

# The role of temporal and dynamic signal components in the perception of syllable-final stop voicing by children and adults<sup>a)</sup>

Susan Nittrouer<sup>b)</sup>

Utah State University UMC 6840, Logan, Utah 84322-6840

(Received 11 September 2003; revised 6 January 2004; accepted 9 January 2004)

Adults whose native languages permit syllable-final obstruents, and show a vocalic length distinction based on the voicing of those obstruents, consistently weight vocalic duration strongly in their perceptual decisions about the voicing of final stops, at least in laboratory studies using synthetic speech. Children, on the other hand, generally disregard such signal properties in their speech perception, favoring formant transitions instead. These age-related differences led to the prediction that children learning English as a native language would weight vocalic duration less than adults, but weight syllable-final transitions more in decisions of final-consonant voicing. This study tested that prediction. In the first experiment, adults and children (eight and six years olds) labeled synthetic and natural CVC words with voiced or voiceless stops in final C position. Predictions were strictly supported for synthetic stimuli only. With natural stimuli it appeared that adults and children alike weighted syllable-offset transitions strongly in their voicing decisions. The predicted age-related difference in the weighting of vocalic duration was seen for these natural stimuli almost exclusively when syllable-final transitions signaled a voiced final stop. A second experiment with adults and children (seven and five years old) replicated these results for natural stimuli with four new sets of natural stimuli. It was concluded that acoustic properties other than vocalic duration might play more important roles in voicing decisions for final stops than commonly asserted, sometimes even taking precedence over vocalic duration. © 2004 Acoustical Society of America. [DOI: 10.1121/1.1651192]

PACS numbers: 43.71.Ft, 43.71.An [RLD]

Pages: 1777–1790

## I. INTRODUCTION

In 1955, Denes reported that vowel duration for the noun *use* (as in *the use*) was shorter than vowel duration for the verb *use* (as in *to use*). In a complementary perceptual test he found that the proportion of voiced judgments increased as vowel duration increased. The finding was revolutionary for its time because, contrary to the consensus opinion at that time, it showed that listeners can make phonetic judgments with information other than temporally discrete pieces of the spectral structure. Since that report, the relation between vowel duration and syllable-final consonant voicing has been well-studied in both speech production and perception.

Regarding speech production, there is no question that syllables with voiced final stops generally have longer vowels than syllables with similar phonetic structures in similar contexts, but with voiceless final stops. This effect is so pervasive that Chen (1970) went so far as to call it a *language-universal* phenomenon. Of course, the phenomenon cannot really be universal across languages, even if for the simple reason that some languages do not permit syllable-final obstruents. However, even among those that do, a few languages have been identified that lack a vowel-length distinction based on syllable-final consonant voicing. For example,

Flege and Port (1981) reported that native Arabic speakers do not differentiate vowel length for Arabic words ending in voiced and voiceless final stops. This absence of a vowel-length effect for even a few languages demonstrates that the effect is not an inevitable consequence of syllable production. Nonetheless, most languages that permit syllable-final obstruents demonstrate a vowel-length distinction based on the voicing feature of the final consonant. In particular, studies of English have consistently demonstrated this effect (e.g., Chen, 1970; Crowther and Mann, 1992, 1994; Flege and Port, 1981; House and Fairbanks, 1953; Peterson and Lehiste, 1960). To be completely accurate, the entire voiced portion of a (stressed) syllable is shorter preceding a voiceless obstruent than preceding a voiced obstruent. In addition to the vowel nucleus, the voiced portion may consist of transitions into and out of the vowel nucleus, as well as sonorant consonants (Raphael, Dorman, Freeman, and Tobin, 1975; Raphael, Dorman, and Liberman, 1980). For this reason, the terms *vocalic length* and *vocalic duration* will be used in this manuscript instead of *vowel length* and *vowel duration*.

It is not clear why vocalic duration should be longer when the final obstruent is voiced than when it is voiceless. One reason that has been suggested is that speakers begin the closure gesture sooner for voiceless consonants because these closures require greater force than that required for voiced consonants, and people tend to begin relatively difficult tasks sooner than easier tasks (Malécot, 1970). However, there are numerous arguments against this suggestion, including the simple fact that the vocalic-length distinction for

<sup>a)</sup>Portions of this work were presented at the 141st meeting of the Acoustical Society of America, Chicago, June 2001.

<sup>b)</sup>Electronic mail: nittrouer@cpd2.usu.edu

final-consonant voicing is not universal. Be that as it may, given that the phenomenon is at least prevalent across languages with syllable-final obstruents we would expect vocalic length to be used by listeners of those languages in perception. And, indeed, numerous reports have shown that vocalic duration influences voicing judgments for syllable-final consonants made by adult speakers of languages with a vocalic-length distinction associated with final consonant voicing (Crowther and Mann, 1992, 1994; Denes, 1955; O'Kane, 1978; Raphael, 1972; Raphael *et al.*, 1975; Raphael *et al.*, 1980). Other acoustic properties of syllables ending with voiced or voiceless final stops have also been found to influence adults' voicing judgments (Hogan and Rozsypal, 1980; Summers, 1988; Wardrip-Fruin, 1982), with spectral characteristics associated with the vocal-tract closing gesture apparently being weighted particularly heavily (Hillenbrand, Ingrisano, Smith, and Flege, 1984). In fact, Hillenbrand *et al.* concluded that the release burst and voicing during closure contribute little to voicing decisions for final stops. This conclusion makes sense *prima facie* because in natural speech final stops may not be released, and speakers do not always voice during closure for obstruents (e.g., Klatt, 1976). However, the precise acoustic correlate of the vocal-tract closing gesture that contributes most to voicing decisions for final stops remains unclear. Fischer and Ohde (1990) tried to separate the contributions of first formant (F1) transition rate and F1 frequency at voicing offset to these decisions, and concluded that of the two, F1-offset frequency was weighted more strongly by adult listeners. Unfortunately, they did not manipulate higher formants, which also vary in frequency at voicing offset as a function of consonant voicing (Nittrouer, Estee, Lowenstein, and Smith, submitted). Furthermore, the changing (i.e., dynamic) nature of formant transitions is generally considered to be critical to speech recognition for both adults and children (e.g., Browman and Goldstein, 1990; Miranda and Strange, 1989; Nittrouer, Manning, and Meyer, 1993; Strange, 1989; Sussman, MacNeilage, and Hanson, 1973), owing in part to demonstrations that listeners can understand signals in which sinusoids are substituted for center formant frequencies (e.g., Remez, Rubin, Pisoni, and Carrell, 1981). In these "sinewave" signals, many acoustic properties traditionally associated with phonetic perception are missing. The current study did not separate the influences of formant transitions from discrete frequencies at voicing offset. Nonetheless, we considered entire transitions to have contributed to voicing decisions for final stops, when they contributed at all, because this is the signal property generally considered to influence voicing decisions in similar studies (e.g., Hillenbrand *et al.*, 1984; Wardrip-Fruin, 1982).

The current study focused on children's weighting of vocalic duration and syllable-offset transitions in decisions of voicing for syllable-final stops. This topic was selected for study because earlier work has suggested that young children in the range of three to eight years of age prefer dynamic signal properties to other sorts of properties for making phonetic decisions (e.g., Morrongiello, Robson, Best, and Clifton, 1984; Nittrouer, 1992; Parnell and Amerman, 1978). In particular, three studies have examined voicing decisions for syllable-final stops by three-year-olds, six-year-olds, and

adults (Greenlee, 1980; Krause, 1982; Wardrip-Fruin and Peach, 1984). Evidence across the three studies supports the suggestion that children fail to weight vocalic duration as strongly as adults, instead relying on formant transitions near voicing offset to make voicing decisions about final stops. While this result matches more general suggestions that young children weight dynamic signal properties (i.e., formant transitions) most strongly, and then gradually learn how other acoustic properties signal phonetic identity in their native language, the question of children's perception of syllable-final stop voicing was worth another look because of some irregularities in stimulus construction in the existing three studies. For example, step size on the vocalic-duration continuum in Greenlee was as large as 70 ms for some steps, instead of the 20–30 ms more commonly used in studies with only adult listeners.

One other study has also looked at children's abilities to use vocalic duration in voicing decisions about syllable-final stops. Lehman and Sharf (1989) created synthetic versions of *beet* and *bead* that differed in vocalic duration only (i.e., formant transitions were consistent across stimuli). Adults and children of the ages of five, eight, and ten years labeled these stimuli. The authors reported that category boundaries were similar across groups, but that the functions became steeper with increasing age. This age-related change in slopes indicates that vocalic duration was being weighted more heavily by older listeners, but unfortunately this study could make no comparison of the relative weighting of vocalic duration and syllable-offset transitions because formants were not manipulated. When only one acoustic property varies across stimuli in a labeling task, listeners must turn their perceptual attention to that property, if they are to do the task at all. Because Lehman and Sharf provide no data on how many children attempted to do the task, but failed, we are unable to estimate how many children were unable to make the required shift in their perceptual attention.

The question of how children weight vocalic duration and formant transitions in their voicing decisions for syllable-final stops is particularly intriguing because of the contradictory predictions that would be made based on children's speech perception capacities and on the nature of input to children. Again, studies of children's speech perception lead to the prediction that children would weight formant transitions at voicing offset as much as or more than adults do, but that children would weight vocalic duration less than adults. In addition to developmental studies, this prediction is supported by cross-linguistic data showing that adult native speakers of languages that either fail to have syllable-final obstruents or fail to make a vocalic-length distinction based on the voicing of the final consonant do not weight vocalic duration as much as adult native speakers of languages that make a vocalic-length distinction based on final-consonant voicing (Crowther and Mann, 1992, 1994; Flege and Wang, 1989). Such language-specific results for perceptual weighting strategies indicate that learning must be involved in the acquisition of the strategies under investigation.

On the other hand, Ratner and Luberoff (1984) showed that language directed to infants and toddlers (9 months to 2

years, 3 months) actually exaggerates the vocalic-length distinction, while often deleting the syllable-final consonant itself.<sup>1</sup> The acoustic characteristics that signal specific phonetic distinctions in a language are commonly reported to shape the listening strategies of infants by the time they are roughly one year of age (e.g., Jusczyk, 1997). Therefore, the Ratner and Luberoff finding leads to the prediction that young children would use vocalic length in making their voicing decisions concerning syllable-final stops as much as adult native speakers of the language they are learning. The current study tested these contradictory predictions.

## II. EXPERIMENT 1: REPLICATING AND EXTENDING CROWTHER AND MANN

Crowther and Mann (1992, 1994) presented synthetic tokens of *pot* and *pod* in which the voicing of the final stop was signaled by the duration of the vocalic syllable portion and the offset frequency of F1. They found that listeners whose native language either did not permit syllable-final stops (Mandarin and Japanese) or did not show a difference in vocalic length as a function of the voicing of the final stop (Arabic) weighted vocalic duration less than native speakers of American English. No group differences were observed in the weighting of F1-offset frequency. Crowther and Mann concluded from their results that native-language experience generally shapes one's strategies for speech perception, but that some acoustic properties may be more malleable by experience than others. For this particular phonetic distinction, vocalic duration seems to have been more malleable than F1-offset frequency. Crowther and Mann offered no suggestions for what might make the weighting of vocalic duration more malleable than the weighting of F1-offset frequency, but the finding is in agreement with results from developmental studies. As described in the Introduction, developmental studies of speech perception have found that children generally weight dynamic components of the signal (i.e., formant transitions) more than other signal properties, including temporal properties (e.g., Morrongiello *et al.*, 1984; Nittrouer, 1992; Parnell and Amerman, 1978). In particular, several studies have reported that children between 3 and 6 years of age weight syllable-offset transitions more than adults, but vocalic duration less, in decisions of final-stop voicing (Greenlee, 1980; Krause, 1982; Wardrip-Fruin and Peach, 1984). In sum, using formant transitions for the purpose of making phonetic decisions seems to be the "default" strategy. Apparently people learn to use other acoustic properties for making phonetic decisions through their experiences with a native language. Thus, the adult, non-native speakers of English who served as listeners in Crowther and Mann (1992, 1994) may have been exhibiting these default strategies in their labeling responses. It follows that children, who have less English experience than adults, might be expected to perform similarly to the non-native listeners in those two studies. This experiment tested that prediction.

### A. Method

#### 1. Listeners

Twenty nine adults between the ages of 19 and 39 years participated, as well as 30 children between 7 years, 11

months and 8 years, 5 months, and 29 children between 5 years, 11 months and 6 years, 5 months. In addition, 12 children between 3 years, 11 months and 4 years, 5 months participated, but seven of these children were unable to label stimuli reliably for any of the three sets of stimuli tested. As a result, testing with 4-year-olds was discontinued, and no data from this age group were included.

All participants were native speakers of American English, and had to meet additional criteria to participate. They had to pass a hearing screening of the pure tones 0.5, 1.0, 2.0, 4.0, and 6.0 kHz presented at 25 dB HL. Children needed to score at or above the 30th percentile on the Goldman-Fristoe Test of Articulation, Sounds-In-Words subtest (Goldman and Fristoe, 1986). Children could have had no more than six episodes of otitis media before their second birthday. Adults were administered the reading subtest of the Wide Range Achievement Test—Revised (Jastak and Wilkinson, 1984), and needed to demonstrate at least an 11th grade reading level.

#### 2. Equipment and materials

Testing took place in a soundproof booth, with the computer that controlled the experiment in an adjacent room. The hearing screening was done with a Welch Allen TM262 audiometer and TDH-39 earphones. Stimuli were stored on a computer and presented through a Creative Labs Soundblaster card, a Samson headphone amplifier, and AKG-K141 headphones. The experimenter recorded responses with a keyboard connected to the computer. Two hand-drawn pictures (8 in. × 8 in.) were used to represent each response label in each experiment, such as a *buck* (a male deer) and a *bug*. Gameboards with ten steps were also used with children: they moved a marker to the next number on the board after each block of test stimuli. Cartoon pictures were used as reinforcement and were presented on a color monitor after completion of each block of stimuli. A bell sounded while the pictures were being shown and served as additional reinforcement.

#### 3. Stimuli

Three sets of stimuli were created for this experiment: synthetic *pot/pod*, synthetic *buck/bug*, and natural *buck/bug*. Both sets of synthetic stimuli were created using the Sensyn Laboratory Speech Synthesizer. The synthetic *pot/pod* stimuli were identical to the stimuli used by Crowther and Mann (1992, 1994), except that Crowther and Mann used three settings for F1-offset frequency and we used only two: the highest and lowest values used by them. The vocalic duration of these stimuli varied from 100 to 260 ms, in 20 ms steps. All vocalic portions were preceded by a 50 ms interval of aspiration noise. Fundamental frequency ( $f_0$ ) started at 138 Hz and fell linearly throughout the stimulus to an offset frequency of 95 Hz. F3 was constant throughout at 2460 Hz. F2 was constant at 1160 Hz until 50 ms before offset, at which time it began rising to an ending frequency of 1425 Hz. In all stimuli, F1 was constant at 675 Hz until 50 ms before offset, at which time it fell to either 555 Hz (most

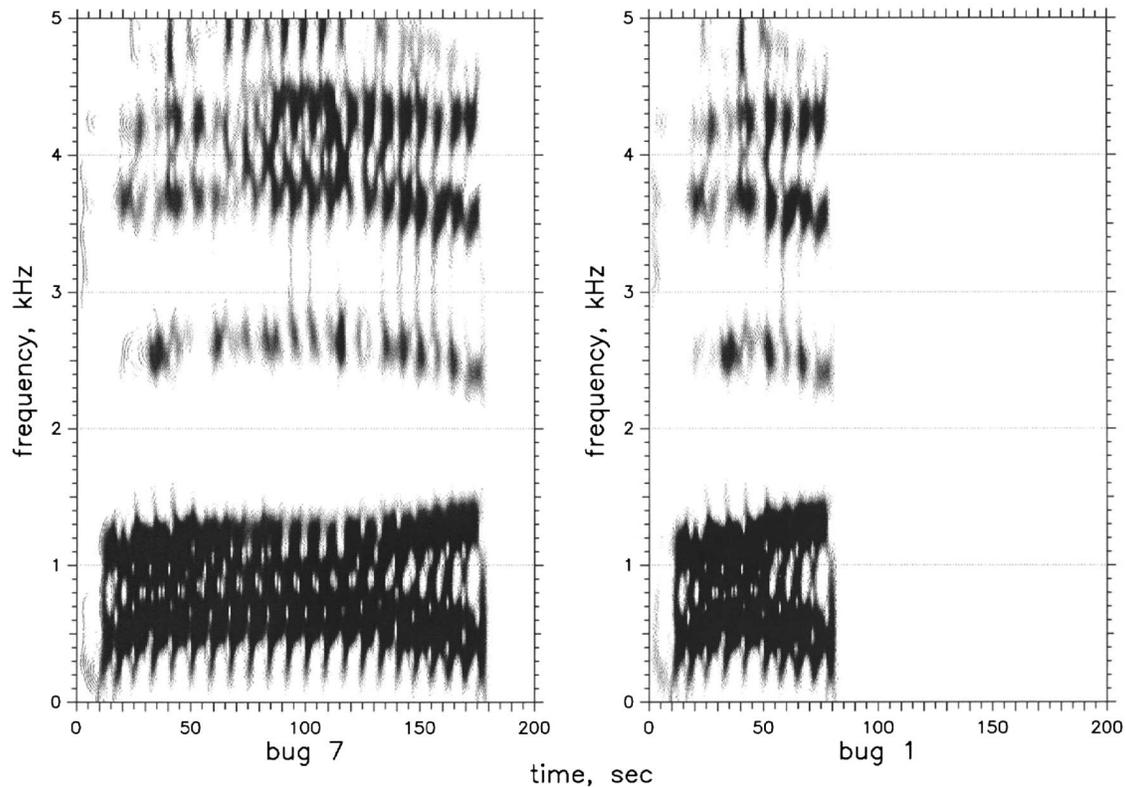


FIG. 1. Spectrograms of the longest (bug 7) and shortest (bug 1) *bug* stimuli created from the same natural token.

*pot*-like) or 355 Hz (most *pod*-like). Thus there were 18 stimuli: two F1-offset frequencies  $\times$  nine vocalic durations.

The synthetic *buck/bug* stimuli were created to be as similar as possible to the *pot/pod* stimuli, while still evoking *buck* and *bug* responses. The vocalic duration of these stimuli also varied from 100 to 260 ms in 20 ms steps. The  $f_0$  began at 138 Hz and fell throughout to 95 Hz. However, both F2 and F3 had offset transitions: F2 was constant at 1000 Hz until 50 ms before offset, at which time it rose linearly to 1800 Hz. F3 was constant at 2700 Hz until 50 ms before offset, at which time it fell to its ending frequency of 2000 Hz. F1 started at 400 Hz, and rose linearly to 600 Hz over the first 50 ms. For half of the stimuli F1 remained at 600 Hz until stimulus offset. For the other half of the stimuli, F1 fell over the final 50 ms to 450 Hz. Thus there were 18 of these stimuli: two F1-offset frequencies  $\times$  nine vocalic durations.

The natural *buck/bug* stimuli were created from natural tokens of an adult, male speaker producing these words in isolation. The speaker was a native speaker of American English. He produced ten tokens of each word, in randomized order. The three tokens of each word that matched each other most closely in vocalic duration and  $f_0$  contour were selected for modification. With each token, the release burst and any voicing during closure was deleted. Vocalic length was then manipulated either by reiterating a single pitch period from the most stable spectral region of the syllable (to lengthen syllables) or by deleting pitch periods from the most stable spectral region of the syllable (to shorten syllables). For both kinds of manipulation, care was taken to align signal portions at zero crossings so no audible clicks resulted. Also, initial and final formant transitions were not disrupted. Seven

stimuli varying in vocalic duration from roughly 85 to 176 ms were created from each token this way. These endpoint values were selected because they match the mean lengths of natural *buck* and *bug*, but clearly the continuum was briefer than that of the synthetic stimuli. The step size was 15 ms (2 pitch periods) on average, but obviously these steps varied slightly according to small differences in  $f_0$  across stimuli. The mean durations of the three stimuli made from *buck* and of the three stimuli made from *bug* at each step were within 5 ms of each other. This procedure for manipulating natural tokens of words ending in voiced and voiceless final stops differs from procedures used in most earlier studies, in that we used both voiced and voiceless tokens, and we did not disrupt offset transitions. Figure 1 shows the longest and shortest stimuli created from the same *bug* token, and shows that this method of modifying stimuli was successful in only affecting the duration of the stable syllable portion. In all, 42 natural stimuli were created: seven vocalic durations  $\times$  two kinds of offset transitions  $\times$  three tokens of each.

#### 4. Procedures

All participants attended two testing sessions. At the first session, screening tasks were completed first. Next, one set of the *buck/bug* stimuli (either synthetic or natural) was presented, with the choice of which set to present randomized across listeners. At the second session, the *pot/pod* and the other set of *buck/bug* stimuli were presented.

The same procedures were followed for each set of stimuli. Practice items were presented before the testing began. Practice items consisted of the best exemplars of each

category, which were the stimuli that should most strongly evoke voiced and voiceless percepts. For example, the best exemplar of *pot* was the stimulus with a 100 ms vocalic portion and the 555 Hz F1 offset. The best exemplar of *pod* was the stimulus with a 260 ms vocalic portion and the 355 Hz F1 offset. For the two synthetic sets of stimuli, there were just two best exemplars (one voiced and one voiceless). Each of these was played five times in random order, and the listener had to respond to at least nine of them correctly to proceed to testing. For the natural *buck/bug* stimuli, there were the six best exemplars (three *buck* and three *bug*). Each of these was played twice, with the 12 stimuli presented in random order. The listener had to respond correctly to at least 11 to proceed to testing.

During testing, ten blocks of stimuli were presented. There were 18 stimuli per block during testing with the two sets of synthetic stimuli. There were 14 stimuli per block during testing with the natural *buck/bug* stimuli. Because there were actually three tokens of each natural stimulus (i.e., each vocalic duration at each level of syllable-offset transitions), the program randomly selected one of the tokens to present during the first block, and then repeated this random selection during the next block without replacement. After the first three blocks, this process was repeated until testing was completed.

Listeners responded by saying the label and pointing to the picture that represented their selection. Having both kinds of responses served as a check that participants were paying attention to the task. In the rare instance that a listener pointed to one picture, but said the other label, the experimenter gave a gentle reminder to pay attention and the stimulus was replayed. To have their data included in the final analysis, participants needed to give at least 80% correct responses to the best exemplars during testing. This requirement was an additional assurance that data were analyzed only from participants who maintained attention to the task.

For children, cartoon pictures were displayed on the monitor and a bell sounded at the end of each block. They moved a marker to the next space on a gameboard after each block as a way of keeping track of how much more time they had left in the test.

Several methods have been used to characterize the weights assigned to various acoustic properties in labeling tasks. Traditionally, each listener's data are plotted as cumulative distributions of the proportion of one response (e.g., *pod* responses) across levels of the acoustic property manipulated in a continuous fashion (vocalic duration in this experiment) for each level of the acoustic property manipulated in a noncontinuous fashion (formant-offset frequencies in this experiment). Best-fit lines are then obtained, often using probit analysis (Finney, 1964), and slopes and distribution means (i.e., phoneme boundaries) computed. This method can extrapolate so that phoneme boundaries outside of the range tested can be obtained. However, we typically impose limits on the values that extrapolated phoneme boundaries can take, restricting them to 3.5 steps beyond the lowest and highest value tested. Because the null hypothesis is that listeners of all ages will show similar differences between pho-

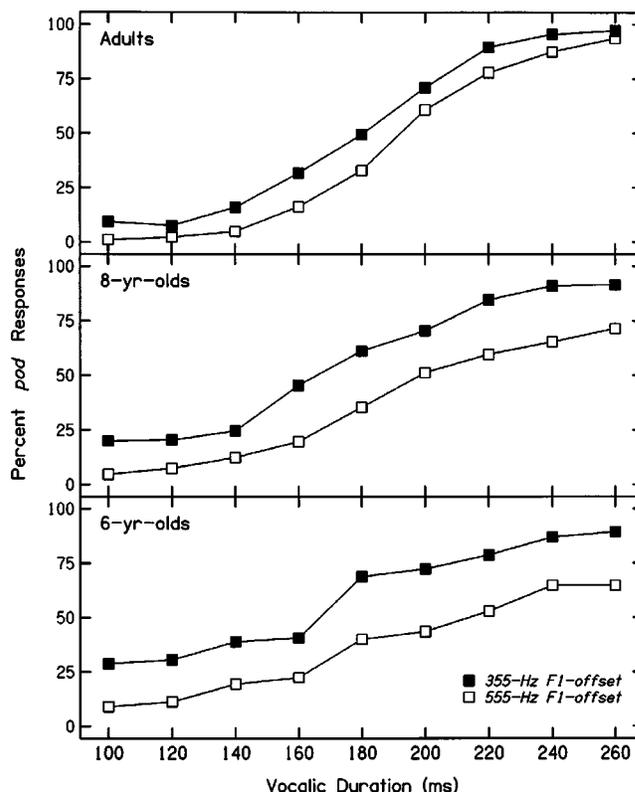


FIG. 2. Mean labeling functions for each age group for synthetic *pot/pod* stimuli, Experiment 1.

neme boundaries, this restriction serves only to constrain the probability of rejecting that null hypothesis. The mean slope of the functions is taken as an indication of the weight assigned to the continuously varied property: the steeper the functions, the more weight that was assigned to that property. The separation between functions at the phoneme boundaries (for each level of the noncontinuous property) is taken as an indication of the weight assigned to that noncontinuous property: the greater the separation, the greater the weight that was assigned. More recently, some investigators (e.g., Turner *et al.*, 1998) have started computing partial correlation coefficients (partial *r*s) for each of the acoustic properties manipulated in the experiment and the proportion of one response option given (i.e., looking at how well each acoustic property predicts responses). Both kinds of metrics were computed in this experiment.

## B. Results

### 1. *Pot/pod*

Data were excluded for twelve 6-year-olds and four 8-year-olds because they either failed to label nine out of ten practice stimuli correctly or did not maintain 80% correct responses during testing itself. Data were included for 29 adults, 26 8-year-olds, and 17 6-year-olds.

Figure 2 shows mean labeling functions from each age group for the synthetic *pot/pod* stimuli. Table I provides two estimates of the weighting of vocalic duration (mean slopes and partial *r*s for vocalic duration) and two estimates of the weighting of F1-offset frequency (separation in phoneme boundaries and partial *r*s for F1-offset frequency). In many

TABLE I. Mean slopes, separations in functions (at phoneme boundaries), and partial  $r$ s (for vocalic duration and F1 offset) for the *pot/pod* labeling task, Experiment 1. *Note:* Standard deviations (SDs) are given in parentheses.

	Adults	8-year-olds	6-year-olds
<i>Weight assigned to vocalic duration</i>			
Mean slope, across functions	0.64 (0.21)	0.43 (0.21)	0.31 (0.18)
Mean partial $r$ for vocalic duration	0.91 (0.04)	0.78 (0.20)	0.72 (0.17)
<i>Weight assigned to F1-offset frequency</i>			
Mean separation in functions	0.93 (0.93)	2.91 (2.99)	3.64 (3.12)
Mean partial $r$ for F1 offset	0.14 (0.14)	0.33 (0.23)	0.41 (0.23)

labeling experiments, slope is given with a physical reference (e.g., change in probit units per ms of change in vocalic duration). In this experiment, however, step size on the vocalic-duration continuum differed for the two synthetic stimulus sets and for the one natural stimulus set (20 ms per step for the synthetic stimuli; 15 ms per step for the natural). For that reason, the slope is given here as the change in probit units per step on the vocalic-duration continuum. Phoneme boundaries represent the step on the vocalic-duration continuum at which the function reaches the 50th percentile. The separation in phoneme boundaries is given here using steps again. Both Fig. 2 and Table I indicate that the weight that was assigned to vocalic duration increased with increasing age: the functions became steeper, and partial  $r$ s for vocalic duration increased. At the same time, the weight that was assigned to the F1-offset frequency diminished with increasing age: the separation in functions decreased, as did partial  $r$ s for F1-offset frequency.

One-way analyses of variance (ANOVAs), with age as the factor, were performed on mean slopes (across the functions with the 555 Hz and 355 Hz F1 offsets), mean separations in phoneme boundaries, and partial  $r$ s for vocalic duration and F1-offset transition. Because the main effect of age was significant for all these measures, post hoc  $t$  tests were also performed. Results of these ANOVAs and  $t$  tests are presented in Table II. Precise results are given for any analysis with  $p$  less than 0.10. When  $p$  is greater than 0.10, results are simply described as nonsignificant (NS). For post hoc  $t$  tests, both actual, computed  $p$  values as well as the Bonferroni significance levels, adjusted for multiple  $t$  tests, are provided. In all cases, the results of the statistical tests generally support impressions of age-related differences gleaned from Fig. 2 and Table I: Children's labeling functions were shallower than those of adults, but were more widely separated. Children's partial  $r$ s were smaller than those of adults for vocalic duration, but were greater for F1-offset. Although it appears from Table I that 8-year-olds performed more like adults than did 6-year-olds, no statistically significant differences were observed between 6- and 8-year-olds on any of the measures.

## 2. Synthetic buck/bug

Fourteen 6-year-olds, 13 8-year-olds, and six adults were unable to reach the criteria for having their data in-

TABLE II. Results of ANOVAs for the *pot/pod* labeling task, Experiment 1. *Note:* Degrees of freedom for the main effect of age are 2, 69. Degrees of freedom for the post hoc  $t$  tests are 69.

	$F$ or $t$	$p$	Bonferroni significance level
<i>Slopes</i>			
Age effect	15.83	<0.001	
6- vs 8-year-olds	-1.80	NS	NS
6-year-olds vs adults	-5.30	<0.001	<0.001
8-year-olds vs adults	-3.92	<0.001	<0.001
<i>Partial <math>r</math>s for vocalic duration</i>			
Age effect	9.69	<0.001	
6- vs 8-year-olds	-1.19	NS	NS
6-year-olds vs adults	-4.07	<0.001	<0.001
8-year-olds vs adults	-3.23	0.002	<0.01
<i>Separation in phoneme boundaries</i>			
Age effect	7.96	<0.001	
6- vs 8-year-olds	0.96	NS	NS
6-year-olds vs adults	3.64	<0.001	<0.01
8-year-olds vs adults	3.01	0.004	<0.05
<i>Partial <math>r</math>s for F1-offset frequency</i>			
Age effect	11.16	<0.001	
6- vs 8-year-olds	1.22	NS	NS
6-year-olds vs adults	4.34	<0.001	<0.001
8-year-olds vs adults	3.51	<0.001	<0.01

cluded on either training or testing. Consequently, data are included for 15 6-year-olds, 17 8-year-olds, and 23 adults.

Figure 3 shows mean labeling functions for each group, and Table III provides the same estimates of the weighting of vocalic duration and F1-offset frequency, as shown in Table I. As found for the *pot/pod* stimuli, children appear to have weighted vocalic duration less than adults, but to have

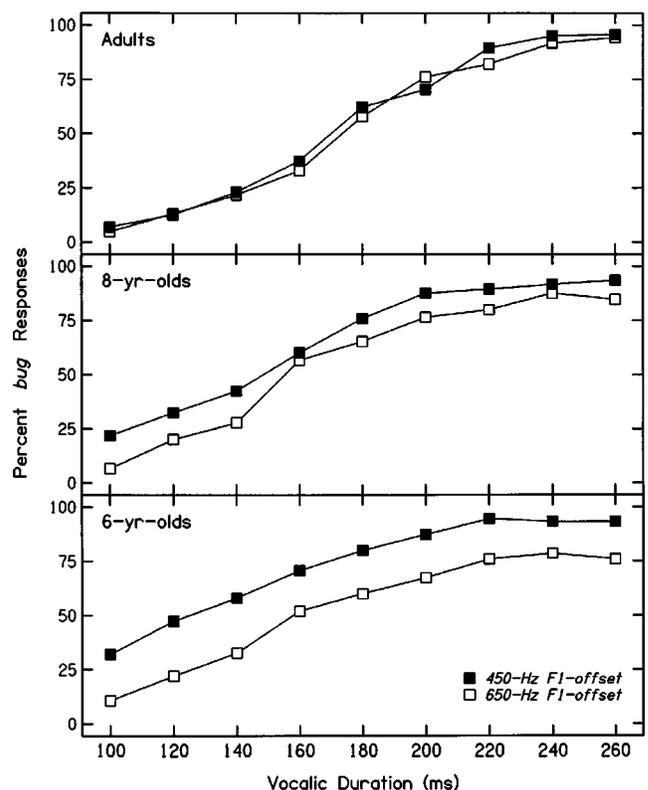


FIG. 3. Mean labeling functions for each age group for synthetic *buck/bug* stimuli, Experiment 1.

TABLE III. Mean slopes, separations in functions (at phoneme boundaries) and partial  $r$ s (for vocalic duration and F1-offset) for the synthetic *buck/bug* labeling task, Experiment 1. Note: Standard deviations (SDs) are given in parentheses.

	Adults	8-year-olds	6-year-olds
<i>Weight assigned to vocalic duration</i>			
Mean slope, across functions	0.53 (0.18)	0.41 (0.19)	0.37 (0.25)
Mean partial $r$ for vocalic duration	0.91 (0.04)	0.83 (0.13)	0.74 (0.13)
<i>Weight assigned to F1-offset frequency</i>			
Mean separation in functions	0.24 (0.82)	1.51 (1.64)	2.80 (1.75)
Mean partial $r$ for F1 offset	0.06 (0.09)	0.18 (0.16)	0.37 (0.16)

weighted F1-offset frequency more. Results for the one-way ANOVAs and post hoc  $t$  tests, shown in Table IV, confirm these impressions. Unlike the *pot/pod* stimuli, however, there were some statistically significant differences in performance between the 6- and 8-year-olds, with the 8-year-olds performing a little more similarly to adults.

### 3. Natural *buck/bug*

One 6-year-old and two 8-year-olds failed to reach criteria on either the training or testing for having their data included. Therefore data were analyzed for 28 6-year-olds, 28 8-year-olds, and 29 adults.

Figure 4 shows mean labeling functions for each age group, and Table V provides the same estimates of the weighting of vocalic duration and formant offsets as provided for the synthetic *pot/pod* and *buck/bug* stimuli. A very different pattern of responding is apparent for these natural stimuli than what was found for the synthetic stimuli. Labeling functions are widely separated, depending on whether

TABLE IV. Results of ANOVAs for the synthetic *buck/bug* labeling task, Experiment 1. Note: Degrees of freedom for the main effect of age are 2, 52. Degrees of freedom for the post hoc  $t$  tests are 52.

	$F$ or $t$	$p$	Bonferroni significance level
<i>Slopes</i>			
Age effect	3.46	=0.039	
6- vs 8-year-olds	-0.53	NS	NS
6-year-olds vs adults	-2.42	=0.019	<0.010
8-year-olds vs adults	-1.93	=0.059	NS
<i>Partial <math>r</math>s for vocalic duration</i>			
Age effect	13.95	<0.001	
6- vs 8-year-olds	-2.61	=0.012	<0.05
6-year-olds vs adults	-5.26	<0.001	<0.001
8-year-olds vs adults	-2.57	=0.013	<0.05
<i>Separation in phoneme boundaries</i>			
Age effect	15.48	<0.001	
6- vs 8-year-olds	2.61	=0.012	<0.05
6-year-olds vs adults	5.53	<0.001	<0.001
8-year-olds vs adults	2.85	0.006	<0.05
<i>Partial <math>r</math>s for F1-offset frequency</i>			
Age effect	22.33	<0.001	
6- vs 8-year-olds	3.80	<0.001	<0.01
6-year-olds vs adults	6.68	<0.001	<0.001
8-year-olds vs adults	2.72	0.001	<0.05

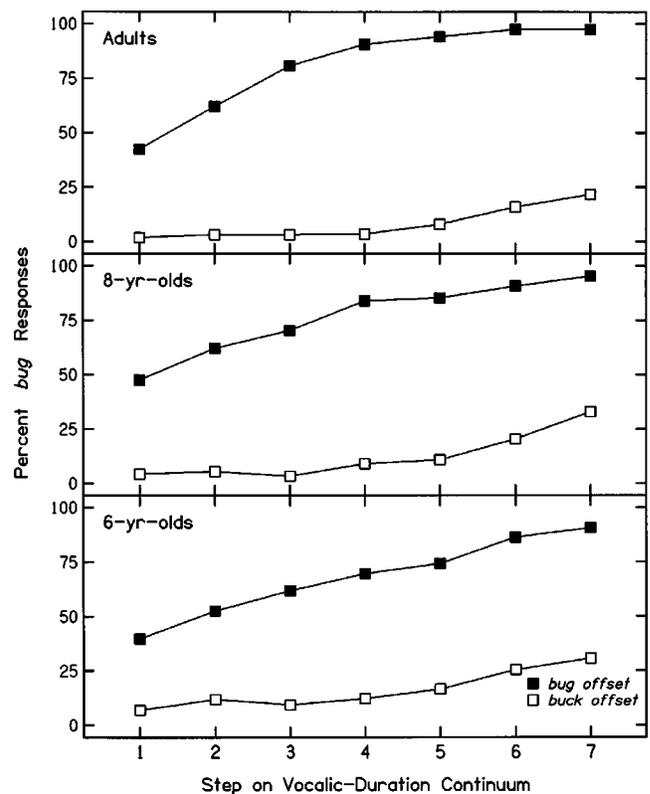


FIG. 4. Mean labeling functions for each age group for natural *buck/bug* stimuli, Experiment 1.

stimuli were created from tokens with voiced or voiceless final stops. This pattern is apparent for listeners in all three age groups.

Table VI shows the results of the one-way ANOVAs conducted on all four measures. Unlike the statistical results for both sets of synthetic stimuli, these results differ for the two measures of each acoustic property. For the measures of weight assigned to vocalic duration (i.e., slopes and partial  $r$ s for vocalic duration), the analysis of slope showed a significant age effect, but the analysis of partial  $r$ s did not. It appears from Fig. 4 that this result is due to adults' function for stimuli with *bug* offsets being steeper than those of either children's group. Functions for stimuli with *buck* offsets appear similar in shape across age groups. One-way ANOVAs conducted on slope for each function separately confirm this

TABLE V. Mean slopes, separations in functions (at phoneme boundaries), and partial  $r$ s (for vocalic duration and F1-offset) for the natural *buck/bug* labeling task, Experiment 1. Note: Standard deviations (SDs) are given in parentheses.

	Adults	8-year-olds	6-year-olds
<i>Weight assigned to vocalic duration</i>			
Mean slope, across functions	0.48 (0.27)	0.33 (0.17)	0.28 (0.21)
Mean partial $r$ for vocalic duration	0.31 (0.18)	0.35 (0.20)	0.40 (0.21)
<i>Weight assigned to formant offsets</i>			
Mean separation in functions	7.89 (2.83)	7.92 (2.61)	7.27 (2.77)
Mean partial $r$ for formant offsets	0.86 (0.15)	0.84 (0.13)	0.74 (0.20)

TABLE VI. Results of ANOVAs for the natural *buck/bug* labeling task, Experiment 1. *Note:* Degrees of freedom for the main effect of age are 2, 82. Degrees of freedom for the post hoc *t* tests are 82.

	<i>F</i> or <i>t</i>	<i>p</i>	Bonferroni significance level
Slopes			
Age effect	5.73	=0.005	
6- vs 8-year-olds	-0.78	NS	NS
6-year-olds vs adults	-3.24	=0.002	<0.01
8-year-olds vs adults	-2.45	=0.017	<0.05
Partial <i>r</i> s for vocalic duration			
Age effect	1.46	NS	
Separation in phoneme boundaries			
Age effect	0.50	NS	
Partial <i>r</i> s for formant offsets			
Age effect	4.75	=0.01	
6- vs 8-year-olds	-2.40	=0.019	<0.10
6-year-olds vs adults	-2.88	=0.005	<0.05
8-year-olds vs adults	-0.46	NS	NS

impression. The main effect of age is significant only for functions for stimuli with *bug* offsets,  $F(2,82)=10.59$ ,  $p < 0.001$ . Post hoc *t* tests done on slopes of functions for stimuli with *bug* offsets reveal the same trends as seen in Table VI: There is no statistically significant difference in slopes for 6- and 8-year-olds' functions, but there are significant differences for 6-year-olds versus adults,  $t(82) = -4.41$ ,  $p < 0.001$  (Bonferroni significance level =  $< 0.001$ ), and for 8-year-olds versus adults,  $t(82) = -3.30$ ,  $p = 0.001$  (Bonferroni significance level =  $< 0.01$ ). In brief, it appears from Fig. 4 that listeners of all ages gave very few *bug* responses when formant offsets failed to signal that the vocal tract was closing (i.e., when stimuli had *buck* formant offsets). However, when formant offsets signaled a closing vocal tract, adults provided *bug* responses more consistently, at shorter vocalic durations than children did.

For the measures of weight assigned to formant offsets (i.e., separations in functions and partial *r*s for formant offsets), the analysis of separations in functions showed no age effect, but the analysis of partial *r*s did. This last result can actually be traced to 6-year-olds weighting formant offsets less than adults or 8-year-olds did.

### C. Discussion

This experiment was undertaken to examine whether children would be found to weight formant transitions more and vocalic duration less than adults in decisions of syllable-final stop voicing, as has been reported by three earlier studies (Greenlee, 1980; Krause, 1982; Wardrip-Fruin and Peach, 1984). For this purpose, the synthetic stimuli of Crowther and Mann (1992, 1994) were replicated, and another set of synthetic stimuli was created that closely matched those stimuli. Stimuli were also created from natural tokens of words ending in voiced and voiceless final stops. Adults' results for both sets of synthetic stimuli closely matched those of Crowther and Mann: vocalic duration explained variation in their responses to a large extent, with F1 transitions explaining little. This result agrees with the traditional view of the role of vocalic duration in decisions of syllable-final consonant voicing. It was found that 6- and 8-year-old

children weighted vocalic duration less than adults, and weighted F1-offset transitions more. This result strongly supports the assertion that children attend largely to dynamic signal components in speech perception, gradually modifying their perceptual strategies to weight other components more as they gain experience with their native language (e.g., Nittrouer, Manning, and Meyer, 1993).

The third set of stimuli included in this experiment was generated from natural tokens of a speaker saying *buck* and *bug*. All information that might normally be available from the closure and release was removed, but differences between stimuli that exist in the syllable nuclei and margins remained. For example, kinematic studies have shown that jaw opening is faster and more extensive for words with voiceless, rather than voiced, final stops (Gracco, 1994; Summers, 1987). It follows that the rate of formant transition at the beginnings of syllables is faster for words with voiceless stops at the end, and maximum F1 frequency is greater for words with voiceless, rather than voiced, final stops. For example, Nittrouer *et al.* (submitted) found that F1 at the syllable center was 58 Hz higher in *buck* than in *bug*. Finally, the rate and final frequency of all formant transitions at the syllable's end can differ, depending on stop voicing. In this experiment, when all these acoustic properties varied naturally, partial *r*s indicated that listeners of all ages weighted vocalic duration very little. Instead, listeners largely weighted the other acoustic properties that correlate with the voicing of the final stop. Based on the studies of others (e.g., Hillenbrand *et al.*, 1984), it seems likely that the property that accounted for most of the variation in listeners' responses was the collective pattern across formants of syllable-final transitions. But there is no way to know that for sure without further experiments. In any event, these results with stimuli created from natural tokens require a tempering of the conclusions reached with synthetic stimuli.

There is some reason to suggest, however, that partial *r*s did not reveal the whole story in this case. When asked to label stimuli created from natural tokens, adults had steeper functions for those stimuli with syllable-offset transitions appropriate for voiced final stops, rather than those with transitions appropriate for voiceless final stops. That is, their *bug* responses approached 100% at briefer vocalic durations, and remained at that level of responding across much of the vocalic-duration continuum. Steep labeling functions are generally taken as an indication that listeners weighted the property represented on the *x* axis strongly, and so it would be reasonable to conclude that adults weighted vocalic duration in their voicing decisions when the final stop indicated complete vocal-tract closure. Children, on the other hand, show no indication of weighting vocalic duration strongly for any stimuli.

Of course, any time results for one set of stimuli are so strikingly different from results for other, similar stimuli, the possibility arises that perhaps those results are spurious. This concern was exacerbated here because the range of vocalic durations used was briefer in the natural stimuli than in either set of synthetic stimuli. Consequently, before the conclusion was firmly reached that vocalic duration actually has little influence on listeners' judgments of syllable-final con-

sonant voicing in natural listening conditions, it seemed important to examine perceptual responses for other stimuli created from natural tokens. For that reason, another experiment was conducted.

### III. EXPERIMENT 2: NATURAL STIMULI

The purpose of this experiment was to test whether adults' and children's labeling responses for a wider range of natural stimuli would reveal similar weighting strategies to those found for the stimuli created from natural tokens in Experiment 1. In this second experiment, four sets of stimuli were created from natural tokens of word pairs ending in voiced or voiceless final stops, using the same procedure as was used to create the natural stimuli in Experiment 1 (i.e., manipulating vocalic duration by reiterating or deleting pitch periods at the most stable spectral region of the syllable).

#### A. Method

##### 1. Listeners

Eleven adults between the ages of 20 and 31 participated. In addition, 13 7-year-olds (between 6 years, 11 months and 7 years, 5 months), 25 5-year-olds (between 4 years, 11 months and 5 years, 5 months) and 17 3½-year-olds (between 3 years, 5 months and 3 years, 11 months) participated. All participants needed to meet the same criteria as those in Experiment 1.

##### 2. Equipment and materials

The same equipment and materials were used in this experiment as in Experiment 1, except that new pictures were created because different stimuli were used.

##### 3. Stimuli

Four sets of stimuli were created, with care taken to vary the place of constriction across vowels and consonants: *cop/cob*, *boot/booed*, *feet/feed*, and *pick/pig*. The same male speaker who provided tokens of the natural *buck* and *bug* used in Experiment 1 provided tokens for use in the creation of these stimuli. The same procedures as used in Experiment 1 were used to modify these stimuli to create seven-step continua, varying in vocalic duration. In particular, care was again taken not to disrupt offset transitions. Of course, the four vowels used in this experiment differ in intrinsic duration, and so the continua differed in range: *cop/cob* varied from 82 to 265 ms; *boot/booed* varied from 97 to 258 ms; *feet/feed* varied from 93 to 255 ms; and *pick/pig* varied from 62 to 178 ms. For each set there were 42 stimuli: seven steps on the vocalic duration continuum × two formant-offset conditions × three tokens of each.

##### 4. Procedures

The procedures for this experiment were essentially the same as for Experiment 1. All participants attended two sessions. However, all four sets of stimuli could be presented to individual listeners only in the adult and 7-year-old groups. This was because only adults and 7-year-olds could complete two sets of stimuli at the first session, after completing the screening procedures. These listeners were then presented

with the final two sets of stimuli at the second session. The order of presentation of sets was randomized across these older participants. The 3½- and 5-year-olds were presented with only three sets of stimuli: one set at the first session (after the hearing and speech screening tasks), and two sets at the second session. The stimuli that would be presented to a child were selected in a serial fashion. That is, the first child tested in each age group was presented with three sets of stimuli, randomly selected. The next child was presented with the fourth set, and the first two sets presented to the first child. The third child tested was then presented with the third and fourth sets, as well as the first, and so on.

One additional training procedure was incorporated into the protocol in this second experiment. Before administering the training with the best exemplars, unaltered stimuli were presented for practice. These unaltered stimuli had whatever voicing during closure was present in the original stimuli, as well as the release bursts. As in practice with the best exemplars, there were six of these stimuli (i.e., three of the word with a voiceless final stop and three of the word with a voiced final stop). Each stimulus was presented two times, and the listener had to respond to 11 of these items correctly in order to receive the training with the best exemplars.

#### B. Results

Data from the 3½-year-olds were not included because not enough children in this age group were able to complete the tasks. Of the 17 children in this age group participating, two failed the hearing screening, and two refused to cooperate at all. Of the remaining 13 children, some became uncooperative on the second day after the presentation of one set of stimuli, and so could not be presented with the other set. In all, these 3½-year-olds were presented with 33 sets of stimuli. As a group they reached the training criterion with the unaltered stimuli for 24 of these sets. Subsequently, the criterion for training with the best exemplars was reached for 17 sets of stimuli. However, 3½-year-olds reached the 80% correct criterion for endpoint stimuli during testing for just five sets of stimuli. That was not enough to provide useful information about the weighting strategies of these young listeners.

One 5-year-old failed the hearing screening, and one failed to reach criterion on the Goldman-Fristoe Test of Articulation. Regarding the training with unaltered stimuli, one 5-year-old failed to reach criterion for *feet/feed* and one failed to reach criterion for both *cop/cob* and *boot/booed*. Several 5-year-olds failed to reach either the best-exemplar training criterion or the testing criterion for having their data included in the analyses. Specifically, four 5-year-olds (out of 18 tested) failed to reach criterion for *pick/pig*; seven (out of 17 tested) failed to reach criterion for *cop/cob*; five (out of 15 tested) failed to reach criterion for *feet/feed*; and eight (out of 18 tested) failed to reach criterion for *boot/booed*.

Only one 7-year-old failed to reach the testing criterion for having data included in the analyses for *feet/feed*, and one failed to reach the testing criterion for *pick/pig*. Data for one adult participant was lost for *cop/cob* because the experimenter neglected to save the data after testing.

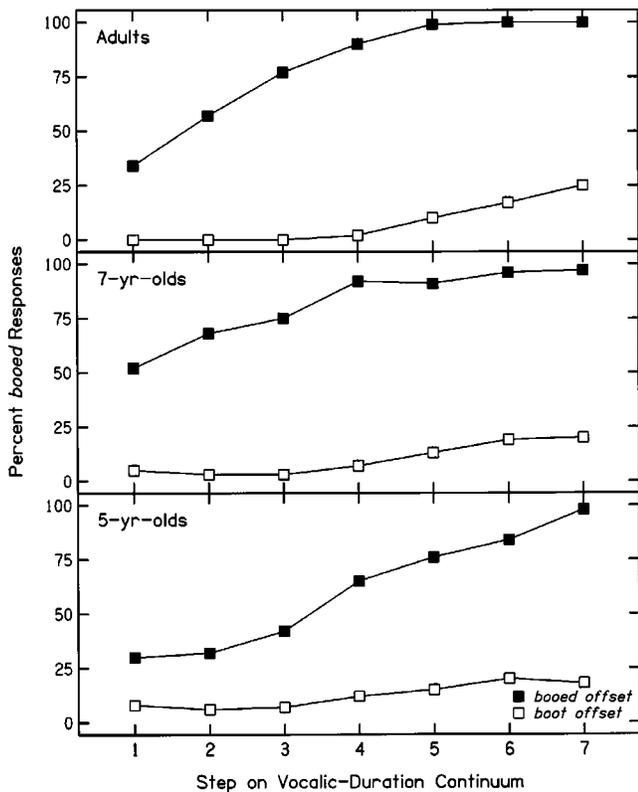


FIG. 5. Mean labeling functions for each age group for natural *boot/booed* stimuli, Experiment 2.

Results for all four sets of stimuli were similar, and are illustrated in Fig. 5, showing mean labeling functions for *boot/booed*. Clearly, listeners in all three age groups showed fairly flat functions with large separations between functions. This pattern of results traditionally is interpreted as indicating that listeners weighted only slightly the property varied continuously (in this case, vocalic duration) and weighted heavily the property varied dichotomously (in this case, syllable-offset transitions).

Two estimates of the weighting of vocalic duration and syllable-offset transitions were used in Experiment 1. Both estimates gave similar results, except where the weighting of vocalic duration in decisions for natural stimuli were concerned. In that case, partial *r*s indicated that adults and children weighted vocalic duration similarly, but slopes showed that adults weighted vocalic duration more than children. This age-related difference was particularly evident when stimuli were created from words with voiced final stops. Be-

TABLE VIII. Mean separations in functions (at phoneme boundaries), Experiment 2. Note: Standard deviations (SDs) are given in parentheses.

	Adults	7-year-olds	5-year-olds
<i>cop/cob</i>	3.75 (2.57)	1.62 (3.84)	2.95 (3.27)
<i>feet/feed</i>	5.01 (2.42)	6.18 (1.35)	6.59 (3.34)
<i>boot/booed</i>	7.02 (2.34)	7.39 (2.51)	6.75 (1.40)
<i>pick/pig</i>	9.42 (2.97)	9.03 (2.42)	9.87 (2.08)

cause of this result, it was concluded that slopes and separations between labeling functions provided more accurately described labeling results. At the same time, however, it is difficult to compare the weighting of the two acoustic properties using slopes and separations between functions because they provide different metrics. Partial *r*s allow us to do this kind of comparison. Consequently, both kinds of estimates are presented for this second experiment, but with an emphasis on slopes and separations between functions.

Table VII shows partial *r*s for each stimulus set. From these estimates it appears that listeners in all age groups performed similarly by weighting syllable-offset transitions strongly while weighting vocalic duration not as much. The one-way ANOVAs (with age as the main effect) done on each set of the partial *r*s provided eight analyses: one for vocalic duration and one for formant offsets for each of the four stimulus sets. Unlike results for natural *buck/bug* in Experiment 1, where a significant age effect was found for partial *r*s of formant offsets, none of these ANOVAs revealed a significant age effect. At least for *cop/cob* and *boot/booed*, both of which had ranges of group means similar to that found for natural *buck/bug* in Experiment 1, this failure to find a significant effect might be due to smaller sample sizes in this second experiment, compared to the first.

Mean separations in functions for each stimulus set are presented in Table VIII. From these estimates it is clear that listeners in all age groups weighted syllable-offset transitions similarly and strongly. The one-way ANOVAs done on separations in phoneme boundaries revealed no significant age effects.

Table IX shows mean slopes across functions, and for stimuli created from words with voiceless and voiced final stops separately. As was found for natural *buck/bug* in Experiment 1, it appears as if there is an age-related increase in

TABLE VII. Partial *r*s for each stimulus set, Experiment 2. Note: Standard deviations (SDs) are given in parentheses.

	Vocalic duration			Formant offsets		
	Adults	7-year-olds	5-year-olds	Adults	7-year-olds	5-year-olds
<i>cop/cob</i>	0.48 (0.20)	0.54 (0.15)	0.55 (0.18)	0.77 (0.17)	0.66 (0.12)	0.66 (0.19)
<i>feet/feed</i>	0.56 (0.20)	0.49 (0.11)	0.52 (0.22)	0.69 (0.18)	0.72 (0.07)	0.69 (0.26)
<i>boot/booed</i>	0.37 (0.22)	0.43 (0.19)	0.46 (0.13)	0.84 (0.15)	0.76 (0.13)	0.71 (0.18)
<i>pick/pig</i>	0.22 (0.18)	0.28 (0.18)	0.22 (0.17)	0.93 (0.08)	0.88 (0.10)	0.90 (0.08)

TABLE IX. Mean slopes, across functions, and for voiceless and voiced stimuli separately, Experiment 2. *Note:* Standard deviations (SDs) are given in parentheses.

	Adults	7-year-olds	5-year-olds
Mean slope, across functions			
<i>cop/cob</i>	0.56 (0.22)	0.37 (0.19)	0.34 (0.16)
<i>feet/feed</i>	0.67 (0.16)	0.44 (0.18)	0.31 (0.24)
<i>boot/bood</i>	0.60 (0.22)	0.36 (0.17)	0.34 (0.12)
<i>pick/pig</i>	0.37 (0.23)	0.26 (0.12)	0.25 (0.14)
Mean slope, voiceless stimuli			
<i>cop</i>	0.49 (0.28)	0.27 (0.31)	0.24 (0.17)
<i>feet</i>	0.45 (0.24)	0.25 (0.22)	0.20 (0.18)
<i>boot</i>	0.35 (0.31)	0.17 (0.21)	0.20 (0.17)
<i>pick</i>	0.22 (0.18)	0.17 (0.17)	0.23 (0.32)
Mean slope, voiced stimuli			
<i>cob</i>	0.62 (0.24)	0.47 (0.16)	0.41 (0.18)
<i>feed</i>	0.88 (0.22)	0.62 (0.23)	0.41 (0.31)
<i>bood</i>	0.85 (0.36)	0.55 (0.37)	0.49 (0.17)
<i>pig</i>	0.52 (0.36)	0.35 (0.21)	0.27 (0.18)

mean slopes, especially for slopes computed across functions and for slopes from stimuli with offset transitions appropriate for voiced final stops. There appears to be less of an age-related increase in slopes for stimuli with offset transitions appropriate for voiceless final stops. Because the analyses of slopes for the natural *buck/bug* in Experiment 1 resulted in a significant age effect for *bug*, but not for *buck*, one-way ANOVAs were computed for stimuli with syllable-offset transitions appropriate for voiced and voiceless final stops separately. As with *buck/bug*, in every case stimuli with syllable-offset transitions appropriate for voiced final stops showed significant, or close to significant, age effects, indicating that adults' functions were steeper than those of children: *bood*,  $F(2,30)=3.79$ ,  $p=0.034$ ; *cob*,  $F(2,29)=3.06$ ,  $p=0.062$ ; *feed*,  $F(2,30)=8.89$ ,  $p<0.001$ ; and *pig*,  $F(2,34)=3.09$ ;  $p=0.058$ . Only one set of stimuli with syllable-offset transitions appropriate for voiceless final stops showed a significant (or close to significant) age effect: *feet*,  $F(2,30)=4.01$ ,  $p=0.029$ . Labeling functions for these stimuli are presented in Fig. 6. Adults showed a slightly stronger inclination to label stimuli with *feet* offset transitions as *feed* at longer vocalic durations than children did. However, this is the only set of stimuli with syllable-offset transitions appropriate for voiceless final stops in which this trend was found, and adults still did not label more than 50% of these stimuli as *feed*, even at the longest durations.

### C. Discussion

This second experiment was undertaken to see if children and adults would demonstrate the labeling pattern ob-

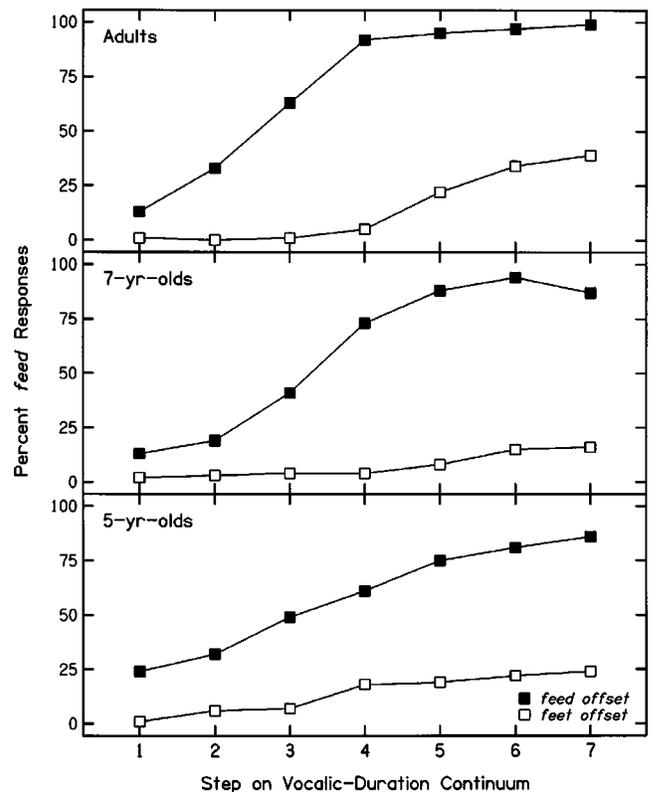


FIG. 6. Mean labeling functions for each age group for natural *feet/feed* stimuli, Experiment 2.

served for stimuli created from natural tokens in Experiment 1: That is, would listeners of all ages weight formant offsets greatly and vocalic duration much less so? Four sets of stimuli were created from natural tokens varying in place of constriction for both the stop and the vowel. For all sets of stimuli, the partial *rs* revealed that adults, 7-year-olds, and 5-year-olds alike weighted formant offsets greatly, and vocalic duration much less so. For all sets of stimuli, adults, 7-year-olds, and 5-year-olds alike showed large separations between functions, depending on whether stimuli were created from words with voiceless or voiced final stops. Because earlier studies have attributed this effect to syllable-offset transitions, we conclude that these transitions are the most likely source of the effect here, as well. For all sets of stimuli, adults, 7-year-olds, and 5-year-olds failed to show particularly steep functions. However, as with the one set of stimuli created from natural tokens in Experiment 1, age-related differences in slope were observed for labeling functions of all stimuli created from words with voiced final stops. Only one of the functions for stimuli created from words with voiceless final stops showed this age effect. In summary, the general pattern of developmental increase in the weighting of vocalic duration was somewhat attenuated for these natural stimuli, but it certainly was not eradicated.

### IV. GENERAL DISCUSSION

The original goal of this study was to examine whether children learning English as their first language would demonstrate the same weighting strategies for words ending in voiced and voiceless stops as non-native English-speaking

adults have demonstrated in experiments by others. Regarding weighting strategies for native English-speaking adults, it was assumed at the outset that the results of other investigators showing that vocalic duration is an important, if not primary, cue to voicing for syllable-final consonants would be easily replicated. Accordingly, the synthetic stimuli of Crowther and Mann (1992, 1994) were used, and a second set of synthetic stimuli created with the same general design. Vocalic durations and F1-offset transitions in these stimuli matched what is found in natural speech samples, but other aspects of acoustic structure were held constant across stimuli. Adults' results for both sets of stimuli did indeed match those obtained by Crowther and Mann from native English-speaking adults: Vocalic duration was weighted strongly, with much less weight given to F1-offset transitions. Children's results for vocalic duration fit predictions derived from studies of non-native English-speaking adults: children weighted vocalic duration less than the native English-speaking adults in this study did. At the same time, children's results for the F1-offset transitions fit predictions derived from other developmental studies: children relied on these dynamic signal components more than the adults did.

If this study had stopped there, longstanding views of human speech perception, and of how it develops, would have been perpetuated. However, modified natural stimuli were also included in this study. With those stimuli, adults and children alike mainly made decisions about final-stop voicing based on some property (or properties) of the natural stimuli other than vocalic duration. The most likely candidate for this other property is syllable-offset transitions—the dynamic components of the signal. This perceptual strategy had been predicted for children. It was adults who performed differently from expectations. Adults' apparently strong reliance on dynamic signal components in these phonetic judgments requires a reconsideration of basic principles regarding speech perception. To be sure, a few investigators in the past have suggested that vocalic duration is not a critical cue to adults' voicing decisions for final consonants (Hillenbrand *et al.*, 1984; Wardrip-Fruin, 1982). Nonetheless, vocalic duration has continued to be studied as a critical cue to syllable-final voicing, and so is presumably considered by many investigators to be just such a cue. In fact, Wardrip-Fruin herself investigated the development of children's abilities to use vocalic duration as a cue to final-consonant voicing just two years after concluding that it was not an adequate or necessary cue in adults' speech perception (Wardrip-Fruin and Peach, 1984). In that later study she concluded that adults use both syllable-offset transitions *and* vocalic duration in voicing decisions for final stops. Thus, the role of vocalic duration in these decisions remains equivocal, and any debate on this question reflects general controversy over theories of speech perception. For most of the history of human speech perception research, predominant theories have held that listeners make phonetic judgments by summing the information provided by several acoustic properties (e.g., Hodgson and Miller, 1992; Hogan and Rozsypal, 1980; Kewley-Port, Pisoni, and Studdert-Kennedy, 1983; Massaro and Oden, 1980). However, experiments showing that listeners can recover a phonetic representation from sine wave

replicas of speech, which lack most of the static spectral properties of natural speech, have led to the suggestion that human speech perception may actually involve the tracking of dynamic changes in the speech wave form (e.g., Remez *et al.*, 1981). Of course, that conclusion was reached with stimuli that are highly unnatural, and therefore its ability to explain the perception of natural speech signals can be questioned.

Many experiments using modified natural stimuli or synthesized formant speech have demonstrated that adults use acoustic information that is not dynamic in their phonetic decisions. For example, adults use static spectral information (such as fricative noises) in decisions about sibilant place of constriction (e.g., Heinz and Stevens, 1961; Kunisaki and Fujisaki, 1977; Nittrouer and Miller, 1997), and use temporal information in decisions about voicing of initial stops and stops in clusters (e.g., Abramson and Lisker, 1967; Best, Morrongiello, and Robson, 1981; Nittrouer, Crowther, and Miller, 1998). At the same time, however, when phonetic information from static spectral and/or temporal properties is constrained by natural conditions, adults increase the weight they assign to dynamic signal components. For example, natural /f/ and /θ/ noises differ from each other spectrally far less than /s/ and /ʃ/ noises, and so adults weight the fricative-vowel formant transitions more in decisions of fricative place for /f/-vowel and /θ/-vowel sequences than for /s/-vowel and /ʃ/-vowel sequences (Harris, 1958; Nittrouer, 2002). The most novel finding reported in the current study is that adults weighted formant transitions at voicing offset greatly in these voicing decisions, even though a temporal property was readily available. This result suggests that experienced language users, as well as less-experienced children, may actually track the dynamic changes of the vocal tract, as suggested by investigators such as Remez *et al.* (1981). While this suggestion is not new, this study elegantly demonstrates the principle for natural signals.

There is, however, one caveat to this suggestion. The conclusion that adults use dynamic signal properties in the perception of natural speech was reached because of structural differences in the synthetic and natural stimuli used in these two experiments. Both sets of synthetic stimuli lacked formant transitions higher than F1, but all five sets of natural stimuli included those higher transitions. As a result of this difference between synthetic and natural stimuli, the conclusion was reached that it must have been the syllable-offset transitions in the natural stimuli that evoked the voiced or voiceless percepts for listeners. But, in fact, there were other attributes of the natural signals that could have accounted for the phonetic decisions. For example, F1 frequency is higher at syllable center in syllables with voiceless, rather than voiced, stops (Nittrouer *et al.*, submitted; Summers, 1987), and this property has some effect on voicing decisions for adults (Summers, 1988). However, the weight assigned to this property compared to other acoustic properties has not been thoroughly studied, and no study has been conducted of the effects of F1 at syllable center on children's voicing decisions. Also, intensity decays more rapidly at syllable offset for voiceless final stops (Hillenbrand *et al.*, 1984). The perceptual weight assigned to this property has not been in-

investigated independently of spectral changes at syllable offset. Investigations specifically manipulating these other properties must be completed before it can be unequivocally concluded that formant transitions at the syllable's offset in natural tokens account largely for voicing decisions by adults and children.

In the same vein, the current study did not manipulate all potential cues to syllable-final stop voicing. In particular, release bursts and voicing during closure were removed from the natural stimuli, and were simply not present in the synthetic stimuli. Although Hillenbrand *et al.* (1984) suggested that the release burst and voicing during closure contribute little to voicing decisions for final stops, those investigators only examined adults' perception. Of course, noise spectra would be predicted to contribute mostly to decisions of constriction place, rather than of voicing. In addition, it has been consistently shown that children do not weight noise spectra greatly in place decisions for syllable-initial voiceless stops (e.g., Parnell and Amerman, 1978) or fricatives (e.g., Nittrouer, 1992; Nittrouer and Miller, 1997). Thus, there is little reason to suspect that release bursts would influence listeners' (especially children's) decisions about the voicing of syllable-final obstruents. Nonetheless, the influence specifically of burst releases, as well as of voicing during closure, on voicing decisions for final obstruents warrants examination.

In summary, this study was undertaken to examine whether young children's weighting of vocalic duration and syllable-offset transitions in decisions of voicing for final stops would match predictions derived from developmental and cross-linguistic studies. As predicted by other developmental studies, as well as by cross-linguistic studies with adults, children were found to weight vocalic duration less and syllable-offset transitions more than adults when synthetic stimuli were used. When edited, natural stimuli were presented, listeners of all ages based voicing decisions for final stops largely on some acoustic property (or properties) other than vocalic duration, although evidence of a developmental increase in the weighting of vocalic duration remained. Most likely, the other acoustic properties weighted heavily were the dynamic syllable-offset transitions.

## ACKNOWLEDGMENTS

This work was supported by Research Grant No. R01 DC00633 from the National Institute on Deafness and Other Communication Disorders, the National Institutes of Health. The author thanks Kathi Bodily, Sandy Estee, Kathy Shapley, and Melanie Wilhelmsen for help with data collection. The comments of Van Summers regarding an earlier draft of this manuscript are gratefully acknowledged.

<sup>1</sup>Unfortunately, Ratner and Luberoff did not precisely define what they meant by the "syllable-final consonant," and so we do not know whether that means only the voicing during closure and the release burst, or if syllable-offset transitions are included in their definition, as well. But it does not really matter because their study is important mainly because it informs us that vocalic duration is exaggerated in child-directed speech.

- Abramson, A. S., and Lisker, L. (1967). "Discriminability along the voicing continuum: Cross-language tests," *Proceedings of the 6th International Congress of Phonetic Sciences*, 569–573.
- Best, C. T., Morrongiello, B., and Robson, R. (1981). "Perceptual equivalence of acoustic cues in speech and nonspeech perception," *Percept. Psychophys.* **29**, 191–211.
- Browman, C. P., and Goldstein, L. (1990). "Gestural specification using dynamically-defined articulatory structures," *J. Phonetics* **18**, 299–320.
- Chen, M. (1970). "Vowel length variation as a function of the voicing of the consonant environment," *Phonetica* **22**, 129–159.
- Crowther, C. S., and Mann, V. (1992). "Native language factors affecting use of vocalic cues to final consonant voicing in English," *J. Acoust. Soc. Am.* **92**, 711–722.
- Crowther, C. S., and Mann, V. (1994). "Use of vocalic cues to consonant voicing and native language background: The influence of experimental design," *Percept. Psychophys.* **55**, 513–525.
- Denes, P. (1955). "Effect of duration on the perception of voicing," *J. Acoust. Soc. Am.* **27**, 761–764.
- Finney, D. J. (1964). *Probit Analysis* (Cambridge University Press, Cambridge, England).
- Fischer, R. M., and Ohde, R. N. (1990). "Spectral and duration properties of front vowels as cues to final stop-consonant voicing," *J. Acoust. Soc. Am.* **88**, 1250–1259.
- Flege, J. E., and Port, R. (1981). "Cross-language phonetic interference: Arabic to English," *Lang Speech* **24**, 125–146.
- Flege, J. E., and Wang, C. (1989). "Native-language phonotactic constraints affect how well Chinese subjects perceive the world-final English /t-/d/ contrast," *J. Phonetics* **17**, 299–315.
- Goldman, R., and Fristoe, M. (1986). *Goldman Fristoe Test of Articulation* (American Guidance Service, Circle Pines, MN).
- Gracco, V. L. (1994). "Some organizational characteristics of speech movement control," *J. Speech Hear. Res.* **37**, 4–27.
- Greenlee, M. (1980). "Learning the phonetic cues to the voiced-voiceless distinction: A comparison of child and adult speech perception," *J. Child Lang* **7**, 459–468.
- Harris, K. S. (1958). "Cues for the discrimination of American English fricatives in spoken syllables," *Lang Speech* **1**, 1–7.
- Heinz, J. M., and Stevens, K. N. (1961). "On the properties of voiceless fricative consonants," *J. Acoust. Soc. Am.* **33**, 589–593.
- Hillenbrand, J., Ingrisano, D. R., Smith, B. L., and Flege, J. E. (1984). "Perception of the voiced-voiceless contrast in syllable-final stops," *J. Acoust. Soc. Am.* **76**, 18–26.
- Hodgson, P., and Miller, J. L. (1992). "Phonetic category structure depends on multiple acoustic properties: Evidence for within-category trading relations," *J. Acoust. Soc. Am.* **92**, 2464.
- Hogan, J. T., and Rozsypal, A. J. (1980). "Evaluation of vowel duration as a cue for the voicing distinction in the following word-final consonant," *J. Acoust. Soc. Am.* **67**, 1764–1771.
- House, A. S., and Fairbanks, G. (1953). "The influence of consonant environment upon the secondary acoustical characteristics of vowels," *J. Acoust. Soc. Am.* **25**, 105–113.
- Jastak, S., and Wilkinson, G. S. (1984). *The Wide Range Achievement Test—Revised* (Jastak Associates, Wilmington, DE).
- Jusczyk, P. W. (1997). *The Discovery of Spoken Language* (The MIT Press, Cambridge, MA).
- Kewley-Port, D., Pisoni, D. B., and Studdert-Kennedy, M. (1983). "Perception of static and dynamic acoustic cues to place of articulation in initial stop consonants," *J. Acoust. Soc. Am.* **73**, 1779–1793.
- Klatt, D. H. (1976). "Linguistic uses of segmental duration in English: Acoustic and perceptual evidence," *J. Acoust. Soc. Am.* **59**, 1208–1221.
- Krause, S. E. (1982). "Vowel duration as a perceptual cue to postvocalic consonant voicing in young children and adults," *J. Acoust. Soc. Am.* **71**, 990–995.
- Kunisaki, O., and Fujisaki, H. (1977). "On the influence of context upon perception of voiceless fricative consonants," *Annual Bulletin of the Research Institute for Logopedics and Phoniatrics* **11**, 85–91.
- Lehman, M. E., and Sharf, D. J. (1989). "Perception/production relationships in the development of the vowel duration cue to final consonant voicing," *J. Speech Hear. Res.* **32**, 803–815.
- Malécot, A. (1970). "The lenis-fortis opposition: Its physiological parameters," *J. Acoust. Soc. Am.* **47**, 1588–1592.
- Massaro, D. W., and Oden, G. C. (1980). "Evaluation and integration of acoustic features in speech perception," *J. Acoust. Soc. Am.* **67**, 996–1013.

- Miranda, S., and Strange, W. (1989). "The role of spectral, temporal, and dynamic cues in the perception on English vowels by native and non-native speakers," *J. Acoust. Soc. Am.* **85**, S53.
- Morrongiello, B. A., Robson, R. C., Best, C. T., and Clifton, R. K. (1984). "Trading relations in the perception of speech by 5-year-old children," *J. Exp. Child Psychol.* **37**, 231–250.
- Nittrouer, S. (1992). "Age-related differences in perceptual effects of formant transitions within syllables and across syllable boundaries," *J. Phonetics* **20**, 351–382.
- Nittrouer, S., Manning, C., and Meyer, G. (1993). "The perceptual weighting of acoustic cues changes with linguistic experience," *J. Acoust. Soc. Am.* **94**, S1865.
- Nittrouer, S., and Miller, M. E. (1997). "Developmental weighting shifts for noise components of fricative-vowel syllables," *J. Acoust. Soc. Am.* **102**, 572–580.
- Nittrouer, S., Crowther, C. S., and Miller, M. E. (1998). "The relative weighting of acoustic properties in the perception of [s]+stop clusters by children and adults," *Percept. Psychophys.* **60**, 51–64.
- 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., Estee, S., Lowenstein, J. H., and Smith, J. (submitted). "The emergence of mature gestural patterns in the production of voiceless and voiced word-final stops," *J. Acoust. Soc. Am.*
- O'Kane, D. (1978). "Manner of vowel termination as a perceptual cue to the voicing status of postvocalic stop consonants," *J. Phonetics* **6**, 311–318.
- Parnell, M. M., and Amerman, J. D. (1978). "Maturational influences on perception of coarticulatory effects," *J. Speech Hear. Res.* **21**, 682–701.
- Peterson, G. E., and Lehiste, I. (1960). "Duration of syllable nuclei in English," *J. Acoust. Soc. Am.* **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," *J. Acoust. Soc. Am.* **51**, 1296–1303.
- Raphael, L. J., Dorman, M. F., Freeman, F., and Tobin, C. (1975). "Vowel and nasal duration as cues to voicing in word-final stop consonants: Spectrographic and perceptual studies," *J. Speech Hear. Res.* **18**, 389–400.
- Raphael, L. J., Dorman, M. F., and Liberman, A. M. (1980). "On defining the vowel duration that cues voicing in final position," *Lang Speech* **23**, 297–307.
- Ratner, N. B., and Luberoff, A. (1984). "Cues to post-vocalic voicing in mother-child speech," *J. Phonetics* **12**, 285–289.
- Remez, R. E., Rubin, P. E., Pisoni, D. B., and Carrell, T. D. (1981). "Speech perception without traditional speech cues," *Science* **212**, 947–949.
- Strange, W. (1989). "Evolving theories of vowel perception," *J. Acoust. Soc. Am.* **85**, 2081–2087.
- Summers, W. V. (1987). "Effects of stress and final-consonant voicing on vowel production: Articulatory and acoustic analyses," *J. Acoust. Soc. Am.* **82**, 847–863.
- Summers, W. V. (1988). "F1 structure provides information for final-consonant voicing," *J. Acoust. Soc. Am.* **84**, 485–492.
- Sussman, H. M., MacNeilage, P. F., and Hanson, R. J. (1973). "Labial and mandibular dynamics during the production of bilabial consonants: Preliminary observations," *J. Speech Hear. Res.* **16**, 397–420.
- Turner, C. W., Kwon, B. J., Tanaka, C., Knapp, J., Hubbart, J. L., and Doherty, K. A. (1998). "Frequency-weighting functions for broadband speech as estimated by a correlational method," *J. Acoust. Soc. Am.* **104**, 1580–1585.
- Wardrip-Fruin, C. (1982). "On the status of temporal cues to phonetic categories: Preceding vowel duration as a cue to voicing in final stop consonants," *J. Acoust. Soc. Am.* **71**, 187–195.
- Wardrip-Fruin, C., and Peach, S. (1984). "Developmental aspects of the perception of acoustic cues in determining the voicing feature of final stop consonants," *Lang Speech* **27**, 367–379.