain of their responses. Imposing higher amplitude at onset, as in /ba/, would only serve to enhance listeners' abilities to hear the formant structure, regardless of whether it was appropriate for /ba/ or /wa/. The finding that children were more affected by the slowed ART led to the prediction in the current study that children would be more negatively affected by the spectral degradation introduced by hearing loss and the signal processing of CIs. Both factors (reduced amplitude and spectral degradation) should diminish the saliency of the FRT cue, thus affecting listeners' abilities to recover formant structure. Age-related differences in findings from the earlier work, however, suggest that adults are better than children at maintaining effective weighting strategies in the face of such diminished cue saliency. That outcome could be predicted on the basis of the fact that learned weighting strategies should be more firmly established in adulthood than in childhood.

A second goal of the current study was to examine the extent to which the amount of perceptual weight given to each acoustic cue could be predicted from listeners' sensitivity to changes in those cues. Investigations of weighting strategies for second-language learners have revealed that their weighting strategies are not strongly dependent on their sensitivities to the cues involved. An especially good example of that finding comes from a study by Miyawaki et al. (1975), who examined the abilities of native Japanese talkers both to discriminate nonspeech spectral glides and to weight third-formant transitions in their decisions of syllable-initial /r/ versus /l/. Although these listeners were just as sensitive as native English talkers to changes in spectral glides in the frequency region of the third formant, the native Japanese talkers failed to use that structure in their decisions about consonant identity. Similar dissociations between sensitivity and weighting have been demonstrated for children (e.g., Nittrouer & Crowther, 1998), but it nonetheless seemed worth exploring this possibility as explanation for any group differences in weighting strategies that might be observed in the current study.

Finally, a third goal of the current study was to examine whether weighting strategies affect word recognition, a more ecologically valid language measure than phonemic labeling. On one hand, it could be argued that it would not matter which cue to a phonemic distinction is weighted most heavily, as long as those cues covary reliably in how they signal that distinction. Thus, weighting spectral and amplitude structure should be equally as facilitative for word recognition. On the other hand, it might be argued that weighting strategies are so consistent across native talkers of a given language as to suggest that the specific strategies demonstrated by those talkers must have some perceptual advantage over any other possible ones.

A report by Moberly et al. (2014) helped shape predictions for the current study. In that experiment, adults

who are deaf and use CIs performed three tasks: labeling the /ba/ and /wa/ stimuli from Nittrouer et al. (2013), discriminating nonspeech signals that preserved the phonemically relevant spectral and amplitude structure of the speech stimuli, and recognizing words in a standard clinical test. Some of those listeners demonstrated weighting strategies that were atypical when compared to those found in earlier experiments (e.g., Nittrouer et al., 2013; Nittrouer & Studdert-Kennedy, 1986; Walsh & Diehl, 1991). Specifically, some of the listeners with CIs weighted the spectral cue (FRT) less and the amplitude cue (ART) more than previously observed for listeners with normal hearing. However, those weighting strategies could not be explained strictly by listeners' sensitivities to the underlying acoustic structure. Furthermore, the extent to which the listeners with CIs weighted FRT was significantly related to their word-recognition abilities; cue weighting for ART was not. Thus, evidence was provided for the suggestion that weighting strategies involve attentional factors independent of signal saliency. In addition, those outcomes supported the idea that the weighting strategies generally converged upon through development by members of a linguistic community are the most efficient strategies for the processing of that language. In other words, where phonemic decisions are concerned, not all cues are created equal.

## Summary

The current study was motivated by a desire to examine the effects on perceptual weighting strategies of spectral degradation, as might be encountered due to sensorineural hearing loss and subsequent implantation. Also examined was the question of the extent to which weighting strategies for spectrally degraded signals are constrained by the limits of sensitivity to the relevant cues. This issue has important clinical significance because if weighting strategies do not strictly reflect the limits of sensitivity, then auditory training could potentially be implemented to facilitate the acquisition of typical weighting strategies, those that support efficient language processing. In the current study, adults and children with normal hearing were recruited as listeners. This was done to examine whether spectral degradation had different effects on perceptual weighting strategies for listeners of different ages. These listeners were asked to make perceptual judgments with stimuli that had been noise-vocoded to diminish spectral structure. Because all listeners had normal hearing and no speech or language problems, they were presumed to have typical weighting strategies, matching the listeners who participated in the Nittrouer et al. (2013) study.

A second goal of this experiment was to examine the extent to which these weighting strategies could be explained by the sensitivities of the listeners to the two relevant acoustic cues: FRT and ART. The first of these cues would be degraded by the noise vocoding; the second would not.

Finally, a third goal was added to the current experiment. It was to examine whether perceptual weighting strategies explained word recognition, which was considered to

be an ecologically valid speech function. Words in the recognition task were spectrally degraded in the same manner as the syllables in the labeling and discrimination tasks.

## Method

### Participants

Twenty-four adults (between the ages of 18 and 33 years), twenty-six 8-year-olds (between the ages of 8;0 [years;months] and 8;11), and twenty-four 10-year-olds (between the ages of 10;0 and 10;6) originally came to the laboratory to participate. These age groups were selected rather than younger children as in the Nittrouer et al. (2013) study because it was discovered through preliminary testing that 8-year-olds were the youngest children able to handle the labeling task with vocoded syllables.

All listeners in the current study were native talkers of American English, and none of the listeners (or their parents, in the case of children) reported any history of hearing or speech disorder. All listeners passed hearing screenings consisting of the pure tones of 0.5, 1, 2, 4, and 6 kHz presented at 20 dB hearing level to each ear separately. Parents reported that their children were free from significant histories of otitis media, defined as six or more episodes during the first 3 years of life. Children were given the Goldman-Fristoe Test of Articulation—Second Edition (GFTA-2; Goldman & Fristoe, 2000) and were required to score at or better than the 30th percentile for their age in order to participate. None of the 10-year-olds had any errors on the task; of the twenty-six 8-year-olds, two had one error and one had two errors. These results indicate that nearly all of the children in this study had excellent articulation. Adults were given the Word Reading subtest of the Wide Range Achievement Test 4 (WRAT4; Wilkinson & Robertson, 2006), and all demonstrated better than a 12th-grade reading level. All listeners were also given the Expressive One-Word Picture Vocabulary Test—Fourth Edition (EOWPVT-4; Martin & Brownell, 2011) and were required to achieve a standard score of at least 92 (30th percentile). The mean EOWPVT-4 standard score for adults was 104 ( $SD = 6$ ), which corresponds to the 61st percentile. The mean EOWPVT-4 standard score for 8-year-olds was 112 ( $SD = 10$ ), which corresponds to the 78th percentile. The mean EOWPVT-4 standard score for 10-year-olds was 116 ( $SD = 14$ ), which corresponds to the 86th percentile. These scores indicate that the adult listeners had expressive vocabularies that were slightly above the mean of the normative sample used by the authors of the EOWPVT-4, and children had vocabularies closer to one standard deviation above the normative mean.

### Equipment and Materials

All testing took place in a soundproof booth, with the computer that controlled stimulus presentation in an adjacent room. Hearing was screened with a Welch Allyn TM262 audiometer using TDH-39 headphones. Stimuli were stored on a computer and presented through a Creative Labs Sound



Blaster sound card, a Samson headphone amplifier, and AKG K141 headphones. This system has a flat frequency response and low noise. Custom-written software controlled the presentation of the stimuli. The experimenter recorded responses with a keyboard connected to the computer. Parts of each test session (the word lists, EOWPVT-4, and WRAT4 or GFTA-2) were video-recorded using a Sony HDR-XR550V video recorder so that scoring could be done later. Participants wore Sony FM microphones that transmitted speech signals directly into the line input of the camera. This ensured good sound quality for all recordings. The word lists were recorded in a sound booth, directly onto the computer hard drive, via an AKG C535 EB microphone, a Shure M268 amplifier, and a Creative Laboratories Sound Blaster sound card.

For the labeling tasks, two pictures (on 8 × 8-in. cards) were used to represent each response label: for /ba/, a drawing of a baby, and for /wa/, a picture of a glass of water. When these pictures were introduced by the experimenter, it was explained to the listener that they were being used to represent the response labels because babies babble by saying /ba-/ba/ and babies call water /wa/.

For the discrimination tasks, a cardboard response card 4 × 14 in., with a line dividing it into two 7-in. halves, was used with all listeners during testing. On one half of the card were two black squares, representing the *same* response choice; on the other half were one black square and one red circle, representing the *different* response choice. Ten other cardboard cards (4 × 14 in., not divided in half) were used for training with children. On six cards were two simple drawings, each of common objects (e.g., hat, flower, ball). On three of these cards, the same object was drawn twice (identical in size and color), and on the other cards, two different objects were drawn. On four 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 game board with 10 steps was also used with children. They moved a marker to the next number on the board after each block of stimuli in the labeling and discrimination tasks (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 vocoded speech stimuli were used: two sets of vocoded synthetic speech stimuli for labeling tasks, two sets of vocoded stimuli for discrimination tasks, and four vocoded word lists. In addition to these test stimuli, one set of natural, unprocessed speech stimuli was used for training.

*Vocoded synthetic speech stimuli for labeling.* The two synthetic /ba-/wa/ stimulus sets from Nittrouer et al. (2013) were vocoded. These stimuli were based on natural

productions of /ba/ and /wa/. They were created with a Klatt synthesizer (Sensyn) using a sampling rate of 10 kHz with 16-bit resolution. The stimuli were all 370 ms in duration, with a fundamental frequency of 100 Hz throughout. Starting and steady-state frequencies of the first two formants were the same for all stimuli, even though the time that was required to reach steady-state frequencies varied (i.e., FRT). F1 started at 450 Hz and rose to 760 Hz at steady state. F2 started at 800 Hz and rose to 1150 Hz at steady state. F3 was kept constant at 2400 Hz. In addition to varying FRT, the time it took for the amplitude to reach its maximum (i.e., ART) could be varied independently. In natural speech, both FRT and ART are briefer for syllable-initial stops, such as /b/, and longer for syllable-initial glides, such as /w/.

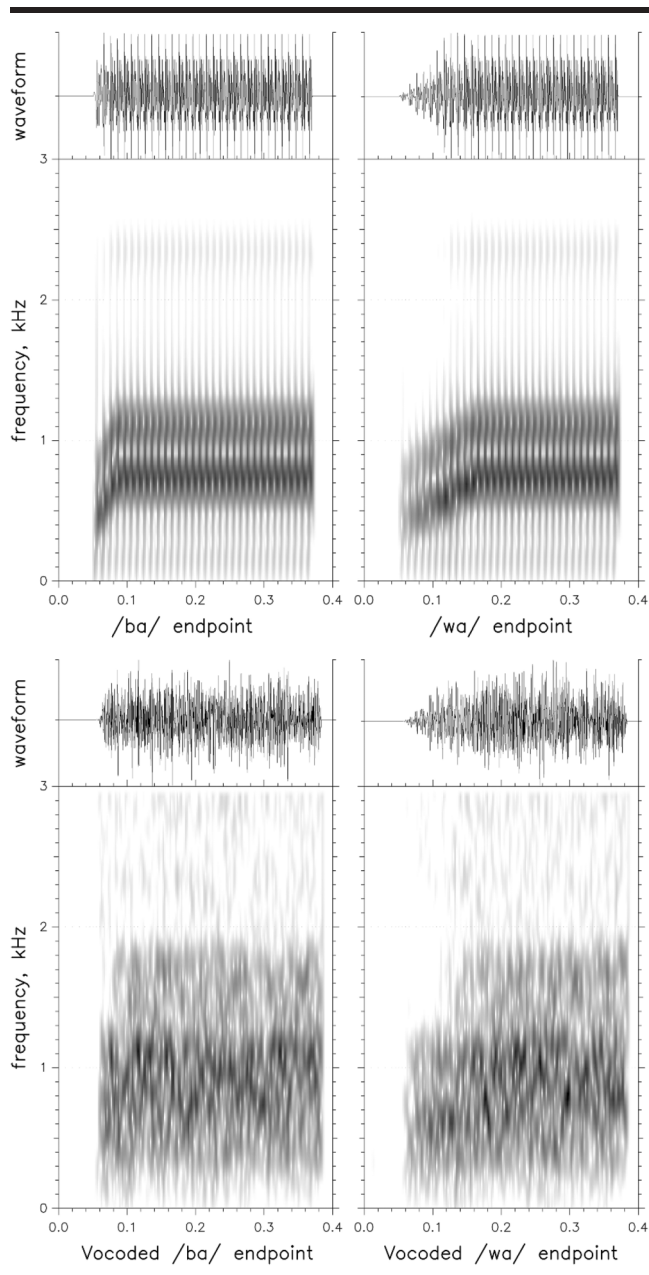
In one set of stimuli in this experiment, FRT varied along a nine-step continuum from 30 to 110 ms in 10-ms steps. Each stimulus was constructed with each of two ART values: 10 and 70 ms. This resulted in 18 FRT stimuli (9 FRTs × 2 ARTs). In the other set of stimuli, ART varied along a seven-step continuum from 10 to 70 ms in 10-ms steps. For these stimuli, FRT was set to both 30 and 110 ms, resulting in 14 ART stimuli (2 FRTs × 7 ARTs).

These stimuli from earlier experiments were vocoded using the same MATLAB routine as used in previous experiments (e.g., Nittrouer & Lowenstein, 2010; Nittrouer, Lowenstein, & Packer, 2009; Nittrouer et al., 2014). All stimuli were first bandpass filtered with a low-frequency cutoff of 250 Hz and a high-frequency cutoff of 8000 Hz. The stimuli were vocoded using eight channels, with cutoff frequencies based on Greenwood's (1990) formula. The cutoff frequencies between channels for vocoding were 437, 709, 1104, 1676, 2507, 3713, and 5462 Hz.

All filtering used in the generation of these stimuli was done with digital filters that had greater than 50-dB attenuation in stop band and had 1-Hz transition bands between pass and stop bands. Each channel was half-wave rectified and filtered below 80 Hz to remove fine structure. The temporal envelopes derived for separate channels were subsequently used to modulate white noise, limited to the same channels as those used to divide the speech signal. The resulting bands of amplitude-modulated noise were combined with the same relative amplitudes across channels as measured in the original speech signals. Root-mean-square amplitude was equalized across all stimuli. Figure 1 shows the most /ba/-like and the most /wa/-like stimuli before and after vocoding.

*Vocoded stimuli for discrimination.* Two sets of synthetic speech stimuli were used to assess listeners' sensitivity to FRT and ART. For assessing sensitivity to FRT, stimuli from along the FRT continuum used in labeling were selected, with ART held constant at 10 ms. Thus only FRT varied across stimuli, and ART was rapid. This FRT continuum consisted of nine stimuli. For assessing sensitivity to ART, stimuli were generated with stable formant frequencies matching the steady-state values used in the labeling stimuli; there were no formant transitions. The ART of the gross temporal envelope of the stimuli varied along an 11-step continuum from 0 to 250 ms, in 25-ms steps. This

**Figure 1.** Spectrograms of the most /ba/-like and /wa/-like stimuli, before vocoding (top) and after vocoding (bottom). Time is represented on the x-axis, in seconds.



ART continuum consisted of 11 stimuli. Both stimulus sets were vocoded in the same way as the stimuli for labeling.

**Vocoded word lists.** The four Central Institute for the Deaf (CID) W-22 word lists were used (Hirsh et al., 1952). This measure consists of four lists of 50 monosyllabic words each, spoken by a male talker. The original recordings of these lists were not used because it was not possible to separate the individual words from the carrier phrase “Say the word.” These lists were newly recorded by a male talker at a 44.1-kHz sampling rate with 16-bit resolution, with the

words and carrier phrase recorded separately. Each word was vocoded in the same way as the stimuli for labeling. The lists were presented via computer, using a MATLAB interface. For each word, the phrase “Say the word” was presented as unvocoded speech, followed by the vocoded word.

**Natural training stimuli.** Five exemplars each of /ba/ and /wa/ from Nittrouer et al. (2013) were used for training for the labeling tasks. These tokens were produced by a male talker and digitized at a 44.1-kHz sampling rate with 16-bit resolution.

### General Procedures

All procedures were approved by the Institutional Review Board of the Ohio State University. All listeners (or their parents, in the case of children) signed the consent form before testing took place. Listeners were tested in a single session that lasted 1 hr 15 min. The hearing screening was administered first, followed by one of the vocoded CID word lists. The specific list presented, out of the four, varied among listeners in each age group, such that equal numbers of listeners heard each list. The two labeling and two discrimination tasks were then presented, in an alternating format. There were eight possible orders for alternating these tasks. Three participants of each age group were assigned to each possible labeling/discrimination order. After these listening tasks, a second vocoded CID word list was presented. Word recognition was evaluated before and after the labeling and discrimination tasks in order to examine whether there would be learning effects of listening to the vocoded stimuli. Finally, listeners were administered the WRAT4 (adults) or GFTA-2 (children), along with the EOWPVT-4.

### Task-Specific Procedures

**Labeling tasks.** For the /ba/-/wa/ labeling tasks, the experimenter introduced each picture separately and told the listener the name of the syllable associated with the picture. Then the listener heard the five natural training exemplars of /ba/ and /wa/. Listeners were instructed to repeat the syllable and point to the picture, which ensured that they were correctly associating the syllable and the picture. Listeners were required to respond correctly to nine of the 10 exemplars to proceed to the next level of training.

Training was provided next with the vocoded endpoints corresponding to the most /ba/-like (30-ms FRT, 10-ms ART) and /wa/-like (110-ms FRT, 70-ms ART) stimuli. These endpoints were the same for both sets of labeling stimuli. Listeners were told they would hear a robot voice saying /ba/ and /wa/. Listeners heard five presentations of each vocoded endpoint, in random order. They had to respond to nine out of 10 of them correctly to proceed to testing. Feedback was provided for up to two rounds of training, and then listeners were given up to three rounds without feedback to reach that training criterion. If a listener was unable to respond to nine of 10 endpoints correctly by

the third round, testing was stopped. If a listener failed training prior to being tested with either stimulus set, no data from that listener were included in the final analysis. This strict criterion, termed the *training* criterion, was established because the labeling data were the focus of the study, so it was critical that these data be available.

During testing, listeners heard all stimuli in the set (18 for the FRT set; 14 for the ART set) presented in blocks of either 18 or 14 stimuli. Ten blocks were presented. Listeners needed to respond with at least 70% accuracy to the endpoints for at least one stimulus set in order to have their data included in the analysis; this was the *endpoint* criterion.

*Discrimination tasks.* For the two discrimination tasks, an AX procedure was used. In this procedure, listeners compare a stimulus that varies across trials (X) to a constant standard (A). For this testing, the comparison (X) stimulus was one of the stimuli from across either the FRT or ART continua. For the FRT stimuli, the A stimulus was the one with a 30-ms FRT. For the ART stimuli, the A stimulus was the one with a 0-ms ART. The standard was always one of the comparison stimuli as well. The interstimulus interval between standard and comparison was 450 ms. After hearing the two stimuli, the listener 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 were used because each served as a check of the reliability of the response.

Before testing with stimuli in each condition, all listeners were presented with five pairs of stimuli that were identical and five pairs of stimuli that were maximally different, in random order. Listeners were asked to report whether the stimuli were the same or different and were given feedback (for up to two rounds). Next, these same training stimuli were presented, and listeners were asked to report if they were the same or different without feedback. Listeners needed to respond correctly to nine of the 10 training trials without feedback in order to proceed to testing and were given up to three rounds without feedback to reach criterion. Unlike criteria for the labeling tasks, listeners' data were not removed from the overall data set if they were unable to meet training criteria for one of the conditions in the discrimination task. However, their data were not included for that one condition.

During testing, listeners heard blocks with all of the comparison stimuli included once. Ten blocks were presented. Listeners needed to respond correctly to at least 14 of these physically same and maximally different stimuli (70%) to have their data included for that condition in the final analysis.

*CID word task.* For the CID word task, listeners were told that they would hear a man's voice say the phrase "Say the word," which would be followed by a word in a robot voice. Listeners were instructed to repeat the word the robot said as accurately as possible. They were given one opportunity to hear each word.

## Scoring and Analyses

*Computation of weighting factors from labeling tasks.* To derive weighting factors for the FRT and ART cues, a single logistic regression analysis was performed on responses to all stimuli across the four continua (two FRT and two ART) for each listener. The proportion of /wa/ responses given to each stimulus served as the dependent measure in the computation of these weighting factors. Settings of FRT and ART served as the predictor variables. Settings for both variables were scaled across their respective continua such that endpoints always had values of 0 and 1. Raw regression coefficients derived for each cue served as the weighting factors, and these FRT and ART weighting factors were used in subsequent statistical analyses. However, these factors were also normalized on the basis of the total amount of weight assigned to the two factors. For example, the normalized FRT weighting factor was (raw FRT weighting factor)/(raw FRT weighting factor + raw ART weighting factor). Negative coefficients were recoded as positive in this analysis. This matches procedures of others (e.g., Escudero, Benders, & Lipski, 2009; Giezen, Escudero, & Baker, 2010), who have termed these new scores *cue ratios*. That term is adopted in this article.

Both types of score—raw weighting factors and cue ratios—were examined in this study because, although related, each provides a slightly different perspective on cue weighting across listener groups. For example, it is possible that listeners in one group could pay less perceptual attention to acoustic structure in the stimuli overall than listeners in another group but nonetheless show similar relative weighting across cues. That difference would be revealed by smaller weighting factors for both cues examined for one group than for another but similar cue ratios across groups.

*Computation of  $d'$  values from discrimination tasks.* The discrimination functions of each listener were used to compute an average  $d'$  value for each condition (Holt & Carney, 2005; Macmillan & Creelman, 2005). The  $d'$  value was selected as the discrimination measure because it is bias free. The  $d'$  value is defined in terms of z-values along a Gaussian normal distribution, which means they are standard deviation units. The  $d'$  value was calculated at each step along the continuum as the difference between the z-value for the "hit" rate (proportion of *different* responses when A and X stimuli were different) and the z-value for the "false alarm" rate (proportion of *different* responses when A and X stimuli were identical). A hit rate of 1.0 and a false-alarm rate of 0.0 required a correction in the calculation of  $d'$  and were assigned values of .99 and .01, respectively. A value of zero for  $d'$  meant the participant could not discriminate the difference between stimuli. A positive  $d'$  value suggests that the hit rate was greater than the false-alarm rate. Using these correction values, the minimum  $d'$  value would be 0 and the maximum would be 4.65. The average  $d'$  value was then calculated across all steps of the continuum, and this value was used in subsequent statistical analyses.

*Scoring of CID word lists.* The CID lists were scored on a phoneme-by-phoneme basis by a graduate research

assistant. In addition, the first author, who is a trained phonetician, scored 25% of listener responses. Words were only scored as correct if all phonemes were correct and no additional phonemes were included. As a measure of interrater reliability, the scores of the graduate research assistant and first author were compared on a phoneme-by-phoneme basis. Percentage agreement was used as the metric of reliability. This procedure was done for data from 8-year-olds, 10-year-olds, and adults separately. For data analyses, the dependent CID measure was the number of words recognized correctly.

## Results

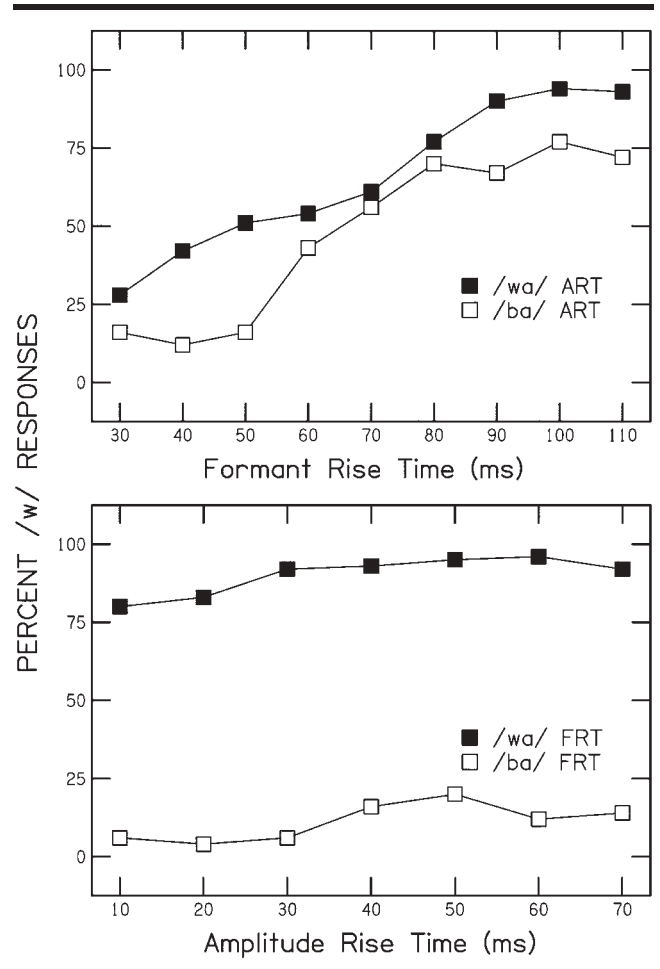
Applying the criteria for participation in the labeling tasks (i.e., at least 90% correct during training for both conditions and 70% correct endpoints during testing for one condition) resulted in data being excluded from eight 8-year-olds, six 10-year-olds, and one adult. Consequently, data were included from eighteen 8-year-olds, eighteen 10-year-olds, and 23 adults. In this report, exact  $p$  values are reported when  $p < .10$ ; otherwise, results are reported as not significant.

### Weighting Factors for Spectrally Degraded Signals

The first question addressed in these analyses concerns how spectral degradation affected weighting strategies for adults and children. Figures 2, 3, and 4 show mean labeling functions for each age group, for both stimulus conditions. The FRT continua are in the top panels and the ART continua are in the bottom panels. Looking first at Figure 2, for adults, it can be seen that the functions are steeper and less separated for the FRT continua than for the ART continua. Those response patterns indicate that listeners attended more to FRT than to ART when listening to stimuli in both conditions. That interpretation arises because steepness of functions is an indicator of the extent to which the cue represented on the  $x$ -axis was weighted (the steeper the function, the greater the weight), and separation between functions is an indicator of the extent to which the binary-set property was weighted (the more separated, the greater the weight). Looking across figures, a developmental trend can be observed in which listeners attended less to FRT and more to ART with decreasing age. In the top panel, this trend appears as shallower functions that are more separated. In the bottom panel, it appears as the opposite: somewhat steeper functions that are less separated.

The top section of Table 2 shows mean raw weighting factors (and standard deviations) for FRT and ART, for each age group separately. These outcomes reveal the developmental trends seen in Figures 2–4: There is a developmental increase in the weight assigned to the FRT cue and a developmental decline in the weight assigned to the ART cue. To test the validity of that impression, a two-way repeated measures analysis of variance (ANOVA) with age as the between-subjects factor and weighting factor as the repeated measure was performed. This analysis revealed

**Figure 2.** Results from the vocoded formant rise time (FRT) continua (top panel) and vocoded amplitude rise time (ART) continua (bottom panel) for adults. Filled symbols indicate when ART or FRT was set to be appropriate for /wa/, and open symbols indicate when ART or FRT was set to be appropriate for /ba/.

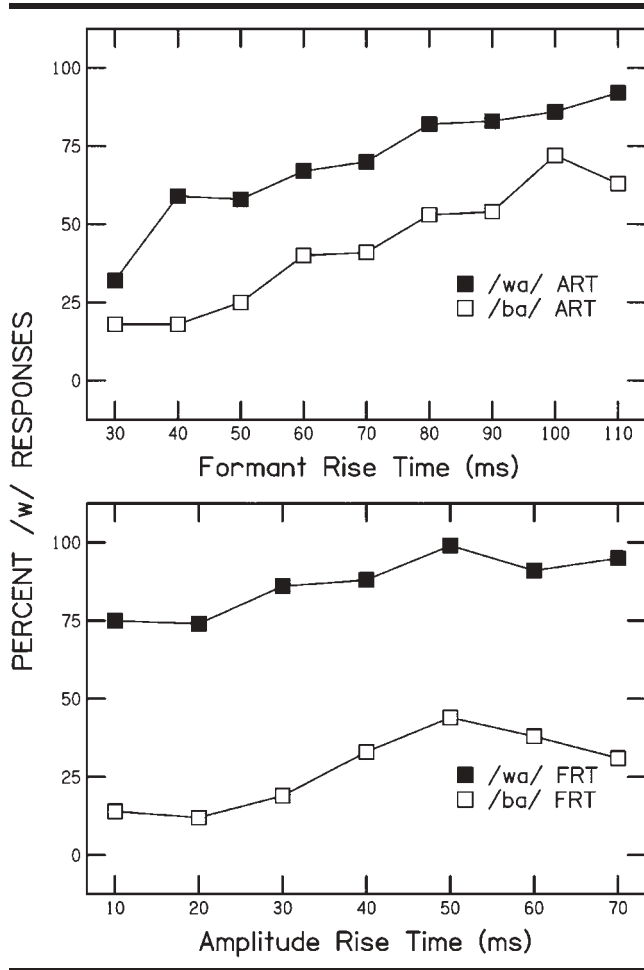


significant main effects of age,  $F(2, 56) = 5.20, p = .008, \eta^2 = .16$ , and weighting factor,  $F(1, 56) = 37.54, p < .001, \eta^2 = .40$ . In addition, the Age  $\times$  Weighting Factor interaction was significant,  $F(2, 56) = 12.09, p < .001, \eta^2 = .30$ . This significant interaction means that the relative amounts of weight given to each factor varied across groups.

To examine that interaction more closely, one-way ANOVAs were performed on the weighting factors for FRT and ART separately. Those analyses revealed that for FRT weighting factors, the main effect of age was significant,  $F(2, 56) = 11.62, p < .001$ . Furthermore, post hoc comparisons with Bonferroni adjustments for multiple comparisons showed that the comparisons of both children's groups to adults were significant: for 8-year-olds versus adults,  $p < .001$ ; for 10-year-olds versus adults,  $p = .008$ . The comparison of 8- versus 10-year-olds was not significant. The one-way ANOVA for ART weighting factors did not reach significance,  $F(2, 56) = 2.49, p = .09$ . Together these results indicate that FRT weighting was the primary difference between



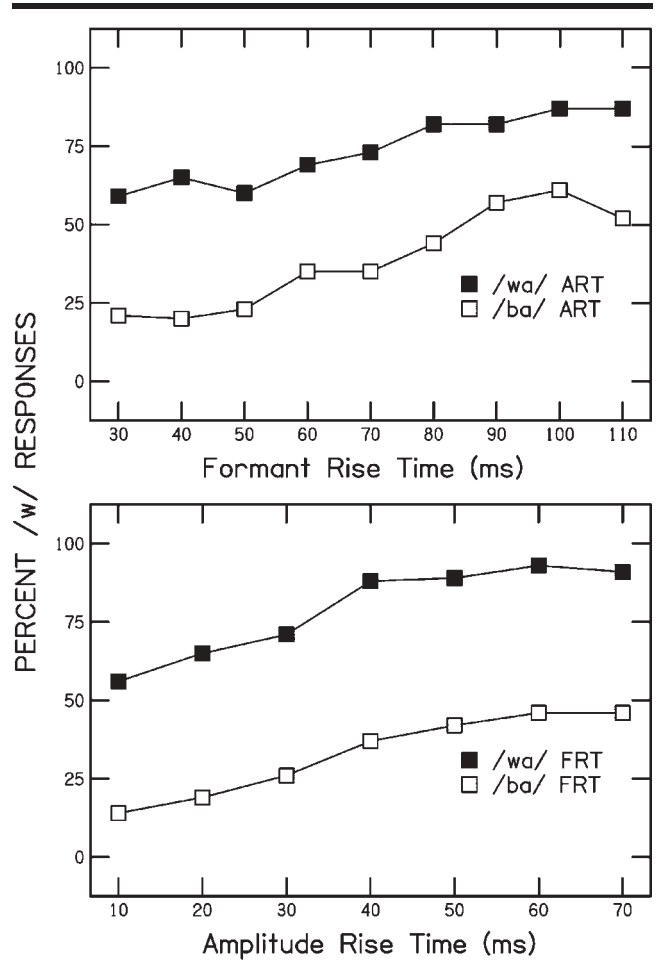
**Figure 3.** Results from the vocoded formant rise time (FRT) continua (top panel) and vocoded amplitude rise time (ART) continua (bottom panel) for 10-year-olds.



children and adults and that there was a developmental increase in that weighting. Both the significant main effect of age and the significant Age  $\times$  Weighting Factor interaction arose due to this developmental trend.

We compared the outcomes of the current study to those from Nittrouer et al. (2013), in which these same stimuli were used without being vocoded. The analyses permitted closer examination of whether adults or children seem to modify weighting strategies when speech signals are spectrally degraded. Looking first at results for adults across the two experiments, it was found that adults in the current study weighted FRT less than adults in the earlier study,  $t(41) = 6.38, p < .001$ . That outcome was predicted on the basis of the spectral degradation of the signals in the current study. Weighting of ART was similar across the two groups. In the comparison of weighting factors from children, only results from the youngest group in this study were used, the 8-year-olds. The reason for this restriction was that all children in the current study were older than those from the earlier study, so it seemed prudent to restrict the comparison to just the youngest ones. As with adults,

**Figure 4.** Results from the vocoded formant rise time (FRT) continua (top panel) and vocoded amplitude rise time (ART) continua (bottom panel) for 8-year-olds.



it was found that these 8-year-olds weighted FRT for these vocoded stimuli less than the 4- and 5-year-olds did for the nonvocoded stimuli in the earlier study,  $t(43) = 7.68, p < .001$ . In this case, however, it was additionally observed that the 8-year-olds in the current experiment weighted ART more than the 4- and 5-year-olds from the earlier one did,

**Table 2.** Mean weighting factors and cue ratios (and standard deviations) for formant rise time (FRT) and amplitude rise time (ART) in the current experiment.

Variable	FRT		ART	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
<b>Weighting factors</b>				
Adults	4.74	2.50	1.02	1.35
10-year-olds	2.93	1.40	1.41	0.88
8-year-olds	2.07	0.93	1.80	.97
<b>Cue ratios</b>				
Adults	0.75	0.20	0.25	0.20
10-year-olds	0.67	0.14	0.33	0.14
8-year-olds	0.53	0.19	0.47	0.19

$t(43) = 3.27, p = .002$ . Thus, there is some evidence that unlike adults, children appear to weight amplitude structure more when spectral structure is degraded.

In order to examine the relative weighting of cues across groups, the cue ratios were examined. These values are shown in the bottom of Table 2 and reveal developmental changes in relative weighting across cues. Eight-year-olds weighted FRT and ART almost equally; 10-year-olds weighted FRT a little more and ART a little less than those younger children; and adults showed an even more unequal weighting strategy. To examine these trends, one-way ANOVAs were conducted. The ANOVA examining FRT cue ratios showed a significant age effect,  $F(2, 56) = 7.41, p = .001$ . In this case, post hoc comparisons with Bonferroni corrections showed a significant difference between adults and 8-year-olds,  $p = .001$ . However, for these cue ratios, 10-year-olds and adults showed similar outcomes; the comparison of their scores was not significant. On the other hand, the comparison of 8-year-olds versus 10-year-olds was close to significant,  $p = .059$ . The ANOVA performed on ART cue ratios produced the same statistical outcomes as the ANOVA on FRT cue ratios, as would be expected given that these scores are complementary proportions derived from a common sum. In a nutshell, there was a developmental decline in the weight given to ART. These findings differ from those observed by Nittrouer et al. (2013) for stimuli that were not vocoded: In that earlier experiment, adults and children showed similar and unequal relative weighting of the cues.

### Sensitivity to FRT and ART

The next question examined was whether weighting strategies were constrained by sensitivity to the two kinds of acoustic structure that served as cues to phonemic decisions. The first step in this process was to examine sensitivity to these cues, using outcomes of the discrimination tasks. For the FRT discrimination task, quite a few listeners failed to meet the criteria to have their data included in the analyses: six 8-year-olds (five failed training, one failed endpoints), six 10-year-olds (four failed training, two failed endpoints), and four adults (all failed training). For the ART discrimination task, the numbers of listeners who failed to meet the criteria were much fewer: three 8-year-olds (two failed training, one failed endpoints), one 10-year-old (who failed endpoints), and no adults. These outcomes immediately present an interesting contradiction that addresses the question of whether sensitivity to the acoustic cues constrained the weighting of those cues: Although it appears that listeners in all age groups were more sensitive to ART than FRT—at least as judged by the fact that more could discriminate stimuli in the ART condition—the weights they assigned to FRT in their phonemic labeling were equal to (8-year-olds) or higher than (10-year-olds and adults) the weights they assigned to ART. This inconsistency makes it hard to propose that sensitivity to the cues could have constrained weighting of those cues.

**Figure 5.** Discrimination results from the vocoded formant rise time (FRT) task (top panel) and vocoded amplitude rise time (ART) task (bottom panel) for adults (open triangles), 10-year-olds (filled squares), and 8-year-olds (open circles).

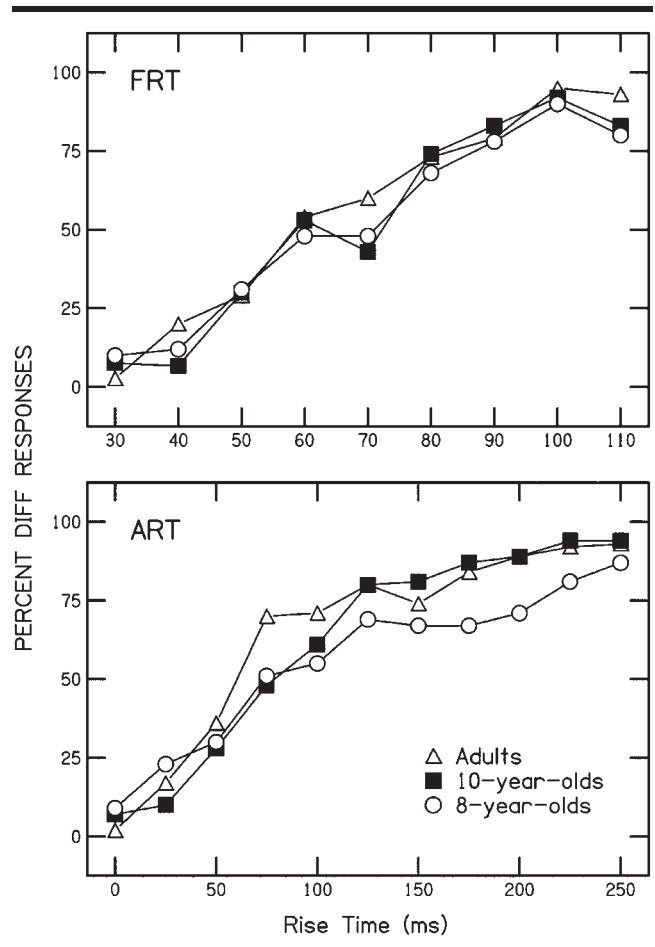


Figure 5 shows discrimination functions for both the FRT and ART continua. Looking first at FRT (top), it appears that functions are similar for all groups, but adults may have been slightly more sensitive than 8-year-olds. Similar trends are seen for ART functions (bottom). Table 3 presents means and standard deviations of  $d'$  values for FRT and ART discrimination. For comparison, a  $d'$  value

**Table 3.** Values of  $d'$  for the formant rise time (FRT) and amplitude rise time (ART) discrimination tasks.

Participants	N	FRT continuum		N	ART continuum	
		M	SD		M	SD
Adults	<b>19</b>	2.64	0.88	<b>23</b>	3.00	0.75
10-year-olds	<b>12</b>	2.34	0.62	<b>17</b>	2.41	0.77
8-year-olds	<b>12</b>	1.81	0.78	<b>15</b>	2.07	1.00

Note. The numbers of participants who were able to meet the criteria for each task are in bold.

of 0 indicates that the listener had no ability to discriminate the cue; a value of 2.33 suggests that the individual could discriminate 50% of different stimuli as different; and a value of 4.65 suggests 100% discrimination of different stimuli as different. With that in mind, it can be seen from Table 3 that adults and 10-year-olds could discriminate more than 50% of the different FRT and ART stimuli, while 8-year-olds correctly recognized as different less than 50% of the stimuli. These  $d'$  values match patterns seen in Figure 5. One-way ANOVAs were performed on data for each discrimination task to test these impressions of trends across age groups. Looking first at the FRT discrimination task, the main effect of age was significant,  $F(2, 40) = 4.08, p = .024$ . However, post hoc comparisons with Bonferroni adjustments showed that only 8-year-olds and adults performed differently from each other,  $p = .020$ . For ART discrimination, the main effect of age was also significant,  $F(2, 52) = 6.00, p = .005$ . Post hoc comparisons with Bonferroni adjustments showed a significant effect for 8-year-olds versus adults,  $p = .005$ . The comparison of 10-year-olds versus adults was not significant, but  $p = .096$ .

In order to evaluate whether sensitivity to either of these cues explained weighting of the cues, Pearson product-moment correlation coefficients were computed. For FRT  $d'$  and FRT cue weighting, a significant correlation coefficient was obtained when all listeners were included in the analysis,  $r(43) = .433, p = .004$ . However, when these correlation coefficients were computed for each age group separately, no significant outcomes were obtained. For ART  $d'$  and ART cue weighting, the correlation coefficient computed with all listeners was not significant. These correlational analyses provide support for the conclusion that sensitivity to these cues did not constrain weighting of the cues for the most part. The one exception is that at the group level, sensitivity to FRT was somewhat related to the weighting of that cue. However, this relationship was not strong enough to be seen when age range was controlled.

One problem with using these correlation analyses to assess whether weighting of the two acoustic cues was constrained by sensitivity to those cues was that not all listeners were able to meet the criteria to have their data included in the analysis of discrimination outcomes. This problem was especially prevalent for the FRT cue, and especially for children. Consequently,  $t$  tests were performed on the weighting factors for each cue, with listeners divided according to whether or not they met the inclusion criteria for the discrimination tasks. Separate tests were done for each age group. For FRT, only 8-year-olds showed a significant difference in cue weighting between those who passed and those who failed the inclusion criteria: The mean weighting factor was 2.50 ( $SD = .66$ ) for those who passed and 1.22 ( $SD = .83$ ) for those who failed,  $t(16) = 3.53, p = .003$ . For ART, there were far fewer listeners who failed to meet the criteria to have their data included for the discrimination task. In fact, a  $t$  test could only be performed on ART weighting factors for 8-year-olds, and no difference was found between those who did and those who did not meet the inclusion criteria for the ART discrimination task.

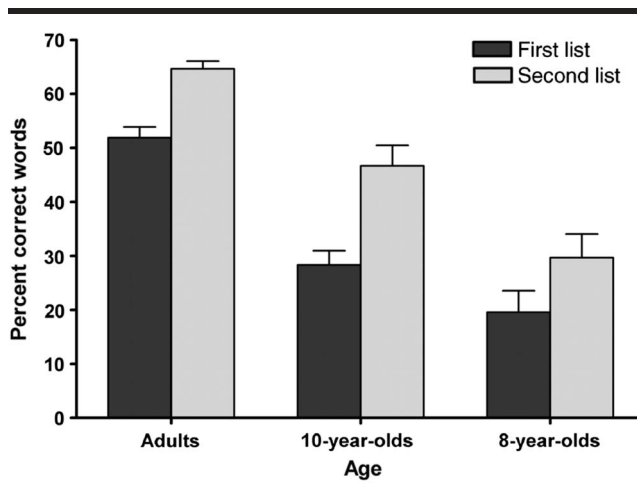
## Word Recognition

The third question addressed in these analyses was whether weighting strategies accounted for how well listeners could recognize words. The reliability metric indicated that phoneme agreement between the two scorers ranged from 94% to 100% across the listeners whose responses they both scored, with a mean agreement of 98%. This was considered very good reliability, and the scores of the first transcriber were used.

In this study, word recognition was evaluated twice: before listeners participated in the labeling and discrimination tasks and after they participated. Figure 6 shows mean percent correct scores for the first and second CID word lists presented, for each age group separately. It is apparent that listeners in all groups recognized more words correctly in the second list they heard, the one administered after testing with the vocoded syllables. Nonetheless, to evaluate that impression, a two-way repeated measures ANOVA was performed with age as the between-subjects factor and presentation number (first or second) as the repeated measure. Significant main effects were found for presentation number,  $F(1, 56) = 157.55, p < .001, \eta^2 = .74$ , and age,  $F(2, 56) = 32.56, p < .001, \eta^2 = .54$ . The Presentation Number  $\times$  Age interaction was also significant,  $F(2, 56) = 7.44, p = .001, \eta^2 = .21$ , an effect that can be attributed to a larger difference between first- and second-list scores for 10-year-olds than for the other two groups.

Next, we examined the extent to which variability in word-recognition scores was explained by the perceptual weight assigned to FRT or ART. To do this, we conducted a stepwise forward regression, with mean word-recognition scores (across the two lists) as the dependent measure and FRT and ART weighting factors as the predictor variables. This analysis was performed with all listeners included. The outcome revealed a significant relationship between FRT weighting factors and average word-recognition scores, standardized  $\beta = .547, p < .001$ . The ART weighting factor

**Figure 6.** Percentage of correct word responses for the vocoded word lists. Error bars indicate standard errors.



did not explain any unique variance in the average word-recognition scores. These results indicate that listeners who weighted FRT more strongly were better at word recognition. However, when this regression analysis was run for each age group separately, a significant effect of FRT was found for 8-year-olds only, standardized  $\beta = .570$ ,  $p < .001$ ; for adults, standardized  $\beta = .367$  and  $p = .085$ . Thus there was some indication of weak within-age-group effects.

## Discussion

One developmental change that typically occurs during childhood is that children acquire strategies for listening to speech that support accurate recognition with a minimum of cognitive effort and interference from unwanted sounds. These optimal strategies require that listeners attend to the acoustic properties in the speech signal that are most informative regarding phonemic structure in their native language. The very fact that these language-specific strategies exist means that children must discover them in their first language, as they interact with the ambient linguistic environment.

A factor that could potentially disrupt a child's ability to acquire the appropriate listening strategies for the language being learned is hearing loss. In particular, the spectral representation of the speech signal is degraded by damaged cochlear functioning. And if a child has hearing loss severe enough to warrant cochlear implantation, then the spectral degradation of the signal being received is even greater because of the signal processing of those devices. These factors can be expected to inhibit access to relevant spectral structure, and that problem in turn could be predicted to negatively affect real-world language functions. The current study was designed to help evaluate these predictions.

With the theoretical framework in mind that was described earlier, three goals were established for the current study. First, perceptual weighting strategies for two acoustic cues to a phonemic labeling task were evaluated, for listeners of three ages: adults, 10-year-olds, and 8-year-olds. Signal processing rendered one of the two cues highly degraded (i.e., rate of formant rise, or FRT), while the other cue was left intact (i.e., rate of amplitude rise, or ART). The extent of degradation across these cues was inversely correlated with the strength of weighting of those cues in typical speech perception. This experimental design allowed an examination of whether listeners can maintain optimal weighting strategies in the face of diminished cue saliency.

The second goal was to examine if limits on auditory sensitivity to either of these cues could explain the weighting of the cue in phonemic decisions. It was expected that sensitivity would be poorer for FRT than for ART, because of the effect that spectral degradation has on the representation of formant structure. Time-varying amplitude structure is left intact by the signal processing employed.

Finally, the third goal of the current study was to see if weighting strategies, as measured by the labeling task, could explain word recognition for spectrally degraded signals. It could be that a weighting strategy strongly dependent

on an atypical cue would facilitate word recognition just as well as the typical strategy, as long as the usual and atypical cues covary reliably in the language.

Outcomes of the current study revealed that children showed disproportionately greater diminishment in weighting of FRT than adults, presumably due to the spectral degradation. Children younger than those in the study reported here participated in a previous study by Nittrouer et al. (2013) and weighted FRT more similarly to adults in that study, where nonvocalized stimuli were presented. Consequently, it seems fair to conclude that the children in this study would show similar FRT weighting factors to adults if the syllables were not vocalized. Thus, the effect of diminished weighting is attributed to that spectral degradation. For ART weighting factors, a small developmental decrease was found. This effect was not observed in the earlier study (Nittrouer et al., 2013), so it likely reflects another consequence of the spectral degradation: Children, but not adults, appear to have increased their perceptual attention to the cue left intact. This conclusion is supported by a comparison across studies (i.e., Table 1 vs. Table 2). Adults in both studies demonstrated similar weighting factors for ART, which would remain intact for vocalized stimuli. Children in the current study, however, showed enhanced weighting of ART compared to children in that earlier study. Although factors other than spectral resolution differed across studies (especially age of the children), the most parsimonious explanation is that this enhanced weighting for amplitude structure reflects children's perceptual adjustment to the degradation in spectral representations. The typical weighting strategy—of attending strongly to FRT and only marginally to ART—would not be as well established for the children as for the adults. Thus these strategies were more malleable in children's speech perception.

A second outcome of the current study was that weighting strategies could not be strictly explained by listeners' sensitivities to the acoustic properties manipulated in stimulus creation. No relationship was observed between sensitivity to ART, as measured by the discrimination task, and weighting of ART in the labeling task. That failure to find a relationship was observed even when all listeners were included in a single analysis so that developmental trends could be observed. However, at that level of analysis where all listeners were included, a relationship was found between sensitivity to FRT and weighting of that cue in labeling. This relationship was moderate in magnitude but disappeared when correlation coefficients were computed for each age group separately. Consequently, it needs to be concluded that the relationship between sensitivity to formant structure (in the face of spectral degradation) and weighting of that structure both become stronger as children get older. Nonetheless, this relationship was not strong enough to explain outcomes within any single age group, although for the youngest children there was a difference in weighting factors as a function of whether or not they could pass the inclusion criteria on the discrimination task. On the whole, however, sensitivity to individual cues did not



account strongly for the amount of weight assigned to those cues. Separate attentional factors seemed to be at work, as well. This conclusion suggests that the weights assigned to acoustic cues even under conditions of spectral degradation could be modified through training, and that conclusion has implications for working with individuals with CIs. The idea that training could enhance listeners' abilities to function with degraded spectral signals also receives support from the demonstration of improved word recognition for spectrally degraded signals over the course of just 1 hr, time spent listening to these degraded signals.

Finally, some evidence was found that weighting strategies that placed perceptual weight on the cue typically attended to most strongly helped to support word recognition. Across all listeners, the standardized  $\beta$  coefficient for FRT was found to be significant in a stepwise regression using word recognition as the dependent measure; the coefficient for ART was not significant. Although this significant effect was replicated only for the youngest age group, it seems fair to suggest that having appropriate weighting strategies for the acoustic cues residing in the speech signal facilitates word recognition. The fact that the relationship was strongest for the youngest listeners in this study—the ones still acquiring language-specific weighting strategies—suggests that the acquisition of weighting strategies typically used by mature talkers/listeners of a language supports the development of speech-recognition abilities.

### ***Clinical Implications and Future Directions***

Outcomes of the current investigation support the possibility that behavioral interventions could help patients with hearing loss tune their perceptual attention to the acoustic properties that best facilitate language processing, even if those properties are not as salient for this population as they are for listeners with normal hearing. As a field, efforts to find ways of enhancing the representation of fine spectral structure in auditory prostheses, including CIs, are ongoing and critically important. Ultimately those efforts could restore typical saliency to the acoustic cues underlying phonemic categorization. At the level of the individual, it is just as important to find ways of designating maps that preserve fine spectral structure as much as possible. However, the work reported here demonstrates that the degree of access to various kinds of acoustic structure in the speech signal (i.e., auditory sensitivity) does not fully account for how well listeners use that structure in phonemic judgments, a conclusion well supported by results from second-language learners. Perceptual attention also plays a role in supporting efficient and accurate speech recognition. In this study, children were found to be not only especially susceptible to the negative consequences of spectral degradation but also especially helped by having appropriate weighting strategies. Therefore, providing appropriate auditory training for these young listeners would be particularly important. That training needs to focus on helping them acquire efficient and accurate listening strategies. Based on

the results of the current study, it is difficult to ascertain whether auditory training would be most effective if it involved simple signals—such as the syllables used in this experiment, which evoked the observed training effects—or if materials more natural in structure would facilitate learning to an even greater extent than was found here. In the current study, only one kind of training material was provided, without the intent of it serving as a training aid. Future research will need to compare outcomes for different types of materials.

### **Summary**

One important chore children must accomplish during childhood is discovering the most effective strategies for processing speech signals. In particular, children must learn which components, or cues, of that signal must be attended to (i.e., weighted) and which can safely be perceptually ignored. Many of the cues most relevant to speech perception involve spectral structure. Consequently, it is reasonable to suspect that the degradation in the spectral structure that accompanies processing by CIs could interfere with children's abilities to discover optimal weighting strategies. In the current experiment, that suspicion was explored.

Earlier work has shown that adults and children extend similar proportions of their perceptual attention to the spectral cue examined in the current study when that cue is not degraded. Thus, any age-related differences found in the weighting of this cue (i.e., FRT) when it is degraded could be attributed to age-related differences in the magnitude of effect of that degradation. To explore these possibilities, three goals were addressed in the current study: First, perceptual weighting factors for the spectral cue (FRT) and another cue that is not spectral in nature (ART) were measured. The first cue would be expected to be diminished by the signal processing employed; the second would not. Second, sensitivity to these cues was measured, using a discrimination task. Finally, word recognition for spectrally degraded signals was measured, to see if the weighting factors derived could explain this word recognition.

Outcomes of the current study showed that indeed spectral degradation had a greater effect on weighting strategies for children than for adults. However, that age-related difference could not be fully explained by sensitivity to the relevant structure. Finally, the amount of perceptual weight assigned to the spectral cue explained significant amounts of variability in word recognition.

In summary, it is clear that listeners of any age can be affected by the signal degradation inherent in CI processing, but children disproportionately more so. The finding that the magnitude of this effect cannot be explained by listeners' sensitivity to the relevant cue means that auditory training could potentially ameliorate the negative effect observed. The finding that cue weighting influences word recognition to even a small extent suggests that putting forth effort to develop appropriate training strategies would be a worthwhile endeavor.

## Acknowledgments

This work was supported by National Institute on Deafness and Other Communication Disorders Grant R01 DC000633, awarded to Susan Nittrouer. We thank Taylor Wucinich for help testing listeners and scoring.

## References

- Aslin, R. N., & Pisoni, D. B. (1980). Some developmental processes in speech perception. In G. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child phonology: Volume 2, perception* (pp. 67–96). New York, NY: Academic Press.
- Beddor, P. S., & Strange, W. (1982). Cross-language study of perception of the oral–nasal distinction. *The Journal of the Acoustical Society of America*, *71*, 1551–1561.
- Bourland Hicks, C., & Ohde, R. N. (2005). Developmental role of static, dynamic, and contextual cues in speech perception. *Journal of Speech, Language, and Hearing Research*, *48*, 960–974.
- Carpenter, A. L., & Shahin, A. J. (2013). Development of the N1–P2 auditory evoked response to amplitude rise time and rate of formant transition of speech sounds. *Neuroscience Letters*, *544*, 56–61.
- Cho, T., & Ladefoged, P. (1999). Variations and universals in VOT: Evidence from 18 languages. *Journal of Phonetics*, *27*, 207–229.
- Crowther, C. S., & Mann, V. (1992). Native language factors affecting use of vocalic cues to final consonant voicing in English. *The Journal of the Acoustical Society of America*, *92*, 711–722.
- Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971, January 22). Speech perception in infants. *Science*, *171*, 303–306.
- Escudero, P., Benders, T., & Lipski, S. C. (2009). Native, non-native and L2 perceptual cue weighting for Dutch vowels: The case of Dutch, German and Spanish listeners. *Journal of Phonetics*, *37*, 452–465.
- Giezen, M. R., Escudero, P., & Baker, A. (2010). Use of acoustic cues by children with cochlear implants. *Journal of Speech, Language, and Hearing Research*, *53*, 1440–1457.
- Goldman, R., & Fristoe, M. (2000). *Goldman-Fristoe Test of Articulation—Second Edition*. Circle Pines, MN: American Guidance Service.
- Gottfried, T. L., & Beddor, P. S. (1988). Perception of temporal and spectral information in French vowels. *Language and Speech*, *31*, 57–75.
- Greenlee, M. (1980). Learning the phonetic cues to the voiced–voiceless distinction: A comparison of child and adult speech perception. *Journal of Child Language*, *7*, 459–468.
- Greenwood, D. D. (1990). A cochlear frequency-position function for several species—29 years later. *The Journal of the Acoustical Society of America*, *87*, 2592–2605.
- Harris, K. S. (1958). Cues for the discrimination of American English fricatives in spoken syllables. *Language and Speech*, *1*, 1–7.
- Hirsh, I. J., Davis, H., Silverman, S. R., Reynolds, E. G., Eldert, E., & Benson, R. W. (1952). Development of materials for speech audiometry. *Journal of Speech and Hearing Disorders*, *17*, 321–337.
- Holt, R. F., & Carney, A. E. (2005). Multiple looks in speech sound discrimination in adults. *Journal of Speech, Language, and Hearing Research*, *48*, 922–943.
- Hoonhorst, I., Colin, C., Markessis, E., Radeau, M., Deltenre, P., & Serniclaes, W. (2009). French native speakers in the making: From language-general to language-specific voicing boundaries. *Journal of Experimental Child Psychology*, *104*, 353–366.
- Jenkins, J. J., Strange, W., & Edman, T. R. (1983). Identification of vowels in “vowelless” syllables. *Perception & Psychophysics*, *34*, 441–450.
- Kuhl, P. K. (1987). Perception of speech and sound in early infancy. In P. Salapatek & L. Cohen (Eds.), *Handbook of infant perception: Vol. 2. From perception to cognition* (pp. 275–382). New York, NY: Academic Press.
- Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user’s guide* (2nd ed.). Mahwah, NJ: Erlbaum.
- Martin, N. A., & Brownell, R. (2011). *Expressive One-Word Picture Vocabulary Test—Fourth Edition (EOWPVT-4)*. Novato, CA: Academic Therapy.
- Maurer, D., & Werker, J. F. (2014). Perceptual narrowing during infancy: A comparison of language and faces. *Developmental Psychobiology*, *56*, 154–178.
- Mayo, C., Scobbie, J. M., Hewlett, N., & Waters, D. (2003). The influence of phonemic awareness development on acoustic cue weighting strategies in children’s speech perception. *Journal of Speech, Language, and Hearing Research*, *46*, 1184–1196.
- Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A. M., Jenkins, J. J., & Fujimura, O. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception & Psychophysics*, *18*, 331–340.
- Moberly, A. C., Lowenstein, J. H., Tarr, E., Caldwell-Tarr, A., Welling, D. B., Shahin, A. J., & Nittrouer, S. (2014). Do adults with cochlear implants rely on different acoustic cues for phoneme perception than adults with normal hearing? *Journal of Speech, Language, and Hearing Research*, *57*, 566–582.
- Nittrouer, S. (2002). Learning to perceive speech: How fricative perception changes, and how it stays the same. *The Journal of the Acoustical Society of America*, *112*, 711–719.
- Nittrouer, S. (2007). Dynamic spectral structure specifies vowels for children and adults. *The Journal of the Acoustical Society of America*, *122*, 2328–2339.
- Nittrouer, S., & Crowther, C. S. (1998). Examining the role of auditory sensitivity in the developmental weighting shift. *Journal of Speech, Language, and Hearing Research*, *41*, 809–818.
- Nittrouer, S., & Lowenstein, J. H. (2010). Learning to perceptually organize speech signals in native fashion. *The Journal of the Acoustical Society of America*, *127*, 1624–1635.
- Nittrouer, S., Lowenstein, J. H., & Packer, R. R. (2009). Children discover the spectral skeletons in their native language before the amplitude envelopes. *Journal of Experimental Psychology: Human Perception and Performance*, *35*, 1245–1253.
- Nittrouer, S., Lowenstein, J. H., & Tarr, E. (2013). Amplitude rise time does not cue the /ba/–/wa/ contrast for adults or children. *Journal of Speech, Language, and Hearing Research*, *56*, 427–440.
- Nittrouer, S., & Studdert-Kennedy, M. (1986). The stop–glide distinction: Acoustic analysis and perceptual effect of variation in syllable amplitude envelope for initial /b/ and /w/. *The Journal of the Acoustical Society of America*, *80*, 1026–1029.
- Nittrouer, S., & Studdert-Kennedy, M. (1987). The role of coarticulatory effects in the perception of fricatives by children and adults. *Journal of Speech and Hearing Research*, *30*, 319–329.
- Nittrouer, S., Tarr, E., Bolster, V., Caldwell-Tarr, A., Moberly, A. C., & Lowenstein, J. H. (2014). Low-frequency signals support perceptual organization of implant-simulated speech for adults and children. *International Journal of Audiology*, *53*, 270–284.
- Ohde, R. N., & German, S. R. (2011). Formant onsets and formant transitions as developmental cues to vowel perception. *The Journal of the Acoustical Society of America*, *130*, 1628–1642.

- 
- Strange, W., Jenkins, J. J., & Johnson, T. L.** (1983). Dynamic specification of coarticulated vowels. *The Journal of the Acoustical Society of America*, *74*, 695–705.
- Walsh, M. A., & Diehl, R. L.** (1991). Formant transition duration and amplitude rise time as cues to the stop/glide distinction. *The Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *43*(A), 603–620.
- Wardrip-Fruin, C., & Peach, S.** (1984). Developmental aspects of the perception of acoustic cues in determining the voicing feature of final stop consonants. *Language and Speech*, *27*, 367–379.
- Werker, J. F., & Tees, R. C.** (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior & Development*, *7*, 49–63.
- Wilkinson, G. S., & Robertson, G. J.** (2006). *The Wide Range Achievement Test 4*. Lutz, FL: Psychological Assessment Resources.