Perception & Psychophysics
1998, 60 (1), 51–64

The relative weighting of acoustic properties in the perception of [s]+stop clusters by children and adults

SUSAN NITTROUER, COURT S. CROWTHER, and MARNIE E. MILLER
Boys Town National Research Hospital, Omaha, Nebraska

We examined the perceptual weighting by children and adults of the acoustic properties specifying complete closure of the vocal tract following a syllable-initial [s]. Experiment 1 was a novel manipulation of previously examined acoustic properties (duration of a silent gap and first formant transition) and showed that children weight the first formant transition more than adults. Experiment 2, an acoustic analysis of naturally produced *say* and *stay*, revealed that, contrary to expectations, a burst can be present in *stay* and that first formant transitions do not necessarily distinguish *say* and *stay* in natural tokens. Experiment 3 manipulated natural speech portions to create stimuli that varied primarily in the duration of the silent gap and in the presence or absence of a stop burst, and showed that children weight these stop bursts less than adults. Taken together, the perception experiments support claims that children integrate multiple acoustic properties as adults do, but that they weight dynamic properties of the signal more than adults and weight static properties less.

Phonetic perception requires the integration of multiple acoustic properties from across the spectral and temporal domains. Consequently, a change in the setting of one property alters the settings of other properties needed to elicit a specific phonetic decision. These reciprocal relations among acoustic properties, known as "trading relations," have been demonstrated by numerous labeling experiments (e.g., Bailey & Summerfield, 1980; Best, Morrongiello, & Robson, 1981; Dorman, Studdert-Kennedy, & Raphael, 1977; Fitch, Halwes, Erickson, & Liberman, 1980; Mann & Repp, 1980; Repp, 1982; Repp, Liberman, Eccardt, & Pesetsky, 1978; Summerfield & Haggard, 1977). Strictly speaking, demonstrating that a trading relation holds between acoustic properties for a phonetic category does not imply "perceptual equivalence" among (or between) those acoustic properties.[1] That is, stimuli that have different combinations of parameter settings across properties could elicit the same category response from listeners while remaining perceptually discriminable, especially given that typical labeling experiments employ only two category labels.

However, several experiments have demonstrated perceptual equivalence among acoustic properties known to trade. Both Fitch et al. (1980) and Best et al. (1981) showed that longer durations of silence between an [s] noise and a vocalic portion were needed for adults to assign [s]+stop labels to stimuli when formant frequencies at voicing onset were high rather than low. That is, these two very different kinds of properties (one spectral, one temporal) trade: The setting of one property that is required to elicit a stop response changes when the other property is manipulated experimentally. To test the discriminability of stimuli assigned the same label in spite of having different parameter settings, Fitch et al. and Best et al. conducted discrimination tests in three conditions: (1) stimuli differed in settings of only one property (the one-cue condition); (2) stimuli differed in settings of both the duration of silence and formant-onset frequencies in such a way that both favored either the presence or the absence of a stop (the two-cues-cooperating condition); and (3) stimuli differed in settings of both properties in such a way that when one favored the presence of a stop the other favored its absence (the two-cues-conflicting condition). Both studies showed that discrimination was best when the two cues cooperated and was worst (near chance) when the two cues conflicted. Therefore, it was concluded that silence and formant transitions immediately following voicing onset are properties that "have their effects on the same perceptual dimension" (Fitch et al., p. 343); that is, they are perceptually equivalent. In this case, perceptual equivalence is consistent with production in the sense that the same articulatory event, namely a vocal tract closing/opening gesture, gives rise to both acoustic properties.

Morrongiello, Robson, Best, and Clifton (1984) demonstrated both a trading relation and perceptual equivalence for children's perception of *say* versus *stay* employing the same acoustic properties used in the work with adults by Best et al. (1981): the silence interval (gap) between the [s] noise and the vocalic portion and the extent of the first formant ($F1$) transition immediately follow-

---

This work was supported by Research Grant 5 R01 DC 00633 to S.N. We thank Nancy Luke and John Schneider for help running subjects and Catherine Best, Joanne Miller, Donna Neff, Michael Studdert-Kennedy, and two anonymous reviewers for helpful comments on an earlier draft of this paper. Correspondence should be addressed to S. Nittrouer, Boys Town National Research Hospital, 555 North 30th St., Omaha, NE 68131 (e-mail: nittrouer@boystown.org).

ing voicing onset. That is, longer gaps were needed to elicit *stay* responses for stimuli with relatively high *F*1-onset frequencies (*F*1 onset).[2] This finding led Morrongiello et al. to conclude that children integrate multiple acoustic properties in a phonetically relevant manner, just as adults do. However, it was also found that children required less silence than adults to compensate for the higher *F*1-onset frequency, and so respond "stay." This finding led the authors to conclude that the perceptual weights assigned to the various properties integrated in this phonetic decision differed for adults and children.

Investigations by others have led to the same conclusion (Greenlee, 1980; Nittrouer, 1992, 1996a; Nittrouer & Studdert-Kennedy, 1987; Wardrip-Fruin & Peach, 1984); this work has led to the generation of a model of developmental changes in speech perception termed the "developmental weighting shift" (Nittrouer, Manning, & Meyer, 1993). This model proposes that mature and immature listeners of a language weight (attend to) various acoustic properties of the signal differently, and that continued experience with the language brings the weighting schemes of the immature listeners in line with those of the mature listeners. Although future work is likely to demonstrate language-specific variations in which properties acquire enhanced weights as children mature, one general trend seems to be that there is a developmental decrease in the weight assigned to dynamic properties (i.e., those involving spectral change over several tens of milliseconds), but a developmental increase in the perceptual weight assigned to phonetically relevant static properties (i.e., those that do not involve spectral change). Specifically regarding the *say–stay* stimuli, Nittrouer (1992) replicated the labeling results of Morrongiello et al. (1984). Table 1 displays results from Nittrouer and shows a developmental increase from 5-year-olds to 7-year-olds to adults in the size of the trading relation between gap duration and *F*1 onset: The difference between category boundaries (given in milliseconds of gap) for stimuli with a 230-Hz *F*1 onset and for stimuli with a 430-Hz *F*1 onset increases with increasing age. Like Morrongiello et al., Nittrouer found that this age-related difference was entirely due to the function for the 430-Hz *F*1 onset being placed at longer gaps by older listeners. Furthermore, Nittrouer found that 27 three-year-olds all assigned the label *stay* to even the most *say*-like stimulus (i.e., the one with a 430-Hz *F*1 onset and a 0-msec gap). Nittrouer interpreted these findings as evidence that younger listeners weight the *F*1 transition more than adults, and so assign the label *stay* to stimuli with even brief (i.e., not very extensive) transitions.

Table 1
Mean Category Boundaries in Milliseconds of Gap Duration for Each *F*1 Onset (Top) and Difference Between These Boundaries (Bottom)

	5-Year-Olds	7-Year-Olds	Adults
*F*1 onset			
230 Hz	18	17	18
430 Hz	26	32	36
Difference	8	15	18

Note—From "Age-Related Differences in Perceptual Effects of Formant Transitions Within Syllables and Across Syllable Boundaries," by S. Nittrouer, 1992, *Journal of Phonetics*, *20*, p. 369. Copyright 1992 by Academic Press. Reprinted with permission.

The main goal of the present study was to examine further age-related differences in the perceptual weighting of the properties involved in *say–stay* decisions. In the first experiment, two different gap durations were paired with vocalic portions varying along a continuum of *F*1-onset frequencies. If children do weight *F*1 transitions more than adults, so that even brief transitions are sufficient to elicit *stay* labels, then category boundaries should be placed at higher *F*1 onsets for children than for adults. The second experiment acoustically analyzed natural tokens of *say* and *stay*, produced both in context and in isolation. Variation in *F*1-onset frequency has been shown to influence stop decisions in such a way that lower *F*1-onset frequencies lead to more stop percepts for burstless [s]-vowel stimuli (see, e.g., Best et al., 1981; Fitch et al., 1980). This perceptual effect reflects the acoustic consequences of vocal-tract opening: As the vocal tract opens, *F*1 frequency rises. Therefore, a lower *F*1-onset frequency specifies a more closed vocal tract. Another acoustic property that has been shown to elicit stop judgments following an [s] noise is the presence of a burst (Bailey & Summerfield, 1980). The second experiment was intended to provide a description of the acoustic characteristics that distinguish natural *say* and *stay* utterances. Experiment 3 manipulated natural syllable portions in a perception experiment; the primary goal was to investigate possible age-related differences in the weighting of bursts (found to be present in all *stay* tokens of Experiment 2) in decisions about stop presence or absence following [s] noise. The combined results of these experiments provide a more complete description than that previously available of differences between children and adults in weighting schemes for the perception of [s]+stop clusters.

EXPERIMENT 1
Age-Related Differences in the Weighting of *F*1 Transitions

The specific purpose of this experiment was to test the hypothesis that children weight the *F*1 transition more than adults in decisions about the presence or absence of a stop. The stimuli were largely the same as those used in earlier *say–stay* experiments (Best et al., 1981; Morrongiello et al., 1984; Nittrouer, 1992), except that only two gap durations were used and *F*1-onset frequency varied along a continuum. In the earlier experiments, only two *F*1-onset frequencies were used, and gap duration varied along a continuum. The hypothesis being tested would be supported if children placed category boundaries at higher *F*1-onset frequencies than adults.

Children who were 3 years of age were purposely selected for this experiment, rather than older children, because 3-year-olds had labeled even the most *say*-like stimulus in the Nittrouer (1992) experiment as *stay*. Two explanations were considered for that trend. One possibility is that 3-year-olds do not distinguish *say* and *stay*, but have a lexical preference for *stay*. A more plausible explanation would be that 3-year-olds weight the acoustic properties for this phonetic decision in such a way that more tokens elicit *stay* responses. Because this experiment used stimuli for which the range of $F1$ onsets extended to higher frequencies than the stimuli of the 1992 experiment, we were able to provide support for one explanation over the other. If 3-year-olds labeled all stimuli as *stay* (as in the 1992 experiment), then support would be bolstered for the suggestion that 3-year-olds simply do not distinguish *say* and *stay*, but have a lexical preference for *stay*. However, if they labeled some stimuli as *say*, but placed the *say–stay* category boundary at a higher $F1$-onset frequency, then support would be bolstered for the suggestion of age-related differences in perceptual weighting of this property.

Method

Subjects. Eleven children between the ages of 3 years 5 months and 4 years 2 months and 8 adults between the ages of 19 and 47 years participated. All listeners passed a hearing screening consisting of pure tones at the frequencies 0.5, 1, 2, 4, and 6 kHz presented at 25 dB HL (American National Standards Institute, 1989) to each ear. None of the subjects had a history of speech or language problems, and no one in their immediate families had ever had such a problem. The children were given the Goldman–Fristoe Test of Articulation (Goldman & Fristoe, 1986) and were required to score at better than the 30th percentile for their age. In particular, all children were able to produce syllable-initial /s/ and /st/. All children were free from significant histories of otitis media, defined as fewer than six episodes during the first 2 years of life. The adults were given the reading subtest of the Wide Range Achievement Test–Revised (WRAT-R; Jastak & Wilkinson, 1984) and were required to demonstrate at least an 11th-grade reading level.

Equipment. All perceptual testing took place in a sound-attenuated booth. Hearing was screened with a Welch Allyn TM262 tympanometer/audiometer using TDH-39 headphones. Recorded stories were presented via a Nakamichi MR-2 audiocassette player with AKG-K141 headphones. Synthetic versions of the stories were generated with DECtalk, a text-to-speech synthesizer. Presentation of stimuli and recording of responses were controlled by a computer. A Data Translation 2801A digital-to-analog converter, a Frequency Devices 901-F filter, a Crown D-75 amplifier, and AKG-K141 headphones were used for stimulus presentation. This system has a flat frequency response. Reinforcement for children consisted of cartoon characters presented on a color graphics monitor.

Stimuli. Eight vocalic portions were synthesized with a Sensimetrics software synthesizer at a sampling rate of 10 kHz with low-pass filtering below 4.9 kHz and were modeled after those of Nittrouer (1992). All vocalic portions were 300 msec in length, with the fundamental frequency ($f0$) falling linearly throughout from 120 to 100 Hz. $F3$ fell through the first 40 msec from 3196 to 2694 Hz, where it remained for the next 120 msec. It then rose to 2929 Hz over the next 90 msec, where it remained for the final 50 msec. $F2$ was constant at 1840 Hz for the first 160 msec and then rose to 2240 Hz over the next 90 msec, where it remained. $F1$-onset frequency varied across stimuli from 211 to 561 Hz in 50-Hz steps. From its starting frequency, $F1$ rose over the first 40 msec to 611 Hz, where it remained for the next 120 msec. It then fell to 304 Hz over the next 90 msec, where it remained until the end of the segment. Each of these vocalic portions was combined with the natural [s] noise used in Nittrouer (1992) at each of two gap durations: 10 and 30 msec. Consequently, there were 16 stimuli (8 vocalic portions × 2 gap durations). Figure 1 displays the most *stay*-like stimulus (with a 30-msec gap and a 211-Hz $F1$ onset) and the most *say*-like stimulus (with a 10-msec gap and a 561-Hz $F1$ onset).

Procedure. First, the screening tasks were done. Next, two kinds of training were provided. In the first of these, listeners heard five tokens each of natural, unedited *say* and *stay* presented in random order. Listeners had to respond correctly to 9 of these 10 tokens before proceeding to the next training task, simply to ensure that they understood the task and knew the response labels. Second, the two best exemplars of stimuli with synthetic vocalic portions were presented five times each. The best exemplar of *say* was the token with a 10-msec gap and a 561-Hz $F1$ onset. The best exemplar of *stay* was the token with a 30-msec gap and a 211-Hz $F1$ onset. Listeners had to respond correctly to 9 of these 10 tokens before proceeding to testing. During testing, each stimulus was presented 10 times in randomized blocks of 16 each (i.e., one stimulus at each level of gap duration combined with each vocalic portion). Listeners indicated their responses by saying the response label (*say* or *stay*) and pointing to one of two pictures (a boy with a cartoon bubble over his head [thus saying something] or a girl signaling to a poodle to "stay"). The experimenter entered the responses into the computer. Listeners had to respond correctly to 8 out of the 10 best exemplars during testing for their data to be included in the final analysis.

Three modifications were made to these procedures for the children. First, children heard recorded stories about *say* and *stay* before the training. These stories were approximately 2 min each and were presented first by live voice and then by synthetic voice. This procedure was meant both to familiarize children with the category labels and to give them an opportunity to listen to synthetic speech. Second, at the end of each block of 16 stimuli, a cartoon drawing appeared on the graphics monitor. Finally, children moved a plastic rabbit to the next number on a game board at the end of each block to keep track of how many blocks were left.

The labeling data for each continuum (i.e., the 10- and the 30-msec gap conditions) from each listener were transformed to probit scores (Finney, 1964). From each probit distribution, we derived the mean (i.e., the category boundary) and the slope.[3] Two-way analyses of variance (ANOVAs) were performed on both the category boundaries and the slopes, with age as the between-subjects factor and gap as the within-subjects factor.

Results and Discussion

Five 3-year-olds met the screening criteria but then failed to meet the criteria for either training or testing. One child did not obtain 90% correct responses with the natural tokens of *say* and *stay*, and 3 children did not obtain 90% correct responses with the best exemplars of the test stimuli during training. Responses of these 4 children did not reveal any particular preference for *say* or *stay*. One child passed both training exercises but then did not obtain 80% correct responses to the best exemplars during testing. These subjects' data were not included in the final analysis.

Figure 2 displays labeling functions for adults (top) and 3-year-olds (bottom). Table 2 shows mean category boundaries for adults and 3-year-olds (top), and the difference in placement of these category boundaries (bot-

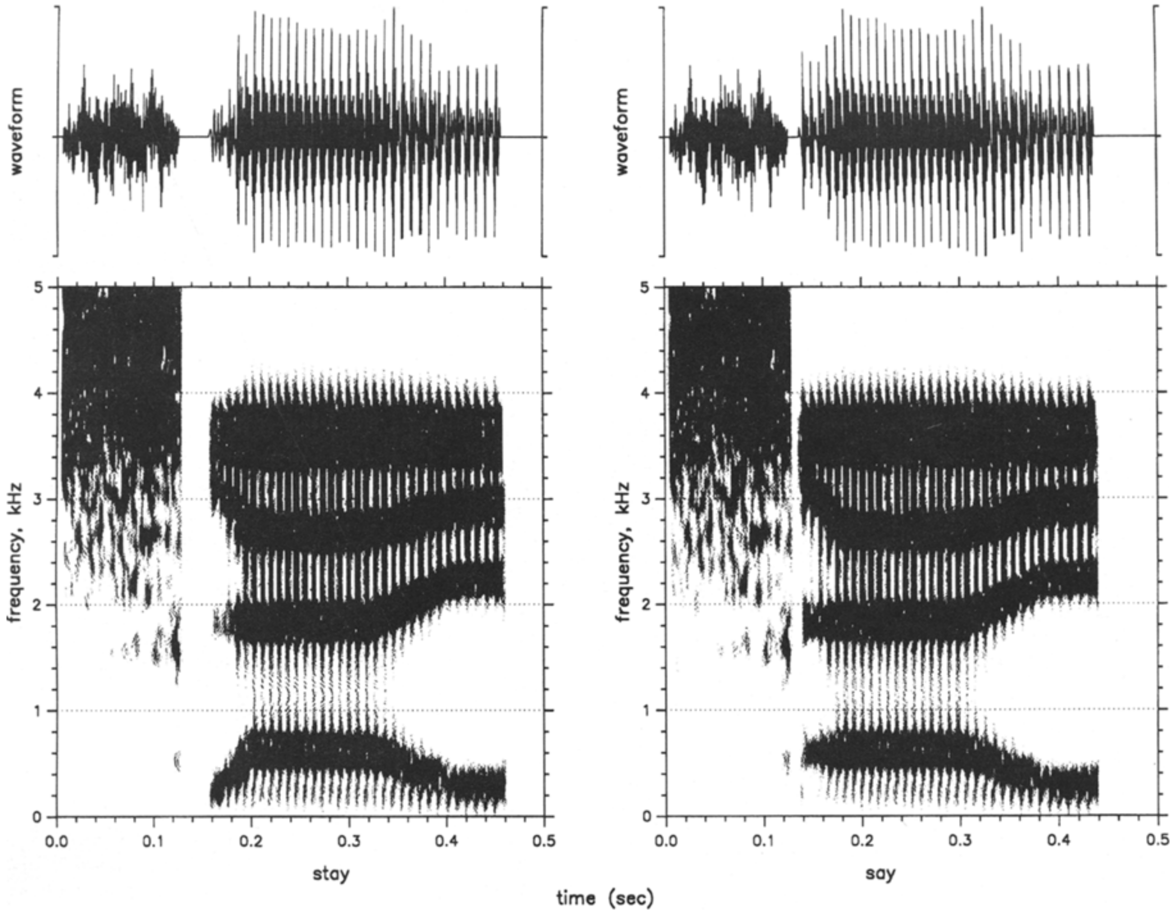

Figure 1. The most *stay*-like and the most *say*-like stimuli in Experiment 1.

tom). The children showed a greater separation in boundaries based on gaps than did adults; this age-related difference appears to have been due primarily to children's mean boundary for the 30-msec gap condition being at a higher $F1$ onset than that of adults. These impressions are supported by the results of the ANOVA, which revealed a significant main effect of gap [$F(1,12) = 140.42$, $p < .001$], as well as a significant gap × age interaction [$F(1,12) = 15.19, p = .002$]. A simple effects analysis revealed a significant age effect only for the 30-msec gap condition [$F(1,12) = 10.84, p = .006$].

Table 3 shows the mean slope of each function for adults and 3-year-olds (top) and mean slope across the two functions (bottom). No statistically significant effects were observed for slopes.

The results of this experiment provide further support for Morrongiello et al.'s (1984) suggestion that children weight the acoustic properties relevant to *say–stay* decisions differently from adults. Support is also provided for Nittrouer's (1992) more specific suggestion that children weight the $F1$ transition more than adults do, so even stimuli with brief $F1$ transitions (i.e., roughly 100 Hz) were sufficient to elicit *stay* responses from these children.

EXPERIMENT 2
Acoustic Analysis of Natural Tokens

The goal of this experiment was to examine the acoustic characteristics of naturally produced *say* and *stay* tokens in greater detail than has previously been reported. It is often not clear how directly parameter settings for synthetic stimuli in perception experiments relate to those of natural speech. This ambiguity is confounded by the possibility that children may not treat synthetic stimuli as much like natural speech as adults do. One untested hypothesis for why children performed differently from adults in previous *say–stay* studies is that the acoustic properties manipulated in the synthetic portions were not those that arise from the natural production of these words, and children are more greatly affected by this deviation from natural speech than are adults. Therefore, investigating the acoustic properties distinguishing the *say–stay* distinction in natural productions seemed an important goal.

Method

Subjects. Twenty adults between the ages of 20 and 46 years (10 men and 10 women) served as speakers. All were native speakers of American English but varied in terms of their dialects. None of the speakers had a history of speech or hearing problems.

Equipment. An AKG C535 EB microphone, a Shure M268 preamplifier, a Frequency Devices 901F filter, and a Data Translation DT2801A analog-to-digital converter were used. CSpeech Version 4 software (Milenkovic, 1993) was used for digitizing and analyzing speech tokens. SPECTO (Neely & Peters, 1989) was used for creating spectrograms.

Stimuli. All 20 speakers produced three tokens each of *say* and *stay* in isolation, in randomized order. In addition, 6 of the men and 6 of the women read a passage containing three tokens of each word. The passage is provided in Appendix A.

Procedure. Each speaker sat roughly 6 in. from the microphone in a sound-attenuated room. Words in isolation and words in context were recorded on separate days. Speech samples were digitized directly to the computer at a sampling rate of 20 kHz, with low-pass filtering below 9.8 kHz.[4] Spectrograms were made of one randomly selected sample of *say* and *stay* from each speaker, simply to get an overall picture of acoustic patterns. Two specific measurements were made for each token. First, the time between the release of closure and the onset of voicing was measured for *stay*. We refer to this measure as "burst duration." Bailey and Summerfield (1980) referred to the same measure as "aspiration duration," although in all likelihood most of this interval was the burst: Stops that are phonemically "voiceless" are realized as unaspirated in [s]+stop clusters (Best et al., 1981; Klatt, 1975). The frequencies of $F1$, $F2$, and $F3$ for the first pitch period were derived from LPC spectra for both *stay* and *say*. The spectra of the burst noises had the shape predicted by the alveolar place of closure: Burst spectra were diffuse with rising spectrum (Stevens & Blumstein, 1978), and so are not reported here.

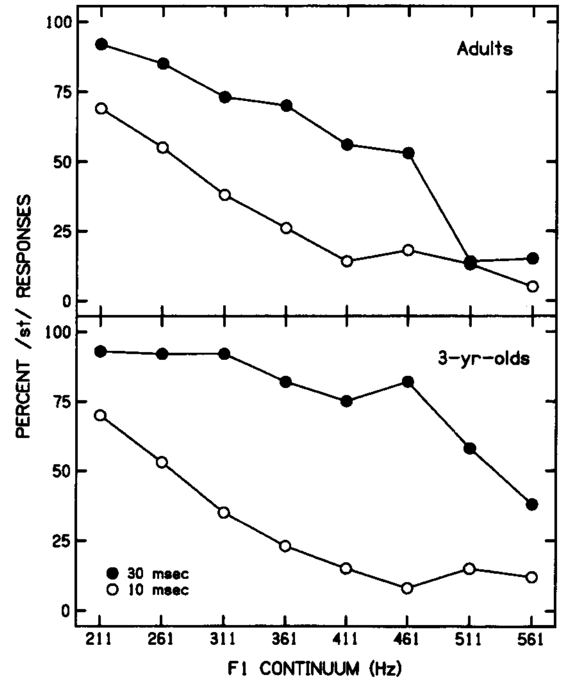

Figure 2. Labeling functions for adults and 3-year-olds in Experiment 1. Gap condition is the curve parameter. Open circles indicate 10-msec gaps; filled circles indicate 30-msec gaps. $F1$-onset frequency is represented on the x-axis.

Table 2
Mean Category Boundaries and Standard Deviations in Hertz of $F1$ Onset for Each Gap Duration in Experiment 1 (Top) and Difference Between These Boundaries (Bottom)

	3-Year-Olds		Adults	
	M	*SD*	*M*	*SD*
Gap = 30 msec	529	61.5	418	62.9
Gap = 10 msec	262	77.3	284	62.9
Difference	267	56.7	134	66.7

Table 3
Mean Rescaled Slopes and Standard Deviations for Each Function (Top) Given in Probit Units per Hertz of $F1$ Onset * 1000 for Experiment 1 and Mean Slope Across Functions (Bottom)

	3-Year-Olds		Adults	
	M	*SD*	*M*	*SD*
Gap = 30 msec	−8.88	7.58	−6.82	7.39
Gap = 10 msec	−6.56	3.66	−9.24	7.49
M	−7.72	5.15	−8.03	3.28

Results and Discussion

Appendix B shows means across the three tokens of *say* and *stay* produced in isolation for each speaker. Appendix C shows means across the three tokens produced in context. Figure 3 displays spectrograms of one *stay* and one *say* produced in isolation by 1 of the male speakers, and Figure 4 displays these words produced in context by this same speaker. The general patterns seen in the spectrograms of this speaker's samples were highly consistent with those from all other speakers, and two results are of interest. First, every *stay* token had a noticeable burst. Second, mean $F1$-onset frequency for *say* and *stay* tokens never differed from each other by more than 1 standard deviation. That is, the stop closure following the [s] in *stay* is not specified by a lower $F1$-onset frequency than that found for *say*. Instead, a short release burst occurs, thus allowing the vocal tract to open sufficiently before voicing onset so that when voicing does start, $F1$ frequency is similar to that of *say* at voicing onset.

One other trend in these data deserves mention. Although inconsistent, there was a tendency for $F3$-onset frequency to be slightly lower and/or for $F2$-onset frequency to be slightly higher in /(st)eɪ/ than in /(s)eɪ/. These acoustic differences suggest that, on average, the vocal-tract closure in /(st)eɪ/ may be slightly more backed than the constriction in /(s)eɪ/. However, as Figures 3 and 4 suggest, there is no rapid and/or extensive change following these onset frequencies. This acoustic pattern is consistent with the articulatory fact that in producing these words, speakers move from a consonant closure or constriction near the front of the mouth to a vowel constriction near the front of the mouth. The largest movement in the production of these consonant–vowel syllables is in the vertical dimension, and thus the first formant exhibits the most spectral change.

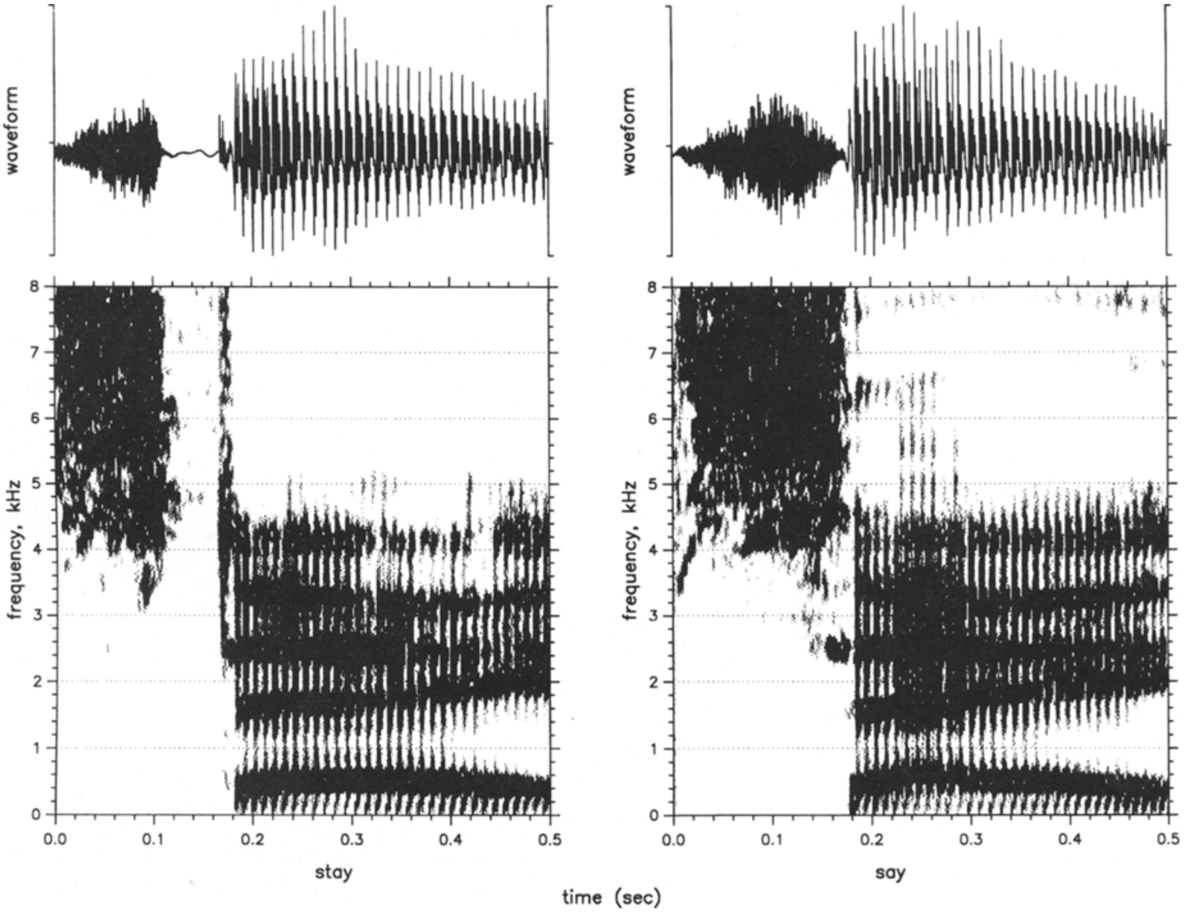

Figure 3. Spectrograms of *stay* and *say* produced by an adult male speaker in isolation.

EXPERIMENT 3
Age-Related Differences in the Perceptual Weighting of Stop Bursts

The primary goal of this experiment was to examine potential age-related differences in the perceptual weighting of stop bursts. Because bursts are considered a static property (i.e., they do not change spectrally over their duration), the prediction was that adults would weight bursts more than children would. This experiment would also shed light on whether children's previous results differed from those of adults because of age-related differences in the processing of synthetic speech.

Method

Subjects. Subjects were 14 four-year-olds (M = 4 years 3 months; range = 4 years 0 months to 4 years 6 months), 13 six-year-olds (M = 6 years 1 month; range = 5 years 11 months to 6 years 5 months), and 13 adults between 20 and 40 years of age. Requirements for participation in this experiment were the same as those in Experiment 1.

Equipment. The equipment was the same as that in Experiment 1.

Stimuli. Ten tokens each of *say* and *stay* produced in isolation by a male speaker were recorded at a 10-kHz sampling rate with low-pass filtering below 4.9 kHz. Of these, five tokens each of *say* and *stay* were randomly selected for use in this experiment. The [s] noise was deleted from the *say* tokens, and the [s] noise, gap, and burst were deleted from the *stay* tokens. Thus only the vocalic portions from each kind of utterance (*say* or *stay*) remained. Mean formant frequencies for the first pitch period and for the (relatively) steady-state [e] portion of the [eɪ] diphthong are shown in Table 4. Mean $F1$ frequencies for both kinds of vocalic portions were similar, and closer to the $F1$-onset frequency of Best et al.'s (1981) weak *day* (i.e., the portion with the 430-Hz $F1$ onset) than to their clear *day* (i.e., the portion with the 230-Hz $F1$ onset). Table 4 also shows mean durations for the *say* and *stay* vocalic portions and shows that there were no substantial differences across portions. These vocalic portions were presented to 5 adult native speakers of American English 10 times each, and were all unambiguously labeled as *day*.

Each of these burstless vocalic portions was combined with the natural [s] noise used in Experiment 1 and in Nittrouer (1992), at each of seven gap durations, varying between 0 and 48 msec in 8-msec steps. Next, a single burst was selected and added to the front of each vocalic portion. This burst was 9.3 msec long and did not appear to incorporate any noise that might more appropriately be considered aspiration noise. These burst+vocalic portions were then combined with the natural [s] noise at each gap duration. Thus there were 140 stimuli (7 gap durations × 2 kinds of vocalic portion × 2 burst conditions × 5 tokens of each). Figure 5 shows the most *stay*-like stimulus (constructed with an /eɪ/ taken from *stay*, a

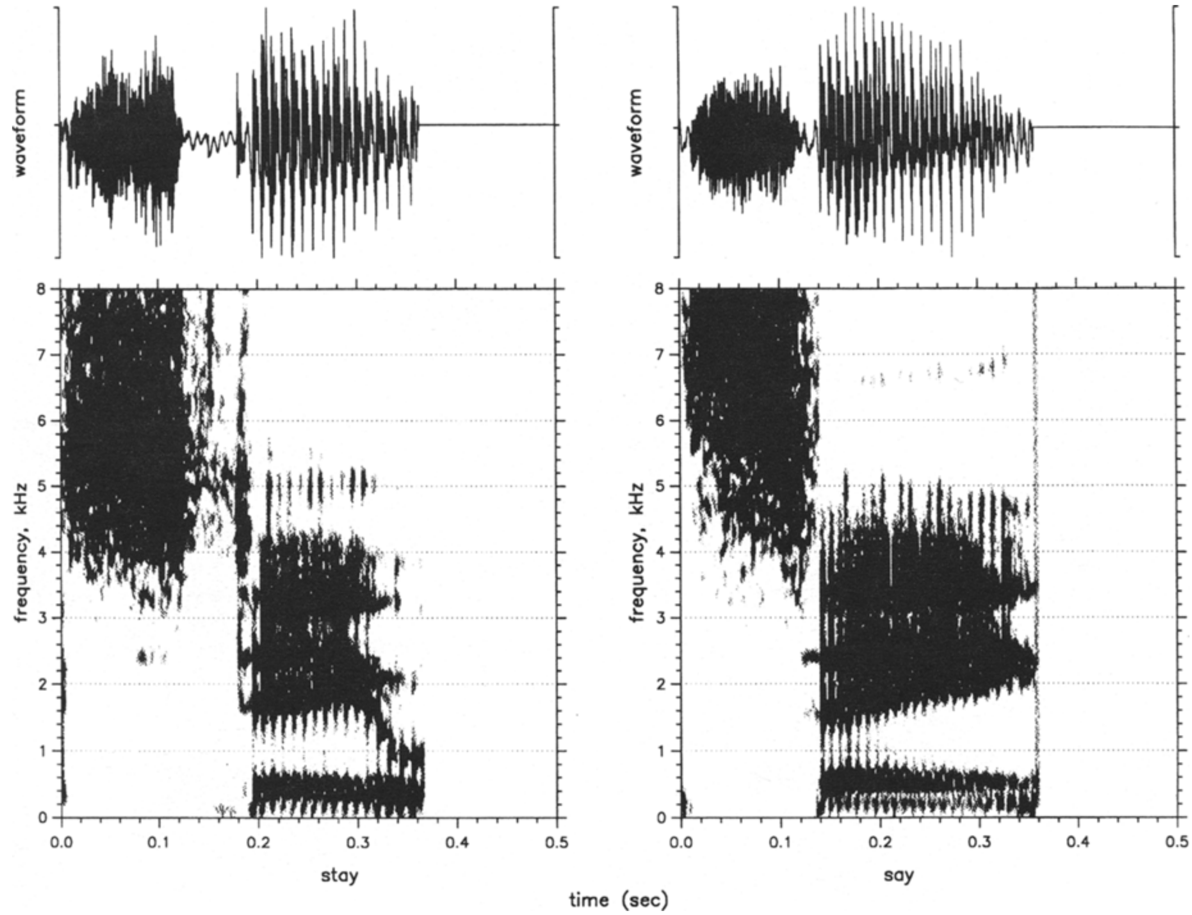

Figure 4. Spectrograms of *stay* and *say* produced by an adult male speaker in context.

burst, and a 48-msec gap) and the most *say*-like stimulus (constructed with an /eɪ/ taken from *say*, no burst, and a 0-msec gap).

Procedure. Procedures were the same as for Experiment 1. During testing, each stimulus was presented twice, yielding a total of 10 responses to each kind of vocalic portion (/(s)eɪ/ or /(st)eɪ/) in each burst condition (burst present or absent) at each gap duration. Stimuli were presented in 10 blocks of 28 each. Specific tokens of the vocalic portions varied within each block. For training, the best exemplars of *say* were the burstless /(s)eɪ/ portions, with a 0-msec gap, and the best exemplars of *stay* were the /(st)eɪ/ portions, with the burst and a 48-msec gap.

The labeling data for each vocalic portion × each burst condition for each listener were transformed to probit scores. A distribution mean (i.e., the category boundary) and a slope were then derived. Three-way ANOVAs were performed on category boundaries and slopes, with age as a between-subjects factor, and vocalic portion and burst condition as within-subjects factors. To examine potential developmental changes more closely, two planned orthogonal comparisons were also performed: adults versus all children and 6-year-olds versus 4-year-olds.

Results and Discussion

Only one potential listener was dismissed—a 4-year-old for failing to meet the training criterion of 9 out of 10 correct. Figure 6 shows mean functions for all three

Table 4
Mean Formant Frequencies in Hertz, Portion Durations in Milliseconds, and Standard Deviations of the Five Vocalic Portions Used in Experiment 3

	/(s)eɪ/				/(st)eɪ/			
	1st PP		Steady-State		1st PP		Steady-State	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
*F*1	398	43.9	490	15.7	392	30.3	503	30.4
*F*2	1654	24.7	1733	47.9	1676	68.3	1774	31.8
*F*3	2496	48.5	2475	23.8	2414	124.8	2485	24.9
Duration	497	29.2			476	46.5		

Note—Formant frequencies are given for the first pitch period (1st PP) and for the steady-state /e/ portions of the diphthong.

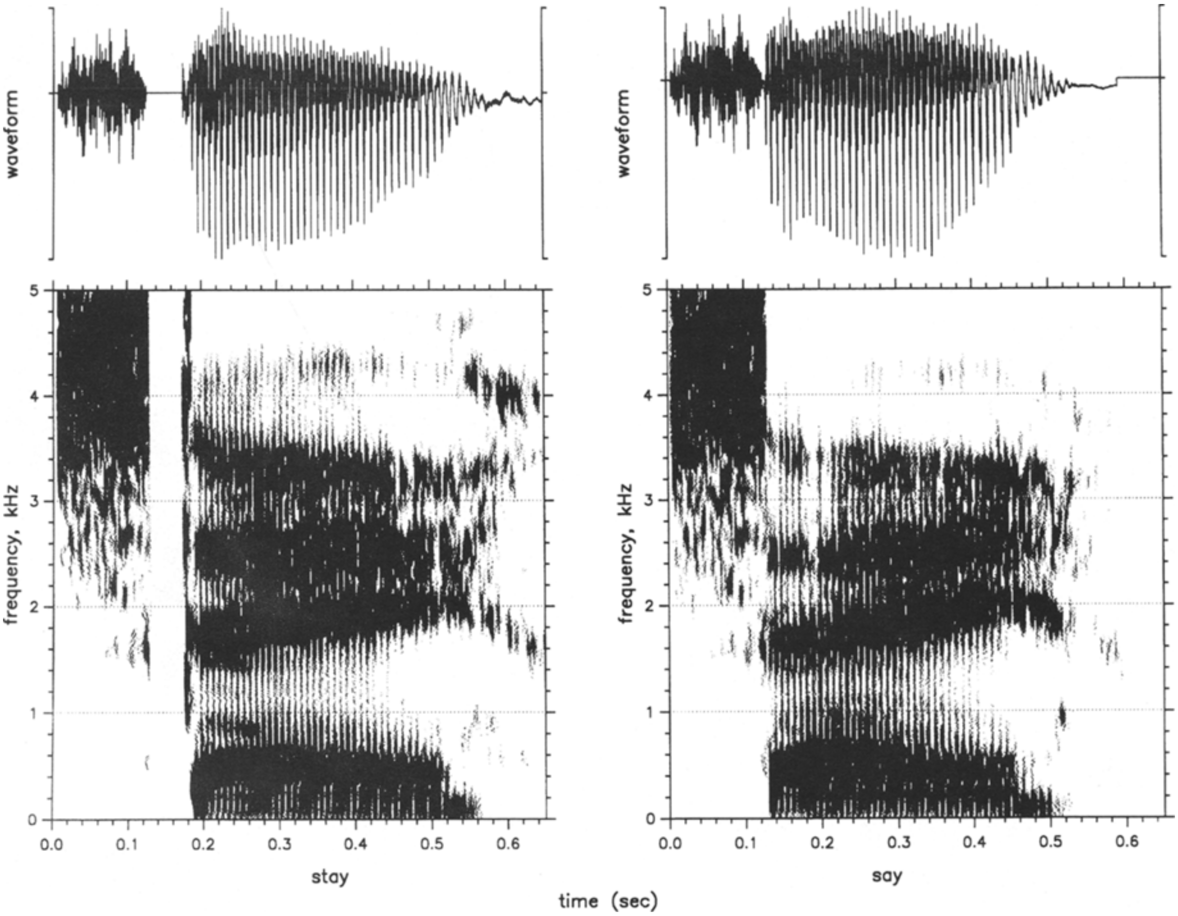

Figure 5. The most *stay*-like and the most *say*-like stimuli in Experiment 3.

groups of listeners. Table 5 displays mean category boundaries for each function for each age group (top), and the effects of the burst condition and vocalic portion (bottom). The burst effect is given by the difference in placement of boundaries for the burst and burstless conditions, for /(st)eɪ/ and /(s)eɪ/ portions separately. The effect of vocalic portion is indicated by the difference in placement of boundaries for /(st)eɪ/ and /(s)eɪ/, for the burst and burstless conditions separately. Clearly, the presence or absence of a burst had a stronger influence on responses than whether the vocalic portion was taken from *say* or *stay*: The separation in functions depending on burst condition (circles vs. squares, whether filled or open) was greater than the separation in functions depending on vocalic portion (open vs. closed symbols, whether circles or squares). Also, there was a developmental increase in the size of the burst effect; that is, the separation between functions depending on burst condition increased with increasing age. This age effect appears to have been largely due to the placement of category boundaries for burstless stimuli: Adults required more silence than did children to label burstless stimuli as *stay*. The effect of whether the vocalic portion was taken from *say* or *stay* was restricted to burstless stimuli, and was not large. However, it was slightly larger for adults than for children.

The ANOVA on category boundaries provided statistical confirmation for the trends described above. The main effect of age was significant [$F(2,36) = 10.49, p < .001$], as was the planned comparison of adults versus all children [$F(1,36) = 20.54, p < .001$]. Across all conditions, children generally had lower category boundaries. Means across all four functions were 21 msec for adults, 15 msec for 6-year-olds, and 14 msec for 4-year-olds. As shown in Table 5, however, this age-related difference in the placement of category boundaries seems to have been largely due to an age-related difference for burstless stimuli.

The main effect of burst was also significant [$F(1,36) = 739.57, p < .001$], as was the age × burst interaction [$F(2,36) = 10.65, p < .001$]. The planned comparison of adults versus all children for the effect of burst was significant [$F(1,36) = 21.22, p < .001$], but the comparison of 6- versus 4-year-olds was not. Thus the effect of burst was greater for adults than for children, but there was no significant difference in the size of the effect for 6- and

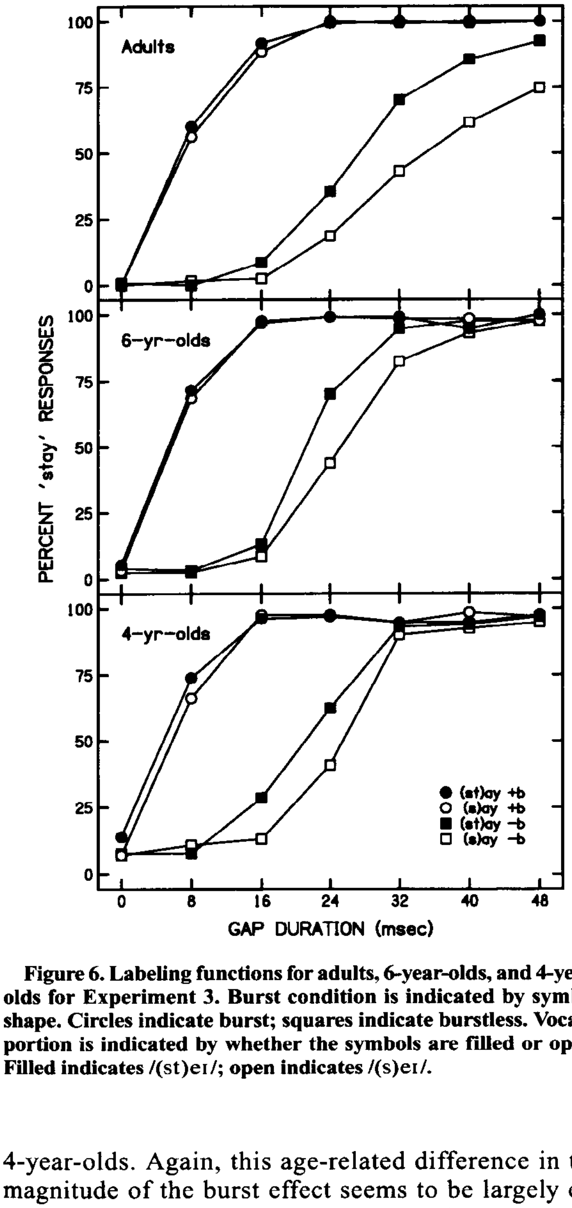

Figure 6. Labeling functions for adults, 6-year-olds, and 4-year-olds for Experiment 3. Burst condition is indicated by symbol shape. Circles indicate burst; squares indicate burstless. Vocalic portion is indicated by whether the symbols are filled or open. Filled indicates /(st)eɪ/; open indicates /(s)eɪ/.

4-year-olds. Again, this age-related difference in the magnitude of the burst effect seems to be largely explained by results for the burstless stimuli.

The main effect of vocalic portion was significant [$F(1,36) = 92.03, p < .001$], as was the age × vocalic portion interaction [$F(2,36) = 4.26, p = .022$]. The planned comparison of adults versus all children was significant for the effect of vocalic portion [$F(2,36) = 8.47, p = .006$], but the planned comparison of 6- versus 4-year-olds was not. Thus the effect of vocalic portion was greater for adults than for children, but there was no difference in the magnitude of this effect for 6- and 4-year-olds. Once again, this main effect seems to be attributable to responses for burstless stimuli. It was only responses to these stimuli that seemed to demonstrate any effect of vocalic portion, and this effect appears slightly larger for adults than for children.

A simple effects analysis was done to see if the impressions that all significant differences in response patterns between children and adults could be accounted for by results for burstless stimuli. This analysis looked at the effects of age and vocalic portion at each level of burst separately. Results showed a significant age effect in the placement of category boundaries for the burstless stimuli only [$F(2,36) = 14.14, p < .001$]. Thus, adults required more silence than did children to compensate for the absence of a burst. Also significant for stimuli with no burst was the main effect of vocalic portion [$F(1,36) = 94.71, p < .001$], and the age × vocalic portion interaction [$F(2,36) = 7.68, p = .002$]. Thus, adults showed a greater effect of the vocalic portion on responses to burstless stimuli than did children.

Returning to results for the ANOVA, the interaction of burst × vocalic portion was significant [$F(1,36) = 37.12, p < .001$]. Moreover, there was a significant three-way interaction of age × burst × vocalic portion [$F(2,36) = 5.68, p = .007$], and the planned comparison of adults versus all children was significant for the burst × vocalic portion interaction [$F(1,36) = 10.32, p = .003$]. Thus, the way in which the effect of vocalic portion varied across burst conditions differed for children and adults, but not for 6- and 4-year-olds.

Table 6 displays mean slopes for /(s)eɪ/ and /(st)eɪ/ stimuli in the burstless condition. Because category boundaries for stimuli with bursts were so close to 0 msec, computation of slopes using probit analysis was difficult. Therefore, those slopes were not considered here. Judging from Figure 5, however, there was no difference in steepness of these functions for children and adults. For the burstless stimuli, it similarly appears that there were no age-related differences in steepness of functions, and the ANOVA supports this impression.

Category boundaries for the burstless stimuli are similar to those observed by Nittrouer (1992) for stimuli with the 430-Hz $F1$ onset, for all age groups. These functions also show the developmental trend reported by Nittrouer: the younger the listener, the lower the category boundary. (See Table 1, showing boundaries from

Table 5
Mean Category Boundaries and Standard Deviations in Milliseconds of Gap Duration for Experiment 3 (Top) and Mean Differences in Category Boundaries for the Burst Effect ([−burst]–[+burst]) Across Each Vocalic Portion and for the Effect of Vocalic Portion [/(s)eɪ/–/(st)eɪ/] Across Each Burst Condition (Bottom)

	4-Year-Olds		6-Year-Olds		Adults	
	M	*SD*	*M*	*SD*	*M*	*SD*
Burst						
/(st)eɪ/	4.8	3.4	6.2	2.4	8.6	4.9
/(s)eɪ/	6.3	4.5	6.6	2.0	9.0	5.1
Burstless						
/(st)eɪ/	20.7	2.9	21.7	3.9	28.9	7.8
/(s)eɪ/	24.0	3.7	25.7	4.2	36.7	8.8
Burst effect						
/(st)eɪ/	15.9	2.6	15.5	3.0	20.3	6.0
/(s)eɪ/	17.7	5.0	19.1	3.0	27.8	7.8
Vocalic-portion effect						
(+burst)	1.5	3.1	0.4	2.4	0.4	1.8
(−burst)	3.3	2.7	4.0	2.0	7.8	4.5

Table 6
Mean Slopes and Standard Deviations for Each Function of Burstless Stimuli (Top) Given in Probit Units per Millisecond of Gap for Experiment 3, and Mean Slope Across Functions (Bottom)

	4-Year-Olds		6-Year-Olds		Adults	
	M	*SD*	*M*	*SD*	*M*	*SD*
/(st)eɪ/	.11	.06	.15	.06	.14	.04
/(s)eɪ/	.11	.06	.13	.05	.11	.05
M	.11	.05	.14	.05	.12	.04

Nittrouer, but note that children's age groups differ slightly between the two studies.) The burstless stimuli, compared with the stimuli with bursts, were spectrally more similar to the 430-Hz *F*1 onset stimuli from that earlier experiment, but these burstless stimuli were natural and the earlier ones were synthetic. Thus, similar results have been observed for stimuli with similar parameter settings, even though stimuli in one set were produced by a software synthesizer and those in the other set were produced by a human vocal tract. This comparable result across experiments supports the suggestion that previous results were not due to age-related differences in bias against synthetic stimulus portions: As long as the filter characteristics were the same for the natural and synthetic stimuli, perceptual effects were the same.

One result does not appear to support the suggestion that children perceptually weight dynamic properties more than adults do: For burstless stimuli, the adults showed a larger effect of whether the vocalic portion was taken from a /(s)eɪ/ or a /(st)eɪ/ token than did the children. The onset frequencies of *F*2 and *F*3 (and thus the *F*2 and *F*3 transitions) differed slightly for /(s)eɪ/ and /(st)eɪ/, so that the closure in /(st)eɪ/ must be slightly more backed, on average, than the constriction in /(s)eɪ/. Apparently, children do not pay attention to this subtle difference in place of closure.

GENERAL DISCUSSION

The results of these three experiments extend our understanding of the integration and weighting of acoustic properties in phonetic decisions for children and adults. Experiments 1 and 3 demonstrated that acoustic properties show evidence of a trading relation (and thus of integration) in the labeling responses of children, just as they do in adults' responses. On the basis of the work of Morrongiello et al. (1984), we conclude that this trading relation reflects perceptual equivalence for gaps and *F*1 transitions for both adults and children. At present, we cannot propose the same perceptual equivalence for the gaps and bursts because the requisite discrimination tests have not been done. However, the weights assigned to both the *F*1 transitions and the bursts were found to differ in children's and adults' labeling decisions. These age-related differences are consistent with other findings showing that a shift in the perceptual weighting of acoustic information occurs as a result of experience with one's native language (e.g., Nittrouer, 1996b; Nittrouer et al., 1993). Specifically in the perception of [s]-stop-vowel syllables, evidence was found in Experiment 1 to support the suggestion that *F*1 transitions are weighted more by children than by adults, so that less extensive transitions are sufficient to elicit [s]+stop responses from children. Experiment 3 showed that the burst is weighted more by adults than by children: to respond "stay," adults required more silence than did children to compensate for the absence of a burst. These developmental patterns match those found in previous developmental work (Greenlee, 1980; Morrongiello et al., 1984; Nittrouer, 1992; Nittrouer & Studdert-Kennedy, 1987; Wardrip-Fruin & Peach, 1984). The *F*1 transition is a dynamic property, and children weighted it more in these phonetic decisions than adults did. The burst is a static property, and children did not assign as much weight to it as adults did. Thus we again see evidence of a developmental decrease in the weight assigned to a dynamic property and of a developmental increase in the weight assigned to a static property.[5] This developmental weighting shift may reflect the child's initial need to learn to recognize syllables (or words, which are often single syllables for young children) in the continuous speech stream (Jusczyk, 1992, 1993, 1994). A speech processing scheme that focuses on acoustic properties specifying general patterns of movement toward or away from a constriction would be most useful in achieving this goal. Later, the child's growing lexicon may provide pressure for individual lexical items to be distinguished from each other more precisely. At this time the child starts needing to recognize syllabic detail, and so it becomes useful to increase attention to the acoustic properties that specify the locations and shapes of those constrictions.

Experiment 2 showed a production pattern for *stay* that differed from expectations based on Best et al. (1981). The 2 speakers in Best et al. did not produce bursts, and voicing started at the instant of closure release. Consequently, *F*1 onset was lower in *stay* than in *say*. All 20 speakers in this study produced bursts in each *stay* token, and voicing onset was delayed by 15–20 msec after the release of closure. Naturally, the vocal tract continued its opening movement during the lag between the release of closure and voicing onset, while the burst was being produced. On the basis of the *F*1 onset measures, it appears that this lag was just long enough that when voicing did begin in *stay*, the degree of vocal-tract opening was similar to that at the release of the [s] constriction in *say*. Consequently, *F*1 onset was similar in *stay* and *say*. Although it is impossible to know at present which pattern is most common for American English [st] (burstless with a low *F*1 onset or containing a burst with an *F*1 onset similar to that of [s]), it is possible that listeners are only rarely exposed to burstless *stay* tokens with *F*1 onsets as low as 230 Hz. Therefore, earlier studies (Best et al.; Morrongiello et al., 1984; Nittrouer, 1992) employing variants of *say–stay* may not have been manipulating acoustic properties that typically distinguish between these words in natural speech.

Nonetheless, listeners in those studies were more likely to respond "stay" when *F*1 onset was extremely low. We suggest that this finding indicates that the set of properties that are relevant to a phonetic distinction, and so can trade with each other, is not arbitrary. Instead, the combined results of those earlier experiments, and of Experiments 1 and 3 in this study, support the suggestion of others (e.g., Fowler & Rosenblum, 1991; Liberman & Mattingly, 1985) that acoustic properties are linked in production and perception because they arise from the same articulatory event. The acoustic properties studied in these experiments all arise from the vocal tract being closed and then opened. There is complete silence (a gap) during the time when the vocal tract is actually closed. Then, while the vocal tract is opening for the production of an open vowel, the lowest resonant frequency of the vocal tract (*F*1) rises. If voicing provides a sound source for that resonance shortly after the release of closure, then *F*1 frequency will be quite low at its onset. However, a burst may occur following closure release, so the start of voicing is delayed and *F*1 is higher at its onset. Results from across the several experiments using [s]+stop-vowel stimuli demonstrate that all these properties specify a vocal-tract closing/opening gesture for listeners of all ages. Furthermore, these acoustic–articulatory relations appear to be transparent to the perceiver. Regardless of how frequently (or infrequently) [s]+stop clusters occur without a burst in natural speech, listeners still hear an extremely low *F*1 onset in burstless stimuli as specifying a stop closure. Children do not learn that some arbitrary set of acoustic properties signals a given phonetic structure. Instead, children hear others producing speech and learn that certain recurrent patterns of gesture correspond to specific phonetic structures (Studdert-Kennedy, 1987). These studies on the perception of [s]+stop clusters demonstrate this notion elegantly: Acoustic properties that specify vocal-tract closure in [s] clusters can be used by children and adults, even though those properties may not consistently specify closure in that context in natural speech. Of course, these claims would be challenged by studies showing that complex nonspeech stimuli can demonstrate trading relations, assuming that there would be no reason for listeners to attribute the various properties involved to the same sound-producing event. In fact, several studies with adults using nonspeech analogues of some acoustic properties that have been shown to trade in phonetic perception have demonstrated perceptual effects similar to those for speech stimuli (Hillenbrand, 1984; Parker, Diehl, & Kluender, 1986). However, no study of nonspeech stimuli has demonstrated similarity in perceptual effect across more than two acoustic properties, as this series of experiments has for gap, *F*1 onset, and stop burst. The only common link among these three diverse properties is that they all specify the same articulatory event. Thus, the best explanation for phonetic trading relations remains that acoustic properties engage in this kind of trade because they arise from the same articulatory event.

REFERENCES

American National Standards Institute (1989). *Specifications for audiometers* (ANSI S3.6-1989). New York: Author.

Bailey, P., & Summerfield, Q. (1980). Information in speech: Observations on the perception of [s]-stop clusters. *Journal of Experimental Psychology: Human Perception & Performance*, **6**, 536-563.

Best, C. T., Morrongiello, B. A., & Robson, R. C. (1981). Perceptual equivalence of acoustic cues in speech and nonspeech perception. *Perception & Psychophysics*, **29**, 191-211.

Dorman, M. F., Studdert-Kennedy, M., & Raphael, L. J. (1977). Stop-consonant recognition: Release bursts and formant transitions as functionally equivalent, context-dependent cues. *Perception & Psychophysics*, **22**, 109-122.

Finney, D. J. (1964). *Probit analysis*. Cambridge: Cambridge University Press.

Fitch, H. L., Halwes, T., Erickson, D. M., & Liberman, A. M. (1980). Perceptual equivalence of two acoustic cues for stop-consonant manner. *Perception & Psychophysics*, **27**, 343-350.

Fowler, C. A., & Rosenblum, L. D. (1991). The perception of phonetic gestures. In I. G. Mattingly & M. Studdert-Kennedy (Eds.), *Modularity and the motor theory of speech perception* (pp. 33-59). Hillsdale, NJ: Erlbaum.

Goldman, R., & Fristoe, M. (1986). *Goldman–Fristoe Test of Articulation*. Circle Pines, MN: American Guidance Service.

Greenlee, M. (1980). Learning the phonetic cues to the voiced–voiceless distinction: A comparison of child and adult speech perception. *Journal of Child Language*, **7**, 459-468.

Hillenbrand, J. (1984). Perception of sine-wave analogs of voice onset time stimuli. *Journal of the Acoustical Society of America*, **75**, 231-240.

Jastak, S., & Wilkinson, G. S. (1984). *The wide range achievement test–revised*. Wilmington, DE: Jastak Associates.

Jusczyk, P. W. (1992). Developing phonological categories from the speech signal. In C. A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development: Models, research, implications* (pp. 17-64). Timonium, MD: York Press.

Jusczyk, P. W. (1993). From general to language-specific capabilities: The WRAPSA model of how speech perception develops. *Journal of Phonetics*, **21**, 3-28.

Jusczyk, P. W. (1994). Infant speech perception and the development of the mental lexicon. In J. C. Goodman & H. C. Nusbaum (Eds.), *The development of speech perception: The transition from speech sounds to spoken words* (pp. 227-270). Cambridge, MA: MIT Press.

Klatt, D. H. (1975). Voice onset time, frication, and aspiration in word-initial consonant clusters. *Journal of Speech & Hearing Research*, **18**, 686-706.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, **21**, 1-36.

Mann, V. A., & Repp, B. H. (1980). Influence of vocalic context on perception of the [ʃ]–[s] distinction. *Perception & Psychophysics*, **28**, 213-228.

Milenkovic, P. (1993). *CSpeech Core user's manual Version 4.0*.

Morrongiello, B. A., Robson, R. C., Best, C. T., & Clifton, R. K. (1984). Trading relations in the perception of speech by 5-year-old children. *Journal of Experimental Child Psychology*, **37**, 231-250.

Neely, S. T., & Peters, J. E. (1989). *SPECTO user's guide* (Technical Memorandum No. 14). Omaha, NE: Boys Town National Research Hospital.

Nittrouer, S. (1992). Age-related differences in perceptual effects of formant transitions within syllables and across syllable boundaries. *Journal of Phonetics*, **20**, 1-32.

NITTROUER, S. (1996a). Discriminability and perceptual weighting of some acoustic cues to speech perception by 3-year-olds. *Journal of Speech & Hearing Research*, **39**, 278-297.

NITTROUER, S. (1996b). The relation between speech perception and phonemic awareness: Evidence from low-SES children and children with chronic OM. *Journal of Speech & Hearing Research*, **39**, 1059-1070.

NITTROUER, S., MANNING, C., & MEYER, G. (1993). The perceptual weighting of acoustic cues changes with linguistic experience. *Journal of the Acoustical Society of America*, **94**, S1865.

NITTROUER, S., & STUDDERT-KENNEDY, M. (1987). The role of coarticulatory effects in the perception of fricatives by children and adults. *Journal of Speech & Hearing Research*, **30**, 319-329.

PARKER, E. M., DIEHL, R. L., & KLUENDER, K. R. (1986). Trading relations in speech and nonspeech. *Perception & Psychophysics*, **39**, 129-142.

REPP, B. H. (1982). Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. *Psychological Bulletin*, **92**, 81-110.

REPP, B. H., LIBERMAN, A. M., ECCARDT, T., & PESETSKY, D. (1978). Perceptual integration of acoustic cues for stop, fricative, and affricative manner. *Journal of Experimental Psychology: Human Perception & Performance*, **4**, 621-637.

SIMON, C., & FOURCIN, A. J. (1978). Cross-language study of speech-pattern learning. *Journal of the Acoustical Society of America*, **63**, 925-935.

STEVENS, K. N., & BLUMSTEIN, S. E. (1978). Invariant cues for place of articulation in stop consonants. *Journal of the Acoustical Society of America*, **64**, 1358-1368.

STUDDERT-KENNEDY, M. (1987). The phoneme as a perceptuomotor structure. In A. Allport, D. Mackay, W. Pruiz, & E. Scheerer (Eds.), *Language perception and production* (pp. 67-84). London: Academic Press.

SUMMERFIELD, Q., & HAGGARD, M. (1977). On the dissociation of spectral and temporal cues to the voicing distinction in initial stop consonants. *Journal of the Acoustical Society of America*, **62**, 435-448.

WARDRIP-FRUIN, C., & PEACH, S. (1984). Developmental aspects of the perception of acoustic cues in determining the voicing feature of final stop consonants. *Language & Speech*, **27**, 367-379.

NOTES

1. We have deliberately chosen to use the term *acoustic property* rather than the more common *acoustic cue* because of the narrow definition traditionally associated with the cue. (See Repp, 1982, for a discussion.)

2. *F*1 *onset* is the term used here to refer to the starting frequency of the *F*1 transition. In all stimuli described in this manuscript, whether natural or synthetic, *F*1 rose over roughly the first 40 msec to the same steady-state frequency for the same vowel. Therefore, higher *F*1-onset frequencies correlated with less extensive *F*1 transitions.

3. As is customary, slope is given as the change in ordinate units per change in unit on the abscissa. Ordinarily, we give this value in probits per unit change on the abscissa. In this experiment, however, these values were so small that we multiplied them by 1,000 to make them more readable.

4. Samples produced in isolation by the 1st male and the 1st female speaker were digitized at a 10-kHz sampling rate with low-pass filtering below 4.9 kHz. However, this difference in sampling rate did not produce any noticeable difference in the two measures of interest (*F*1-onset frequency and burst duration.

5. The conclusion that the *F*1 transition decreases in perceptual weight with development is contrary to the suggestion of Simon and Fourcin (1978). Those authors concluded that attention to (i.e., weighting of) the *F*1 transition in decisions of initial-stop voicing is something that must be learned. This conclusion was based on the finding that *F*1 transition influenced the labeling responses of English-learning children, but not those of French-learning children. Because "voiced" stops in French are prevoiced, while "voiceless" stops have a short-lag voice onset time, *F*1-onset frequency does not differ much between the two conditions (according to Simon and Fourcin). Consequently, Simon and Fourcin suggested, French children never need to learn how *F*1 varies with voicing. However, there are numerous difficulties with the construction of stimuli in that report, most of which have been described by Hillenbrand (1984). We add only that the vowels used by Simon and Fourcin were close vowels, and so not much rise in *F*1 frequency would be expected following release of closure. In other words, *F*1 would be expected to start low and stay low.

APPENDIX A

The following carrier phrase was used for eliciting *say* and *stay* tokens produced in context.

> "When you travel, it's funny how you pick up the habits and **sayings** of the people you **stay** with. Recently I was **staying** with friends in the Boston area. They would **say** things like "bubbler" for water fountain and "tonic" for pop. I found that I would be **saying** these same phrases after I had **stayed** there only a short time."

APPENDIX B

For Tokens Produced in Isolation, Mean Formant Frequencies (Hertz) and Burst Durations (Milliseconds) for Women (W1–W10) and Men (M1–M10), Group Means for Women (WM) and for Men (MM), the Grand Mean Collapsed Across Sex (GM), and Standard Deviations

	/(s)eɪ/			/(st)eɪ/			
Speaker	*F*1	*F*2	*F*3	*F*1	*F*2	*F*3	Burst
W1	498	1983	3125	436	2204	3063	17.1
W2	404	1881	2715	456	1882	2741	15.4
W3	449	1856	2471	491	1921	2471	17.6
W4	547	1855	2565	521	1875	2611	23.9
W5	449	1699	3001	482	2025	3021	22.0
W6	540	2005	2806	488	2214	2813	24.1
W7	514	1933	2773	495	2012	2884	19.0
W8	540	2142	2969	459	2270	3151	21.7
W9	365	1816	2598	410	2181	2897	18.4
W10	514	1901	2891	488	1936	2884	14.5
WM	482	1907	2802	472	2052	2867	19.4
SD	68	141	252	41	152	197	4.6
M1	436	1536	2507	426	1621	2546	17.8
M2	540	1615	2520	482	1764	2539	13.7
M3	397	1536	2545	368	1710	2470	14.4
M4	430	1491	2253	404	1543	2305	18.4
M5	436	1582	2480	423	1654	2435	13.8
M6	430	1680	2513	456	1826	2526	18.4
M7	410	1537	2422	423	1875	2572	11.8
M8	443	1589	2370	397	1751	2474	18.4
M9	443	1497	2474	430	1576	2448	14.9
M10	469	1556	2305	449	1595	2415	16.2
MM	443	1562	2439	426	1692	2473	15.8
SD	39	57	99	32	112	79	2.4
GM	463	1737	2620	449	1872	2667	17.6
SD	62	207	267	46	227	251	4.3

(Continued on next page)

APPENDIX C

For Tokens Produced in Context, Mean Formant Frequencies (Hertz) and Burst Durations (Milliseconds) for Women and Men, Group Means for Women (WM) and for Men (MM), Grand Mean Collapsed Across Sex (GM), and Standard Deviations

	/(s)eɪ/			/(st)eɪ/			
Speaker	*F*1	*F*2	*F*3	*F*1	*F*2	*F*3	Burst
W1	482	2064	3184	462	2122	3095	18.7
W2	430	2095	2735	462	2109	2897	20.5
W4	482	1901	3034	462	1975	3008	24.1
W6	534	1934	2774	475	2070	2747	17.6
W9	456	2077	2920	423	2103	2911	16.9
W10	462	1986	2949	462	2057	2832	21.2
WM	474	2009	2933	458	2073	2915	19.8
SD	32	74	152	16	49	113	2.4
M3	430	1543	2493	449	1634	2510	26.4
M4	407	1439	2454	339	1439	2409	14.5
M5	404	1588	2422	410	1641	2441	17.7
M6	417	1667	2715	430	1849	2656	21.7
M8	469	1719	2643	430	1934	2754	21.4
M9	404	1758	2650	449	1693	2389	9.7
MM	422	1619	2563	418	1698	2526	18.6
SD	23	109	111	38	160	134	5.4
GM	448	1814	2752	438	1886	2721	19.2
SD	49	235	245	45	234	247	5.4

(Manuscript received May 2, 1996;
revision accepted for publication January 30, 1997.)