Article

Amplitude Rise Time Does Not Cue the /bɑ/–/wɑ/ Contrast for Adults or Children

Susan Nittrouer,[a] Joanna H. Lowenstein,[a] and Eric Tarr[a]

Purpose: Previous research has demonstrated that children weight the acoustic cues to many phonemic decisions differently than do adults and gradually shift those strategies as they gain language experience. However, that research has focused on spectral and duration cues rather than on amplitude cues. In the current study, the authors examined *amplitude rise time* (ART; an amplitude cue) and *formant rise time* (FRT; a spectral cue) in the /bɑ/–/wɑ/ manner contrast for adults and children, and related those speech decisions to outcomes of nonspeech discrimination tasks.
Method: Twenty adults and 30 children (ages 4–5 years) labeled natural and synthetic speech stimuli manipulated to vary ARTs and FRTs, and discriminated nonspeech analogs that varied only by ART in an AX paradigm.

Results: Three primary results were obtained. First, listeners in both age groups based speech labeling judgments on FRT, not on ART. Second, the fundamental frequency of the natural speech samples did not influence labeling judgments. Third, discrimination performance for the nonspeech stimuli did not predict how listeners would perform with the speech stimuli.
Conclusion: Even though both adults and children are sensitive to ART, it was not weighted in phonemic judgments by these typical listeners.

Key Words: speech perception, children, adults, amplitude rise time

W hen a person speaks, the actions of the larynx and vocal tract shape all forms of structure in the resulting speech signal: spectral, temporal, and amplitude. It would seem that all those forms of structure would contribute equally to phonemic decisions, but decades of speech perception research have shown otherwise. Listeners attend to, or *weight*, the various kinds of structure in the signal differently depending on many factors, including what their first language experience was; what the decision is that needs to be made; and what the other phonetic, syntactic, and lexical properties of the signal are.

Children's Weighting Strategies

Another factor that affects how acoustic structure in the speech signal gets weighted is the age of the listener. For example, it has been found that young children (ages 3–7 years) weight formant transitions more than adults when making judgments about the sibilants /s/ and /ʃ/ in the syllable-initial position but weight static fricative noises

less than adults in these same decisions (Mayo, Scobbie, Hewlett, & Waters, 2003; Nittrouer, 1992; Nittrouer & Miller, 1997; Nittrouer & Studdert-Kennedy, 1987; Siren & Wilcox, 1995). Complementary to those results, research also has found that children weight formant transitions more in decisions concerning the voicing of syllable-final stops, but they weight the duration of the vocalic syllable portion less than adults (Greenlee, 1980; Krause, 1982; Nittrouer, 2004; Wardrip-Fruin & Peach, 1984). Finally, children rely on formant transitions to the same extent as adults in decisions regarding place of articulation for syllable-initial, voiceless stops but are poorer at using the burst or aspiration noise in those decisions (Parnell & Amerman, 1978).
In this case, formant transitions are weighted strongly by adults in place decisions regarding stop consonants; thus, there likely is no way for that weighting to be any stronger. In a similar manner, adults weight formant transitions strongly in decisions regarding place for weak fricatives (Harris, 1958), and it has been found that children weight formant transitions in those decisions to a similar extent as adults (Nittrouer, 2002). Taken together, the evidence across these studies reveals that children generally pay more attention to the time-varying spectral structure of formant transitions than do adults and pay less attention to durational cues and static spectral structure. In other words, children's weighting strategies differ from those of adults. These strategies change for children—at least, for those children who are learning language typically—as they gain experience with language, a trend that has been termed

[a]The Ohio State University, Columbus

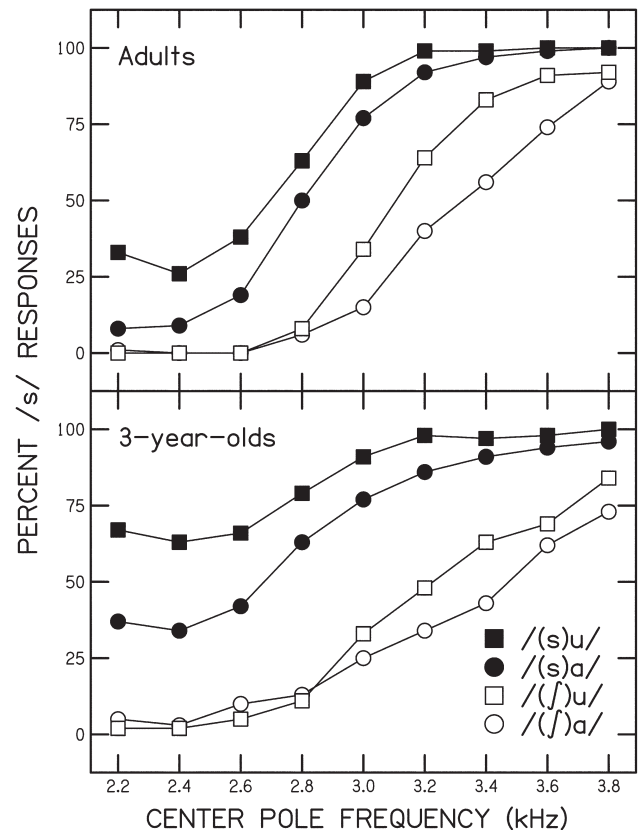Correspondence to Susan Nittrouer: nittrouer.1@osu.edu

the *developmental weighting shift* (Nittrouer, Manning, & Meyer, 1993).

Researchers have collected evidence supporting the developmental weighting shift largely employing the methods used in studies of categorical perception, but with two acoustic cues varying across stimuli instead of just one. In this type of study, one cue is set to vary in equal-sized acoustic steps along a continuum. In sibilant labeling experiments, for example, researchers have used single-pole noises to model the sibilants, and the center frequencies of those noises have spanned a range from one appropriate for an /ʃ/ noise to one appropriate for an /s/ noise. Each of those noises was then combined with each of two vocalic segments: one with formant transitions appropriate for a syllable-initial /ʃ/ and one with transitions appropriate for a syllable-initial /s/. All stimuli resulting from those combinations were presented to listeners multiple times for labeling, and the proportion of one labeling category was plotted as a function of noise frequency, separately for each vocalic segment. Outcomes for one such experiment (Nittrouer, 1992) are shown in Figure 1, for adults (top panel) and 3-year-olds (bottom panel). Estimates of the perceptual weighting of the sibilant noises and formant transitions can be gathered from the shapes of the functions and the separation between them: The steeper the functions, the more weight that was assigned to the sibilant noises. The greater the separation between functions, the more weight that was assigned to the formant transitions. One can infer from Figure 1 that children paid less attention to the fricative noises than did adults (the labeling functions are shallower) and paid more attention to formant transitions than did adults (the functions are more separated). This method of collecting data was used in the present study.

Weighting Strategies or Auditory Sensitivity

In spite of the robust evidence supporting the suggestion that children weight acoustic cues to phonemic contrasts differently than do adults, the following question can be asked: Are children simply less sensitive to some acoustic properties in the speech signal than are adults? Perhaps children's weighting strategies actually reflect their underlying auditory capacities; perhaps children make linguistic decisions on the basis of what acoustic structure is most salient to them. That specific question has already been addressed in two studies that tried to relate children's labeling of speech stimuli to their discrimination thresholds for nonspeech analogs of those speech stimuli. One experiment compared outcomes for adults and for 3-year-olds (Nittrouer, 1996), and the other compared outcomes for adults and for 5- and 7-year-olds (Nittrouer & Crowther, 1998). Similar results were reported in both cases: Slightly greater acoustic differences were needed between two stimuli in order for children to report that those stimuli were different, but those outcomes could not explain labeling results. One reason was that children's difference limens were greater than those of adults for all acoustic properties examined, even the spectral glides used to model the formant transitions that children so strongly weight in speech

perception. In addition, children's difference limens were sufficiently small for each property that it seemed unlikely that sensitivity could be a factor in children's weighting strategies. For example, 5- and 7-year-olds in Nittrouer and Crowther's (1998) experiment had a mean difference limen of 83 Hz, which was larger than the 39 Hz found for adults. However, in the sibilant labeling experiments, step size for the sibilant noises was 200 Hz, so children's sensitivity to the frequency of those noises was considered adequate to dismiss a lack of sensitivity as a possible explanation for their weaker weighting of the noises. Overall, no correspondence was found between the specific properties for which children had diminished sensitivity (compared to adults) and their weighting strategies.

A similar lack of correspondence between auditory sensitivity to acoustic properties and processing of speech signals has been observed for populations other than typically developing children. For example, in several studies,

researchers have investigated potential relationships between sensitivity to acoustic properties and recognition of phonemic segments related to those properties for children with dyslexia (e.g., Bishop, Carlyon, Deeks, & Bishop, 1999; Mody, Studdert-Kennedy, & Brady, 1997; Nittrouer, 1999). None of these studies found a clear connection. A particularly elegant demonstration of this kind of finding was provided by Rosen and Manganari (2001). Those investigators expanded on a report by Wright and colleagues (1997), which showed that children with dyslexia demonstrated enhanced backward masking compared with typically reading peers. Using that report as a starting point, Rosen and Manganari were able to replicate the finding of greater backward masking for nonspeech stimuli in teenagers with dyslexia, compared with typically reading peers. However, no evidence was found that the enhanced backward masking deleteriously affected recognition of syllable-initial phonemes, as would have been expected if that auditory deficit accounted for the phonological problems of children with dyslexia. Thus, evidence of poor performance on a psychoacoustic task using nonspeech stimuli could not predict recognition of a phonemic contrast related to that task.

Another line of investigation has consistently failed to find evidence of a link between auditory sensitivity to acoustic properties in the speech signal and the processing of the speech signal itself. Empirical evidence from second-language learners shows that listeners may be sensitive to change in an acoustic property but nonetheless fail to base phonemic decisions on that property if linguistic experience has not promoted the use of that particular property. For example, Miyawaki et al. (1975) examined sensitivity to third formant glides, which on their own are not heard as speechlike, and labeling of stimuli along a /ɹɑ/-to-/lɑ/ continuum, which is based on those third formant glides. Three groups of listeners participated: (a) native English speakers, (b) native Japanese speakers who were experienced second-language users of English, and (c) native Japanese speakers who were inexperienced second-language users of English. Japanese does not have a /ɹɑ/–/lɑ/ distinction. The results showed that the Japanese listeners were just as sensitive to the third formant glides as the English speakers, but they did not pay attention to that property when it was integrated into speech signals to the extent that native English speakers did. Consequently, they were unable to label the speech stimuli appropriately. Thus, a disassociation was again demonstrated between auditory sensitivity for an acoustic property and how that property is weighted in speech perception. Some effect of English experience was found, however, suggesting that adults can modify their weighting strategies for speech.

In the current experiment, sensitivity to the acoustic properties manipulated in the speech stimuli was explored to determine whether it can explain patterns of perceptual weighting for those speech stimuli. Although findings regarding how well listeners can discriminate acoustic cues in nonspeech signals generally cannot predict how well those listeners integrate those cues into phonetic decisions, it still must be the case that sensitivity to a cue is a basic requisite for being able to use that cue. As in earlier studies that addressed this question, nonspeech stimuli were used in the current study because it is extremely difficult to assess sensitivity to acoustic cues in speech signals, if speech signals are used. It is a quintessential characteristic of speech perception that once a cue is integrated into a speechlike percept, listeners can no longer isolate that cue for the purpose of judging its auditory qualities (see, e.g., Mann & Liberman, 1983; Remez, Pardo, Piorkowski, & Rubin, 2001).

The Case for Examining the Weighting of Amplitude Cues

The work on developmental shifts in weighting strategies for speech have thus far examined only temporal and spectral cues, but there is good reason to want to explore potential shifts in the weighting of amplitude cues, as well. Cochlear implants provide a veridical representation of amplitude structure in the speech signal, but degraded spectral structure. Thus, it would be useful to examine whether listeners—both adults and children—attend to amplitude structure when it cues a phonemic contrast. In addition, the possibility has been raised that children with reading problems might be poorer at recovering amplitude structure from the speech signal than their normal-hearing peers (Goswami, Fosker, Huss, Mead, & Szucs, 2011; Goswami et al., 2002), sparking further interest in how strongly typically developing children weight amplitude structure in their phonemic decisions. For these reasons, we selected for study a phonological contrast that involves a distinction in the rate of amplitude change—namely, the stop–glide manner contrast of /bɑ/ versus /wɑ/.

For both of these syllables, the vocal tract is initially closed (or tightly constricted) at the lips. As the syllable progresses, that lip constriction need only be opened for the syllable to be produced. Rate of change in lip opening is responsible for the acoustic differences between stops and glides, with two properties primarily affected: (a) the rate of formant change and (b) the rate of amplitude change at syllable onset. The first and second formants increase in frequency as the vocal tract is opened, and the overall amplitude of the signal increases. The spectral cue involving formant transitions is different in this case from that examined in earlier studies of developmental shifts in cue weighting because it does not involve variation in direction of change; instead, it involves variation only in rate of change. In earlier studies, the starting or ending frequencies of formant transitions, as well as direction of change, varied across stimuli as a function of consonant place. For /bɑ/ and /wɑ/, formant frequencies are similar at syllable onset and steady-state regions, but the time required to move from one to the other differs. Consequently, the nature of the cue is different than in earlier studies, so it is not possible to predict beforehand whether children will weight this cue more, similarly, or less than adults do.

Several studies have already examined adults' weighting strategies for this manner contrast. For example, Nittrouer and Studdert-Kennedy (1986) manipulated the

amplitude rise time (ART) of natural /bɑ/ and /wɑ/ syllables such that tokens of each were given the structure of the other while preserving original formant structures. The results showed that adults based their phonemic decisions almost entirely on the rate of formant change; the effect of ART was small. However, there was one confound in using natural speech samples to examine weighting strategies for these syllables that may have influenced outcomes, and that was fundamental frequency (f0): It also differs at syllable onset for these tokens. Because the vocal tract is completely closed prior to the start of a /bɑ/ syllable, subglottal pressure is greater at voicing onset in these syllables than in /wɑ/ syllables, where air continues to flow through the narrow lip constriction. As a result, f0 is higher in frequency at constriction release in /bɑ/ than in /wɑ/. That factor was not controlled in Nittrouer and Studdert-Kennedy's experiment, but it was in a subsequent experiment. Walsh and Diehl (1991) used synthetic speech tokens in which f0 remained consistent and replicated Nittrouer and Studdert-Kennedy's findings, thus minimizing the probability that the outcomes of that earlier study were due to differences in f0. Nonetheless, in the current experiment, we examined potential effects of f0 by using both natural and synthetic stimuli. In addition, in the current experiment, we made use of new methods of switching gross amplitude envelopes among natural tokens and of imposing envelope structure on synthetic syllables in order to ensure that these envelopes were what they were described to be.

The Current Experiment

In all, the current experiment had three objectives. The first objective was to examine how adults and children weight ART and rate of formant transitions (hereafter referred to as *formant rise time* [FRT]) in this stop-manner contrast. To meet this objective, we selected adults and children as participants. The second objective of this study was to determine whether f0 influences these decisions. We accomplished this objective by including two kinds of speech stimuli: (a) natural tokens that retained natural variability in f0 and (b) synthetic tokens in which f0 was held constant across stimuli. The third objective of this study was to measure sensitivity to ART for nonspeech stimuli for the same listeners from whom the phonemic labeling results were obtained in order to evaluate the extent to which auditory sensitivity to this property accounts for the extent to which it is weighted in the stop–glide decision.

Method

Listeners

Participants were 20 adults between 18 and 40 years of age and 30 children ranging in age from 4;3 (years;months) to 5;11. The mean age of children was 5;2 (*SD* = 0;7). None of the listeners (or, in the case of children, their parents) 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 25 dB HL to each ear separately. Parents of the 4- and 5-year-olds 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 (Goldman & Fristoe, 2000) and were required to score at or better than the 30th percentile for their age in order to participate. Children's scores ranged from the 30th percentile to greater than the 87th percentile (*M* = 59, *SD* = 18). Adult participants were given the Reading subtest of the Wide Range Achievement Test (4th edition; Wilkinson & Robertson, 2006), and all demonstrated better than a 12th grade reading level.

Equipment and Materials

We recorded stimuli using a Shure KSM studio microphone, a Tube MPStudio V3 amplifier, and an Echo Gina 3G digital audio converter, using Adobe Audition software. All testing took place in a soundproof booth, with the computer that controlled stimulus presentation being located 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 Soundblaster 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.

For the labeling tasks, two drawings (on 8″ × 8″ cards) were used to represent each response label: for /bɑ/, a picture of a baby, and for /wɑ/, a picture of the ocean (water). When introducing these pictures, the experimenter explained to the listener that the pictures were being used to represent the response labels because babies babble by saying /bɑ/–/bɑ/ and babies call water /wɑ/.

For the discrimination tasks, a 4″ × 14″ cardboard response card 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, which represented the ''same'' response choice; on the other half were one black square and one red circle, which represented the ''different'' response choice. Ten other cardboard cards (4″ × 14″, 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. We used these cards to ensure that all children knew the concepts of ''same'' and ''different.''

A game board with 10 steps was also used with children. The children 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 created: for labeling tasks, one set of natural speech stimuli and two sets of synthetic speech stimuli; for discrimination tasks, two sets of nonspeech stimuli.

Natural speech stimuli. Three types of stimuli were created from recordings of /bɑ/ and /wɑ/ syllables: (a) syllables with original, unprocessed temporal envelopes; (b) syllables with temporal envelopes transposed within the same category of syllable (/bɑ/ or /wɑ/); and (c) syllables with temporal envelopes switched between the two types of syllable (/bɑ/ and /wɑ/).

Recording and measurements. A male speaker was recorded producing 10 tokens each of /bɑ/ and /wɑ/ in random order, and these tokens were digitized at a 44.1-kHz sampling rate with 16-bit resolution directly onto a hard drive. Five tokens of each were selected, matching duration as closely as possible. Acoustic measurements were made of each token, using TF32 software (Milenkovic, 2004). For these measurements, the vowel was defined as the whole vocalic portion of the syllable, from the release of closure for /bɑ/ or the release of constriction for /wɑ/. Six measurements were made:

1. F0 was measured for the first three pitch periods after vowel onset.

2. F1, F2, and F3 were measured for the first two pitch periods after vowel onset.

3. F1, F2, and F3 were measured at the start of vowel steady state, defined as the point where F2 no longer rose more than 10 Hz from one pitch period to the next, and was calculated using 26-pole linear predictive coding (LPC) analysis at each pitch period.

4. Vowel duration was measured from vowel onset to the end of the vowel, which was defined as the zero crossing where energy from the formants higher than F1 was fully attenuated.

5. FRT was measured from vowel onset to the start of the vowel steady state.

6. ART was measured, using the following procedure: The amplitude peak of the syllable was found and root-mean-square (RMS) amplitude was calculated over the five pitch periods with the amplitude peak as the center, using WavEd software (Neely & Peters, 1992). RMS was then computed for individual pitch periods preceding the amplitude peak. The first pitch period with an RMS value ≥ 80% of the peak value was labeled as the end of the amplitude rise. ART was calculated as the duration between vowel onset and the end of the amplitude rise. Table 1 lists the values for each of these measures.

Temporal envelope manipulation. Before the gross temporal envelope (GTE) was interchanged (either transposed or switched) between tokens, prevoicing was removed from the /w/ tokens and the burst was removed from the /b/ tokens to eliminate those as cues for the stop–glide distinction. These tokens are referred to as "unprocessed" in this article. There were three subsequent steps in the process of interchanging the GTE from a model token onto a target token in order to create transposed and switched stimuli.

First, the GTE was removed from the target token so that a new envelope could be applied without interaction from the envelope of that target. A process of pitch period normalization was used in which every individual sample in one pitch period was scaled by the same amount so the maximum amplitude peak across all pitch periods was uniform. To find the individual pitch periods of the target, autocorrelation functions with 40-ms windows starting at the beginning of each pitch period were used, in a three-step process, as follows: First, measuring from the start of one pitch period, the start of the next pitch period was estimated to be between 5.6 and 14.3 ms later, assuming f0 to be between 70 and 180 Hz, which is typical for a male speaker.

Table 1. Fundamental frequency (f0), onset, and steady-state formant frequencies (F1, F2, F3), vowel duration, formant rise time (FRT), and amplitude rise time (ART) for the natural, unprocessed /bɑ/ and /wɑ/ tokens.

Token	f0 (Hz)	Onset F1 (Hz)	Onset F2 (Hz)	Onset F3 (Hz)	Steady-state F1 (Hz)	Steady-state F2 (Hz)	Steady-state F3 (Hz)	Vowel duration (ms)	FRT (ms)	ART (ms)
/bɑ/ 1	126	493	944	2215	777	1162	2301	370	66	7.6
/bɑ/ 2	126	535	977	2187	778	1205	2344	372	66	0.0
/bɑ/ 3	115	522	949	2173	749	1134	2287	367	54	7.8
/bɑ/ 4	125	535	977	2215	792	1176	2301	370	40	23.8
/bɑ/ 5	126	479	935	2130	763	1134	2201	373	49	0.0
M	*123*	*513*	*956*	*2184*	*772*	*1162*	*2287*	*370*	*55*	*7.8*
SD	*5*	*26*	*19*	*35*	*16*	*30*	*53*	*2*	*11*	*9.7*
/wɑ/ 1	99	400	707	2514	749	1091	2358	375	78	39.4
/wɑ/ 2	102	394	664	2529	749	1119	2358	373	128	38.1
/wɑ/ 3	97	379	635	2600	735	1091	2429	381	123	40.2
/wɑ/ 4	98	379	635	2685	763	1105	2400	383	134	68.4
/wɑ/ 5	101	380	692	2500	749	1091	2386	375	94	38.6
M	*99*	*386*	*667*	*2566*	*749*	*1099*	*2386*	*377*	*111*	*44.9*
SD	*2*	*10*	*33*	*77*	*10*	*13*	*30*	*4*	*24*	*13.1*

Second, the peak of the autocorrelation function within that window was used to constrain further where the start of the next pitch period would be. Third, the nearest positive-going zero crossing was identified as the exact start. After every pitch period was identified, each was scaled separately.

The second step in the GTE-interchanging process was extracting the GTE from a second, or model, token. Because of variations of f0 and utterance length across tokens, there could be no temporal alignment of individual pitch periods in the separate model tokens. Therefore, the GTE from each model token was measured by half-wave, rectifying the original signal and low-pass filtering using a cutoff frequency of 20 Hz.

The third step in the GTE-interchanging process was overlaying the extracted GTE of the model onto the normalized target using a pointwise multiplication. Prior to the multiplication, the longer of the two signals was truncated at the end to match the length of the shorter signal. Because tokens had been selected to be similar in length, truncation involved only a pitch period or two, when necessary. After doing this, we measured the ART of the target with the new envelope using procedures already described to ensure that the envelopes of the transposed or switched tokens matched those of the tokens that served as models. For two tokens, slight deviations were found. In those cases, individual pitch periods were adjusted so the ART of the target matched the model ART precisely.

The transposed stimuli were created in round-robin fashion: /bɑ/ 1 was the model imposed on /bɑ/ 2 as the target, /bɑ/ 2 was the model imposed on /bɑ/ 3 as the target, and so forth. The same procedure was used with the /wɑ/ stimuli. The switched stimuli were created so that /bɑ/ 1 was the model and /wɑ/ 1 was the target and vice versa, /bɑ/ 2 was the model and /wɑ/ 2 was the target and vice versa, and so forth.

These manipulations resulted in a total of 30 stimuli of six types: Five unprocessed /bɑ/ tokens, five transposed /bɑ/ tokens (with envelopes from different /bɑ/ tokens), five switched /bɑ/ tokens (with /wɑ/ envelopes), five unprocessed /wɑ/ tokens, five transposed /wɑ/ tokens (with envelopes from different /wɑ/ tokens), and five switched /wɑ/ tokens (with /bɑ/ envelopes). During every block of testing, two tokens of each type were played, creating blocks of 12 stimuli. Ten blocks were presented. Each time the software selected a token of a particular type to play, it did so randomly, but it did not replace that token into the pool until each token in the pool had been played. This process resulted in each stimulus being played a total of four times during testing, so 20 responses were collected to each type. In all, 120 stimuli were presented (6 types × 5 tokens × 4 repetitions).

Synthetic speech stimuli, two sets. Onset and steady-state formant values averaged across the /bɑ/ and /wɑ/ tokens were used to determine parameter settings for a nine-step /bɑ/–/wɑ/ continuum using a Klatt synthesizer (Sensyn). The tokens were all 370 ms in duration, with an f0 of 100 Hz throughout. Starting and steady-state frequencies of the first two formants were the same for all stimuli, even though the time to reach steady-state frequencies varied. 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. FRT varied along a nine-step continuum from 30 ms to 110 ms, in 10-ms steps.
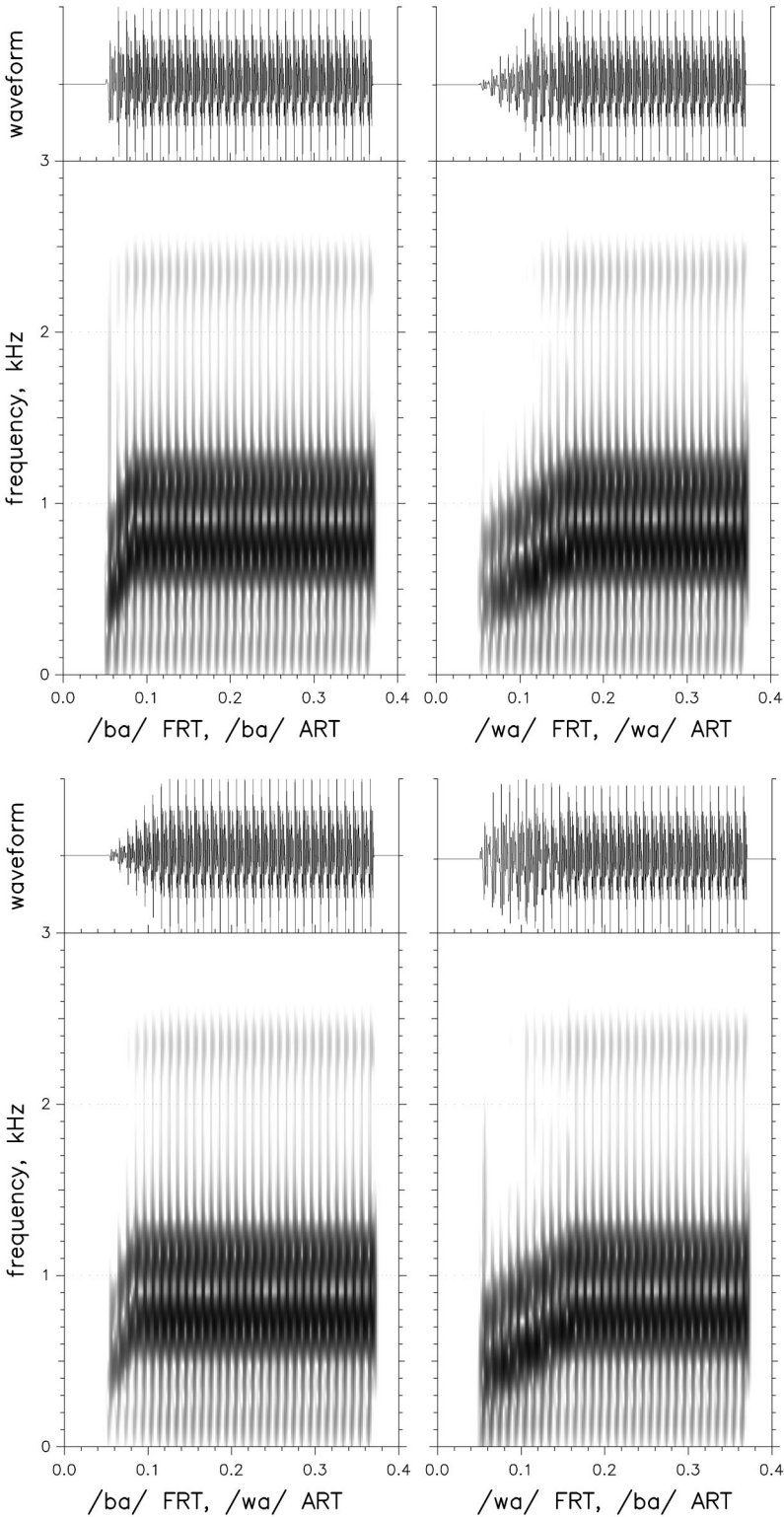
Controlling ART required that the signals be processed in MATLAB because there is no way to reliably control amplitude in Klatt-based speech synthesizers. The amplitude of voicing (AV) parameter is designed to simulate the amplitude of the voicing source before the signal is filtered by the vocal tract. However, that parameter interacts with filter parameters (i.e., those associated with formants) such that there is no direct correspondence between the AV setting and signal level at output. In this experiment, AV was set to a constant value of 60, and MATLAB was used to overlay ART after synthesis. The envelope created by MATLAB started at 0 dB and rose to the maximum amplitude at the end of the rise time, which varied along a seven-step continuum from 10 ms to 70 ms in 10-ms steps. Still one more adjustment needed to be made because Sensyn imposes a brief rise time of its own, even if AV is set to be constant across the stimulus. That unavoidable rise time would interact with the envelope created in MATLAB, if not corrected. Thus, an extra 10-ms frame was appended to the start of the file with identical parameters to those of the first frame, and then that frame was deleted from each file after stimuli were created.

The procedures described above were combined to make two sets of stimuli, one that varied in FRT and one that varied in ART. For the FRT stimulus set, the most /bɑ/-like ART (10 ms) and the most /wɑ/-like ART (70 ms) were applied to each stimulus along the nine-step /bɑ/–/wɑ/ FRT continuum, resulting in 18 FRT stimuli (9 FRTs × 2 ARTs). For the ART stimulus set, stimuli were created with the most /bɑ/-like FRT (30 ms) and most /wɑ/-like FRT (110 ms). Then the seven ARTs were applied to each stimulus, resulting in 14 ART stimuli (2 FRTs × 7 ARTs). Figure 2 shows synthetic stimuli for which FRT and ART signaled phonemic identity in a consistent manner (top panel) and stimuli for which they were set to signal phonemic identity in a contradictory manner. Inspection of the waveforms confirms that ART was implemented as described. During testing, stimuli were played 10 times each in blocks of however many stimuli there were, so that listeners heard a total of 180 FRT and 140 ART stimuli in two separate conditions.

Nonspeech stimuli for discrimination, two sets. For the discrimination task, two types of stimuli were created: (a) one set that was more speechlike in quality and (b) one set that was not speechlike at all, but nonetheless served as an analogue of the first set. As in Miyawaki et al.'s (1975) study, it was considered possible that listeners would be sensitive to ART when it is not heard as part of a speech signal, but fail to attend to it in making the phonemic judgment. Evidence for that position would be obtained if listeners recognized smaller differences in ART with the completely nonspeech signals than they did with the more speechlike signals.

The more speechlike set of stimuli was synthesized with Sensyn, using steady-state formants. The frequencies were 500 Hz for F1, 1000 Hz for F2, and 1500 Hz for F3. These

Figure 2. Spectrograms of synthetic /bɑ/ and /wɑ/ stimuli. The top panel displays stimuli in which FRT and ART are consistent in terms of the manner class signaled (/bɑ/ FRT and /bɑ/ ART; /wɑ/ FRT and /wɑ/ ART), and the bottom panel displays stimuli in which FRT and ART are contradictory in which manner class they signal (/bɑ/ FRT and /wɑ/ ART; /wɑ/ FRT and /bɑ/ ART). Time is represented on the *x*-axis, in seconds.

frequencies are typical values for modeling the resonances of a male vocal tract with a quarter wavelength resonator, although they do not represent any English vowel. The f0 was 100 Hz. Total duration was 370 ms. Onset amplitude envelopes (similar to the envelopes used for the other synthetic speech stimuli) were overlaid onto these signals. ART ranged from 0 to 250 ms in 25-ms steps, resulting in 11 stimuli. The other set of stimuli consisted of sine waves synthesized using TONE (Tice & Carrell, 1997), with the same duration, formant frequencies, and ARTs as the formant stimuli, resulting in 11 stimuli. The stimulus with the 0-ms ART was always the standard (A), and every stimulus (including the standard) was played as the comparison (X). During testing, each stimulus was compared to the standard 10 times, so that listeners heard a total of 110 formant stimuli and 110 sine wave stimuli.

Summary. Five sets of stimuli were presented in this experiment, in five separate tasks: one labeling task with natural speech stimuli, two labeling tasks with synthetic speech stimuli (FRT and ART), and two discrimination tasks (one with formant stimuli and one with sine wave stimuli).

Procedure

All procedures were approved by the The Ohio State University Institutional Review Board. Adults were tested in a single session of 45 min, and 4- and 5-year-olds were tested in two sessions of 45 min each over 2 days. The screening procedures (hearing screening and the Wide Range Achievement Test or Goldman Fristoe Test of Articulation) were administered first. The five test conditions were ordered so that one of the synthetic /bɑ/–/wɑ/ labeling tasks was first, followed by one of the discrimination tasks. The labeling task with the natural /bɑ/–/wɑ/ stimuli was always presented third, followed by the other discrimination task and, finally, the other synthetic /bɑ/–/wɑ/ labeling task. As a result, there were four possible orders of presentation. Adults completed all tasks in the single session, and children completed the first three tasks in the first session and the last two tasks in the second session.

Labeling tasks. For the /bɑ/–/wɑ/ labeling tasks, the experimenter introduced each picture separately and told the listener the name of the syllable associated with that picture. Ten live-voice practice trials were presented in which the listener pointed to the picture and named it after the experimenter said a syllable (five times for each syllable). Having listeners both point to the picture and say the syllable ensured that they were correctly associating the syllable and the picture. Then the listener heard the five unprocessed exemplars of /bɑ/ and /wɑ/ over headphones and was instructed to respond in the same way. Listeners were required to respond to nine of the 10 exemplars correctly to proceed to testing. For synthetic /bɑ/–/wɑ/ stimuli, training was also provided for the endpoints corresponding to the most /bɑ/-like (30 ms FRT, 10 ms ART) and /wɑ/-like (110 ms FRT, 70 ms ART) stimuli. Listeners heard five presentations of each endpoint and had to respond to nine out of the 10 correctly in order to proceed to testing.

Feedback was provided for up to two rounds of training, and then listeners were given up to two rounds without feedback to reach criterion. If listeners were not able to respond to nine of 10 endpoints correctly by the third round, testing for that condition was not done. During testing in each of the three stimulus conditions, listeners needed to respond with 80% or better accuracy to the endpoints in order to have their data included in the analysis.

Different dependent measures were analyzed for the natural and synthetic stimuli. For natural stimuli, the percentage of stimuli of each type given the label of the original (target) syllable served as the dependent variable. Arcsine transformations were used for statistical analysis because results were close to 100%. For synthetic stimuli, the plan was to use each listener's labeling responses to construct cumulative normal distributions of the proportion of /wɑ/ responses across each continuum. Lines could then be fit using probit analysis (Finney, 1971). From these probit functions, slopes and distribution means (i.e., phoneme boundaries) can be computed. These procedures were followed for the FRT continua, which showed typical distributions. However, responses to the ART continua did not show the usual cumulative normal distributions. Thus, instead of fitting probit functions, the percentage of /wɑ/ responses across each continuum served as the dependent variables.

Discrimination tasks. For the discrimination tasks, an AX procedure was used. In this procedure, listeners compare a stimulus, which varies across trials (X), with a constant standard (A). For both the sine wave and formant stimuli, the A stimulus was the one with a 0-ms ART. The interstimulus interval between standard and comparison was 450 ms. The listener responded by (a) pointing to the picture of the two black squares and saying "same" if he or she judged the stimuli as being the same or (b) pointing to the picture of the black square and the red circle and saying "different" if he or she judged the stimuli as being different. Both pointing and verbal responses were used because each served as a check on the reliability of the other.

Before any testing with the acoustic stimuli was done with children, they were shown the drawings of the six same and different objects and were asked to report whether 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 were asked to report whether 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 were 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 were asked to point to *same* and to *different*. 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. Next, these same training stimuli were presented, and listeners were asked to report whether they were the

same or different, only without feedback. Listeners needed to respond correctly to nine of the 10 training trials without feedback in order to proceed to testing. During testing in each of the two stimulus conditions, listeners 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 listener formed cumulative normal distributions, and probit functions were fit to these distributions. From these fitted functions, distribution means were calculated and served as difference thresholds. These thresholds were the 50% points on the discrimination functions.

Results

All 20 adults were able to meet the training and testing criteria to have their data included in the study for all five tasks. All 30 children met the inclusion criteria for the labeling task with natural syllables. For the two labeling tasks with synthetic syllables, 27 of the 30 children met the inclusion criteria for the FRT continua, and 29 of the 30 children met the inclusion criteria for the ART continua. The two discrimination tasks were more difficult for children: Only 12 of the 30 children met the inclusion criteria for the formant stimuli. Those 12 and an additional five children met the inclusion criteria for the sine wave stimuli. These

attrition rates suggest that young children have difficulty discerning acoustic stimuli that differ only on amplitude structure, and that is especially true if the stimuli are speechlike.

Labeling Task With Natural Stimuli

Figure 3 shows the percentage of original responses for adults and children for the unprocessed, transposed, and switched /bɑ/ and /wɑ/ stimuli. Adults and children responded to the unprocessed and transposed stimuli with the original response (/bɑ/ for /bɑ/ and /wɑ/ for /wɑ/) 99.33% to 100% of the time. The switched /wɑ/ stimuli (/wɑ/ with /bɑ/ envelopes) were also heard as /wɑ/ nearly all of the time: 100% for adults and 99% for children. Only one child was affected by the change in ART for the switched /wɑ/ stimuli, hearing 20% of them as /bɑ/. Switched /bɑ/ stimuli (/bɑ/ with /wɑ/ envelopes) showed many more /wɑ/ responses, especially from children. Two of the 20 adults (10%) and 13 of the 30 children (43%) responded to switched /bɑ/ with more than 10% /wɑ/ responses.

Table 2 shows results from two-way analyses of variance (ANOVAs), with age as the between-subjects factor and stimulus type (unprocessed, transposed, or switched) as the within-subject factor, for /bɑ/ and /wɑ/ stimuli separately. For stimuli created from the original /bɑ/ syllables, the main effects of stimulus type and age were statistically significant, as was the Stimulus Type × Age interaction.

Figure 3. Percentage of original responses for the natural unprocessed, transposed, and switched /bɑ/ and /wɑ/ stimuli. Error bars indicate standard errors.
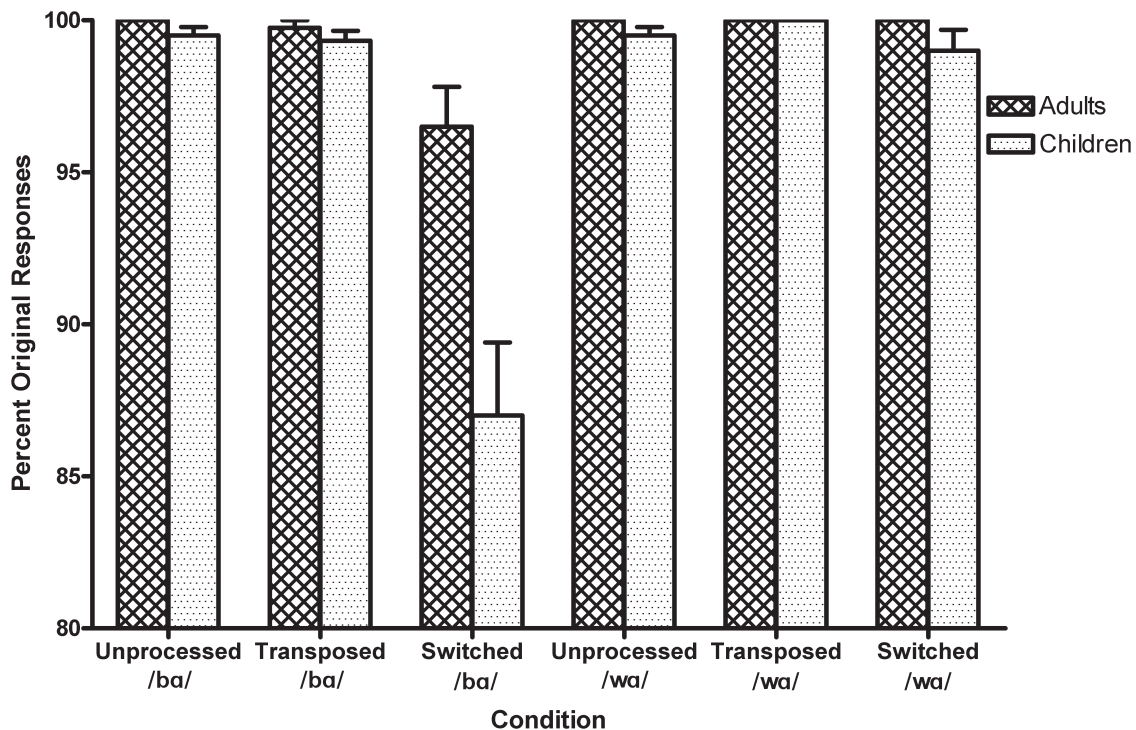
Table 2. Analysis of variance results for the natural unprocessed, transformed, and switched /bɑ/ and /wɑ/ stimuli.

Effect	df	F	p	Partial η²
Original /bɑ/				
Stimulus type	2, 96	30.28	< .001	.39
Age	1, 48	9.77	.003	.17
Stimulus Type × Age	2, 96	5.71	.005	.11
Original /wɑ/				
Stimulus type	2, 96	1.09	ns	
Age	1, 48	3.33	.074	.07
Stimulus Type × Age	2, 96	1.09	ns	

Note. Precise *p* values are shown if they are less than .10; *ns* (not significant) is shown for values greater than .10.
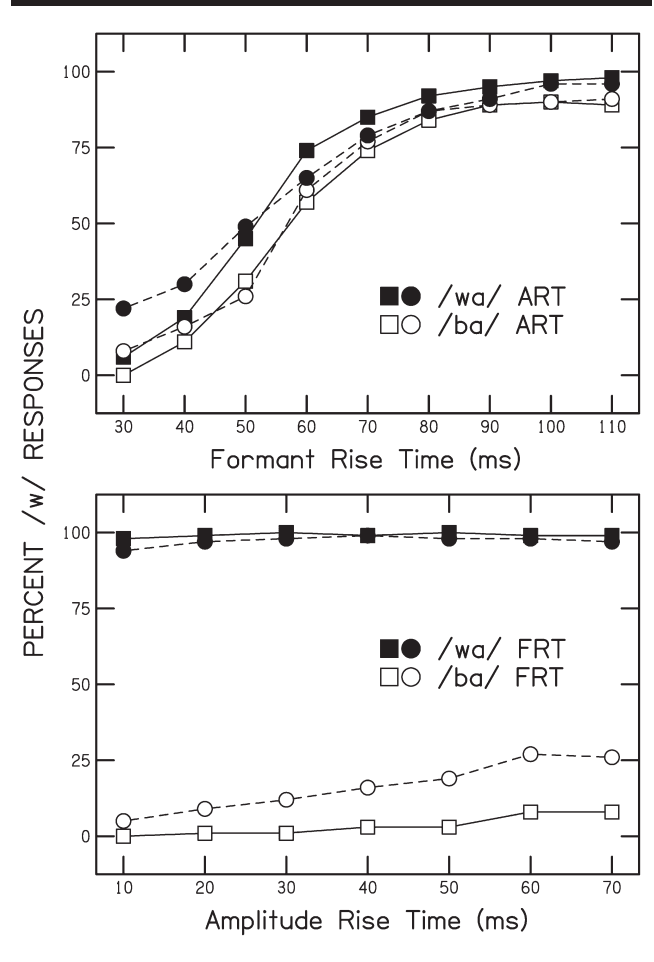
The largest amount of variance was explained by stimulus type ($\eta^2 = .39$). Because a significant interaction was found, we conducted one-way ANOVAs with age as the factor for each of the three stimulus types separately to locate the source of that interaction. Only switched /bɑ/ showed a significant effect, $F(1, 48) = 9.14$, $p = .004$. These statistical findings support the claim that the only natural stimuli that were not labeled entirely according to formant trajectories were the switched /bɑ/ syllables. Labeling of these stimuli was influenced slightly by ART, and that influence was greater for children than for adults. Stimuli created from the original /wɑ/ syllables were labeled entirely according to formant trajectories, by adults and children alike.

Labeling Task With Synthetic Stimuli

Figure 4 shows results for the FRT continua (top panel) and the ART continua (bottom panel) for adults (squares, solid lines) and children (circles, dashed lines). Looking first at results for the FRT continua, on the top panel, it appears that adults and children responded similarly. When these labeling functions are compared with those in Figure 1, which is reprinted from a sibilant-vowel study, responses for both groups resemble those of adults in that earlier study, where functions were steep. This pattern indicates that listeners responded largely on the basis of the cue manipulated along the continuum represented on the *x*-axis. In this case, that was FRT. However, functions from this experiment were less separated on the basis of the second cue (ART in the current experiment) than were those of the adults in the experiment depicted in Figure 1. This pattern indicates that the second cue in this experiment was not weighted strongly.

To examine results for this FRT continuum more closely, we performed two-way ANOVAs on the slopes and phoneme boundaries, with age and ART as main effects. For slopes, the main effects of age and ART were not significant, but the Age × ART interaction was, $F(1, 45) = 4.71$, $p = .035$, $\eta^2 = .10$. This outcome reflects the fact that adults' labeling functions had similar slopes across ARTs, and children's labeling function for the /wɑ/ ART had a slope similar to that of adults. However, children's labeling function for the /wɑ/ ART (filled circles) is slightly shallower, and that likely

Figure 4. Results for the FRT continua (top panel) and ART continua (bottom panel) for adults (squares, solid lines) and children (circles, dashed lines). Filled symbols indicate when either ART (top panel) or FRT (bottom panel) was set to be appropriate for /wɑ/, and open symbols indicate when either ART (top panel) or FRT (bottom panel) was set to be appropriate for /bɑ/.

accounts for the significant interaction. As with natural stimuli, children were slightly affected by the /wɑ/ ART, even when stimuli had clear /bɑ/ FRTs. For phoneme boundaries, the effect of ART was significant, $F(1, 45) = 17.74$, $p < .001$, $\eta^2 = .28$, but neither age nor the Age × ART interaction was significant. These outcomes mean that listeners responded differently to stimuli with /bɑ/ and /wɑ/ ART, but the effect was similar for children and adults. It was not large for either group.

The results for the ART continua (see bottom panel of Figure 4) are striking: Listeners appear to have assigned no weight at all to ART when the formants were /wɑ/-like (filled symbols). Adults also did not weight ART strongly when the formants were /bɑ/-like (open symbols), but children did, to a small degree. About 25% of the stimuli with /bɑ/ FRT and the longest ART were labeled as /wɑ/ by children. This matches the result found for natural tokens, where some children labeled some switched /bɑ/ syllables as /wɑ/, as well

as the result found for the FRT continuum with the /wɑ/ ART, where children's function was shallower than others because some stimuli at the /bɑ/ FRT endpoint were labeled as /wɑ/.
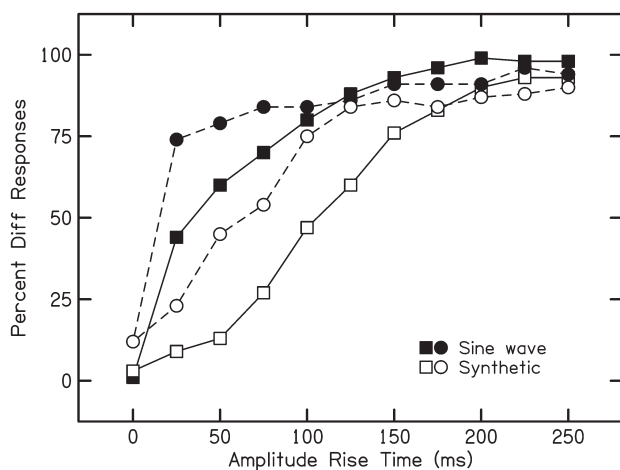
Because responses were all so close to 0% and 100% for these synthetic ART stimuli, it was not possible to compute probit functions. Instead, we computed the percentage of /wɑ/ responses across all steps on the ART continuum for the /bɑ/ and /wɑ/ FRT continua separately. We conducted a two-way ANOVA with age as the between-subjects factor and FRT as the within-subject factor. Arcsine transforms were used. The effect of age was significant, $F(1, 47) = 6.87$, $p = .012$, $\eta^2 = .13$, as was the effect of FRT, $F(1, 47) = 1,591.26$, $p < .001$, $\eta^2 = .97$. The Age × FRT interaction was also significant, $F(1, 47) = 25.93$, $p < .001$, $\eta^2 = .36$. These results confirm that listeners in both groups weighted FRT heavily in their phonemic decisions and that children showed some small weighting of ART, which adults did not do.

From these labeling results, it is clear that all listeners weighted FRT strongly in their stop-glide decisions. Neither adults nor children weighted ART strongly at all. Of course, the question arises as to whether listeners are sensitive to this acoustic structure or not.

Discrimination Tasks

Figure 5 shows discrimination functions for sine wave (filled symbols) and synthetic (open symbols) stimuli for adults (squares) and children (circles). It appears that both adults and children were sensitive to ART. One striking aspect of these results is that children appear to have been more sensitive to changes in ART than adults, but that impression needs to be tempered by the fact that many children were not able to meet criteria to participate at all, indicating they were not able to judge ART in these nonspeech stimuli. Thus, the children included in the data

Figure 5. Discrimination results for the sine wave (filled symbols) and synthetic (open symbols) nonspeech stimuli for adults (squares) and children (circles).

shown here represent only those children who could perform the task with these stimuli and, on average, they appear to have been slightly more sensitive to ART than were adults.

We performed a two-way ANOVA with age as the between-subjects factor and stimulus type as the within-subject factor on distribution means from adults and the 12 children who met criteria for participation with both types of stimuli. We found that the effect of age was significant, $F(1, 30) = 4.95$, $p = .034$, $\eta^2 = .14$, as was the effect of stimulus type, $F(1, 30) = 48.63$, $p < .001$, $\eta^2 = .62$. The Age × Stimulus type interaction was not significant. Thus, we can conclude that listeners were more sensitive to ART for sine wave stimuli than for formant stimuli, the more speechlike sounds. The children who met criteria for participating with these nonspeech stimuli were more sensitive to ART than were adults.

Finally, we compared labeling results for children who could and could not meet criteria to participate in the discrimination tasks with nonspeech stimuli to determine whether there were any differences in speech labeling between these groups. We conducted two-way ANOVAs on dependent measures for each set of speech stimuli, with group (children who could or could not meet criteria with the nonspeech stimuli) as the between-subjects factor and stimulus type as the within-subject factor. For no set of speech stimuli was either a significant effect of group or a Group × Stimulus type interaction found. Therefore, it may be concluded that children who were unable to perform the discrimination tasks with nonspeech stimuli performed indistinguishably from children who could perform those tasks on speech labeling. Mean scores for the Goldman Fristoe Test of Articulation were also submitted to statistical analysis, and no significant difference was observed.

Discussion

The purpose of this study was to examine the labeling of stimuli differing in manner class (stops vs. glides) by adults and children. This purpose was motivated by three factors: (a) a desire to extend work on developmental changes in perceptual weighting strategies for speech, (b) an interest in constructing a base from which the question can be explored of how listeners with cochlear implants might process the amplitude structure provided by those implants, and (c) suggestions that children with dyslexia might have difficulty processing amplitude structure in the speech signal (Goswami et al., 2002, 2011). Three specific objectives were addressed by the current study. The first objective was to examine whether adults and young children differ in terms of how they weight the acoustic cues to the stop–glide contrast. Outcomes revealed that adults and children alike based their decisions about whether syllables started with stops or glides almost entirely on the rate of formant change. That similarity across age groups matches earlier findings showing that when adults heavily weight formant transitions in phonemic decisions, children demonstrate the same strategies. In general, children weight formant transitions strongly in phonemic decisions, so when adults do the same, their

strategies match. However, earlier findings exploring age-related changes in weighting strategies for speech signals typically involved phonemic contrasts in which the direction of formant transitions differed across categories. The current experiment extends those findings by revealing that when the primary acoustic difference is rate of formant change, adults and children alike weight that property greatly.

One age-related difference was observed, however—and that happened when the cues conflicted, as was the case for switched /bɑ/ syllables. In that case, adults were able to attend to the formant cue only. Children, on the other hand, were slightly affected in their labeling decisions by ART. The suggestion has previously been made that children are obliged to integrate acoustic structure for speech signals, even when that structure provides conflicting information about phonemic identity (Nittrouer & Crowther, 2001). That suggestion may explain outcomes for this current study because the children were deleteriously affected when rate of formant transition and rate of amplitude change cued different phonemic decisions. Even at that, however, the effect was found only if the rate of formant change was rapid; it was not found for the switched /wɑ/ syllables, where the rate of formant change was gradual. In any event, responses of listeners in both age groups were overwhelmingly based on the rate of formant frequency change.

The second objective of the current study was to determine whether f0 in natural stimuli might be used by listeners in making decisions about this manner contrast. This objective was accomplished by using both natural, modified stimuli in which f0 retained its original value and synthetic stimuli that shared a consistent f0. Response patterns matched for the two types of stimuli, so no evidence was found to indicate that listeners give any weight at all to f0 in this particular phonemic decision.

The third objective was to determine whether sensitivity to the acoustic property of ART, as measured for nonspeech sounds, could explain the extent to which listeners weight that property in making the stop–glide distinction. A disassociation between sensitivity and weighting of acoustic properties has been observed in experiments with adults who are second-language learners and in experiments with children. In the current experiment, a similar disassociation was observed: Adults and children—at least, the children who were able to perform the task—showed mean discrimination thresholds of less than 50 ms, so they were highly sensitive to variation in ART, yet they did not use this acoustic property in their phonemic decisions. Thus, further evidence was found to support the position that measured sensitivity to acoustic properties reveals little about the role that those properties may play in phonemic decisions. Here, as elsewhere, evidence of separate organizational strategies for speech and nonspeech stimuli is seen. Both kinds of signals require some perceptual organization of sensory elements, but the extent to which that organization is explained by primitive versus schema-based principles varies for speech and nonspeech signals (Bregman, 1990; Remez, 2008). Furthermore, the specific schemas underlying perception differ for speech and nonspeech signals

(Nittrouer & Tarr, 2011; Remez, Rubin, Berns, Pardo, & Lang, 1994). As a result, the ability to make decisions about the auditory qualities of nonspeech signals predicts little about how the properties of those signals will be used in the perception of speech. In the current study, even more evidence of the disassociation between the perception of nonspeech and speech signals is provided by the fact that the children who could not perform the discrimination task with nonspeech signals showed the same outcomes in their labeling of speech stimuli as the children who could perform the discrimination task. Thus, it may be concluded that a demonstration that a specific property is used in speech perception predicts little about whether listeners will necessarily be able to judge its auditory qualities when it is presented in a nonspeech context.

In sum, the current experiment demonstrates that neither adults nor children use ART in making decisions about whether they heard /bɑ/ or /wɑ/. Instead, these listeners weighted FRT strongly in making this decision. Although formant transitions specify this manner distinction differently from the way formant transitions specify place of consonant constrictions (i.e., rate of change vs. direction/extent of change), it should perhaps not be surprising that adults and children weighted the cue similarly. In either case, formant transitions involve the relatively slowly modulating changes of vocal tract resonances. Evidence from experiments with sine wave speech has revealed that adults and children are comparable in their abilities to comprehend those signals, which replicate the slowly modulating changes of vocal tract resonances (Nittrouer & Lowenstein, 2010; Nittrouer, Lowenstein, & Packer, 2009). This evidence is complementary to the idea that it is precisely formant transitions that first capture children's attention when perceiving speech (e.g., Nittrouer, 2006). According to this account, children's perceptual attention gradually shifts away from being highly and almost singly focused on formant transitions in the speech signal and starts to incorporate other phonemically informative kinds of acoustic structure. Where children with dyslexia are concerned, several studies have demonstrated that they depend even more than age-matched control subjects on formant transitions in making phonemic judgments (Boada & Pennington, 2006; Johnson, Pennington, Lowenstein, & Nittrouer, 2011). That outcome is reflected in the results of Goswami et al. (2011), who found that the children with dyslexia placed their /bɑ/–/wɑ/ phoneme boundaries at briefer FRTs than did children who read typically. The common interpretation of this sort of finding is that children with dyslexia are delayed in shifting their perceptual weight away from formant transitions and toward other kinds of acoustic structure (see, e.g., Boada & Pennington, 2006; Nittrouer, 1999).

Of course, finding that samples of some listener populations fail to weight ART in making decisions regarding this manner contrast does not necessarily mean that all other populations of listeners will similarly disregard it. The current study demonstrates that adults and children with age-appropriate speech perception failed to attend to ART, and research by Goswami and colleagues (2011)

showed that children with dyslexia do not attend to this property either. However, there is one other population of listeners for whom the question needs to be explored of how well amplitude structure supports phonemic decisions: listeners with cochlear implants. The processing strategies of cochlear implants affect the quality of the signal properties delivered. Where formants are concerned, change over time is represented only when frequencies cross channels, so small changes in formant frequencies are missing. Because children typically depend so strongly on the time-varying spectral structure of the speech signal, it is important to determine how well they might be able to use amplitude structure instead when spectral structure is impoverished in this way. Cross-linguistic studies of weighting strategies have revealed that listeners depend on acoustic cues that are most relevant in their native language. For example, adults whose native language is English rely on the length of the vocalic portion before vocal tract closure in decisions regarding the voicing of final obstruents (Chen, 1970; Peterson & Lehiste, 1960; Raphael, 1972). However, listeners whose native language either does not include syllable-final obstruents or does not make a length distinction based on voicing fail to weight this acoustic property strongly (Crowther & Mann, 1992, 1994; Flege & Wang, 1989). Nonetheless, individuals can modify their weighting strategies as they gain experience with a new, second language (Miyawaki et al., 1975). Thus, the hypotheses could be posed that perhaps adults who receive cochlear implants modify existing weighting strategies once they get those implants, and perhaps children with implants develop strategies that involve weighting ART strongly. Future investigations will need to explore those hypotheses.

In summary, the current study was undertaken to reexamine the acoustic and perceptual bases of labeling decisions for stimuli differing according to a stop–glide manner distinction. The results showed that adults and children alike based these decisions on the rate of formant frequency change, and hardly at all on ART.

Acknowledgments

References

Bishop, D. V., Carlyon, R. P., Deeks, J. M., & Bishop, S. J. (1999). Auditory temporal processing impairment: Neither necessary nor sufficient for causing language impairment in children. *Journal of Speech, Language, and Hearing Research, 42,* 1295–1310.

Boada, R., & Pennington, B. F. (2006). Deficient implicit phonological representations in children with dyslexia. *Journal of Experimental Child Psychology, 95,* 153–193.

Bregman, A. S. (1990). *Auditory scene analysis.* Cambridge, MA: MIT Press.

Chen, M. (1970). Vowel length variation as a function of the voicing of the consonant environment. *Phonetica, 22,* 129–159.

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.

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

Finney, D. J. (1971). *Probit analysis.* Cambridge, United Kingdom: Cambridge University Press.

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

Goldman, R., & Fristoe, M. (2000). *Goldman Fristoe Test of Articulation—Second Edition.* Circle Pines, MN: AGS.

Goswami, U., Fosker, T., Huss, M., Mead, N., & Szucs, D. (2011). Rise time and formant transition duration in the discrimination of speech sounds: The ba–wa distinction in developmental dyslexia. *Developmental Science, 14,* 34–43.

Goswami, U., Thomson, J., Richardson, U., Stainthorp, R., Hughes, D., Rosen, S., & Scott, S. K. (2002). Amplitude envelope onsets and developmental dyslexia: A new hypothesis. *Proceedings of the National Academy of Sciences USA, 99,* 10911–10916.

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.

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

Johnson, E. P., Pennington, B. F., Lowenstein, J. H., & Nittrouer, S. (2011). Sensitivity to structure in the speech signal by children with speech sound disorder and reading disability. *Journal of Communication Disorders, 44,* 294–314.

Krause, S. E. (1982). Vowel duration as a perceptual cue to postvocalic consonant voicing in young children and adults. *The Journal of the Acoustical Society of America, 71,* 990–995.

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

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.

Milenkovic, P. (2004). TF32 [Computer software]. Madison, WI: University of Wisconsin—Madison.

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.

Mody, M., Studdert-Kennedy, M. G., & Brady, S. (1997). Speech perception in poor readers: Auditory processing or phonological coding? *Journal of Experimental Child Psychology, 64,* 199–231.

Neely, S. T., & Peters, J. E. (1992). *WavEd user's guide* (Technical Memo No. 15). Omaha, NE: Boys Town National Research Hospital.

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

Nittrouer, S. (1996). Discriminability and perceptual weighting of some acoustic cues to speech perception by three-year-olds. *Journal of Speech and Hearing Research, 39,* 278–297.

Nittrouer, S. (1999). Do temporal processing deficits cause phonological processing problems? *Journal of Speech, Language, and Hearing Research, 42,* 925–942.

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. (2004). The role of temporal and dynamic signal components in the perception of syllable-final stop voicing by children and adults. *The Journal of the Acoustical Society of America, 115,* 1777–1790.

Nittrouer, S. (2006). Children hear the forest. *The Journal of the Acoustical Society of America, 120,* 1799–1802.

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., & Crowther, C. S. (2001). Coherence in children's speech perception. *The Journal of the Acoustical Society of America, 110,* 2129–2140.

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. (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., Manning, C., & Meyer, G. (1993). The perceptual weighting of acoustic cues changes with linguistic experience. *The Journal of the Acoustical Society of America, 94,* 1865.

Nittrouer, S., & Miller, M. E. (1997). Predicting developmental shifts in perceptual weighting schemes. *The Journal of the Acoustical Society of America, 101,* 2253–2266.

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. (2011). Coherence masking protection for speech signals in children and adults. *Attention, Perception, & Psychophysics, 73,* 2606–2623.

Parnell, M. M., & Amerman, J. D. (1978). Maturational influences on perception of coarticulatory effects. *Journal of Speech and Hearing Research, 21,* 682–701.

Peterson, G. E., & Lehiste, I. (1960). Duration of syllable nuclei in English. *The Journal of the Acoustical Society of America, 32,* 693–703.

Raphael, L. J. (1972). Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English. *The Journal of the Acoustical Society of America, 51,* 1296–1303.

Remez, R. E. (2008). Perceptual organization of speech. In D. B. Pisoni & R. E. Remez (Eds.), *The handbook of speech perception* (pp. 28–50). Oxford, United Kingdom: Blackwell.

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

Remez, R. E., Rubin, P. E., Berns, S. M., Pardo, J. S., & Lang, J. M. (1994). On the perceptual organization of speech. *Psychological Review, 101,* 129–156.

Rosen, S., & Manganari, E. (2001). Is there a relationship between speech and nonspeech auditory processing in children with dyslexia? *Journal of Speech, Language, and Hearing Research, 44,* 720–736.

Siren, K. A., & Wilcox, K. A. (1995). Effects of lexical meaning and practiced productions on coarticulation in children's and adults' speech. *Journal of Speech and Hearing Research, 38,* 351–359.

Tice, B., & Carrell, T. (1997). TONE: Tone-analog waveform synthesizer [Computer software]. Lincoln, NE: University of Nebraska—Lincoln.

Walsh, M. A., & Diehl, R. L. (1991). Formant transition duration and amplitude rise time as cues to the stop/glide distinction. *Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 43,* 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.

Wilkinson, G. S., & Robertson, G. J. (2006). *The Wide Range Achievement Test* (4th ed.). Lutz, FL: Psychological Assessment Resources.

Wright, B. A., Lombardino, L. J., King, W. M., Puranik, C. S., Leonard, C. M., & Merzenich, M. M. (1997, May 8). Deficits in auditory temporal and spectral resolution in language-impaired children. *Nature, 387,* 176–178.