

# Formant onsets and formant transitions as developmental cues to vowel perception

Ralph N. Ohde<sup>a)</sup> and Sarah R. German

*Department of Hearing and Speech Sciences, Vanderbilt Bill Wilkerson Center for Otolaryngology and Communication Sciences, Vanderbilt University Medical Center, 1215 21st Avenue South, Room 8310, Nashville, Tennessee 37232-8242*

(Received 2 August 2010; revised 19 January 2011; accepted 8 May 2011)

The purpose of this study was to determine whether children give more perceptual weight than do adults to dynamic spectral cues versus static cues. Listeners were 10 children between the ages of 3;8 and 4;1 (mean 3;11) and ten adults between the ages of 23;10 and 32;0 (mean 25;11). Three experimental stimulus conditions were presented, with each containing stimuli of 30 ms duration. The first experimental condition consisted of unchanging formant onset frequencies ranging in value from frequencies for [i] to those for [a], appropriate for a bilabial stop consonant context. The second two experimental conditions consisted of either an [i] or [a] onset frequency with a 25 ms portion of a formant transition whose trajectory was toward one of a series of target frequencies ranging from those for [i] to those for [a]. Results indicated that the children attended differently than the adults on both the [a] and [i] formant onset frequency cue to identify the vowels. The adults gave more equal weight to the [i]-onset and [a]-onset dynamic cues as reflected in category boundaries than the children did. For the [i]-onset condition, children were not as confident compared to adults in vowel perception, as reflected in slope analyses. © 2011 Acoustical Society of America. [DOI: 10.1121/1.3596461]

PACS number(s): 43.71.An, 43.71.Es, 43.71.Ft [JES]

Pages: 1628–1642

## I. INTRODUCTION

### A. A theoretical overview of vowel perception

Early theories of vowel perception (Peterson and Barney, 1952) assumed that the acoustic identity of a (monophthongal) vowel was established by a static set of target formants, an ideal steady state or “canonical,” context-free version of the vowel. These target formants were used to explain the problem of “speaker normalization,” the fact that even infants and young children can discriminate or identify vowels spoken by different speakers of various ages and genders (Kuhl, 1979; Kuhl, 1983; Kubaska and Aslin, 1985). However, several facts have brought this theory into question. One is the problem of “segmentation.” Within a syllable, there may not be an identifiable representative steady-state set of formant frequencies, due to coarticulation, or influences from the surrounding phonetic environment (cf. Strange, 1989b, for review). More problematic is the fact that phonetically unsophisticated listeners can easily identify coarticulated vowels that do not contain formants that are identical to the vowel target. If the “simple target” model was correct, then more ambiguous coarticulated vowels should be less easily identified; however, this is not the case. Even in studies (Verbrugge *et al.*, 1976; Strange *et al.*, 1976) which did not replicate the above results (cf. Strange, 1989b), “in no study using natural (as opposed to synthetic) stimuli were isolated vowels identified more accurately than coarticulated vowels, as would be predicted if static targets were the primary source of information for vowel identity” (Strange, 1989b).

Evidently, vowel perception is more complex than simply identifying a specific target. The results obtained from studies (Jenkins *et al.*, 1983; Strange *et al.*, 1983; Strange, 1989a) in the 1980s for listeners identifying the edited waveforms allowed an evaluation of the acoustic properties used in vowel perception. By systematically deleting sections of the waveform, they found that perceptual cues to the vowel were located in the syllable nucleus (the “quasi-steady state” portion of the vowel); the intrinsic vowel durational differences; and the formant transitions both into and out of the vowel. For example, Strange and colleagues replaced most of the steady-state center of the vowel with either noise or silence creating a “silent center” vowel containing mostly the formant transitions as cues to vowel identity. Although slightly lower than for the entire unaltered syllable, accuracy of identification was often comparable for the formant transitions alone and when only the vowel center was provided (Nearey, 1989). In other words, no single cross section of the vowel is entirely representative of the “target” vowel; rather, the trajectory shape of formants and changes over the duration of the vowel, in relation to the phonetic context, are all intrinsic aspects of the identity of the vowel and are used as such by listeners. One confound of research on vowel transitions is that these acoustic properties also contain the vowel onset (within the first glottal pulse of transitions), which includes important acoustic cues of the vowel place feature (Sussman, 1990). In addition, studies have found that when the dynamic formant transitions including the formant onset are spliced onto the wrong end of the vowel, vowel identification is extremely poor (Kuwabara, 1985). This implies that identification of the vowel is contingent on the correct order of each component of the waveform.

<sup>a)</sup>Author to whom correspondence should be addressed. Electronic mail: ralph.n.ohde@vanderbilt.edu

Although recent research has established fairly well that there are important acoustic correlates to vowel identification other than simple target formant frequencies, Nearey (1989) points out that there are still questions left to answer. For example, although it is commonly accepted that the first formant (F1) and the second formant (F2) are “the primary determinants of vowel quality,” these frequencies often overlap between vowels, between speakers and throughout the speech of a single speaker. If the “simple target” theory does not hold, then another explanation needs to be found to explain speaker normalization (i.e., discriminating vowels regardless of age, gender, etc., of the speaker). Nearey (1989) suggests that a combination of “extrinsic” frame-of-reference information (e.g., vocal tract size) in conjunction with “intrinsic” vowel-internal information (e.g., F0 and formant ratios) may be used. F3 may also be used to identify a speaker’s vocal tract shape and size (Perry *et al.*, 2001).

## B. Development of vowel perception: Formant onsets and formant transitions

Developmental theories of vowel perception are complicated because children appear to attend to different acoustic correlates to a greater extent than adults. Although research on acoustic correlates of formant onsets and formant targets has established that both of these properties carry important information on place of articulation of vowels (Sussman, 1990), there is no developmental evidence indicating that children and adults weight these perceptual correlates of place of articulation in a similar manner. Most researchers have long accepted the idea that children’s perception of speech is somehow different from that of adults, though whether this difference is due to immaturely developed auditory perception or to what part of the signal they attend to is still debated (Bourland Hicks and Ohde, 2005). Overall the evidence seems to support both hypotheses: while children’s perception based on auditory cues may be immature compared to adults, they also appear to attend more to different parts of the speech signal than adults do. While adults tend to give perceptual weight to both static and dynamic acoustic properties of speech (Strange, 1989b), some researchers have speculated that children give more weight than adults do to the dynamic spectrally changing (Nittrouer, 1992) formant transitions. It is clear that if children attend to dynamic formant transitions more than static formant targets, then there would be little developmental support for the simple target model as the perceptual mechanism involved in the perception of place of articulation of vowels. However, contrary to this view, some prior developmental results might be interpreted to support a formant target model (Sussman, 2001).

Ohde *et al.* (1996) speculated that “vowels may be apprehended as unique perceptual units early in development,” because even young infants can perform speaker normalization. Infants have demonstrated an ability to discriminate vowel signals regardless of age, gender, etc. of the speaker and can even link acoustic to facial cues by four months of age (Kuhl, 1979, 1983). Only a few other studies have been conducted specifically on the development of vowel perception (Parnell and Amerman, 1978; Murphy

*et al.*, 1989; Ohde *et al.*, 1996; Ohde and Haley, 1997; Malech and Ohde, 2003). Of these studies, none have examined potential differences in cue weighting of vowels in children compared to adults.

One early landmark study was published by Parnell and Amerman (1978). These researchers compared the ability of 4- and 11-yr-old children, and adults to identify vowels and consonants based on isolated static and dynamic cues, and found that young children uniformly performed more poorly on identification tasks than older children or adults. However, they also found that the magnitude of the difference was less between young children and older children and adults in the tasks that gave the participants dynamic information such as formant transitions plus vowel center, as opposed to static information such as aperiodic stop-consonant bursts. These findings seem to indicate that young children may pay more attention than adults to dynamic information in assessing speech sound identities. Even the older children, who appeared to do almost as well as the adults, tended to use durational cues more than the adults. Parnell and Amerman concluded, however, that young children’s perceptual abilities are not as well developed as those of older participants.

Other investigators have also concluded that young children appear to use dynamic cues such as formant transitions relatively more than do adults in identifying fricatives (Nittrouer, 1996), glides (Bourland Hicks and Ohde, 2005), and vowels (Murphy *et al.*, 1989; Nittrouer, 1996; Ohde *et al.*, 1996; Ohde and Haley, 1997; Malech and Ohde, 2003), especially the latter. Nittrouer (1996) termed this perceptual process the Developmental Weighting Shift. More recent interpretations of these developmental changes in perception indicate two levels of formant transition bias (Mayo and Turk, 2004, 2005). Changes in how formant transitions are processed have been labeled as the developmental bias and the acoustic bias hypotheses and refer to the processes whereby children weight transitions more heavily than adults or weight transitions more heavily than other acoustic cues, respectively. Children’s accuracy of vowel perception has been shown to be affected more than adults’ by consonantal context and stimulus duration (Ohde *et al.*, 1996), as well as for vowel duration (Ohde and Haley, 1997). Although Ohde *et al.* (1995) did not find formant transitions or durational components to be important to young children in identifying place of articulation for stop consonants, the evidence from other studies indicates that children’s vowel perception may be slightly different from consonant perception. Ohde and Haley (1997) found that short stimulus onset information for stops and vowels was sufficient for 3- to 4-yr-old children and adults to identify speech sounds, even without dynamic formant transitions. However, dynamic formant motion generally improved vowel identification across all ages and consonant context. For children, the accuracy of vowel identification increased with increases in stimulus duration. They concluded that for vowels, formant target frequencies and transitions were the most important cues. In the Ohde and Haley study, formant onsets were not tested in trading relation conditions that systematically controlled both formant onsets and transitions.

One study by Sussman (2001) did appear to contradict Nittrouer's (1996) Developmental Weighting Shift hypothesis, indicating that children may not give more perceptual weight to dynamic cues than adults do. Sussman's study involved the use of whole syllables, silent-center syllables, steady-state formant vowels, and conflicting syllables containing transitions from one vowel with steady-states from another. Results showed that adults were better than children at identifying vowels using only dynamic information, and children relied more on steady-state information than adults. However, Sussman acknowledged that the steady-state information presented was "longer [and] more powerful" than the dynamic information. Sussman's use of powerful in this context appears to relate to amplitude. The implication is that the longer the stimulus the greater its amplitude. Therefore, it is possible that the children, with less mature perceptual systems than adults, weighted this cue more heavily simply because it was so much more acoustically dominant, rather than because of a weighting strategy that emphasized static cues. Sussman did not employ a trading relations type paradigm so specific differences in cue weighting between children and adults could not be determined.

It has been theorized that children's development of speech perception proceeds from whole syllables to smaller units such as phonemes (Jusczyk and Derrah, 1987; Nittrouer and Miller, 1997; Mayo and Turk, 2004). This would at least partially explain why, if the developmental weighting shift theory is valid, segmental information (e.g., fricative spectrum) is less useful to younger children than to adults, and why young children's responses were based on the multisound, within-syllable dynamic formant transition information (Nittrouer, 1992; Ohde *et al.*, 1995).

Some researchers have also discussed the possibility that neurological immaturity or immature auditory mechanisms are the developmental basis of children's identifications, rather than perceptual weighting differences (Parnell and Amerman, 1978; Sussman, 2001). This possibility is partially negated as well as supported by Nittrouer (1996). Nittrouer found that 3-yr-old children's difference limens (DL) for both fricative-noise spectra and F2-onset frequencies were larger than those for adults. Children were not as sensitive as adults to these spectra. If auditory sensitivity was greater for adults than children, then it would be predicted that they would weight the fricative noise spectrum and F2 formant onsets more than children do in labeling tasks. Adults' identification functions should be steeper and more widely separated than children's functions (Ohde and Sharf, 1988; Sussman and Carney, 1989). Contrary to the prediction, children's functions were more widely separated than those for adults indicating a greater weighting of the formant transitions by children than adults. In sum, children's auditory development is certainly more immature than adults', but this immaturity cannot entirely account for what still appears to be different perceptual weighting for children than adults. It is possible that children weight dynamic cues more because this is what they are able to process better.

Based on existing studies, it appears that very little research has been conducted on vowel perception in children. Moreover, there appears to be little research on the

development of vowel perception assessed according to cue weighting strategies in children compared to adults. Most developmental research points to dynamic information, whether stimulus duration or formant transitions and trajectories, as the most important cues for vowel perception in children. However, further investigation is necessary to verify this and to determine why it might occur more in children than adults.

The general purpose of the current research was to determine if 3- to 4-yr-old children give more perceptual weight to the dynamic properties of vowels when compared to adults. Specifically, it was predicted that when given conflicting formant onsets and dynamic information in the same synthetically produced syllables children would attend less to formant onsets and more to dynamic cues as compared to adults. For example, if the F2 onset frequency in a labial stop-consonant syllable cues [i], but the following formant transition motion cues [a], then it would be predicted that children should shift sooner from [i] identification to [a] identification than adults if they attend more to the dynamic formant transitions. In a CV syllable, consonant bursts also cue place of articulation of stops. However, in the current research, the synthesized CV syllable did not contain a stop burst. Thus, place of articulation for the vowel and the consonant was represented by formant onset frequencies. There are several possible reasons why children may differ from adults in processing these formant onsets. First, children may not have completely established or learned the acoustic correlates of vowels for onset frequencies associated with a particular consonant place of articulation. Second, the relatively short duration onsets may not be processed by children in the same manner as adults due to immature auditory mechanisms and/or skills. Third, the spectral cues in formant onsets may be lower in amplitude resulting in weaker cues in formant onsets. Fourth, overall spectral processing of features for some vowels may be difficult for children to identify. Sussman (2001) found the +high/+front [i] vowel was difficult to perceive by some young children perhaps due to a concentration of high frequency energy. Presumably, vowels with high frequency cues may be difficult to identify due to a poorly developed auditory system for these frequencies.

In sum, the absolute onset frequency for [bi] versus [ba] should be identified differently if onset frequency is salient. If on the other hand stimulus duration is too short for perception by young children, vowel identification from these onset frequencies should be similar. The specific questions and hypotheses addressed by this study were as follows: (1) Are children's identification responses to stimuli relative to adult responses, in the region of the identification function where cues are in conflict, influenced more by onset frequency or formant transition cues? It is hypothesized that children will be more influenced by formant transitions than formant onsets (Nittrouer, 1992; Ohde and Haley, 1997; Mayo and Turk, 2004; Mayo and Turk, 2005). (2) Is there evidence supporting that formant onset frequencies provide stronger perceptual cues to the back/low than the front/high feature distinction of the [i] versus [a] vowel? It is hypothesized that the acoustic correlates of place of articulation will be more salient in the [a] vowel than [i] vowel (Sussman, 2001). (3) Is variability greater for children than adults as evidence

from perceptual slope analyses? It is hypothesized that perceptual variability will be greater in children than adults (Thibodeau and Sussman, 1979; Ohde *et al.*, 1996; Ohde and Haley, 1997). (4) Do formant onsets and formant transitions integrate in perceptual development? It is hypothesized that formant onsets and formant transitions follow a developmental course that includes a mechanism of perceptual integration (Ohde and Camarata, 2004; Stevens, 2000).

## II. METHOD

### A. Participants

The participants were ten adults and ten children. The children's ages ranged from 3;8 to 4;1, with a mean age of 3;11. Nine of the children were 3 years old and one was 4 years old. Three- to four-year-old children were tested in this research because they represented the lower limits in age at which reliable data can be obtained. The mean age for the adults was 25;11, ranging from 23;10 to 32;0. All participants were monolingual speakers of American English. The children exhibited normal articulation, defined as performance within normal limits on the *Arizona Articulation Proficiency Scale, Third Revision* (Fudala, 2000) and age appropriate language abilities based on performance within normal limits on the *Test of Early Language Development*, 2nd ed. (Hresko *et al.*, 1991). The adults used a standard American dialect and exhibited no evidence of speech or language impairment based on informal assessment. A pure-tone hearing screening at 20 dB hearing level for the octave frequencies between 500 and 4000 Hz was performed on both children and adults prior to each testing session, and tympanometry was performed for each child. All participants passed the hearing screenings, and they had normal tympanograms. All participants were paid for their participation and children also received a piece of candy as a prize after each testing session. Four 3-year-olds were unable or unwilling to complete the perceptual testing. These children were replaced by new participants.

### B. Test stimuli

The stimuli were generated using the Klatt cascade/parallel formant synthesizer (Klatt, 1980; Klatt and Klatt, 1990) at a sampling rate of 10 kHz. The output was low-pass filtered with a cutoff frequency of 4 kHz. Two vowels, [i] and [a], were used as the bases for the construction of stimuli. These two vowels were chosen to be very different in terms of their first (F1) and second (F2) formant frequencies (for onset and target frequencies). The frequencies were derived from syllables containing a bilabial stop-consonant onset. The stimuli were synthesized with source and resonance parameters appropriate for an adult male vocal tract. Five different pairs of control stimuli and three experimental continua were generated. This section describes the natural control productions and the synthesis of the control and experimental stimuli.

#### 1. Control stimuli

The first pair of control stimuli consisted of naturally produced syllables [bi] and [ba] of about 300 ms duration

each, recorded by an adult male speaker. Ten syllables (five of each [bi] and [ba]) were presented in randomized order at the beginning of each testing session. The second pair of control stimuli, also presented at the beginning of each session, consisted of synthesized full-duration (300 ms) [bi] and [ba] syllables. These synthesized full duration stimuli were based on our previous research (Ohde *et al.*, 1996; Ohde and Haley, 1997) and were appropriate for standard adult male values. Each of these natural and synthetic stimuli ([bi] and [ba]) was reiterated five times. These stimuli were also arranged in a group of 10 in randomized order, with five examples of each syllable. Each of the endpoint control and experimental stimuli was presented before its corresponding experimental continuum to provide practice to the participants in identifying these sounds. The order of presentation of these control stimuli was as follows: natural full-duration [bi] and [ba] syllables, synthesized full-length 300 ms duration [bi] and [ba] syllables, and synthesized 30 ms duration endpoint stimuli from the three experimental conditions (static onset, [i]-onset, and [a]-onset).

#### 2. Experimental stimuli

As indicated above, the last three pairs each of 30 ms duration experimental stimuli contained five examples of each of the two endpoint stimuli from one of the three experimental continua. These stimuli would be the most [bi]-like or [ba]-like syllable. Thus, one of these experimental pairs consisted of five of the [bi] endpoints (S1; see Fig. 1) and five of the [ba] endpoints (S10; see Fig. 1) from the static-onset stimulus continuum. Another corresponded to endpoints from the [i]-onset condition, and the third experimental pair contained endpoints from the [a]-onset condition.

The experimental stimuli had durations of 30 ms. There were three conditions based on the following formant information: (1) static-onset, (2) [i]-onset + transition, and (3) [a]-onset + transition. The static-onset condition contained only stationary F2-F4 information (consonant spectral onset frequencies). The first formant (F1) did change from 190 Hz for all stimuli and then moved to the target F1 frequency varying from 330 to 720 Hz, going from [i] to [a]. Two conditions contained static and dynamic vowel transition information. These three spectral conditions will be referred to as static-onset, [i]-onset, and [a]-onset, as illustrated for F1-F4 in Figs. 1, 2, and 3, respectively.

The static-onset stimuli contained ten static F2 transition onsets, ranging from 1800 Hz (corresponding to an [i] endpoint) to 900 Hz (corresponding to an [a] endpoint) were used. Each F2 value in the range differed from the previous one by 100 Hz (e.g., stimulus 1: 1800 Hz, stimulus 2: 1700 Hz, etc). The third and fourth formants simultaneously varied from endpoint [i] values to endpoint [a] values as well (F3 from 2600–2000 Hz, and F4 from 3200–3500 Hz). The F2, F3, and F4 endpoints for [i] and [a] are 1800, 2600, and 3200 Hz, and 900, 2000, and 3500 Hz, respectively. To simulate natural stop consonant + vowel syllables as closely as possible, the F1 onset was 190 Hz for all stimuli and moved to a target F1 frequency ranging from 330 ([i]-like) to 720

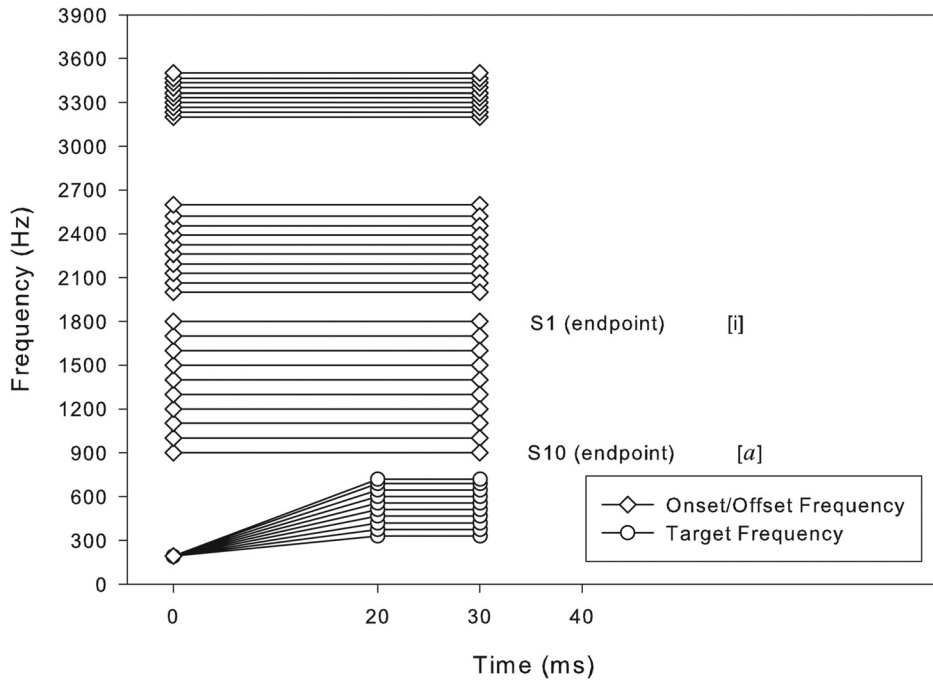


FIG. 1. This figure shows the schematic representation of the first four formants (F1, F2, F3, and F4) for static-onset stimuli. The four formants contain ten stimulus values and each set of ten values represent the formants from low (F1) to high (F4). S1 and S10 are the [i] and [a] endpoints, respectively. See the Appendix for detailed F1–F4 stimulus values.

Hz ([a]-like). All signals were 30 ms in duration, since previous research has shown that young children can accurately process vowels of this duration (Ohde *et al.*, 1996; Ohde and Haley, 1997).

The two dynamic ([i]-onset and [a]-onset) conditions contained an initial onset frequency appropriate for a labial stop and a transition toward a target frequency which was not reached (only 30 ms of the typical 40 ms transition duration was presented). The F1–F4 formant onset frequencies were held constant across the stimulus continua in these two dynamic conditions. For example, F2 onset was held constant at either 1800 Hz for the first condition ([i]-onset) or 900 Hz for the second condition ([a]-onset) and the typical F2 target varied from 2200 Hz

([i]) to 1200 Hz ([a]) at 100 Hz intervals (i.e., 2200, 2100, etc). In these conditions, the short stimulus duration (i.e., 30 ms) ensured that the F2 vowel target frequency was not actually reached, thereby giving only one true static cue, the onset; and one dynamic cue, the transition trajectory. See the Appendix for specific F1 to F4 onset and F1 to F4 target frequencies for the static onset, [i] onset, and [a] onset conditions in Tables III, IV, and V, respectively.

Full-duration (300 ms) synthesized vowel exemplars ([ba] and [bi]) were included at randomly determined points in the experimental conditions to monitor children’s attention. Listeners needed to maintain an identification accuracy of at least 75% correct to continue in the experiment.

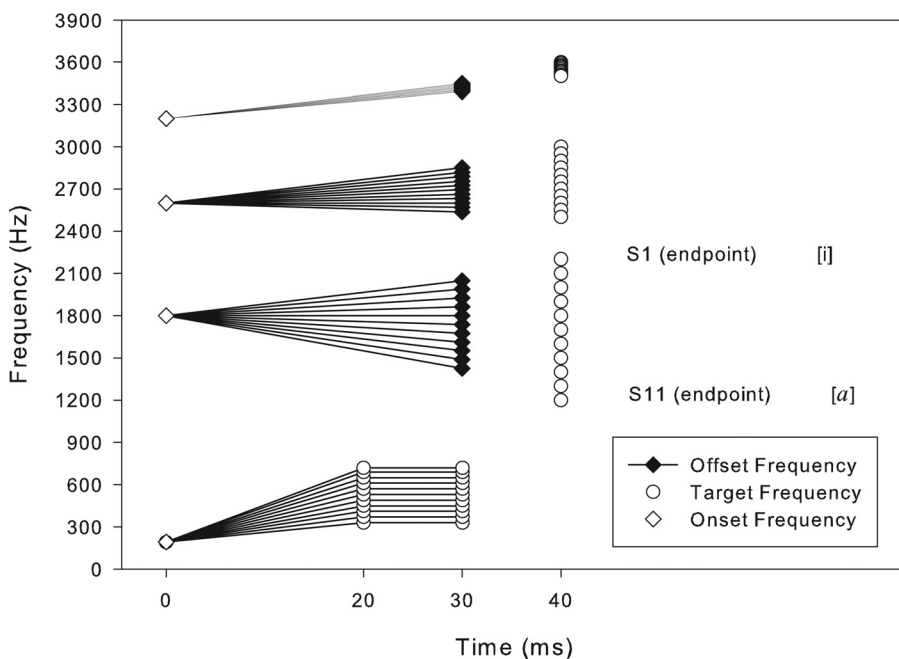


FIG. 2. This figure shows the schematic representation of the first four formants (F1, F2, F3, and F4) for [i]-onset stimuli. The four formants contain 11 stimulus values and each set of 11 values represent the formants from low (F1) to high (F4). S1 and S11 are the [i] and [a] endpoints, respectively. See the Appendix for detailed F1–F4 stimulus values.

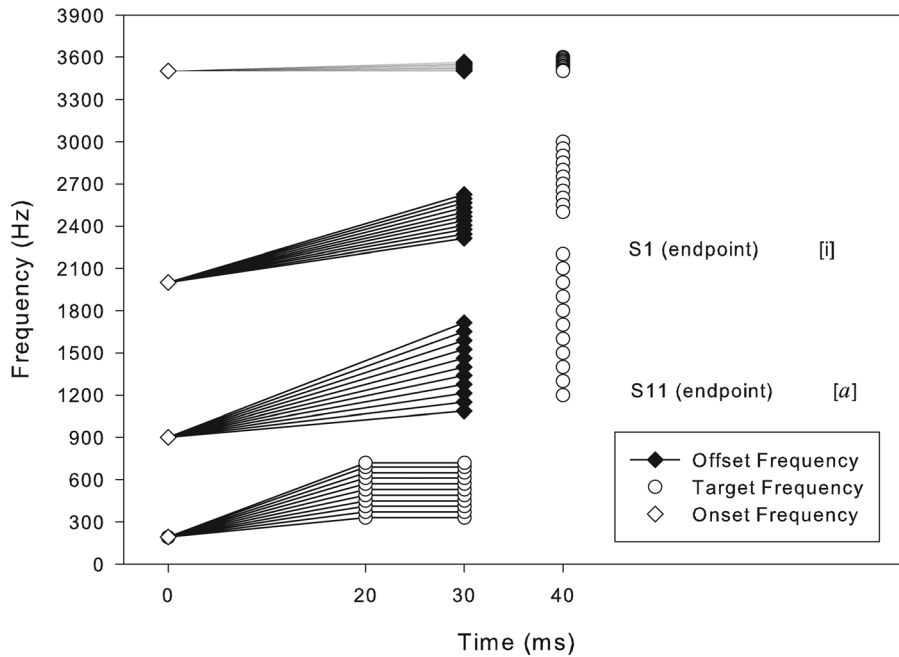


FIG. 3. This figure shows the schematic representation of the first four formants (F1, F2, F3, and F4) for [a]-onset stimuli. The four formants contain 11 stimulus values and each set of 11 values represent the formants from low (F1) to high (F4). S1 and S11 are the [i] and [a] endpoints, respectively. See the Appendix for detailed F1–F4 stimulus values.

### C. Procedure

Each listener heard 48 (4 presentations  $\times$  10 stimuli + 8 full length syllables) static-onset stimuli, and 52 (4 presentations  $\times$  11 stimuli + 8 full length syllables) each of the dynamic-onset condition. The difference in numbers of stimuli in the static (10) versus the dynamic (11) conditions was partly based on our interest in using the static condition as a control for comparison and replication of our previous research (Ohde *et al.*, 1996; Ohde and Haley, 1997) on static onsets of 30 ms duration. This static-onset condition also provided information regarding any potential effects of a stimulus continuum on identification of the [bi] and [ba] endpoints particularly for children. With the exception of the 10 Hz F1 onset frequency differences indicated above, the spectral and duration properties of the current end point values were identical to those of our previous research (Ohde *et al.*, 1996; Ohde and Haley, 1997). For the static-onset condition, a 10 step continuum was appropriate.

The vowels were identified using a two-alternative forced-choice testing paradigm (2AFC). The participants were asked to decide whether the syllable they heard was [bi] or [ba]. Participants indicated their choice by pointing to one of two puppets for children, or by pressing buttons on a response box marked [i] and [a] for adults. The children received a brief practice period to learn to associate each syllable with a different puppet and to make syllable identification choices. The adult participants received practice in making identification choices and using the response box. After the participant responded to a stimulus by choosing one syllable over the other, the next stimulus was presented. At the beginning of each session, the natural and synthesized full-length syllable practice pairs were presented, and each endpoint practice pair was presented before its corresponding experimental stimulus condition. The static-onset control condition was presented first to listeners, and the [i]-onset and [a]-onset conditions

were counterbalanced across participants. Because the static-onset condition lacked dynamic formant transition trajectories with the exception of F1 and was included as a control condition, it was presented first to all participants.

### D. Statistical analyses

Two repeated-measures analyses of variance (ANOVAs) on category boundaries and slopes were performed on the data, as well as a *t* test of the difference score between the [i]-onset and [a]-onset boundaries. These analyses included the between factor of age (adult and child) and the within factor of spectral condition (static-onset, [i]-onset, and [a]-onset). In addition, linear contrasts were computed on several main and interaction effects. The phonetic boundaries (category boundaries), which mark the 50% point between sound categories, were determined by assigning to each stimulus a number ranging from either 1 to 10 (static-onset condition) or 1 to 11 (onset + transition conditions) and then performing a linear regression analysis of the Z transformations of the proportions, also known as probit units (Finney, 2009). Linear regression was performed on Z scores (“y”) versus stimulus values (“x”). The slopes were derived from the least-mean-squares analysis. Significant results at the  $p < 0.05$  or  $p < 0.01$  levels are reported.

## III. RESULTS AND DISCUSSION

### A. Category boundaries

The repeated measures analysis of variance of the category boundaries revealed a significant main effect of spectral condition [ $F(2,36) = 20.21$ ;  $p < 0.01$ ], and a significant interaction of spectral condition by age [ $F(2,36) = 6.93$ ;  $p < 0.01$ ]. Linear contrasts revealed that adults’ category boundaries for the [i]-onset condition and [a]-onset condition were significantly different from the static-onset condition. As illustrated in Figs. 4, 5, and 6, this result may not be

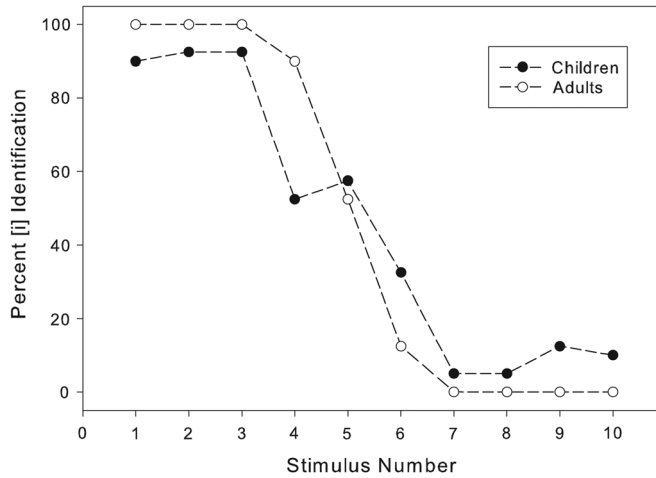


FIG. 4. Mean percent [i] responses of children and adults for the static-onset condition.

surprising because the static-onset continuum contained only 10 stimuli, whereas the onset-transition continua had 11 stimuli each. Thus, the differences in the number of stimuli among conditions may have influenced the placement of category boundaries. However, the adult category boundaries for the [i]-onset condition and [a]-onset condition were not statistically different in terms of numbers of stimuli heard as [i] or [a]. These same analyses for children revealed that the static-onset was significantly different from the [i]-onset condition, and the [i]-onset was significantly different from the [a]-onset condition. This shows that the children's relative categorization of the [i]-[a] contrast is not only different for static versus dynamic conditions, but also between dynamic conditions. The mean category boundaries, slopes, and measures of variability [(standard deviation – SD) and (standard error – SE)] are reported in Table I. It is clear from Table I that slopes are shallower for children than adults.

In comparing the absolute category boundaries of children and adults, linear contrast tests revealed a significant age difference for the [a]-onset category boundaries. These results show that adults heard more of the stimuli as [i] than the children did for this condition. To determine if the difference

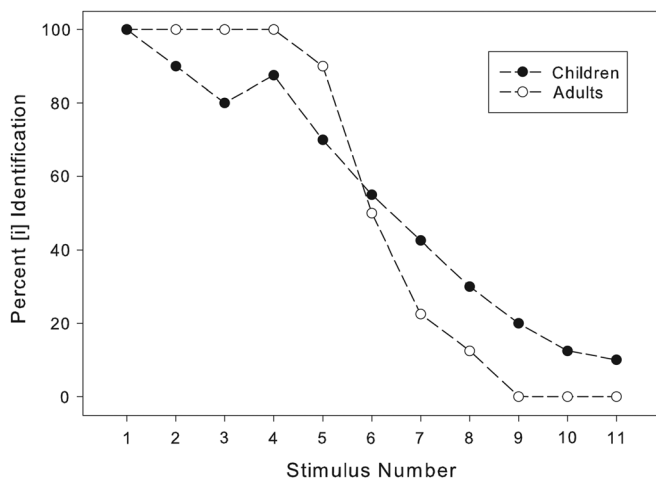


FIG. 5. Mean percent [i] responses of children and adults for the [i]-onset condition.

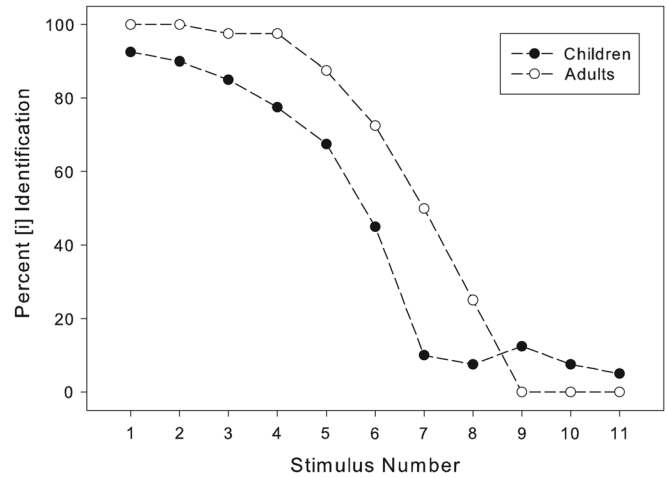


FIG. 6. Mean percent [i] responses of children and adults for the [a]-onset condition.

between [i] and [a] category boundaries differentiated children from adults, a two-tailed independent samples *t* test was computed for these difference scores. The findings revealed a significant between group difference [ $t(18) = 3.20; p < 0.01$ ], further indicating that young children and adults perform different operations in the perception of two vowel sound categories.

The results of this research indicate that children differ in their perceptual identification of vowels from adults, but these findings do not support the Developmental Weighting Shift hypothesis as presented by Nittrouer (1996). Rather than utilizing the dynamic formant information in a vowel more than adults did, the child group relied more heavily than the adult group on the [a] formant-onset frequency cue. This was evidenced by the fact that children's category boundaries were significantly skewed more toward the onset vowel frequency as compared to adults, whose boundary shift occurred more toward the target frequency, as previously predicted (Blumstein and Stevens, 1980). In other words, in the [i]-onset condition, children perceived more vowels as [i] compared to the [a]-onset condition, while adults did not show this preference. This shift indicates that

TABLE I. The mean (stimulus number), standard deviation (SD), and standard error (SE) for category boundary (CB) and slope of children and adults across the three experimental conditions of static-onset, [i]-onset, and [a]-onset.

	Child			Adult		
	Mean	SD	SE	Mean	SD	SE
Static-onset						
CB	4.8583	1.0249	0.3241	5.0660	0.5313	0.1680
Slope	1.1927	0.9258	0.2927	1.6855	0.9161	0.2897
[i]-onset						
CB	6.5854	0.9024	0.2853	6.2595	0.7260	0.2295
Slope	0.4889	0.2969	0.0938	1.5153	1.0316	0.3262
[a]-onset						
CB	5.3806	0.8478	0.2681	6.8055	0.7277	0.2301
Slope	0.5822	0.2747	0.0868	1.1202	0.8443	0.2669

the children paid more attention to the formant-onset frequency of the vowel and less attention to the dynamic formant transition when identifying a vowel. Adults, in contrast, gave more comparable weight to the [i]-onset and [a]-onset spectral information as illustrated in their category boundaries than children did.

The static-onset stimuli served in part in this study as a control condition. The child and adult groups did not differ significantly in their category boundaries for this condition, which contained only static information with the exception of F1 as illustrated in Fig. 1. This provided further support that the difference between children's and adults' phonetic boundaries in the onset + dynamic conditions was not due to attentional problems or misunderstanding of the task on the part of the children. The children were equally able to assign vowel identities to steady-state information as the adults were, and their perception based on the static onset frequency only did not differ from adults' perception.

## B. Category variability

Differentiating sound categories demands an ability to not only conceptualize acoustic phonetic detail into a category whole, but also an ability to decide relatively abruptly when one sound category ends and another begins. Phonetic boundaries indicate an ability to conceptualize category wholes, whereas slopes indicate how confidently one decides where one category ends and another begins. One general prediction

is that children would be less confident and would more gradually change (more shallow slope values) from the [i] category to the [a] category than adults (Thibodeau and Sussman, 1979; Ohde and Sharf, 1988). The repeated measures analysis of variance of the slopes revealed a significant main effect of spectral condition [ $F(2,36) = 3.82; p < 0.05$ ] and a significant between factor effect of group [ $F(1,18) = 8.35; p < 0.01$ ]. Linear contrast analyses were computed and revealed that children were significantly different from adults in their slope values for the [i]-onset condition [ $F(1,18) = 9.14; p < 0.01$ ] and marginally significantly different [ $F(1,18) = 3.67; p < 0.07$ ] from adults in their slope values for the [a]-onset condition. Thus, as clearly shown in Fig. 5 and Tables I and II, children were less definitive in their boundary change than adults for the [i]-onset condition. The results of these analyses show that children's perception of the front/high and back/low vowel features are more variable and less well developed than adults' perception of these features. Previous studies have found that children are more variable in perception of vowels and consonants than adults (Thibodeau and Sussman, 1979; Sussman and Carney, 1989; Ohde *et al.*, 1996; Ohde and Haley, 1997).

## C. Acoustic basis of front/high and back/low vowel feature identification

Two important properties of formants may play unique roles in the perceptual development of vowel features. First,

TABLE II. The mean (percent [i] identification), standard deviation (SD), and standard error (SE) for continua stimuli of children and adults across the three experimental conditions of static-onset, [i]-onset, and [a]-onset.

	1	2	3	4	5	6	7	8	9	10	11 <sup>a</sup>
Static-onset											
Child											
Mean	90.00	92.50	92.50	52.50	57.50	32.50	5.00	5.00	12.50	10.00	*
SD	12.91	12.08	16.87	46.32	26.48	33.44	10.54	10.54	24.30	12.91	*
SE	4.08	3.82	5.34	14.65	8.37	10.57	3.33	3.33	7.68	4.08	*
Adult											
Mean	100.00	100.00	100.00	90.00	52.50	12.50	0.00	0.00	0.00	0.00	*
SD	0.00	0.00	0.00	12.91	38.10	17.68	0.00	0.00	0.00	0.00	*
SE	0.00	0.00	0.00	4.08	12.05	5.59	0.00	0.00	0.00	0.00	*
[i]-onset											
Child											
Mean	100.00	90.00	80.00	87.50	70.00	55.00	42.50	30.00	20.00	12.50	10.00
SD	0.00	17.48	19.72	17.68	28.38	30.73	16.87	22.97	25.82	17.68	12.91
SE	0.00	5.53	6.24	5.59	8.98	9.72	5.34	7.26	8.16	5.59	4.08
Adult											
Mean	100.00	100.00	100.00	100.00	90.00	50.00	22.50	12.50	0.00	0.00	0.00
SD	0.00	0.00	0.00	0.00	12.91	39.09	24.86	17.68	0.00	0.00	0.00
SE	0.00	0.00	0.00	0.00	4.08	12.36	7.86	5.59	0.00	0.00	0.00
[a]-onset											
Child											
Mean	92.50	90.00	85.00	77.50	67.50	45.00	10.00	7.50	12.50	7.50	5.00
SD	12.08	17.48	17.48	27.51	28.99	22.97	17.48	16.87	13.18	16.87	10.54
SE	3.82	5.53	5.53	8.70	9.17	7.26	5.53	5.34	4.17	5.34	3.33
Adult											
Mean	100.00	100.00	97.50	97.50	87.50	72.50	50.00	25.00	0.00	0.00	0.00
SD	0.00	0.00	7.91	7.91	17.68	14.19	37.27	23.57	0.00	0.00	0.00
SE	0.00	0.00	2.50	2.50	5.59	4.49	11.79	7.45	0.00	0.00	0.00

<sup>a</sup>For cells marked with an asterisk, there were only 10 stimuli in the static-onset condition.



the effects of formant onset frequencies on vowel category perception were different in children than adults for the [a]-onset condition. These formant onset frequencies were perceptually more dominant in children's perceptions than in those of adults. The brief onset frequencies represent a coarticulated cue that provides acoustic information not only for place of articulation of consonants but also for the distinction of front/high ([i]) versus back/low ([a]) vowels. Vowel onset frequencies and formant transitions may not be independent and must be integrated by perceptual mechanisms that are not clearly understood. What is clear is that children appear to weight and/or integrate vowel onset frequencies and formant transitions differently from adults.

Second, the extent of the F2 frequency change from stimulus onset frequencies to the frequency value at stimulus offset may play a role in reducing perceptual variability.

As shown in Fig. 7, the absolute frequency change of F2 is less for the [i]-onset condition [Fig. 7(A)] than for the [a]-onset condition [Fig. 7(B)]. Figure 7 represents the endpoint stimuli (stimuli 1 and 11 from Figs. 2 and 3, respectively) for the [i]-onset and the [a]-onset conditions in terms of the formant trajectories in the F1-F2 plane. For example, the greatest frequency change in F2 trajectory was 813 Hz from [a]-onset (900 Hz) to [a]-offset (1713 Hz). The magnitude of frequency change in F2 trajectories may account for the finding in the current study of no difference in slopes for children and adults in the [a]-onset condition.

In summary, these results support the hypothesis that children's vowel perception depends on at least two processes that are different between children and adults for the vowels examined in the current study. First, children's category boundaries were significantly different from adults'

category boundaries for the [a] vowel. Second, children's slopes were significantly shallower than adults' slopes for the [i] vowel. It appears that children's general ambiguity for the perception of the [i] vowel may relate to perceptual development, and specifically to an immaturity in vowel perception. This may support the hypothesis that perceptual development of vowels follows a course beginning with general immaturity as reflected in shallower slopes, followed by category boundaries that are displaced from adult category boundaries but are similar to adults as reflected by slopes. Children's perceptual systems may indeed be more immature than adults'. However, the fact that children's category boundaries significantly differed between two conditions suggests that children are not simply "guessing" at a vowel's identity but actually using a slightly different (from adults), but still active, cognitive process to identify vowel sounds. This vowel perceptual process appears to utilize static information (e.g., vowel onset frequency) more than dynamic (e.g., transition trajectory) information at least for perception of the front/high and back/low vowel features. In addition, children may use the extent of formant trajectory change as a way to reduce perceptual variability.

#### IV. GENERAL DISCUSSION

##### A. Formant onsets and formant transitions in perceptual development of vowels

Based on the findings of this research, young children do attend differently to the formant onset frequency than do adults for the [a]-onset condition. These findings support Sussman's (2001) research, which suggested that children rely more on steady-state (static) information and adults are

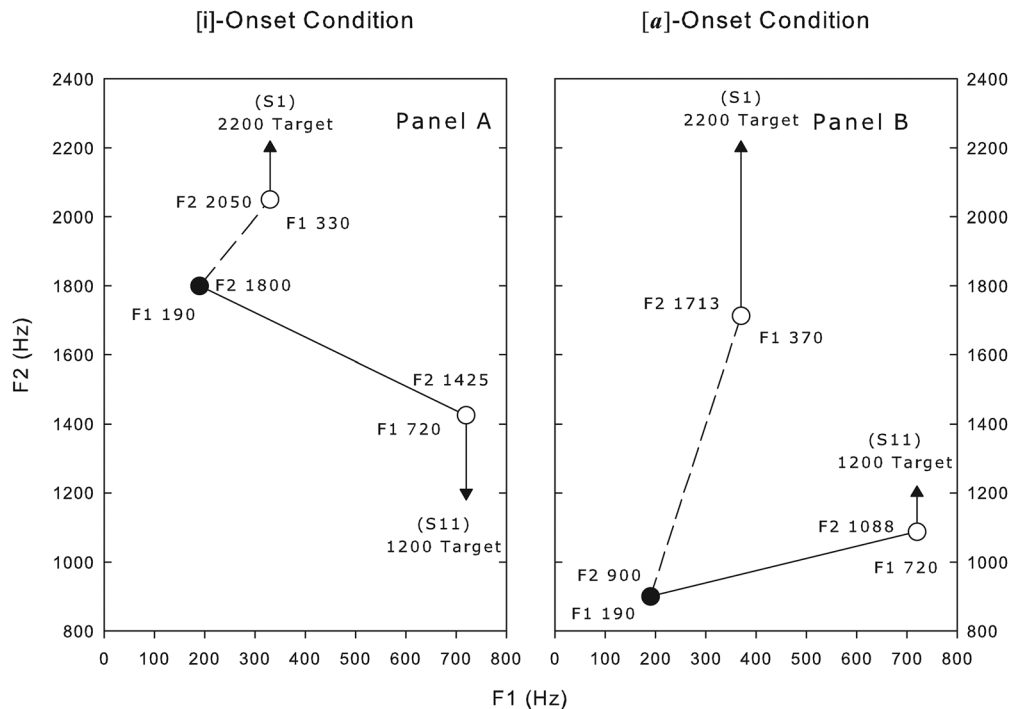


FIG. 7. This shows (A) a representation of the [i]-onset and (B) [a]-onset endpoint (S1 and S11) stimuli in terms of the formant trajectories in the F1-F2 plane. The onset values are the same across stimuli and are depicted with a filled circle. The open circles represent the formant values reached at 30 ms. The arrows indicate the F2 targets not reached by these stimuli. The numerical values within (A) and (B) are in Hz.

more able than children to use dynamic cues in perception. By controlling formant onsets across stimulus continua ranging in formant transitions appropriate for [i] and [a], the role of formant transition onsets could be examined for children and adults. By presenting the vowel information more naturally in the current research regarding durational cues (i.e., the static information was only presented in the first glottal pulse for the [i]-onset and [a]-onset conditions similar to previous research; Ohde and Haley, 1997), the issue of whether the static information was more salient due to duration could be assessed. The onsets were controlled in the current study by holding them constant across onset-dynamic conditions. Thus, these short static onsets can be examined relative to their independent effects on vowel perception. In fact, if duration was the deciding factor, children should have chosen a vowel identity based more on the 25 ms transition, which was not the case at least for the [a]-onset condition. That is, the rapid spectral changes in the transition, dynamic cue, may not be resolved by the child's auditory system in the same manner as an adult's auditory system.

If children have a physiologically and anatomically immature hearing mechanism, this may account for the differences between children's and adults' perception of rapid spectral changes in dynamic formant transitions. Vowel identification as reflected in category boundary was significantly different for children and adults for the [a]-onset condition. However, the children's and adults' category boundaries were not significantly different for the [i]-onset condition. These findings suggest that if an immature hearing mechanism accounts for these above perceptual differences in the [a]-onset and [i]-onset conditions for children and adults, it is a complicated process. An equally plausible explication of the [a]-onset category boundary difference between children and adults would be differences in the integration of formant onsets and formant transitions. The category boundaries for the [i]-onset and [a]-onset conditions were not significantly different for adults. These category boundaries were significantly different for children. For the [a]-onset condition, children's category boundary was about 1.5 steps different from the adults' boundary. Thus, onsets may integrate differently with transitions for [i] and [a]. As the perceptual system matures, it is hypothesized that listeners begin to weight and integrate formant onset frequencies and formant transitions in assigning an identity to the vowel.

An alternative explanation relates to the process of learning the salient acoustic-phonetic properties of sound contrasts, for example, [i] versus [a]. Typically developing children apparently need to learn these acoustic-phonetic relations of a given property, and its variation across phonetic context. For example, young children seem to perceive speech based on a longer sound unit such as the syllable. Formant transitions serve to delimit these syllable sized units (Nittrouer *et al.*, 2000). Children appear to learn the importance of formant transitions and weight them more heavily than adults or more heavily than they weight other acoustic cues, e.g., fricative noise (Nittrouer, 2002; Mayo and Turk, 2004). In the case of [bi] and [ba], the relevant acoustic-phonetic formant onset frequency relation is a relatively high and a relatively low F2 onset value, respectively. However,

when the [bi] and [ba] syllables also contain dynamic formant motion in the direction of a vowel target, that is, different from the onset frequency specification, for example, [i]-onset and [a]-target (see Figs. 2, 3, and 7), then these inappropriate acoustic-phonetic relations must be perceptually determined early in development. If formant transitions are dominant in perceptual development, then children should show a transition bias in these conflicting cue stimuli. If, however, children show a bias for transition onsets occurring within the first 5 ms of the syllable, then there should be an onset bias for these conflicting cues. Listeners must be able to perceptually compensate for these conflicting cues with perhaps adults better at compensation than children. Thus, one hypothesis is that the differences observed between children and adults in these two-cue conflicting conditions relates to variations in segmental perception based on different perceptual criteria due to learning acoustic-phonetic relations across phonetic contexts.

Strange *et al.* (1983) described two theories of vowel perception, the target normalization theory (elaborated target model) and the dynamic specification theory that emphasize different degrees of support for static and dynamic perceptual cues. The first states that there are canonical "target" formants, or transformations thereof, for each vowel, which are unique and are the only information used to identify a vowel by a listener. Thus, the rest of the vowel, including the onset and transition trajectory, is not taken into account by this theory. The second theory attempts to look at the vowel as a whole, which is a sum of all of its parts including dynamic information resulting from a number of articulatory gestures, which a listener will incorporate into his/her final identification of the vowel. Thus, the "changing spectro-temporal configuration" of the acoustic pattern of the vowel "provides sufficient information for the identification of the phonetic units" (Strange *et al.*, 1983).

The implication of the dynamic specification theory is that temporally dynamic cues such as formant transitions are essential for accurate identification of vowels even in the absence of the attenuated vowel nucleus. If children pay more attention to dynamic formant transitions, as indicated by the developmental weighting shift hypothesis, then it would be predicted that these cues should be dominant relative to formant targets and formant onsets. However, Sussman (2001) found that vowels were more accurately identified by typically developing young children from vowel steady-state centers than formant transitions. The current research examined the role of formant onsets and transitions as developmental cues to vowel perception.

## B. Theories in perception and perceptual development

The current developmental research provides support for the dynamic specification theory in that it shows that both adult and child listeners take static and dynamic information into account when identifying a vowel. In fact, if the target normalization theory was correct, the stimuli in this study would have been too ambiguous to classify because they lacked canonical target frequencies entirely. Moreover,

previous findings show that vowel identification from synthetic CV syllables containing either dynamic formant transitions or static formant onsets were higher for the former than the latter for both adults and children (Ohde and Haley, 1997). Thus, the data from the current study as well as previous research (Ohde and Haley, 1997) indicate that this target is not the only property contributing to the identity of a vowel.

The developmental information derived from the current study also relates to the target normalization and dynamic specification theories. Although participants gave some weight to both static and dynamic information, children did give different perceptual weight to static onsets than adults did when this cue was in the context of changes in the [a]-onset dynamic transition condition. The static onset was not the target per se, but it was canonical in that it corresponded exactly to the expected set of formants within that consonantal context. Since the children were influenced by this aspect of the vowel, it is possible that they have an immature perceptual process which relies more heavily than adults' on static onsets and less than adults' on dynamic information.

The Developmental Weighting Shift hypothesis described by Nittrouer (1996) contains two tenets. The first tenet asserts that children's perceptual weighting of various acoustic parameters differs from adults, becoming more adult-like with experience in the child's native language. This tenet implies that children's perception differs from that of adults not because of auditory immaturity but because children use a different strategy or lack the strategy that adults have for perceptual weighting. Nittrouer's research supported this tenet, showing that 3-yr-olds demonstrated perceptual strategies, which were different from adults and partially but not entirely attributable to an immature auditory system. In two studies (Nittrouer, 1996; Nittrouer and Crowther, 1998), age related differences in auditory processing were examined. In the first study, Nittrouer tested 3-yr-old children's auditory processing abilities for acoustic cues relevant to fricative vowel syllables. The findings of this study revealed that although 3-yr-old children were less sensitive to these cues than adults, auditory sensitivity could not account entirely for age differences in perceptual weighting. In the second study (Nittrouer and Crowther), difference limens (DLs) were obtained for 5- and 7-yr-old children and adults for non-speech dynamic-spectral, static-spectral, and temporal acoustic properties. The hypotheses that children will be more sensitive than adults to dynamic-spectral cues but less sensitive than adults to static cues, and that children will be more sensitive to dynamic-spectral than to static-spectral cues, but adults will be more sensitive to static-spectral than to dynamic cues were tested. The Nittrouer and Crowther study suggested that children would be more sensitive than adults to dynamic spectral cues and least sensitive than adults to static cues. Results of their research did not support the hypotheses. Overall, children had larger DLs than adults in both static and dynamic conditions, and both groups had larger DLs for dynamic-spectral than to static-spectral cues.

The second tenet of the Developmental Weighting Shift hypothesis indicates that children utilize dynamic transi-

tional information more than adults do in their identification of phones. This perceptual bias of children to formant transitions may be due to either a developmental effect where they just attend to transitions more than adults or due to an acoustic prominence where they weight transitions more than other cues to a sound feature (Mayo and Turk, 2004, 2005). The current study does not support tenet 2 of the Developmental Weighting Shift hypothesis because the perceptual bias to dynamic formant transitions by children was contradicted in the [a]-onset condition. In addition, children and adults were not significantly different in the identification of the static-onset condition. The current research does corroborate tenet 1 of this hypothesis in that children used a different strategy from adults in their perceptual weighting as reflected in the [a]-onset condition, but used static cues most readily.

Previous research (Parnell and Amerman, 1978; Murphy *et al.*, 1989; Nittrouer, 1996; Ohde *et al.*, 1996; Ohde and Haley, 1997) revealed that children used dynamic more than static cues in identification, relative to adults. The present study showed the opposite, that children used dynamic information less than adults did, and relied more on static information agreeing in a general way with Sussman's (2001) findings. The static cues in the current research were quantitatively and qualitatively different from the static cues in Sussman's research. The major differences between the static cues in the current study and those in Sussman's study were their short duration (30 ms) and synthetic type of stimulus used. It should be noted that the 30 ms static onset cue in the current study was very accurately perceived by children and adults. Although Sussman's static cue was longer in duration than those in the current study, identification was high for the static cues in both studies.

Sussman (2001) pointed out that most of the research that Nittrouer (1996) used to support her hypothesis focuses on consonants, not vowels. Ohde *et al.* (1995), focusing on stop consonant identification, found that children were generally as accurate with static as they were with dynamic cues. Thus, this research agrees indirectly with the current findings. Ohde *et al.* (1996) and Ohde and Haley (1997) found that accuracy of vowel perception was affected more in children than adults by consonantal context and stimulus duration. These findings may be reconcilable with the current data because they also imply a different perceptual weighting strategy for children; the current study eliminated both of these cues indicated by Ohde and Haley (1997) as factors by making duration and consonantal context the same for all stimuli. If participant age influences cue weighting of vowels, then the dynamic to static cue emphasis shift may occur at a younger age than the 3- to 4-yr-old children in the current study. Overall, the current study supports the concept of the Developmental Weighting Shift hypothesis by showing that 3- to 4-yr-old children have a different strategy of weighting various acoustic cues than adults do but differs from this hypothesis regarding the type of signal that appears to be perceived first.

Although it seems clear from these research findings that children employ a perceptual weighting strategy, that is, different from adults' in identifying vowels based on spectral

information, several questions remain regarding the specifics of this process. One important issue relates to why these children paid greater attention to static than to dynamic cues. The relevance of static cues in development is supported by generally similar perception of the static-onset condition by children and adults, and the dissimilar perception of the [i]-onset and [a]-onset conditions by children and adults. The findings show that children and adults are significantly different in their placement of category boundaries for the [a]-onset condition. It is possible that static attributes for vowels are simply more prominent to children by nature, or it is also possible that the saliency of these cues occurred because of the temporal primacy of the information. In other words, did the children pay more attention to the static attributes because they occurred first in the presentation of the dynamic-onset conditions? The temporal order of cues to consonants is important. Studies of consonant features have shown that place of articulation is more prominently represented in consonant-vowel (CV) syllables than vowel-consonant (VC) syllables (Redford and Diehl, 1999; Ohde *et al.*, 2006). The strength of the cues to consonant identification in the CV syllable appears to be based on the spectral discontinuity occurring at the boundary between the consonant and vowel. This boundary cue is rich in phonetic content at the spectral discontinuity and forms an acoustic landmark where acoustic parameters are examined and cues are perceptually extracted (Chang and Blumstein, 1981; Stevens, 2002).

It is also possible that there was an influence of specific formants on the overall perception of the vowel. For example, Sussman (2001) states that either the low F1 (Elliott and Katz, 1980; Schneider *et al.*, 1986) or patterns of the low-frequency components (Syrdal and Gopal, 1986) may have created difficulties for language-impaired children in perceiving the [i] vowel. Previous research by Ohde and Camarata (2004) also shows that language-impaired children have greater difficulty perceiving the front [i] vowel than the back [a] vowel. The experimental tasks used in the Ohde and Camarata study were identical to those used in the current study. The current findings also show that children performed differently from adults in the identification and categorization of the [i] vowel. For example, the children's slopes were significantly shallower than adults' slopes for the [i]-onset condition. These results show that children were not as confident in deciding if a stimulus was [i] or [a] in the [i]-onset condition and to some extent the [a]-onset condition. Thus, the observed perceptual variations between children and adults for [i]-onset stimuli could relate, as indicated by Sussman (2001), to differences in perception of specific formants or patterns of formants, but more evidence supporting this view is needed.

Finally, the differences found between children and adults in cue weighting of vowels may be due to different cue integration strategies for these populations. In fact, it has been proposed that integration strategies are learned in childhood (Recasens and Marti, 1990; Ohde and Perry, 1994). Previous research shows substantial differences between children and adults in the perceptual integration of vowel and consonant cues from the burst + aspiration + vowel transition segment (Parnell and Amerman, 1978; Alaskary,

2002). Young 4- to 5-yr-old children were significantly poorer than older children and adults at identifying vowels from the short beginning segment. The percent vowel identification was 55, 85, and 83 for the 4- to 5-yr-olds, 11-yr-olds, and adults, respectively. The 11-yr-olds and adults were able to integrate vowel cues from the consonant burst and vowel transitions, which resulted in about a 30% increase in vowel identification for 11-yr-old children and adults compared to 4- to 5-yr-old children. Thus, the young children were unable to integrate vowel information from different acoustic sources such as the static noise and dynamic formant transition similar to older participants. If formant onsets provide important cues to vowel perception as predicted from findings for acoustic vowel onsets (Sussman, 1990), then these onsets may play a unique role in the development of vowel perception as supported in the present study. However, if later occurring transition information is weighted more heavily than formant onsets as predicted by the Developmental Weighting Shift hypothesis, then vowel formant transitions may actually compete with vowel cues in formant onsets over a developmental period. The results of the present research support this role of competition or perceptual calibration in the perceptual development of vowels, as well as the prediction that perceptual integration develops across age. In the case of vowel perception, children appear to treat onsets and formant transitions separately and do not integrate these cues into a unitary vowel percept similar to adults' vowel perception. The findings showing that children were more influenced by stimulus onsets than adults support an early process of perceptual integration of formant onsets and dynamic formant transitions, that is, different for children and adults. As addressed in our second research question, both children and adults use formant onset frequencies as important perceptual cues to the feature distinction of front/high versus back/low vowels.

## V. CONCLUSIONS

In conclusion, the current research focused on four questions that addressed aspects of developmental vowel perception based on static and dynamic spectral properties. This research supported hypotheses two through four. As restated for clarity, hypotheses two, three, and four indicate that the acoustic correlates of place of articulation will be more salient in the [a] vowel than the [i] vowel (Sussman, 2001), perceptual variability will be greater in children than adults (Thibodeau and Sussman, 1979; Ohde *et al.*, 1996; Ohde and Haley, 1997), and formant onsets and formant transitions follow a developmental course that includes a mechanism of perceptual integration (Ohde and Camarata, 2004; Stevens, 2000). Results indicated (1) that place of articulation was more salient for the [a] vowel than [i] vowel, (2) perceptual variability based on slopes was greater in children than adults for the [i] vowel, and (3) formant onsets and formant transitions follow a developmental course that may include a mechanism of perceptual integration as indicated below. Contrary to hypothesis one that predicted a greater perceptual influence of formant transitions than formant onsets, children's category boundaries for the [a]-onset condition

were significantly different from adults' boundaries. These boundary differences were due to a perceptual effect of formant onsets in the [a]-onset condition. As the formant onsets became more [a]-like than [i]-like, children shifted their responses from [i] to [a] before adults did indicating a greater weighting of formant onsets by children than adults. The fact that boundaries were significantly different between children and adults in the [a]-onset condition, supports a fundamentally different integration and/or weighting process for children than adults. In addition, the results for the static onset condition were also contrary to hypothesis one because children and adults were identical in their perception of static formant onsets. Apparently, under some stimulus conditions, children are in no more need of formant transitions to derive veridical perception than adults.

The above results may represent important developmental issues of cue integration. If formant onsets and transitions perceptually integrate, then it is reasonable to predict that perception should not be substantively different for adults and children. Since perception of category boundaries was different for children and adults in the [a]-onset condition, it is important to determine what type of perceptual process or mechanism accounts for these differences. In the [a]-onset condition, formant onsets and formant transitions were perceptually different, and this difference may reflect an immature integration mechanism. This mechanism must develop over time, but may not be adult-like by even 7 years of age (Stevens, 2000).

Finally, there is at least another process that appears to impact perceptual development of vowels, at least in an indirect way, and, that is, variability. As the current results show, a vowel containing a predominance of high frequency energy like [i] affected children's confidence in judging these sounds and resulted in significantly shallower slopes for children than adults. This means that young children at least in the [i] vowel context could not respond as definitively to the [i] to [a] continuum as adults did. This may be an acoustic effect that delays perceptual learning, i.e., difficulty in perceiving high frequencies. This may not affect category boundaries until features are learned in high frequency contexts such as [i]. Once the features are learned in the [i]

context, variability may decrease, that is, children's shallow identification slopes become more adult-like with categories shifting until integration becomes adult like.

## ACKNOWLEDGMENTS

This research was supported in part by a NIH Grant DC00523-08. The authors express their appreciation to Dan Ashmead and Edward Conture for comments on earlier drafts of this paper, and to Ben Hornsby for assistance in the preparation of figures. We extend our sincere thanks to the adults and the children and their parents who participated in this research, without whose cooperation, help, and patience such study could not have been completed.

## APPENDIX: ONSET, OFFSET, AND TARGET FREQUENCIES (IN Hz) FOR STIMULI IN EACH CONDITION

This appendix contains the frequency values for the three experimental conditions entitled static-onset (Table III), [i]-onset (Table IV), and [a]-onset (Table V).

TABLE III. Static-onset condition.<sup>a</sup>

	F1 onset	F1 offset/F1 target	Static F2	Static F3	Static F4
S1	<b>190</b>	330/330	1800	2600	3200
S2	"	375/375	1700	2520	3235
S3	"	420/420	1600	2455	3265
S4	"	465/465	1500	2390	3300
S5	"	510/510	1400	2325	3335
S6	"	555/555	1300	2260	3365
S7	"	600/600	1200	2195	3400
S8	"	645/645	1100	2130	3435
S9	"	690/690	1000	2065	3465
S10	"	720/720	900	2000	3500

<sup>a</sup>In the static-onset condition, onset frequency values range from those for [i] to those for [a]. For static F2, F3 and F4, onset, offset and target frequencies are the same. F1 target changed in the static condition in order to give the resulting vowel a stop-like quality (Ohde and Haley, 1997; Ohde and Abou-Khalil, 2001). For each stimulus, F1 offset and F1 target frequencies were the same because F1 duration was set at an appropriate 20 ms for a labial stop. This resulted in identical values for F1 offset and F1 target.

TABLE IV. [i]-onset condition.

	F1 onset	F1 offset/F1 target	F2 onset	F2 offset/F2 target	F3 onset	F3 offset/F3 target	F4 onset	F4 offset/F4 target
S1	<b>190</b>	330/330	<b>1800</b>	2050/2200	<b>2600</b>	2850/3000	<b>3200</b>	3450/3600
S2	"	370/370	"	1988/2100	"	2819/2950	"	3444/3590
S3	"	410/410	"	1925/2000	"	2788/2900	"	3438/3580
S4	"	450/450	"	1863/1900	"	2756/2850	"	3431/3570
S5	"	490/490	"	1800/1800	"	2725/2800	"	3425/3560
S6	"	530/530	"	1737/1700	"	2694/2750	"	3419/3550
S7	"	570/570	"	1675/1600	"	2663/2700	"	3413/3540
S8	"	610/610	"	1612/1500	"	2631/2650	"	3406/3530
S9	"	650/650	"	1550/1400	"	2600/2600	"	3400/3520
S10	"	690/690	"	1487/1300	"	2568/2550	"	3394/3510
S11	"	720/720	"	1425/1200	"	2537/2500	"	3388/3500

TABLE V. [a]-onset condition.<sup>a</sup>

	F1 onset	F1 offset/F1 target	F2 onset	F2 offset/F2 target	F3 onset	F3 offset/F3 target	F4 onset	F4 offset/F4 target
S1	<b>190</b>	330/330	<b>900</b>	1713/2200	<b>2000</b>	2625/3000	<b>3500</b>	3563/3600
S2	"	370/370	"	1650/2100	"	2594/2950	"	3556/3590
S3	"	410/410	"	1588/2000	"	2563/2900	"	3550/3580
S4	"	450/450	"	1525/1900	"	2531/2850	"	3544/3570
S5	"	490/490	"	1463/1800	"	2500/2800	"	3538/3560
S6	"	530/530	"	1400/1700	"	2469/2750	"	3531/3550
S7	"	570/570	"	1338/1600	"	2438/2700	"	3525/3540
S8	"	610/610	"	1275/1500	"	2406/2650	"	3519/3530
S9	"	650/650	"	1213/1400	"	2375/2600	"	3513/3520
S10	"	690/690	"	1150/1300	"	2344/2550	"	3506/3510
S11	"	720/720	"	1088/1200	"	2313/2500	"	3500/3500

<sup>a</sup>In the dynamic conditions (Tables IV and V), onset information (either values for [i] or values for [a]) is combined with dynamic transition information. The target (ranging between values for [i] and values for [a]) is not reached because total stimulus duration was 30 ms, whereas the target would be obtained in 40 ms. For Table IV ([i]-onset condition), onset frequencies are for [i]. For Table V ([a]-onset condition), onset frequencies are for [a].

- Alaskary, H. (2001). "The role of static and dynamic cues in the identification of voiceless stop consonants by children and adults," Doctoral dissertation, Vanderbilt University, Nashville, Tennessee.
- Blumstein, S. E., and Stevens, K. N. (1980). "Perceptual invariance and onset spectra for stop consonants in different vowel environments," *J. Acoust. Soc. Am.* **67**, 648–662.
- Bourland Hicks, C., and Ohde, R. N. (2005). "Developmental role of static, dynamic, and contextual cues in speech perception," *J. Speech Lang. Hear. Res.* **48**, 960–974.
- Chang, S., and Blumstein, S. E. (1981). "The role of onsets in perception of stop place of articulation: Effects of spectral and temporal discontinuity," *J. Acoust. Soc. Am.* **70**, 39–44.
- Elliott, L. L., and Katz, D. (1980). "Children's pure-tone detection," *J. Acoust. Soc. Am.* **67**, 343–344.
- Finney, D. J. (2009). *Probit Analysis: A Statistical Treatment of the Sigmoid Response Curve* (Cambridge University Press, Cambridge), pp. 272.
- Fudala, J. B. (2000). *Arizona Articulation Proficiency Scale, Third Revision* (Western Psychological Services, Los Angeles, CA), pp. 60.
- Hresko, W. P., Reid, D. K., and Hammill, D. D. (1991). *Test of Early Language Development*, 2nd ed. (Pro-Ed, Austin, TX), 68 pp.
- Jenkins, J. J., Strange, W., and Edman, T. R. (1983). "Identification of vowels in 'vowelless' syllables," *Percept. Psychophys.* **34**, 441–450.
- Klatt, D. H. (1980). "Software for a cascade/parallel formant synthesizer," *J. Acoust. Soc. Am.* **67**, 971–995.
- Klatt, D. H., and Klatt, L. C. (1990). "Analysis, synthesis, and perception of voice quality variations among female and male talkers," *J. Acoust. Soc. Am.* **87**(2), 820–857.
- Kubaska, C. A., and Aslin, R. N. (1985). "Categorization and normalization of vowels by 3-year-old children," *Percept. Psychophys.* **37**, 355–362.
- Kuhl, P. K. (1979). "Speech perception in early infancy: Perceptual constancy for spectrally dissimilar vowel categories," *J. Acoust. Soc. Am.* **66**, 1668–1679.
- Kuhl, P. K. (1983). "Perception of auditory equivalence classes for speech in early infancy," *Infant Behav. Develop.* **6**, 263–285.
- Kuwabara, H. (1985). "An approach to normalization of coarticulation effects for vowels in connected speech," *J. Acoust. Soc. Am.* **77**, 686–694.
- Malech, S. R., and Ohde, R. N. (2003). "Cue weighting of static and dynamic vowel properties in children versus adults," *J. Acoust. Soc. Am.* **113**, 2257(A).
- Mayo, C., and Turk, A. (2004). "Adult-child differences in acoustic cue effects are influenced by segmental context: Children are not always perceptually biased toward transitions," *J. Acoust. Soc. Am.* **115**(6), 3184–3194.
- Mayo, C., and Turk, A. (2005). "The influence of spectral distinctiveness on acoustic cue weighting in children's and adults' speech perception," *J. Acoust. Soc. Am.* **118**(3), 1730–1741.
- Murphy, W. D., Shea, S. L., and Aslin, R. N. (1989). "Identification of vowels in 'vowelless' syllables by 3-year-olds," *Percept. Psychophys.* **46**, 375–383.
- Nearey, T. M. (1989). "Static, dynamic, and relational properties in vowel perception," *J. Acoust. Soc. Am.* **85**, 2088–2113.
- Nittrouer, S. (1992). "Age-related differences in perceptual effects of formant transitions within syllables and across syllable boundaries," *J. Phon.* **20**, 351–382.
- Nittrouer, S. (1996). "Discriminability and perceptual weighting of some acoustic cues to speech perception by 3-year-olds," *J. Speech Hear. Res.* **39**, 278–297.
- Nittrouer, S. (2002). "Learning to perceive speech: How fricative perception changes, and how it stays the same," *J. Acoust. Soc. Am.* **112**, 711–719.
- Nittrouer, S., and Crowther, C. S. (1998). "Examining the role of auditory sensitivity in the developmental weighting shift," *J. Speech Hear. Res.* **41**, 809–818.
- Nittrouer, S., and Miller, M. E. (1997). "Predicting developmental shifts in perceptual weighting schemes," *J. Acoust. Soc. Am.* **101**, 2253–2266.
- Nittrouer, S., Miller, M. E., Crowther, C. S., and Manhart, M. J. (2000). "The effect of segmental order on fricative labeling by children and adults," *Percept. Psychophys.* **62**, 266–284.
- Ohde, R. N., and Abou-Khalil, R. (2001). "Age differences for stop-consonant and vowel perception in adults," *J. Acoust. Soc. Am.* **110**(4), 2156–2166.
- Ohde, R., and Camarata, S. (2004). "Cue weighting of static and dynamic vowel properties by adults and children with normal language and specific language impairment," *Conference Proceedings – From Sound to Sense: 50+ Years of Discoveries in Speech Communication* (Research Laboratory of Electronics, MIT, Cambridge, MA), pp. 191–196.
- Ohde, R. N., and Haley, K. L. (1997). "Stop consonant and vowel perception in 3- and 4-year-old children," *J. Acoust. Soc. Am.* **102**(6), 3711–3722.
- Ohde, R. N., Haley, K. L., and Barnes, C. M. (2006). "Perception of the [m]-[n] distinction in consonant-vowel (CV) and vowel-consonant (VC) syllables produced by child and adult talkers," *J. Acoust. Soc. Am.* **119**, 1697–1711.
- Ohde, R. N., Haley, K. L., and McMahon, C. W. (1996). "A developmental study of vowel perception from brief synthetic consonant-vowel syllables," *J. Acoust. Soc. Am.* **100**(6), 3813–3824.
- Ohde, R. N., Haley, K. L., Vorperian, H. K., and McMahon, C. W. (1995). "A developmental study of the perception of onset spectra for stop consonants in different vowel environments," *J. Acoust. Soc. Am.* **97**(6), 3800–3812.
- Ohde, R. N., and Perry, A. H. (1994). "The role of short-term and long-term auditory storage in processing spectral relations for adult and child speech," *J. Acoust. Soc. Am.* **96**, 1303–1313.
- Ohde, R. N., and Sharf, D. J. (1988). "Perceptual categorization and consistency of synthesized /r-/ continua by adults, normal children, and /r/-misarticulating children," *J. Speech Hear. Res.* **31**, 556–568.
- Parnell, M. M., and Amerman, J. (1978). "Maturational influences on perception of coarticulatory effects," *J. Speech Hear. Res.* **21**, 682–701.
- Perry, T. L., Ohde, R. N., and Ashmead, D. H. (2001). "The acoustic bases for gender identification from children's voices," *J. Acoust. Soc. Am.* **109**, 2988–2998.
- Peterson, G. E., and Barney, H. E. (1952). "Control methods used in a study of vowels," *J. Acoust. Soc. Am.* **24**, 175–184.
- Recasens, D., and Marti, J. (1990). "Perception of unreleased final nasal consonants," *J. Acoust. Soc. Am.* **3**, 287–299.

- Redford, M. A., and Diehl, R. L. (1999). "The relative perceptual distinctiveness of initial and final consonants in CVC syllables," *J. Acoust. Soc. Am.* **106**, 1555–1565.
- Schneider, B., Trehub, S., Morrongiello, B., and Thorpe, L. (1986). "Auditory sensitivity in preschool children," *J. Acoust. Soc. Am.* **79**, 447–452.
- Stevens, K. N. (2000). "Diverse acoustic cues at consonantal landmarks," *Phonetica* **57**, 139–151.
- Stevens, K. N. (2002). "Toward a model for lexical access based on acoustic landmarks and distinctive features," *J. Acoust. Soc. Am.* **111**, 1872–1891.
- Strange, W. (1989a). "Dynamic specification of coarticulated vowels spoken in sentence context," *J. Acoust. Soc. Am.* **85**(5), 2135–2153.
- Strange, W. (1989b). "Evolving theories of vowel perception," *J. Acoust. Soc. Am.* **85**(5), 2081–2087.
- Strange, W., Jenkins, J. J., and Johnson, T. L. (1983). "Dynamic specification of coarticulated vowels," *J. Acoust. Soc. Am.* **68**, 1622–1625.
- Strange, W., Verbrugge, R. R., Shankweiler, D. P., and Edman, T. R. (1976). "Consonantal environment specifies vowel identity," *J. Acoust. Soc. Am.* **60**, 213–224.
- Sussman, H. M. (1990). "Acoustic correlates of the front/back distinction: A comparison of transition onset versus steady state," *J. Acoust. Soc. Am.* **88**, 87–96.
- Sussman, J. E. (2001). "Vowel perception by adults and children with normal language and specific language impairment: Based on steady states or transitions?" *J. Acoust. Soc. Am.* **109**(3), 1173–1180.
- Sussman, J. E., and Carney, A. E. (1989). "Effects of transition length on the perception of stop consonants by children and adults," *J. Speech Hear. Res.* **32**, 151–160.
- Syrdal, A. K., and Gopal, H. S. (1986). "A perceptual model of vowel recognition based on the auditory representations of American English vowels," *J. Acoust. Soc. Am.* **79**, 1086–1100.
- Thibodeau, L. M., and Sussman, H. M. (1979). "Performance on a test of categorical perception of speech in normal and communication disordered children," *J. Phon.* **7**, 375–391.
- Verbrugge, R. R., Strange, W., Shankweiler, D. P., and Edman, T. R. (1976). "What information enables a listener to map a talker's vowel space?" *J. Acoust. Soc. Am.* **60**, 198–212.