

THE ROLE OF COARTICULATORY EFFECTS IN THE PERCEPTION OF FRICATIVES BY CHILDREN AND ADULTS

SUSAN NITTROUER* MICHAEL STUDDERT-KENNEDY**
Haskins Laboratories, New Haven, CT

Adult listeners are sensitive to the acoustic variations that result from a speaker's coarticulation (or coproduction) of phonetic segments. The present study charted the development of such sensitivity in young children by examining their responses to coarticulatory effects in fricative-vowel syllables. Children, at each of the ages 3, 4, 5, and 7 years, and adults identified tokens from a synthetic /f/-s/ continuum followed by one of four natural vocalic portions: /i/ and /u/, produced with transitions appropriate for either /f/ or /s/. Children demonstrated larger shifts in fricative phoneme boundaries as a function of vocalic transition than did adults, but relatively smaller shifts as a function of vowel quality. Responses were less consistent for children than for adults, and differences between children and adults decreased as children increased in age. Overall, these results indicate that perceptual sensitivity to certain coarticulatory effects is present at as young as 3 years of age. Moreover, the decrease in the sensitivity to vocalic transitions with age suggests that, contrary to a commonly held view, the perceptual organization of speech may become more rather than less segmental as the child develops.

Coarticulation refers to overlapping movements in the production of neighboring or near-neighboring phonetic segments. The acoustic consequences of coarticulation are clearly evident on spectrograms. As early as 1948, Joos described the two major effects. First, lines, vertical to the time axis, cannot be drawn clearly between individual segments. Although the center of a segment may be discernible in the acoustic pattern, regions exist between segments that do not seem to belong uniquely to one or the other. Joos termed this difficulty the problem of perceptual invariance (e.g., Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Pisoni, 1985).

Despite many dozens of studies in the years since Joos wrote, we still have no generally agreed upon solutions to these problems, and no firm understanding of the function of coarticulation (if any) in listening. Is coarticulation necessary and intrinsic to production, and must a listener therefore draw on the contextually variable information that it carries to recover the phonetic message? Or is coarticulation simply a result of a speaker becoming rapid and skillful? If so, are the acoustic consequences of coarticulation merely noise that a listener filters out?

One way of gaining insight into the perceptual function of coarticulatory effects might be to trace its development in children. This was the goal of the present study. Two different outcomes, with different implications, initially seemed possible. If very young children, whose phonetic and phonological skills were still developing, proved sensitive to the acoustic consequences of coarticulation,

we might suspect that coarticulatory information contributed to the recovery of phonetic form. Alternatively, if young children proved relatively insensitive to the acoustic consequences of coarticulation and if their sensitivity increased with age, then we might infer, following Stevens and Blumstein (1978), that speech perception is based initially on invariant characteristics of the signal, and that the ability to use aspects of the signal that vary across contexts is learned through extensive association of this varying information with specific invariant properties.

Researchers interested in listeners' sensitivity to the acoustic consequences of coarticulation have usually varied the phonetic context in which segments were presented for identification. The assumption was that, if identification varied with phonetic context, then listeners were using their knowledge of coarticulation and its acoustic consequences for perception. Although several contextual effects have been investigated, the present experiment focused on just one: the influence of vocalic segments on voiceless fricative identification. We chose fricatives for study, partly because fricative spectra are often said to provide relatively steady-state and invariant perceptual cues, and partly because contextual effects on fricative perception have been quite thoroughly studied in adults.

Adult Studies

Kunisaki and Fujisaki (1977) investigated listeners' abilities to use the acoustic consequences of coarticulation by measuring the effects of vowel context on fricative perception. They prepared a 10-step synthetic fricative continuum (from /f/ to /s/), followed by synthetic /a/ and /u/. These vocalic portions contained first and second formant transitions, but it is unclear from the description whether the transitions were appropriate for /f/ or /s/.

*Current affiliation: Boys Town National Institute, Omaha, NE.

**Also with: Queens College and Graduate Center, City University of New York.

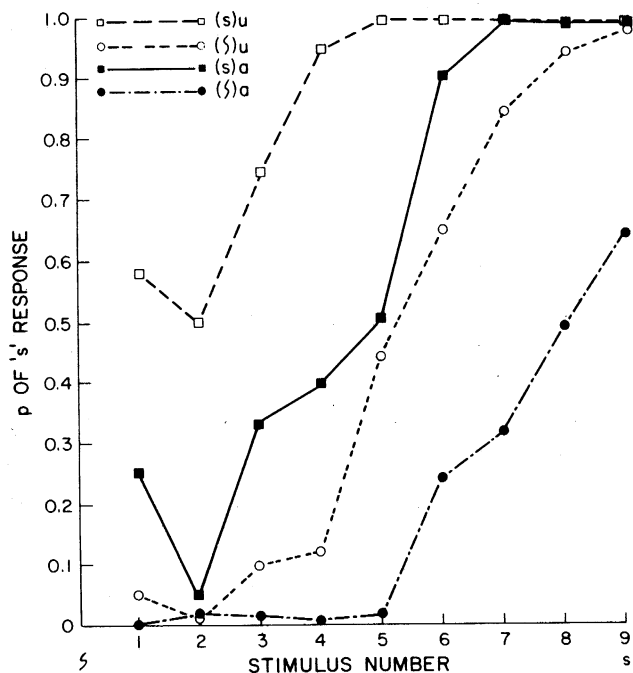


FIGURE 1. Mean identification functions for adult subjects: the probability estimate of an "s" response to fricative-vowel syllables as a function of fricative pole values in a synthetic series ranging from /f/ to /s/, followed by naturally spoken vowels. Parameters of the curves are vowels and the fricatives (in parentheses) after which they were originally spoken (Data from Mann & Repp, 1980).

Nonetheless, these authors demonstrated that the phoneme boundary shifted to a lower value of fricative spectrum (i.e., listeners perceived more stimuli as /s/) when the fricative noise was followed by the rounded vowel /u/ than when it was followed by the unrounded vowel /a/. The explanation proposed by Kunisaki and Fujisaki was that, in natural speech, anticipatory liprounding lowers the pole values of the fricative, so that listeners come to expect that /s/ will have lower pole values when the following vowel is rounded than when it is unrounded. In other words, listeners "know" the acoustic results of liprounding and use this tacit knowledge in their judgments of a synthetic series. The contextual effect in perception is thus linked to coarticulation in production.

An alternative explanation for this effect might be based strictly on properties of general audition. For example, fricative noises with relatively low frequency values might be perceived as relatively higher (i.e., more /s/-like) in the /u/ context than in the /a/ context, due to auditory contrast induced by the lower spectrum for /u/ than for /a/.

Mann and Repp (1980) replicated the vowel context effect of Kunisaki and Fujisaki (1977). They also demonstrated an effect of the initial vocalic transition by appending natural vocalic segments, /a/ or /u/, originally spoken after either /f/ and /s/, to members of a synthetic /f/-/s/ continuum. Identification functions drawn from data obtained in this experiment are shown in Figure 1. It

can be seen that phoneme boundaries shifted as a function of both vowel context and coarticulated formant transition: the likelihood of an "s" response increased both with shifts in the vowel from /a/ to /u/ and with shifts in the fricative from /f/ to /s/ after which the vowel was originally spoken.

However, in a further experiment with flat vocalic formants following the fricative noise, Mann and Repp (1980) found that the vowel context effect virtually disappeared. Thus the question arose as to the actual source of the effect. Did the transition merely provide the coherence between fricative and vowel segments necessary for them to be heard as a single syllable to which both fricative and vowel contribute? Or, was the vowel effect the result of the listener using information in the vocalic transitions rather than vowel quality per se? Or, are transition and vowel context effects separate phenomena?

Evidence existed to suggest that the first possibility provided at least a partial explanation for the vowel context effect disappearing when transitions were removed. Work by Darwin and Bethell-Fox (1977) and by Dorman, Raphael, and Liberman (1979) had demonstrated the importance of signal continuity in providing phonetic coherence. For example, in the Darwin and Bethell-Fox experiment, the perception of a liquid-vowel syllable was changed into the perception of a stop-vowel syllable when fundamental frequency shifted in the region of the segmental boundary. Moreover Mann and Repp (1980) noted that, in the transitionless stimuli, the fricative and vowel segments sounded segregated.

The second possibility, that the transition was solely responsible for the context effect, was eliminated both by the first experiment of Mann and Repp (1980) and by one of Whalen (1981). Both studies demonstrated a vowel context effect when all stimuli provided transitional information appropriate for the same fricative. This effect can be seen in Figure 1.

Whalen (1981) offered support for the third possibility, that both transitions and vowel quality independently affect the placement of the phoneme boundary. Using synthetic fricative noises from an /f/-/s/ continuum and vocalic segments appropriate for /i,u,u,y/ with both /f/ and /s/ transitions, Whalen demonstrated the transition and the vowel context effects with both synthetic and naturally produced vowel segments. Thus, any concern that these effects were artifacts of synthetic speech or of a quirk in the particular natural samples used could be dismissed. The fact that these effects were obtained for English speakers using non-English vowels (/u/ and /y/ suggests that the effects are not related to specific linguistic experience, but rather to some more basic characteristic of auditory or phonetic perception, such as sensitivity to the effects of liprounding on fricative spectra.

Child Studies

Only one study has investigated a phonetic context effect in children (Mann, Sharlin, & Dorman, 1985). In this study, the /f/-/s/ continuum of Mann and Repp (1980)

was combined with naturally produced vocalic segments taken from utterances of /ferv/ and /su/. Consequently, although the vowel context effect could be studied, the transition effect could not. Subjects were adults, 5-year-olds, 7-year-olds with age-appropriate speech production skills, and 7-year-olds judged to misarticulate fricatives. The resulting identification functions (pooled for each subject group) were similar in one respect for the adults and all 3 groups of children: More "s" responses were given in the /u/ context. The sizes of the shifts in mean phoneme boundary as a function of vowel context were also similar for all groups (a shift of approximately one stimulus on the continuum). Although values for slopes were not given, visual inspection of the identification functions suggests that they were equally steep for both groups of 7-year-olds and for adults, but may have been somewhat less steep for the 5-year-olds: In other words, the 5-year-olds may have been less consistent in their responses.

These apparent differences in slopes for the 5-year-olds and for the 7-year-olds and adults are of particular interest. The simplest explanation for these differences is that they reflect the variability usually associated with a partially learned (or not fully established) response. However, the degree of control exerted by formant transitions on responding was not assessed because the same formant transitions (those appropriate to /s/) were used for all stimuli. If the responses of the 5-year-olds were more strongly controlled by formant transitions than the responses of 7-year-olds and adults, the fixed transition stimuli of this study would probably have made these stimuli particularly ambiguous for 5-year-olds. Thus, the age-related slope differences may have reflected a substantive difference in perceptual processing rather than simple variability of response. It would therefore be of interest to investigate the transition effect in children's speech perception to see if children do indeed learn to use different kinds of coarticulatory information at different ages. The chance of discovering such differences would presumably be increased by extending the age range of the study to include children younger than 5 years.

One other study deserves mention. Morrongiello, Robson, Best, and Clifton (1984) investigated the ability of 5-year-olds to use multiple cues in recognizing a stop between an initial consonant and a following vowel. Briefly, a natural /s/ segment was placed before each of two synthetic patterns identical in all aspects except for the extent of an initial F1 transition. One pattern, heard by adults equivocally as either "ay" or "day," had a short F1 transition of 181 Hz; the other pattern, always heard by adults as "day," had a long F1 transition of 381 Hz. An acoustic continuum from "say" to "stay" was then constructed by inserting a silent interval, varying in duration from 0 ms to 104 ms, between the /s/ and the synthetic patterns. For adults, the crossover point between the identification of "say" and of "stay" varied with the extent of the F1 transition: Significantly more "stay" responses were given to tokens with the longer transition (Best, Morrongiello, & Robson, 1981). Morrongiello et al.

found that, although children also gave more "stay" responses to tokens with the longer F1 transition, the shift in phoneme boundary was smaller for children than for adults. This decreased boundary shift was entirely due to the children giving more "stay" responses to tokens with the shorter vocalic transition. Morrongiello et al. proposed that the children may have weighted the transitional information relatively more heavily, and the temporal information relatively less heavily, than adults do. That is, perhaps the children were more sensitive to transitional cues than adults, a possibility encouraged by the finding that any transition—even a brief one—was sufficient to elicit some "stay" responses from the children. This suggestion provides a third possible outcome for the present experiment: that children would prove even more sensitive to some kinds of coarticulatory effects than adults.

Present Study

The main purpose of the present experiment was to trace the development of perceptual sensitivity to the acoustic consequences of coarticulation by comparing children's shifts in phoneme boundaries as a function of phonetic context with those of adults. If the same sized shifts were seen, then this could imply that perceptual use of contextually variable properties does not develop gradually from an initial sensitivity to putatively invariant properties of the speech signal, but is, in some sense, intrinsic to the process of speech perception. If children showed smaller shifts or no shifts, this would imply that the ability to use contextually variable information is an acquired skill (Stevens & Blumstein, 1978)—perhaps of particular utility in following the "reduced" patterns of casual adult speech.

Finally, if children demonstrated relatively greater sensitivity to coarticulatory effects than adults, this could imply that they are more dependent on information specifying syllable coherence: Perhaps young children are not as adept as adults at recovering the individual phonemes from the syllable, but instead tend to perceive syllables as relatively undifferentiated wholes. If children do not recover individual segments to the same extent as adults, it might be predicted that responses would be less consistent for children than for adults, indicating that relatively less weight is being assigned to the spectrum of the fricative noise per se and more weight to the syllabic structure specified, for example, by formant transitions.

METHOD

Subjects

One group of 12 adults and 4 groups of 8 children of the ages 3, 4, 5, and 7 years participated in this experiment. All adults were between 20 and 35 years old, and all children were within -1 and +5 months of their designated

age. No age group had greater than a 5 to 3 ratio between sexes. All subjects were native speakers of American English, speaking a dialect typical of the Middle Atlantic States, and all the children had age-appropriate articulation skills, as judged independently by two speech-language pathologists from recordings of spontaneous speech.

Each subject also passed audiometric and tympanometric screenings. Pure tones of 500 Hz, 1000 Hz, 2000 Hz, and 4000 Hz were presented at 20 dB HL for the audiometric evaluation. Tympanometric results for all subjects displayed normal pressure peaks between +100 daPa and -150 daPa.

Finally, all subjects were required to identify accurately the endpoints of the /f/-s/ continuum, followed by /i/ and /u/ with appropriate vocalic transitions, at greater than 90% accuracy during an initial practice trial. In addition, any subject failing to show greater than 80% correct identification for these good exemplars during the actual testing was eliminated. This requirement served as a partial check on the attentiveness of the subjects. There was no way to measure subjects' attentiveness (either for children or for adults) to the more ambiguous stimuli, and so it was necessary to assume that attention remained constant across stimuli.

One 3-year-old was eliminated because he failed to identify the endpoint stimuli with appropriate transitions at greater than 90% accuracy during the practice session. One 4-year-old was eliminated because she did not respond with greater than 80% accuracy to these tokens during the actual testing. Five children were eliminated because their speech was judged to be less well developed than expected for their ages (three 3-year-olds, and one child at each of the ages of 4 and 5).

Stimuli

For the present study, we needed stimuli that would permit us to observe the effects on fricative perception of fricative spectrum, vowel context, and vocalic transition. We therefore prepared hybrid syllables composed of synthetic fricative noises and natural vowels, analogous to those of Mann and Repp (1980) and of Whalen (1981). The fricatives varied along a 9-step continuum from /f/ to /s/, and the vocalic portions were excerpted from naturally spoken /fi/, /si/, /ju/, and /su/ syllables, so that formant transitions were appropriate for either /f/ or /s/.

The stimuli were constructed on the basis of pilot tests and consisted of tokens from an /f/ to /s/ continuum created on the serial software synthesizer at Haskins Laboratories. Each token contained 210 ms of fricative noise with a single pole value specified. There were 9 tokens on this continuum, with pole values varying from 2200 Hz to 3800 Hz in 200-Hz steps. Each token also had a zero at $0.75 \times$ the pole value, creating a dip in the spectrum. Thus, we synthesized, in effect, two-pole fricatives, more similar to natural speech than might have sufficed for the perception of /f/ and /s/. For example, Heinz and Stevens (1961) demonstrated that a single-pole stimulus is sufficient for listeners to perceive fricatives;

TABLE 1. Formant transition characteristics of vocalic portions used in the present study and by Whalen (1981).

Vocalic portions	F2			F3		
	onset (Hz)	dur. (ms)	steady-state (Hz)	onset (Hz)	dur. (ms)	steady-state (Hz)
<i>Present study</i>						
/f/i/	2070	0	2070	2530	60	2580
/s/i/	1870	80	2050	2430	140	2600
/f/u/	1940	300	1000	2350	60	2150
/s/u/	1620	300	1000	2340	170	2130
<i>Whalen, 1981</i>						
/f/i/	2100	130	2300	2750	130	2900
/s/i/	1700	130	2300	2500	130	3000
/f/u/	1800	300	850	2100	0	2100
/s/u/	1600	300	850	2750	130	2100

however, we wanted stimuli as close to natural speech as possible because children were serving as subjects. Thus, any concern that children may have responded differently from adults because they had difficulty hearing synthetic stimuli as speech was mitigated. The amplitude of these tokens rose gradually by 20 dB over the first 170 ms, remained constant for the next 20 ms, then fell by 5 dB over the final 20 ms. Fricative noises were synthesized with a sampling rate of 10 kHz and were low-pass filtered at 4.9 kHz.

The vocalic portions of the stimuli were the vowels /i/ and /u/ produced by a male speaker in the syllables /fi/, /si/, /ju/, and /su/. The unrounded vowel /i/ was chosen instead of /a/ because second formant trajectories are more distinct between /i/ and /u/ than between /a/ and /u/ (Soli, 1981). The speaker originally produced 5 samples of each syllable, but because Mann and Repp (1980) had found little token-to-token variability in results, it seemed sufficient to use a single sample of each. Tokens were selected so that they matched one another as closely as possible in duration and intonation contour. These syllables were digitized on a VAX computer using a 10-kHz sampling rate and low-pass filtering with an upper cut-off of 4.9 kHz. The vowels, including the vocalic transitions, were isolated from the fricative noise with a waveform editing program. The vocalic segments were between 350 ms and 370 ms in duration. Spectrographic displays and linear predictive code (LPC) analysis were used to obtain frequency values for formant transitions. These values are given in Table 1 along with the transition values from Whalen (1981) for comparison. Thus, there were nine fricative noises combined with four vocalic segments, resulting in a total of 36 unique stimuli. Each stimulus was presented 10 times, making a total of 360 test tokens.

Four test tapes were prepared: two using only the /i/ context, and two using only the /u/ context. Ten practice items began each tape (the two endpoints with appropriate vocalic segments presented five times each in random order). The 90 test items per tape were then presented in five randomized blocks of 18. Before each group of 10 stimuli, a female voice announced the number of the next group of stimuli to be heard (numbers 1 through 9). The interstimulus interval was 2 s. Tapes were made on

Scotch Audio Tape 208 at a tape speed of 19 cm per second (7.5 ips) and at a -5 VU recording level.

Materials and Equipment

Stimuli were presented over a Uher Model 4200 tape recorder using Sennheiser earphones. A group listening station permitted the adults to listen in groups of four and one of the experimenters to listen simultaneously with the child being tested. A Beltone Model 9D portable audiometer and a Grason-Stadler Model 27 portable tympanometer were used to screen pure-tone threshold and middle ear function.

For the children, four hand-drawn pictures were prepared to correspond to each of the four possible responses. One picture was of a shoe; one was of a girl named "Sue"; one was of a boy pointing and saying "see"; and the last was of a girl referred to by the pronoun "she." Several board games were also prepared. They consisted of brightly colored squares and circles (nine on each board) with the numbers 1 through 9 written on the spaces. A small plastic animal served as a marker that was moved to the next space each time the female voice was heard announcing the number of the next group of 10 stimuli. Toy stickers, selected before the start of each game, were given as prizes at the end of the game.

Procedure

Each subject took the hearing screening before testing. Next the children were shown the two pictures to be used in the first perceptual test (either "see" and "she," or "Sue and "shoe"). Practice was provided using live voice to familiarize the child with the labeling procedure. This procedure consisted of having the child point to the picture illustrating the stimulus perceived and say the label associated with that picture. If there was a discrepancy between what the child seemed to be saying and the picture, clarification was requested by the experimenter. After the child correctly responded to 10 practice items with live voice, the 10 taped practice items were presented.

If the child correctly responded to 9 out of 10 of these correctly, then the board game was introduced, and the actual testing started. Stimuli were presented over earphones at approximately 75 dB SPL. Two experimenters were needed during the testing of the 3- to 4-year-olds: one to work with the child, listening and watching for discrepancies between his/her pointing and verbal responses as well as just maintaining attention with constant eye contact and occasional verbal praise, and one to stop the tape between stimulus presentations when necessary and to record the child's responses. The experimenter working with the child was not able to hear the stimulus presentations. The experimenter recording children's responses wore the earphones around her neck so that she could hear when a stimulus had been pre-

sented but could not hear it well enough to form her own impression of whether it was /s/ or /ʃ/. Thus, there was no danger of the response recorded being influenced by her own perception. Only one experimenter was needed for testing 5- and 7-year olds.

Testing was divided into three sessions for the children: two sessions of 20–30 min duration, and one of 10–15 min duration. Activities during the sessions were arranged as follows:

Session 1: Hearing screening and first half of the first perceptual test (one test tape).

Session 2: Second half of first perceptual test and first half of second perceptual test (two test tapes).

Session 3: Second half of second perceptual test (one test tape).

All testing was done in one day with breaks between sessions. Testing took place either in the home or in a daycare facility. The order of presentation of the perceptual tapes was counterbalanced across subjects within each age group. Adults simply responded by writing whether they heard "s" or "sh." In addition, adults listened in groups of four, and all tapes were presented during one session, with the /i/ and /u/ tapes alternated.

RESULTS

Probit analyses were done on individual and group response proportions (probability estimates) for each context. (The general term *context* is used here to refer to the four vocalic portions used: /ʃ/i/, /s/i/, /ʃ/u/, and /s/u/.) Probit analysis fits a cumulative normal curve to probability estimates as a function of stimulus level by the method of least squares (Finney, 1971), estimating the mean (phoneme boundary) and standard deviation for each distribution. The slope is obtained by taking the reciprocal of the standard deviation and, therefore, serves as an index of consistency. Figures 2 through 6 display the identification functions obtained for each age group from probit analyses of mean probability estimates. Table 2 provides mean phoneme boundaries for each age group in each context. The smaller (or lower) the phoneme boundary value, the more "s" responses were given. Results should be compared (a) between different vocalic transitions for the same vowel [/ʃ/i/ vs. /s/i/ and /ʃ/u/ vs. /s/u/], and (b) between different vowel contexts for the same vocalic transition [/ʃ/i/ vs. /ʃ/u/ and /s/i/ vs. /s/u/]. From the adult studies reviewed above, we would expect lower phoneme boundaries for tokens with /s/ transitions in the first case, and lower boundaries for /u/ contexts in the second.

Inspection of the mean phoneme boundaries given in Table 2 reveals that the predicted transition effect is apparent within both vowel contexts (lower phoneme boundaries for tokens with /s/ transitions), but the size of this effect appears greater for /u/ than for /i/. The expected vowel effect occurred only for tokens with /s/ transitions [a lower phoneme boundary for the /s/u/ context than for the /s/i/ context]. For tokens with /ʃ/ transitions, the obtained vowel effect was opposite to what would be

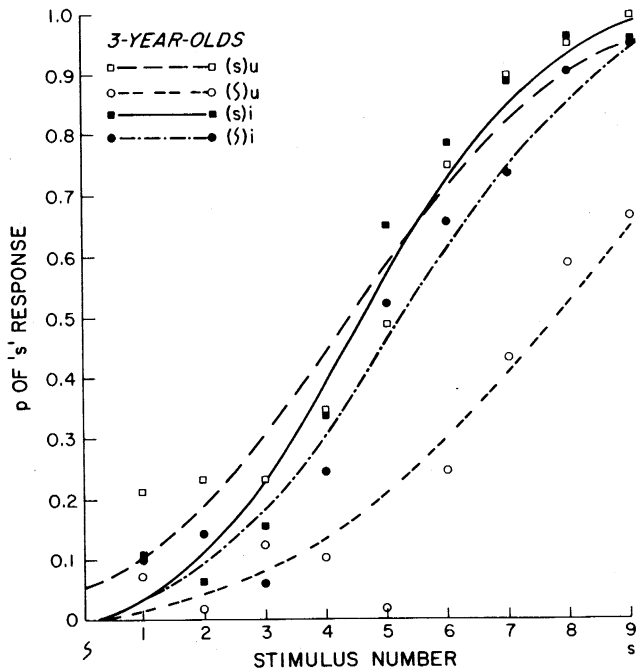


FIGURE 2. Mean identification functions for 3-year-olds: Symbols represent mean probability estimates at each stimulus step for each context (i.e., the vowels /i/ or /u/, with formant transitions appropriate for /ʃ/ or /s/). Lines are the cumulative normal curves fitted to these mean probability estimates with probit analysis.

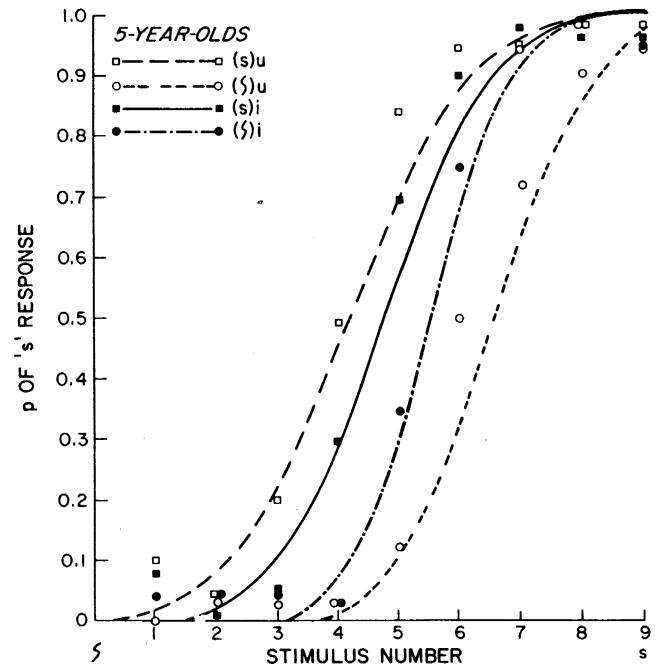


FIGURE 4. Mean identification functions for 5-year-olds. See Figure 2 caption for additional details.

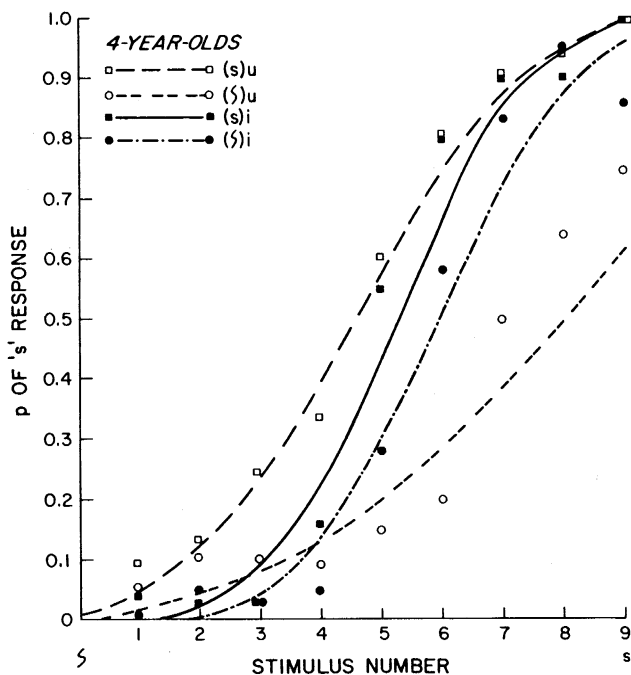


FIGURE 3. Mean identification functions for 4-year-olds. See Figure 2 caption for additional details.

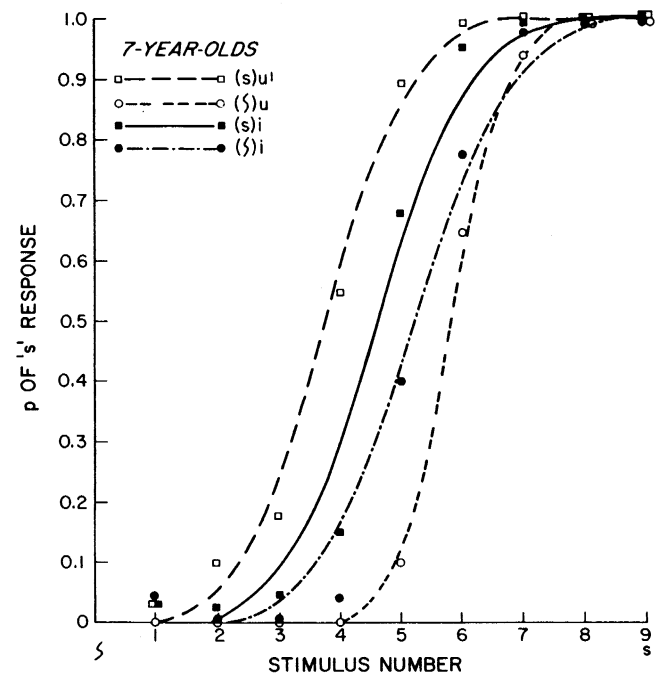


FIGURE 5. Mean identification functions for 7-year-olds. See Figure 2 caption for additional details.

predicted from previous studies: A lower phoneme boundary was obtained in the /ʃi/ context than in the /ʃu/ context. In fact, all age groups demonstrated the highest phoneme boundary in the /ʃu/ context, indicating more "sh" responses in this context than in any other.

A 3-way ANOVA (Age \times Vowel \times Transition) performed on individual phoneme boundaries revealed a significant main effect of transition [$F(1,39) = 171.64, p < .001$], and significant interaction effects for Transition \times Vowel [$F(1,39) = 75.30, p < .001$], for Transition \times Age [$F(4,39) = 3.15, p = .02$], and for Vowel \times Age [$F(4,39) = 2.61, p = .05$]. Planned comparisons indicated a signifi-

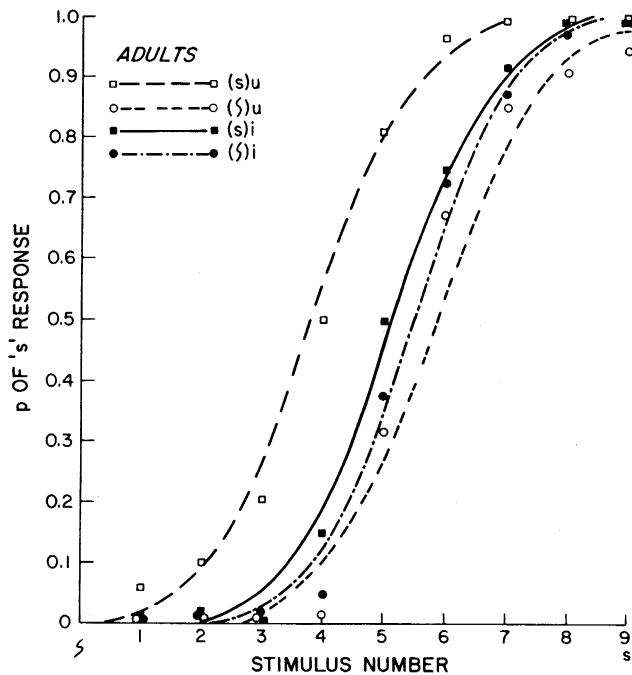


FIGURE 6. Mean identification functions for adults. See Figure 2 caption for additional details.

TABLE 2. Mean phoneme boundaries for each age group (computed from individual phoneme boundaries), with group standard deviations in parentheses.

Groups	/f/i/	/s/i/	/f/u/	/s/u/
Adults	5.46 (0.98)	5.14 (1.03)	5.94 (1.23)	3.84 (1.04)
7-year-olds	5.21 (0.61)	4.63 (0.66)	5.81 (0.39)	3.74 (0.90)
5-year-olds	5.53 (0.57)	4.85 (0.48)	6.26 (0.99)	4.32 (1.53)
4-year-olds	5.95 (0.55)	5.22 (0.75)	7.89 (1.70)	4.51 (0.58)
3-year-olds	5.20 (0.83)	4.60 (0.55)	7.87 (2.16)	3.91 (2.38)

cant, overall effect of age on the placement of phoneme boundaries for the 3-, 4-, and 5-year-olds versus 7-year-olds and adults [$F(1,39) = 4.67, p = .04$].¹ The comparison between 3-, 4-, and 5-year-olds versus 7-year-olds and adults also showed significant results for the vowel effect

¹The following orthogonal, planned comparisons were performed on data for each factor (i.e., Age, Vowel, and Transition):

Group 1	vs.	Group 2
7-year-olds	vs.	Adults
3-year-olds	vs.	4-year-olds
3- and 4-year-olds	vs.	5-year-olds
3-, 4-, & 5-year-olds	vs.	7-year-olds and Adults

These comparisons were intended to locate, if possible, breaks in the developmental sequence from child to adult. Only those comparisons that resulted in statistical significance are reported here.

TABLE 3. Mean slope values^a for each age group (computed from individual slope values).

Groups	/f/i/	/s/i/	/f/u/	/s/u/
Adults	1.37	1.33	1.28	1.24
7-year-olds	1.30	1.48	1.73	1.58
5-year-olds	1.40	1.09	1.01	1.08
4-year-olds	0.86	1.00	0.42	0.60
3-year-olds	0.47	0.50	0.37	0.52

^aSlope values are the reciprocals of the standard deviations of the fitted normal curves. They estimate the changes in probability of an "s" response per unit change in the fricative noise.

[$F(1,39) = 6.75, p = .01$], and for the transition effect [$F(1,39) = 6.04, p = .02$]. Thus, there was a difference between the younger children as compared to the 7-year-olds and adults in the relative placement of phoneme boundaries. Specifically, phoneme boundaries were affected differently by both vowel and transition for the younger children than for the 7-year-olds and adults. Finally, the planned comparison between 3- and 4-year-olds versus 5-year-olds for the transition effect showed a significant difference [$F(1,39) = 5.82, p = .02$]. The size of the transition effect decreased with increasing age.

Inspection of Table 3, which lists mean slopes for each age group, reveals that slope values were generally greater for older subjects. Adults and 7-year-olds demonstrated mean slopes above 1.00; slopes for 5-year-olds were closer to 1.00; and, with one exception, slopes for 4- and 3-year-olds were below 1.00. Looking across contexts within age groups it can be seen that slope values were fairly consistent. Only 4-year-olds seem to display different slope values as a function of context, with the values for /i/ higher than the values for /u/. However, this difference is not statistically significant. A 3-way ANOVA (Age \times Vowel \times Transition) performed on individual slope values revealed a significant main effect of age [$F(4,39) = 13.79, p < .001$]. Planned comparisons indicated significant differences in slope values for 3-, 4-, and 5-year-olds versus 7-year-olds and adults [$F(1,39) = 38.30, p < .001$], and for 3- and 4-year-olds versus 5-year-olds [$F(1,39) = 14.31, p < .001$]. Therefore, consistency of responses increased with increasing age for these subjects.

DISCUSSION

The "Reversed" Vowel Effect

The first finding to be discussed is an apparent discrepancy between results of this study and those of earlier studies. The subjects in the present experiment gave more "s" responses to the unrounded vowel than to the rounded vowel for tokens with /f/ transitions. Not only does this finding differ from those of earlier studies, it is contrary to the suggestion made by Kunisaki and Fujisaki (1977) that more "s" responses would be expected before

rounded than before unrounded vowels, regardless of transition, because fricative poles are lower before rounded vowels.

This “reversed” vowel effect found for tokens with /f/ transitions is probably related to the inequality in the relative size of transition effects found between vowel contexts. The predicted transition effect for these stimuli was that more “s” responses would be given to tokens with /s/ vocalic transitions. This effect was obtained, but to a greater extent for /u/ than for /i/. Both of these findings (i.e., the “reversed” vowel effect for tokens with /f/ transitions and unequal transition effects for /i/ and /u/) may be explained by comparing the acoustic characteristics of the formant transitions used in this and previous studies.

The extents of F2 transitions are more similar for /a/ and /u/ than for /i/ and /u/ (Soli, 1981). This follows from the relatively shorter distance that the tongue must travel from the point of fricative constriction to the front vowel /i/ than to the back vowels /a/ and /u/. The pattern of results obtained by Mann and Repp (1980) for tokens with /f/ transitions may, therefore, have been largely an effect of the following vowel. That is, to say, because /a/ and /u/ are similar in the extent of their transitions from the fricative constriction, differences between vocalic portions with transitions appropriate for the same fricative would be entirely due to vowel quality.

The /i/ portions used by Whalen (1981) exhibited greater transitional differences than the /i/ portions used in this study: Differences in F2 onset values were twice as large between /(f)i/ and /(s)i/, and the extents of /i/ transitions were greater. Moreover, the acoustic differences between the /(f)u/ and /(s)u/ portions of Whalen’s study were not as great as the corresponding differences between the /u/ portions of this study. In particular, the /(f)u/ portion used by Whalen does not provide as much transitional information as the /(f)u/ portion used here because F3 shows no change in Whalen’s /(f)u/. In short, the extent of the transitions for /i/ and /u/ tokens seems to have been more similar in Whalen’s syllables than in those of the present study. Therefore, differences in placement of /(f)i/ and /(f)u/ boundaries in Whalen’s study would have been a function more of vowel context than of differences in the extents of the transitions in the two contexts.

We thus hypothesize that the perceptual weight given to transitions is proportional to their extents. In fact, precisely this result has been reported in a study of the relative perceptual weights assigned to bursts and transitions in naturally spoken stop-vowel syllables (Dorman, Studdert-Kennedy, & Raphael, 1977). In the present study, the transitions of the /(f)i/ portion displayed very little formant movement and may therefore have roughly represented a null condition; /(s)i/ had weak transitions favoring “s”; /(s)u/ had somewhat more perceptually salient transitional information favoring “s”; but /(f)u/ provided very strong transitional information favoring “sh.” Superimposed on this pattern seems to be an independent vowel effect that tends to move the two /u/ functions

in the direction favoring “s” responses—at least for adults, who demonstrate sensitivity to vowel quality.

Clearly, the relation between the size of transitional differences found in acoustic measurements and the size of boundary shifts observed in perception warrants more systematic investigation. In most previous studies of phonetic context effects, vocalic transitions were manipulated as nominal categories: that is, they were treated as either /s/ or /f/ transitions. This measure may be too coarse. Vocalic transitions might better be regarded as continuous variables, relatively more /s/-like or more /f/-like. In any event, the finding that subjects in the present experiment gave more “s” responses to the /(f)i/ context than to the /(f)u/ context argues against explanations of the vowel effect based primarily on auditory contrast. If the predicted vowel effect were a consequence of the fricative noise sounding higher when the vowel spectrum is lower, as in /u/ compared to /i/, this effect should have been found regardless of transition.

Comparisons Between Children and Adults

In considering whether children demonstrated sensitivity to coarticulatory effects to a greater or lesser extent than adults did, three factors must be discussed: the vowel effect, the transition effect, and the extent to which fricative identification was based on the fricative spectrum itself.

Vowels and transitions. The expected vowel effect for these stimuli was that more “s” responses would be given to /u/ tokens than to /i/ tokens due to the lowered fricative poles associated with rounded vowels. Children in this study demonstrated a weaker effect of this sort than did the adults. Visual inspection of the mean identification functions for each age group clearly shows a greater separation between the (s)i and (s)u functions for adults than for children, and for 7-year-olds than for younger age groups.

The predicted transition effect (more “s” responses to tokens with /s/ vocalic transitions) was obtained for all age groups for both vowels, but the size of this effect was larger for /u/ than for /i/. There was also an effect of age on the size of the transition effect: Younger subjects demonstrated greater transition effects than did older subjects. This age effect appears to be due largely (though not entirely) to younger subjects giving substantially more “sh” responses to the /(f)u/ context than older subjects did.

For these /(f)u/ tokens, two opposing factors presumably influenced the locations of phoneme boundaries: Perceptually salient transitional information biased responses toward “sh,” while the rounded vowel biased responses toward “s.” The fact that younger subjects gave substantially more “sh” responses to /(f)u/ tokens than did older subjects seems to be another indication that younger subjects were more influenced by transitional information than older subjects were, and they were also less influenced by vowel quality. While /(s)u/ tokens provided perceptually salient transitional information

concerning fricative identity, vowel quality biased responses in the same direction (toward "s" responses). Therefore, we cannot assess the relative contributions of vowels and transitions to responses in this context.

These results suggest that vowel effects and transition effects are additive.² The negative boundary shift obtained for the /*(f)*i-*(f)*u/ difference evidently reflects the overwhelming cancellation of the vowel effect by the /*(f)*u-*(s)*u/ transition effect: Vowel and transition effects are evidently orthogonal. The fact that the 3- and 4-year-olds demonstrate a large negative vowel effect for /*f*/ tokens is thus a result of their demonstrating a very large transition effect for /u/.

Fricative spectrum. Another purpose of the present investigation was to study the extent to which fricative identification was associated with the fricative spectrum itself. To make this determination, we must look at the slopes of the identification functions. The finding that 3-, 4-, and 5-year-olds exhibited shallower slopes than did 7-year-olds and adults demonstrates that they were less consistent in responding. This decrement in consistency exhibited by young children might simply reflect an inability to pay attention to the task at hand. However, the fact that all subjects responded with greater than 80% accuracy to the end points presented with the appropriate vocalic transition argues against this interpretation. An alternative account is that the lack of consistency reflected the children's relatively lower sensitivity to the fricative spectrum. One adult subject who appeared to attend selectively to the fricative noise, and to make decisions about category based primarily on its spectrum, was Subject A8, whose identification functions are shown in Figure 7. These functions show very consistent responses and no effect of vocalic transition. Yet they do exhibit a vowel effect.

This last result is important because it suggests that attention to the fricative spectrum is associated with a vowel effect. Vowel context effects are found because the fricative spectrum differs as a function of the quality of the following vowel. If listeners did not attend to the fricative spectrum to some extent, then they would not be able to determine whether this spectrum differed as a consequence of the following vowel.

This relation between attending to the fricative spectrum and showing vowel effects might explain why young children did not demonstrate vowel effects to the same extent as adults did in this study. Morrungiello et al. (1984) reported that the children in their study appeared to weigh the transitional information relatively more heavily than the silent interval, as compared to adults. That is, children seemed to be relatively more sensitive to the transitional pattern than to the silent interval. The

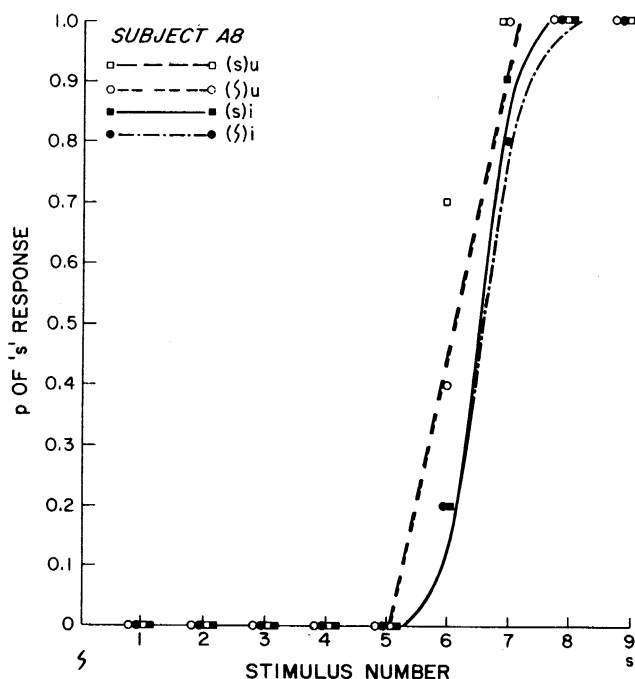


FIGURE 7. Identification functions for Subject A8. See Figure 2 caption for additional details.

results obtained in the present experiment show a similar pattern: Children were apparently more sensitive to the transitional pattern than to the fricative spectrum, as compared to adults.

The shapes of the identification functions also indicate the increased reliance on transitional information of the children in this study. If categorical decisions were weighted more heavily by vocalic transitions than by fricative spectrum, we would expect responses to be biased in favor of the fricative associated with that transition and many of the stimuli to be ambiguous. Therefore, phoneme boundaries would be either extremely low or extremely high, and slope values would be low.

An example of a 3-year-old subject whose responses to /u/ tokens seem to have been controlled primarily by transitional information is Subject 32, whose identification functions are shown in Figure 8. Phoneme boundaries are extreme for /*(f)*u/ and /*(s)*u/, and slope values are low for these contexts. Similar patterns (extreme phoneme boundaries and low slope values) can be found in results obtained for several of the other 3- and 4-year-olds but only for /u/ contexts. No subject displayed a pattern of results indicating strong influence of transitions in either of the /i/ contexts, probably because /i/ transitions provided less salient information about fricative identity than did /u/ transitions. However, many of the 3-year-olds displayed very low slope values, simply indicating inconsistent responses for the /i/ contexts. Apparently, even when transitional information was greatly reduced, 3-year-olds were not able to increase their use of the cues provided by the fricative spectrum for fricative identification.

Four-year-olds seem to have been able to switch response patterns to some extent depending on how much

²On first consideration, the statistically significant Vowel \times Transition interaction may belie this statement; however, the statistical analysis treats values on these factors as nominal categories. In fact, the /u/ transitions provided more information concerning fricative identity than did the /i/ transitions, and it is precisely this inequality of information that we believe helps support the claim of orthogonality of effects.

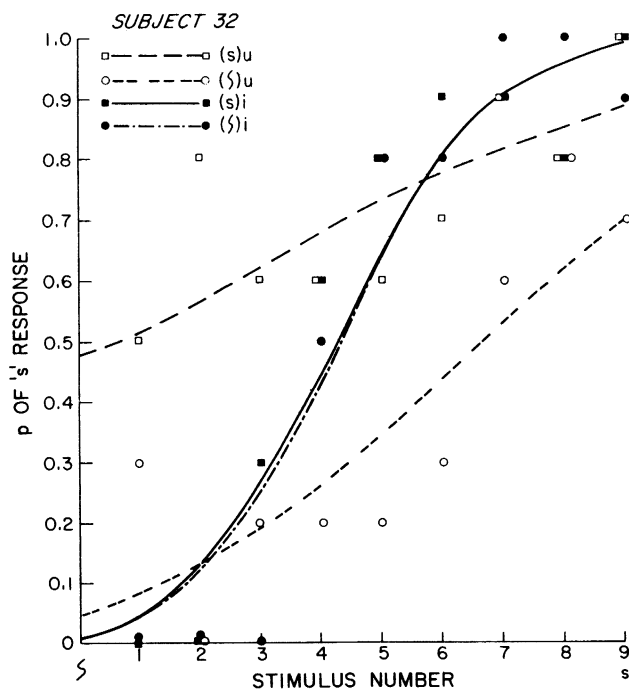


FIGURE 8. Identification functions for Subject 32. See Figure 2 caption for additional details.

information was provided by the transitions. For the /u/ contexts in which the transitions provided perceptually salient information concerning fricative identity, responses seem to have been based largely on vocalic transitions, as indicated by the low slope values and extreme phoneme boundaries. For the /i/ contexts in

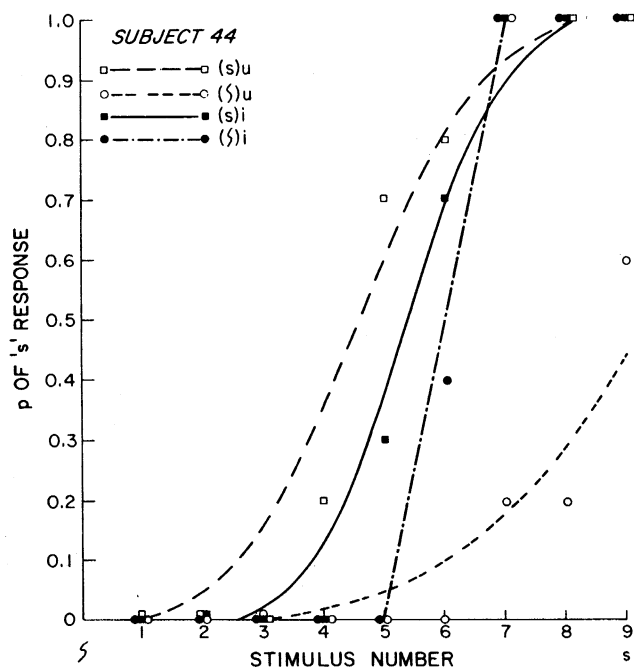


FIGURE 9. Identification functions for Subject 44. See Figure 2 caption for additional details.

which transitions provided little information, four-year-olds seem to have been able to use the fricative spectrum to make decisions about fricative identity to at least some extent, as indicated by higher slope values.

A good example of this response pattern was provided by Subject 44, whose identification functions are shown in Figure 9. Responses for the /*(f)u*/ context are strongly influenced by the vocalic transitions, as indicated by the shallow slope and extremely high phoneme boundary, but responses for the /*(f)i*/ context are clearly based on the fricative spectrum, as indicated by the very steep slope.

Another developmental trend found in these data is that 7-year-olds performed similarly to adults in most respects. Their responses were as consistent as those of adults, indicating strong reliance on the fricative spectrum for fricative identification, and the magnitude of the transition effect for individual 7-year-olds was equivalent to that obtained for individual adults. The only difference between adults and 7-year-olds was in the magnitude of the vowel effect: 7-year-olds demonstrated slightly reduced vowel effects. In general, though, 7-year-olds appear to have developed speech perception strategies that are similar to those used by adults.

Conclusions

Contextual effects on young children's speech perception were initially studied with the expectation that they would be less than or equal to the effects observed in older children and adults. For vowel quality, the effects were indeed reduced: Young children seemed less able than older children and adults to attend to those portions of a syllable primarily associated with the individual segments (that is, with the steady state friction and vocalic formants). On the other hand, for the vocalic transitions, the reverse was found: Younger children were more sensitive than older children and adults to those portions of a syllable that ensure its perceptual coherence.

These results suggest that the younger children were still listening for whole words or syllables, much as (we may infer from their productions) do the 1- to 2-year-olds described by Menyuk and Menn (1979). Of course, it is unlikely that each of the many hundreds of entries in a 3- to 4-year-old's lexicon is still a whole word, quite lacking in segmental structure. But it is not implausible to suppose that 3- to 4-year-olds are still discovering how segments are packaged in the syllable and are not yet as adept at recovering them as they will soon become. In fact, this suggestion receives support from many studies indicating that preliterate children do not have access to the phonemic structure of speech (e.g., Fox & Routh, 1975). However, this interpretation runs contrary to the commonly held view (e.g., Kent, 1983) that the perceptual (and a fortiori articulatory) organization of speech becomes less rather than more segmental as the child develops.

Such an interpretation also runs counter to the claim of Stevens and Blumstein (1978) that sensitivity to coartic-

ulation in adult speech perception is a secondary effect, learned by association with a primary invariant. Our results suggest rather that perceptual sensitivity to certain forms of coarticulation is present from a very early age and, therefore, may be intrinsic to the process of speech perception. The child does not use segments to discover coarticulation, but rather coarticulation to discover segments. The adult, though more skilled at segmental recovery than the child, may still do much the same.

ACKNOWLEDGMENTS

The work reported here was part of a thesis submitted by the first author to the City University of New York in partial fulfillment of the requirements for the doctoral degree in speech and hearing sciences. The authors are grateful to Arthur Boothroyd and Lawrence Raphael, who served as thesis advisors. Bruno Repp and D. H. Whalen offered valuable suggestions at every stage of this work; Carol Fowler reviewed an earlier draft of the paper; Denise Wright, Melanie Campbell, Lise Jensen, Karen Selvaggi, Patricia Nittrouer, and Gloria Schlisselberg helped with various aspects of the data collection and analysis. We thank each of these individuals. In addition, we are grateful to the children who participated in this study, and to their parents who permitted their participation. This work was supported in part by NICHD Grant HD-01994 to Haskins Laboratories. The writing was completed while the first author was a postdoctoral fellow at Haskins Laboratories on NIH Grant NS-07237 and while the second author was on sabbatical leave as a Fellow at the Center for Advanced Study in the Behavioral Sciences, Stanford, CA. The support of the City University of New York and the Spencer Foundation is gratefully acknowledged.

REFERENCES

- BEST, C. T., MORRONGIELLO, B. A., & ROBSON, R. C. (1981). Perceptual equivalence of acoustic cues in speech and nonspeech perception. *Perception and Psychophysics*, 28, 191-211.
- DARWIN, C. J., & BETHELL-FOX, C. E. (1977). Pitch continuity and speech source attribution. *Journal of Experimental Psychology: Human Perception and Performance*, 3, 665-672.
- DORMAN, M. F., RAPHAEL, L. J., & LIBERMAN, A. M. (1979). Some experiments on the sound of silence in phonetic perception. *Journal of the Acoustical Society of America*, 65, 1518-1532.
- 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 and Psychophysics*, 22, 109-122.
- FINNEY, D. J. (1971). *Probit analysis*. Cambridge: Cambridge University Press.
- FOX, B., & ROUTH, D. K. (1975). Analyzing spoken language into words, syllables, and phonemes: A developmental study. *Journal of Psycholinguistic Research*, 4, 331-342.
- HEINZ, J. M., & STEVENS, K. N. (1961). On the properties of voiceless fricative consonants. *Journal of the Acoustical Society of America*, 33, 589-596.
- JOOS, M. (1948). Acoustic phonetics. *Language*, 48(Suppl. 2), 1-136.
- KENT, R. D. (1983). The segmental organization of speech. In P. F. MacNeilage (Ed.), *The Production of Speech* (pp. 57-89). New York: Springer-Verlag.
- KUNISAKI, O., & FUJISAKI, H. (1977). On the influence of context upon perception of voiceless fricative consonants. *Annual Bulletin*, 11, 85-91. (Tokyo: Research Institute for Logopedics and Phoniatrics.)
- LIBERMAN, A. M., COOPER, F. S., SHANKWEILER, D. P., & STUDDERT-KENNEDY, M. (1967). Perception of the speech code. *Psychological Review*, 74, 431-461.
- MANN, V. A., & REPP, B. H. (1980). Influence of vocalic context on perception of the /s/-/s/ distinction. *Perception and Psychophysics*, 28, 213-228.
- MANN, V. A., SHARLIN, H. M., & DORMAN, M. (1985). Children's perception of sibilants: The relation between articulation and perceptual development. *Journal of Experimental Child Psychology*, 39, 252-264.
- MENYUK, P., & MENN, L. (1979). Early strategies for the perception and production of words and sounds. In P. Fletcher & M. Garman (Eds.), *Language acquisition* (pp. 49-70). Cambridge: Cambridge University Press.
- 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.
- PISONI, D. B. (1985). Speech perception: Some new directions in research and theory. *Journal of the Acoustical Society of America*, 78, 381-388.
- SOLI, S. D. (1981). Second formants in fricatives: Acoustic consequences of fricative-vowel coarticulation. *Journal of the Acoustical Society of America*, 70, 976-984.
- 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.
- WHALEN, D. H. (1981). Effects of vocalic transitions and vowel quality on the English /s/-/f/ boundary. *Journal of the Acoustical Society of America*, 69, 275-282.

Received February 21, 1986
Accepted December 30, 1986

Requests for reprints should be sent to Susan Nittrouer, Boys Town National Institute, 555 North 30th Street, Omaha, NE 68131.