

Trading Relations in the Perception of Speech by 5-Year-Old Children

BARBARA A. MORRONGIELLO AND RICK C. ROBSON

University of Massachusetts—Amherst

CATHERINE T. BEST

Teachers College—Columbia University

AND

RACHEL K. CLIFTON

University of Massachusetts—Amherst

Five-year-old children were tested for perceptual trading relations between a temporal cue (silence duration) and a spectral cue (F_1 onset frequency) for the "say-stay" distinction. Identification functions were obtained for two synthetic "say-stay" continua, each containing systematic variations in the amount of silence following the /s/ noise. In one continuum, the vocalic portion had a lower F_1 onset than in the other continuum. Children showed a smaller trading relation than has been found with adults. They did not differ from adults, however, in their perception of an "ay-day" continuum formed by varying F_1 onset frequency only. The results of a discrimination task in which the two acoustic cues were made to "cooperate" or "conflict" phonetically supported the notion of perceptual equivalence of the temporal and spectral cues along a single phonetic dimension. The results indicate that young children, like adults, perceptually integrate multiple cues to a speech contrast in a phonetically relevant manner, but that they may not give the same perceptual weights to the various cues as do adults.

The authors thank the parents and children who participated in this research; their interest, patience, and good humor made this research possible and contributed to making the conduct of this study a most enjoyable experience. We also thank Bruno Repp and Chuck Clifton for helpful comments on an earlier version of this manuscript and Terry Halwes for assistance with stimulus preparation. This research was supported by NIH Grant HD06753-06 and NIMH Grant MH 00332 awarded to Rachel Clifton, by NICHD Grant HD01994 awarded to Haskins Laboratories, and by postdoctoral fellowship Grant NS3085 awarded to Catherine Best. Please address reprint requests to: Barbara Morrongiello, Department of Psychology, University of Toronto/Erindale, Mississauga, Ontario L5L 1C6, Canada.

In the developmental literature on speech perception, there are several reports that children differ from adults in their responses to variations in single acoustic cues for phonetic contrasts. Zlatin and Koenigsnecht (1975), studying the perception of the stop consonant voicing contrast in 2-year-old, 6-year-old, and adult listeners, found that the magnitude of voice-onset-time (VOT) difference necessary for distinguishing between prevocalic stop cognates decreased as a function of age. Simon and Fourcin (1978) varied both VOT and first-formant (F_1) transition steepness in an investigation of 2- to 14-year-old English and French children's perception of voicing oppositions. The authors were particularly interested in studying French speakers' perception of voicing, since the VOT boundary differs from English and the F_1 transition is a more salient cue in French than in English. Their results revealed a linear improvement in labeling accuracy with age for children of both language environments, with an adultlike categorical pattern occurring at 5 to 6 years for the English and 7 to 8 years for the French listeners. Moreover, English-speaking children showed no evidence of utilizing the F_1 transition cue before about 5 years of age. The phoneme boundary between voiced and voiceless percepts also showed a systematic shift until 11 or 12 years of age when it reached a value corresponding to adult performance.

While these differences between children's and adults' phonetic perception of *single* acoustic cues are interesting, evidence is accumulating in the adult speech perception literature that *multiple* acoustic cues often interact to specify a single phonetic contrast. For example, voicing distinctions for initial stop consonants can be cued by changes in VOT, F_1 onset frequency, F_0 contour, or aspiration energy (Haggard, Ambler, & Callow, 1970; Lisker, 1975; Lisker, Liberman, Erickson, Dechovitz, & Mandler, 1977; Repp, 1979); each of these acoustic properties is a consequence of the laryngeal timing variations underlying the production of stop voicing (Ambramson & Lisker, 1965). Multiple acoustic correlates of articulatory contrasts have also been found to serve as cues for the perception of place of articulation (Dorman, Studdert-Kennedy, & Raphael, 1977; Harris, Hoffman, Liberman, Delattre, & Cooper, 1958) and manner of articulation (Dorman, Raphael, & Liberman, 1979; Miller & Liberman, 1979; Repp, Liberman, Eccardt, & Pesetsky, 1978).

Whenever several distinct acoustic cues provide listeners with functionally equivalent information about a single phonetic category contrast, then perceptual "trading relations" can be demonstrated. That is, altering the value of one acoustic cue so as to enhance perception of a phonetic contrast can be said to compensate perceptually for the value of another acoustic cue that specifies the *same* phonetic contrast but to a lesser degree. To the extent that the cues are integrated by the listener into a single phonetic percept, perception of any one cue is not independent of other values of other relevant cues (Repp, 1982). Thus, strengthening

the value of one cue can offset the weakening of another in listeners' perception of the specified phonetic contrast. Such trading relations have been found for voicing (e.g., Summerfield & Haggard, 1977), place (e.g., Bailey & Summerfield, 1980), and manner of articulation (e.g., Dorman, Raphael, & Isenberg, 1980) distinctions. The investigation of trading relations in speech perception has become increasingly more important as investigators seek to find evidence to differentiate between popular processing theories. The finding that two acoustic cues can engage in a trading relation in adults' perception of phonemes suggests that the processes involved in the perception of speech versus nonspeech sounds differ from one another (e.g., Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). Presumably, if identical processes were involved then acoustic differences would preclude listeners integrating the two acoustic cues into a single phonemic category (see Repp, 1982 for a complete discussion). Furthermore, recent evidence indicates that *only* when listeners perceive the two acoustic cues as phonetic information are the cues perceptually integrated, as evidenced by a trading relation (see Experiment 3 of Best, Morrongiello, & Robson, 1981).

In a recent series of experiments, we examined the perceptual equivalence of acoustic cues in adults' perception of speech and related nonspeech sounds (Best et al., 1981). Using a "say-stay" (/sei-/stei/) contrast, we systematically manipulated two acoustic cues that specify the presence or absence of the alveolar stop following the word-initial /s/: F_1 onset frequency (which produced corresponding differences in the 40-msec transition following F_1) and the duration of the silent closure interval. The average trading relation obtained from listeners' identification performance was evident in a "say-stay" boundary shift of 24.6 msec (Experiment 1). In other words, in order to be perceived as "stay," a stimulus with a high F_1 onset frequency (430 Hz) required approximately 25 msec additional silence between the /s/ and the vocalic portion than did a stimulus token having a low F_1 onset frequency (230 Hz).

To provide a more stringent test of whether these two acoustic cues were truly equivalent in perception (cf. Fitch, Halwes, Erickson, & Liberman, 1980), discrimination performance was assessed for stimulus comparisons in which the parameter values for closure duration and F_1 onset frequency were either "cooperating" (i.e., complementing one another phonetically) or "conflicting" (i.e., cancelling each other). Since the Cooperating Cues and the Conflicting Cues conditions differed only in the combination of cue values but not in the magnitudes of differences on each cue dimension, performance in the two conditions should have been equal if listeners discriminated the stimuli by their auditory properties alone. In contrast, listeners performed near chance in the Conflicting Cues condition but at a much higher level in the Cooperating Cues condition. Thus the results supported the hypothesis that the two acoustic

cues provide perceptually indistinguishable ("perceptually equivalent") information along a single phonetic dimension.

In the present research we extended our investigation to children's speech perception. By using the same stimuli as in the Best et al. (1981) study, we sought to determine whether children show a phonetic trading relation and perceptual equivalence of acoustic cues to the /sei/-/stei/ contrast in the same manner as adults do. Children 5 years of age were tested, since this was the age at which Simon and Fourcin (1978) claimed to first find evidence of perceptual use of F_1 transition distinctions in perception of stop voicing contrasts. Children's identification performance was assessed by using a standard forced-choice procedure. However, Wolf (1973) reported that 5- and 7-year-old children have difficulty with the ABX discrimination procedure, and pilot testing in our laboratory confirmed this observation. Consequently, discrimination data were obtained using a 2IAX paired-comparison procedure, in which children judged the pair members as being the "same" or "not the same" (Wolf, 1973).

Since there was some evidence to indicate developmental changes in perception of VOT (Bernstein, 1979; Simon & Fourcin, 1978; Zlatin & Koenigsknecht, 1975) and in the location and stability of various phoneme boundaries in perception and production (Kewley-Port & Preston, 1974; Strange & Broen, 1981; Zlatin & Koenigsknecht, 1976), we expected that children might differ from adults in performance on our multiply-cued stimulus continuum, which involved variations in F_1 onset frequency and in a temporal cue (as in VOT). The developmental literature, however, did not support a particular hypothesis as to the nature of these potential age-related differences (e.g., better utilization of the spectral than of the temporal cue or vice versa), although evidence that young children are less sensitive than adults to small differences in formant frequency information (Eguchi, 1976) suggested that 5-year-olds might be less responsive to F_1 onset manipulations than adults.

Although Simon and Fourcin (1978) claim that English-speaking children begin to make perceptual use of a temporal cue to stop voicing earlier than they make use of a spectral cue, there are some methodological problems with their study.¹ Insofar as Simon and Fourcin's findings generalize to children's perceptual integration of slightly different temporal

¹ Simon and Fourcin did not test the English-speaking and French-speaking children on the same voicing contrasts, and the contrasts were chosen such that neither group was tested on all three places of stop articulation. The English-speaking children were tested with "coat-goat" (3- to 14-year-olds) and "Paul-ball" (2-year-olds), whereas the French children were tested with "toto-dodo." Moreover, the children were given only three presentations of each stimulus from a continuum, which is an extremely low number of repetitions (most adult studies use 10-20 presentations per token) and could artificially inflate the children's variance in performance, especially at younger ages.

and spectral cues for a different phonemic contrast, they suggest that the children in our study might attend more to the temporal than the spectral cue and hence show a smaller trading relation than the adults in Best et al. (1981). However, even if the children do show a reduced trading relation, there is no indication in the developmental literature as to whether a discrimination test would reveal the same perceptual equivalence of the two cues along a single phonetic dimension as was found in adults. The present study was undertaken to assess whether 5-year-olds make perceptual use of multiple acoustic cues for a single phonetic dimension, as adults do. Alternatively, if children attend primarily to the acoustic properties of the stimuli, then one would expect that they would fail to integrate perceptually the temporal and spectral cues as information about a unified phonetic category. In that case, they presumably would hear the auditory differences between differently cued stimuli even within a phonetic category, and would thereby discriminate the Conflicting Cues contrasts as well as they discriminate the Cooperating Cues contrasts. Although this second possibility was less likely on the basis of the adult findings, it could not be dismissed a priori because no studies of trading relations in children existed in the literature.

METHOD

Subjects

Eight children (3 male, 5 female) approximately 5-years-old at the onset of testing (mean age: 60.4 months; range: 57.3–64.9 months) participated in the present experiment. An average of 3½ months elapsed between the first and final testing sessions. Children were reported by parents to have normal hearing and they did not have colds, ear, or throat disturbances on test days. The data from two additional children were excluded from the final analysis because of incomplete test sessions. Parents were paid \$3 for transportation costs, and children selected a prize for each day of participation.

Stimuli

Two sets of synthetic stimuli were used. They were based upon two 290-msec, three-formant syllables created on the Haskins parallel-resonance synthesizer (see Fig. 1), as stylized versions of the vocalic portions of natural utterances of "say" and "stay" produced by a male speaker. They differed from one another only in F_1 onset frequency (230 Hz vs 430 Hz) and the initial 40-msec transition following F_1 . The syllables were identical in formant amplitudes and overall amplitude envelopes, in F_2 and F_3 , and in the F_1 steady-state frequency (611 Hz) beyond the initial 40-msec transition difference (see Best et al., 1981, for complete stimulus descriptions).

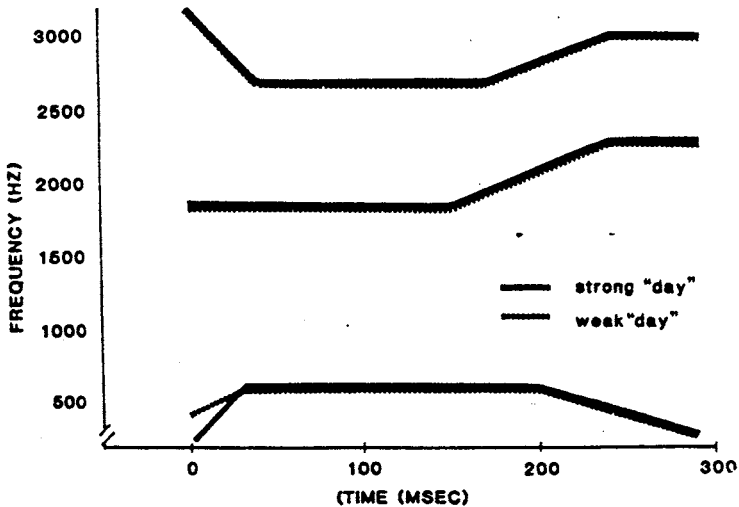


FIG. 1. Schematic diagram of F_1 , F_2 , and F_3 frequencies for synthetic "weak day" and "strong day."

One set of stimuli was an "ay-day" continuum² spanning 14 different syllables. It was created by varying the F_1 onset frequency in approximately 33-Hz steps between 160 and 611 Hz, and included the 230 and 430 Hz F_1 onset syllables described above. In a previous identification test using the "ay-day" continuum (Best et al., 1981), adults identified the 230-Hz syllable as "day" 100% of the time. This syllable will hereafter be referred to as the "strong day," abbreviated D. In contrast, adults identified the 430-Hz syllable as "day" only approximately 50% of the time; therefore, it will be called the "weak day," abbreviated d. To test whether the two test syllables would also differ in children's perception, a stimulus tape was constructed for obtaining the children's identification functions on the "ay-day" continuum. The tape contained 10 presentations of each of the 14 syllables in a randomized sequence. Within each block, the intertrial interval was 4 sec.

The second set of stimuli consisted of two different "say-stay" continua, constructed by preceding the D and d syllables with a natural 120-msec /s/ noise derived from a male speaker's utterance of "say" (see Experiment 2 of Best et al., 1981). The /s/ and the synthetic syllable were separated by silent intervals ranging from 0 to 104 msec, in 8-msec increments. Thus, each continuum comprised 14 tokens.

Two stimulus tapes were constructed. The first was designed to obtain

² In American English, the phonetic and articulatory properties of /t/, /p/, or /k/ following /s/ are actually more characteristic of their voiced cognates /d/, /b/, and /g/, respectively. Thus /steil/ with the /s/ noise removed sounds like "day" rather than "tay."

children's identification functions. This tape consisted of 20 blocks of 14 single-item trials each. Every two successive blocks comprised a randomized sequence of all 14 tokens from each of the two continua, for a total of 10 repetitions per token. Within each block, the intertrial interval was 4 sec.

The second tape constructed from the "say-stay" stimuli was used to test discrimination. A 2IAX discrimination task ("same-not same") was employed. This test included four types of stimulus pairings for discrimination judgments: Physically Same, One Cue, Conflicting Cues, and Cooperating Cues (see Table 1). There were eight different Physically Same pairs, four from each of the two "say-stay" continua. These four pairs were based on the two extreme endpoints of each continuum, which were clear instances of "say" or "stay." There were also three different pairs for the One Cue comparisons. Within each One Cue pair, the tokens were identical in silent gap duration, but differed in the spectral cue (d vs D). These three pairs were selected so that the silent gap durations spanned the adult "say-stay" boundaries (lower panel of Fig. 2), as determined by Experiment 1 of Best et al. (1981). In both the Cooperating and the Conflicting Cues comparisons, also referred to as the Two Cue comparison types, members of each discrimination pair differed on both the spectral and the temporal dimension. In the Cooperating Cues comparisons, the D member of the pair had a 24-msec *longer* silent gap duration than the d member (as in Experiment 1 of Best et al., 1981); thus the temporal and spectral cue values for each pair member "co-operated" in that they both favored the same phonetic category. In the Conflicting Cues comparisons, the D member of a pair had a 24-msec *shorter* silent gap duration than the d member. Here, the value of the temporal cue was designed to cancel the phonetic effect of the spectral cue for each pair member. In both the Two Cue comparison types, a 24-msec difference in silent gap duration was used because this was the magnitude of the trading relation shown by adults for identifications of the two stimulus continua (Experiment 1, Best et al., 1981). There were four different pairs in each of the Two Cue comparison types, selected so as to span the "say-stay" boundaries for adults.

The discrimination tape contained 240 trials organized into 16 blocks of 15 trials each. The 19 different stimulus pairs (8 Physically Same, 3 One Cue, 4 Cooperating Cues, 4 Conflicting Cues) were randomly sequenced within each successive pair of blocks. Within each pair of blocks, each of the "not same" pairs (One cue, Cooperating Cues, Conflicting Cues comparisons) was presented twice, whereas each of the Physically same pairs was presented once. Thus, 16 judgments were obtained for each of the "not same" pairs, and 8 for each of the Physically Same pairs. The interstimulus interval within each pair was 1 sec, and the intertrial interval between successive pairs was 4 sec.

TABLE I
STIMULUS PAIRINGS FOR THE FOUR DISCRIMINATION CONDITIONS

One Cue ^a	Cooperating Cues	Conflicting Cues	Physically Same ^b
s-24-D vs s-24-d	s-32-D vs s- 8-d	s- 8-D vs s-32-d	s- 0-d/D
s-32-D vs s-32-d	s-40-D vs s-16-d	s-16-D vs s-40-d	s- 8-d/D
s-40-D vs s-40-d	s-48-D vs s-24-d	s-24-D vs s-48-d	s- 96-d/D
	s-56-D vs s-32-d	s-32-D vs s-56-d	s-104-d/D

^a "s" stands for the /s/ portion of the syllable; the subsequent number is the number of milliseconds of silence between the /s/ and the vocalic portion of the syllable; "D" stands for the "strong day" syllable and "d" stands for the "weak day" syllable.

^b Because the members of a pair here are physically the same, only one member of each pair type is shown; d/D indicates there was one "weak day" pair and one "strong day" pair.

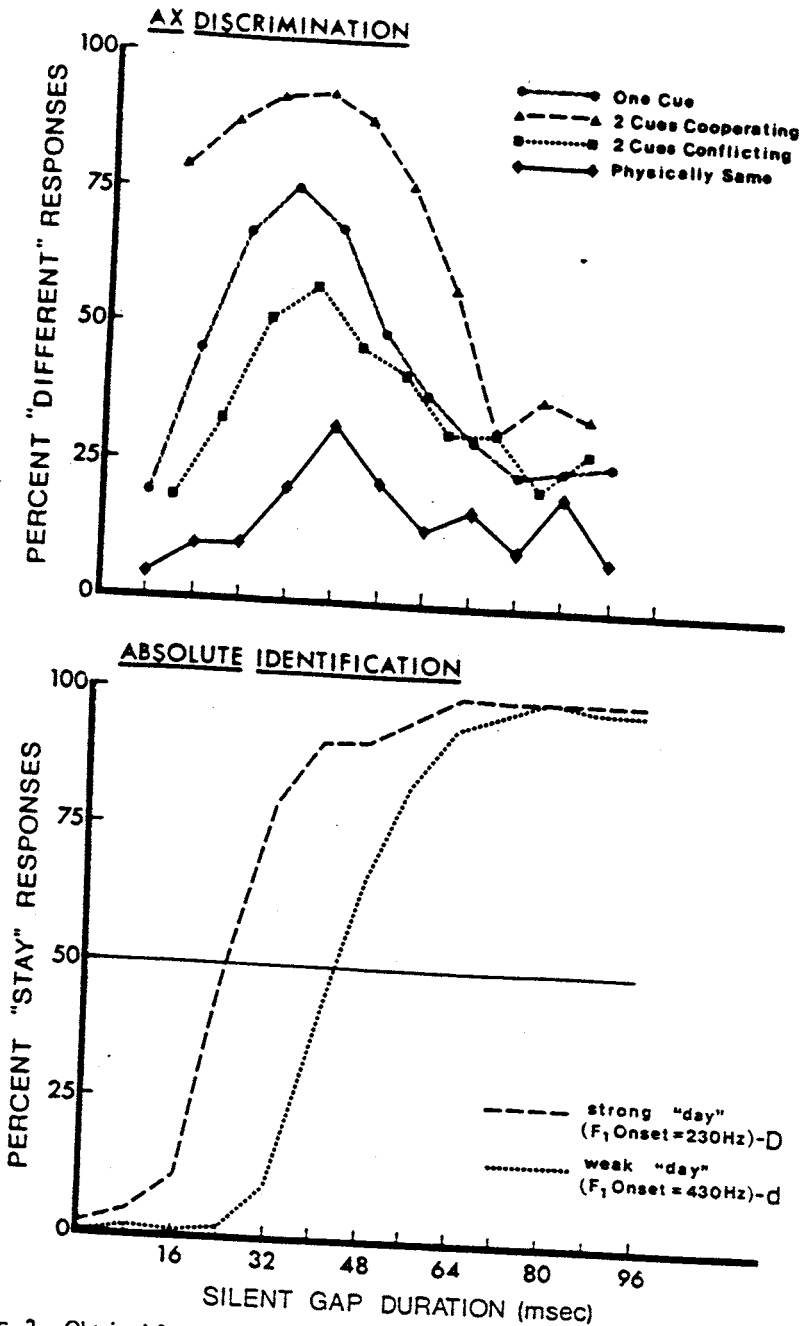


FIG. 2. Obtained functions for the three-way 2IAX discrimination test ("same-different;" upper panel) and the forced-choice identification test on the two "say-stay" stimulus continua (lower panel) for the adults tested in Experiment 2 of Best et al. (1981). [Reprinted, with permission, from Best et al. (1981).]

Apparatus and Procedure

Each child participated in five 50-min sessions conducted within a few weeks of one another. The first and second halves of the "say-stay" identification test were given in sessions 1 and 3, and the two halves of the 2IAX "say-stay" discrimination test were given in sessions 2 and 4. In session 5, the randomized forced-choice "ay-day" identification test was given. Testing was conducted in a sound-attenuated room with the parent and an experimenter (Experimenter 1) present. The stimuli were played on a Revox reel-to-reel tape recorder running at 7.5 ips at a sound pressure level of 60 dB re .0002 dynes cm² (calibrated using the A scale of a General Radio sound level meter) over loudspeakers (Acoustic Research, AR 7) located approximately 1 m to the child's left and right, at a 90° angle to the child's midline.

Upon entering the testing room, children were given 5 min to become accustomed to the new situation. During this time Experimenter 1 encouraged the child to play with two small mechanical robots. Once rapport had been established, the child was told that a big robot in the adjacent equipment room was learning how to speak and that she/he could help the robot learn to talk better. Most children were enthusiastic about participating. After showing a child a robot that had been constructed around the tape recorder and having her/him listen and repeat the words that the robot said (i.e., taped versions of clear endpoint "say" and "stay"), children were taught to use a two-button box in the testing room to indicate their responses "to the robot" in the equipment room. An Esterline-Angus event recorder in the equipment room recorded the child's responses on the two-button box. Throughout the test session, a second experimenter (Experimenter 2) tallied the child's responses directly from the Esterline-Angus recorder and indicated interblock intervals on the permanent paper record. After the test session, the tally completed by Experimenter 2 was checked by a naive observer against the permanent paper tape record.

During the "say-stay" identification tests, the child pressed either of two horizontally adjacent buttons on the button box to indicate whether "say" or "stay" was heard on each presentation. A picture adjacent to each button was a continuous reminder of which button was for "say" (i.e., a picture of a woman talking and the word "say" printed) and which button was for "stay" (i.e., a picture of a woman motioning for her dog to stay and the word "stay" printed). For the "ay-day" test, the pictures used were of a large letter "A" for "ay" and a sun rising over the horizon for "day." The right-left button designation for each word was randomized across test sessions and children.

During the 2IAX discrimination test, two strips of colored tape were substituted for each picture on the button box. For one button the two colors were the "same" (both red) and for the other button the two

colors were "not the same" (red and green). During the 2IAX discrimination test the children were instructed to listen to each pair of words and press a button to indicate whether the pair members were exactly the "same" or "not the same." Again, the right-left button designation was randomized across sessions and children.

On each day of testing the child was reminded of how to use the response box, and was given a block of practice trials to ensure that she/he understood the task and could work through an entire block of trials without difficulty. Experimenter 1 remained with the child throughout each test session and provided verbal encouragement and support, as necessary. In addition, throughout the testing sessions two low-watt blue spotlights provided the child with intermittent feedback, which proved to be particularly effective in motivating the child to perform the task and continue to listen closely. The lights were positioned approximately 1 m in front of the child. On one light a happy face signaled that the child's previous response had been correct. A sad face on the other light indicated an incorrect response. Experimenter 2 controlled the operation of these lights according to the correctness of the child's responses on a sample of trials. During the "say-stay" identification sessions, one of each of the endpoint stimuli for the two continua was randomly selected during the course of two trial blocks for reinforcement. During the discrimination session, one of each of the four types of trials was selected, and for the "ay-day" identification series, one of each of the endpoint stimuli for the continuum received reinforcement.

Between trial blocks in all five sessions, children were allowed to select colored stars that they pasted on a personalized game board. On successive blocks they selected an increasing number of stars and after the last trial they were allowed to select a prize. For most of the children the time during which they selected and pasted stars was sufficient to serve as a rest interval. However, when necessary for maintaining the child's motivation this interblock interval was lengthened and the child was allowed to engage in another play activity for a few minutes.

RESULTS

Identification: "Say-Stay"

The category boundary between "say" and "stay" was defined as that silent interval at which there were 50% "stay" responses. There were no significant test block effects (session 1 vs 3) in the children's identification responses. As can be seen in Fig. 3, the mean category boundary for the D continuum was at 26.4 msec (range: 16.0-32.0 msec). In contrast, the mean category boundary for the d continuum was at 37.5 msec (range: 33.6-43.3 msec). The average difference in category boundaries of 11.1 msec (range: 5.9-17.6 msec) was highly significant

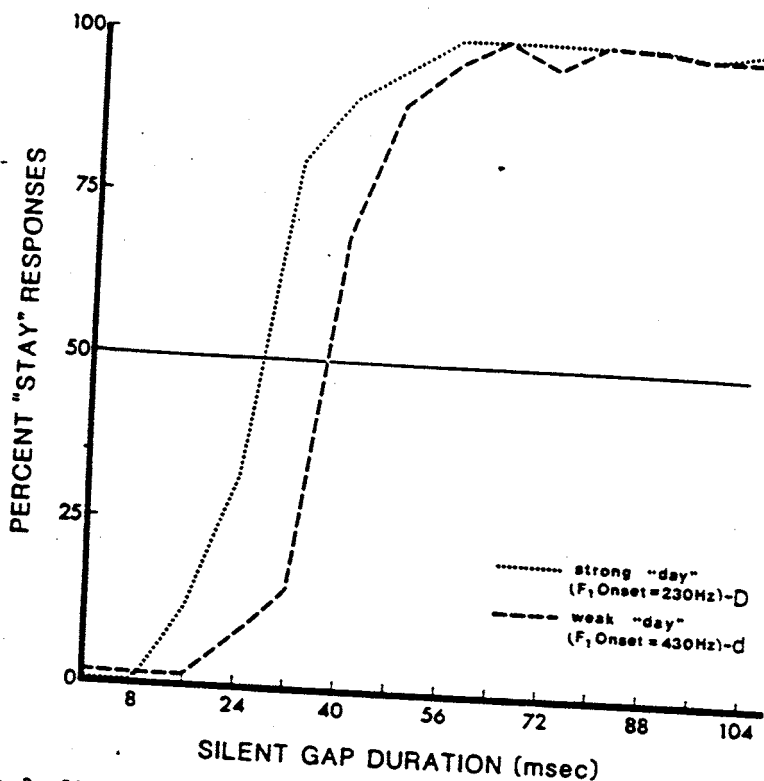


FIG. 3. Identification functions of the children for the "strong day" and "weak day" stimulus continua.

($r(7) = 8.5, p < .001$). In fact, there was no overlap whatsoever in the distribution of category boundaries for the D and d continua.

These results support previous findings, obtained with adults, of a trading relation between spectral and temporal acoustic cues in the perception of stop consonants. In children, "weak day" stimulus tokens required approximately 11.1 msec more silence after /s/ to be heard as "stay" than did "strong day" stimulus tokens (see Fig. 3). The magnitude of this trading relation differs between children and adults ($r(20) = 5.3, p < .001$).³ This difference between children and adults is due exclusively

³ For the two "say-stay" continua in Experiment 1 of Best et al. (1981), the /s/ and the synthetic syllable were separated by silent gaps ranging between 0 and 136 msec, in 8-msec increments, resulting in 18 stimuli per continuum. As mentioned in the introduction, the average trading relation for adult listeners in Experiment 1 of Best et al. (1981) was 24.6 msec. In Experiment 2 of Best et al. (1981) a truncated "say-stay" continuum containing 13 stimuli each was used: stimuli containing gaps greater than 96 msec were eliminated, since the adults in Experiment 1 had identified these as "stay" nearly 100% of the time. The average trading relation for adults tested with this truncated "say-stay"

to a difference in their identification of stimulus tokens from the d continuum (compare Fig. 3 to the bottom panel of Fig. 2). For the d continuum, the mean 50% crossover point for adults in Experiment 2 of Best et al. (1981) was 43.8 msec, whereas that for children was 37.5 msec ($t(20) = 2.2, p < .05$). For the D continuum the respective points were 25.3 and 26.4 msec ($t(20) = .3, n.s.$).

Identification: "Ay-Day"

The results from the "ay-day" identification task may provide some insight into the basis for the difference between children and adults in the magnitude of the trading relation. Children were apparently not less sensitive than adults to the perceptual use of the F_1 spectral cue for the alveolar stop, since as a group they did not differ significantly from adults in the location of the 50% crossover point for the "strong day" continuum. Rather, it was the 50% crossover point for the "weak day" continuum that differentiated the children and adults. One possibility is that children were *more* sensitive than adults to F_1 onset spectral information, in the sense that for children a relatively high F_1 onset supported perception of an alveolar stop, following /s/, more readily than it did for adults. Conversely, the children could be said to be *less* sensitive than adults to the spectral difference between the 230 Hz vs 430 Hz F_1 onsets. Since the "ay-day" identification task involved changes *only* in this spectral cue, it is useful for examining the possibility that the "weak day" vocalic syllable was perceived to be more "day"-like by children than by adults.⁴

The identification functions for the children, and for a sample of 18 adult listeners (Best et al., 1981), are shown in Fig. 4. The 50% crossover point for the children did not differ significantly from that of the adults ($t(3) = .3, n.s.$). The "ay-day" continuum contained the two vocalic syllable tokens used in generating the two "say-stay" continua ("weak day" continuum—430-Hz F_1 onset frequency; "strong day" continuum—230-Hz F_1 onset frequency). Children and adults did not differ in percent of "day" identification for either of these tokens: "strong day" token—adults 99%, children 100%; "weak day" token—adults 46%, children

continuum was 18.5 msec. Because children were tested with this truncated continuum only, our statistics in the present study compared the size of their trading relation relative to the adult trading relation of Experiment 2 in which the adults were tested with the same truncated continuum as the children (see Figs. 2 and 3). However, because the children's discrimination data were obtained prior to completion of testing adults in Experiment 2 of Best et al. (1981), the children's discrimination test was set up based on the adult trading relation of Experiment 1 of Best et al., which was 24.6 msec.

⁴ It is interesting, however, that when the "ay-day" data for individual children were compared to the magnitudes of their "say-stay" trading relations, there was a tendency for children with larger magnitude trading relation to also show larger differences in percent "day" identifications between the "weak day" and "strong day" syllables.

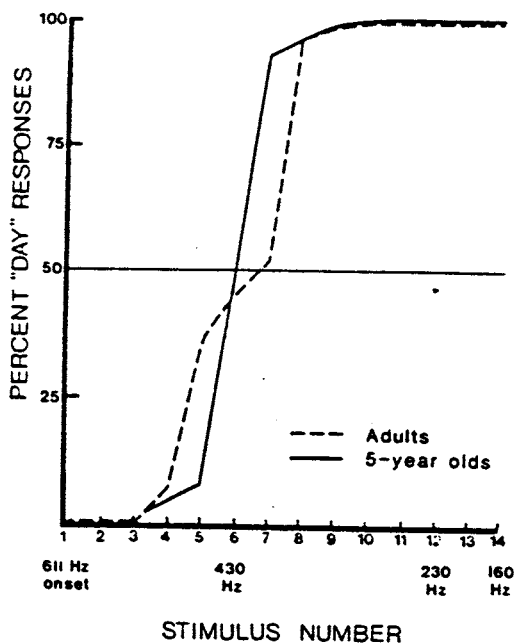


FIG. 4. Children's and adult's identification functions for the "ay-day" stimulus series (stimulus numbers refer to steps of approximately 33 Hz in onset frequency of F_1 ; stimulus 6 is the "weak day" vocalic base of the continua used in the "say-stay" conditions, and stimulus 12 is the "strong day" stimulus base).

54%. These results suggest that children's and adult's perception of the F_1 onset spectral cue was not primarily responsible for the obtained difference in the size of the trading relation.

2IAX Discrimination Test

The discrimination data were compared with discrimination performance predicted from the identification data for the strong and weak "say-stay" continua. For a given discrimination comparison type, the probability of a "correct" response was computed in the following manner (see Best et al., 1981): $p(\text{"correct"}) = \{2 + [p(\text{"say"} \text{ on first member of comparison}) - p(\text{"say"} \text{ on second member})]^2 + [p(\text{"stay"} \text{ on first member}) - p(\text{"stay"} \text{ on second member})]^2\} \div 4$. Since there were no significant effects involving blocks (i.e., testing session 2 vs 4), only results totalled over blocks 1 and 2 will be reported. The results for Physically Same comparison types showed that there was no significant general response bias; the average observed proportion of "not same" responses was 4% and the average predicted proportion was 1%.

There are two aspects of discrimination performance that will be discussed: (1) observed vs predicted performance for each discrimination

type; and (2) the relative rank ordering of discrimination performance across discrimination types. With regard to the latter, it is important to remember that in selecting stimulus pairs for the Conflicting Cues and Cooperating Cues discrimination types, a trading relation typical of adults was assumed (Experiment 1 of Best et al., 1981). Since the children in the study showed a significantly smaller trading relation than adults, however, the discrimination pairs used were not in fact appropriate for providing the most clear and dramatic contrast in the children's performance between the Conflicting and the Cooperating conditions. Specifically, instead of using a 24-msec silent gap difference between the members of Two Cue discrimination pairs, a difference of 11 msec would presumably have been more appropriate.

Nonetheless, the data can provide a test of the perceptual equivalence hypothesis if predicted and obtained discrimination performance were to vary in a similar manner as a function of discrimination condition, particularly if peak performance in the Cooperating Cues condition was still predicted to be higher than performance in the Conflicting Cues condition. To determine whether this was the case, an analysis of variance on predicted peak discrimination levels was performed for the Cooperating Cues, Conflicting Cues, and One Cue conditions. Peak performance was defined as performance on those comparisons in which the pair members straddled the "say-stay" boundary; that is, the second comparison for the One Cue condition, and the average of the second and third comparison in each of the other two discrimination conditions. There was a significant difference among the conditions for the predicted discrimination data ($F(2, 14) = 14.3, p < .001$). Predicted performance was significantly higher for the Cooperating Cues than for the Conflicting Cues conditions, ($t(7) = 4.9, p < .01$), although the difference between the Conflicting Cues and the One Cue conditions was not significant. The observed vs predicted scores for each test condition appear in Fig. 5.

Analysis of variance on the observed performance levels also revealed significant differences among the conditions ($F(2, 14) = 11.3, p < .005$). The pattern of differences among the discrimination conditions conformed to the predicted order, supporting the notion that children, like adults, perceived the diverse acoustic cues as equivalent information along a single phonetic dimension. Peak discrimination was significantly higher for the Cooperating Cues condition than the Conflicting Cues condition, ($t(7) = 3.6, p < .01$). There was no significant difference between the Conflicting Cues and One Cue conditions.

Although the order of the observed peaks across the three discrimination conditions matched the order of the predicted peaks, there was some discrepancy between observed and predicted levels of performance. There was no performance difference between observed and predicted scores across the One Cue comparisons, but there was a significant main effect

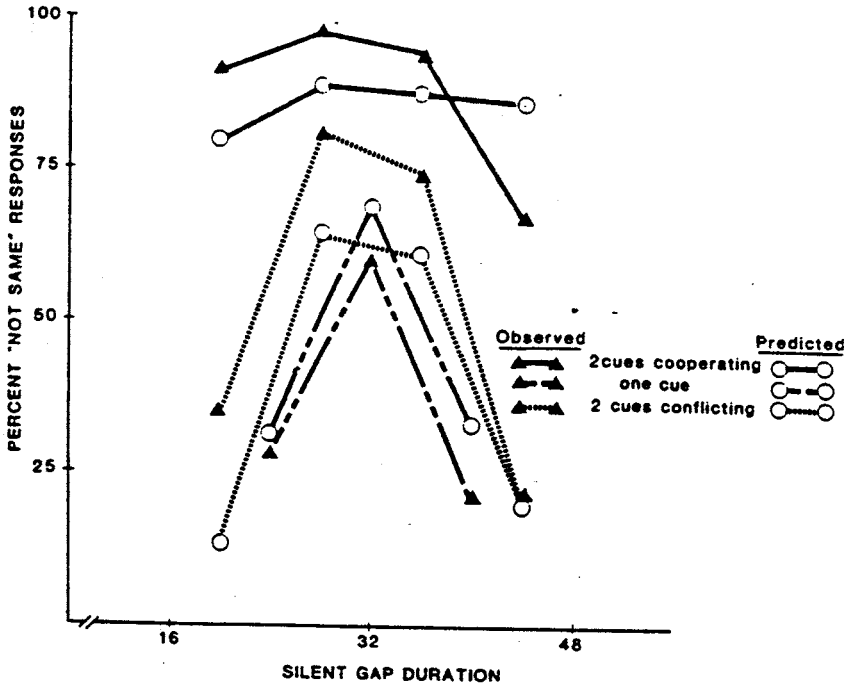


Fig. 5. Children's discrimination functions for 2IAX comparisons.

for observed vs predicted scores across the Conflicting Cues comparisons, $F(1, 17) = 18.7, p < .005$, and across the Cooperating Cues comparisons, $F(3, 21) = 5.8, p < .005$; generally, obtained performance was better than predicted. These differences in obtained vs predicted performance levels indicate some ability to discriminate acoustic differences between stimuli beyond differences in phonemic identity. However, this is not particularly damaging to the phonetic perceptual equivalence hypothesis since the observed-predicted differences are similar in magnitude to those found in adults by Best et al. (1981), and in fact are common in studies on categorical perception of speech segment contrasts.

DISCUSSION

Investigation of trading relations among acoustic cues in phonetic perception can provide valuable insights into how information from diverse acoustic dimensions is integrated in the perception of speech. The present investigation examined children's integration of spectral and temporal cues for the perception of a stop consonant in an /s/ + stop cluster in syllable-initial position. Generally, to perceive the stop consonant children needed approximately 11 msec more silence to compensate for a weak

spectral cue than when a strong spectral cue was present. This trading relation of 11 msec was significantly less than that obtained for a group of adult listeners tested with the same stimuli (Best et al., Experiment 2, 1981). Children and adults did not differ, however, in their perception of the "ay-day" continuum, which was formed by varying only the spectral cue. This suggests that children and adults differed either in their perception of the temporal cue alone, or in their relative weighting of the temporal and spectral cues for phonetic integration in /s/ + stop cluster perception. This former possibility seems less likely given previous reports that children (e.g., Wolf, 1973) and even infants (e.g., Eimas, Siqueland, Jusczyk, & Vigorito, 1971) show the same VOT boundary (a temporal cue) as adults in perception of stop voicing.

The pattern of results obtained in the discrimination conditions support the notion that the two acoustic cues are truly equivalent along a single phonetic dimension in children's perception of speech, even though the stimulus pairings used were not ideally suited to the magnitude of the children's trading relation. For children, both the expected and observed discrimination performances were significantly better when the spectral and temporal cue values "cooperated" phonetically to enhance discrimination along the phonetic dimension, than when the cues "conflicted" phonetically to reduce discriminability along the phonetic dimension. Since the Cooperating and Conflicting Cues conditions involved comparisons that differed by equal amounts along the two acoustic dimensions, the pattern of discrimination findings indicated that the children were not focusing on the acoustic differences as separate or distinct parameters. Instead, like adults, they perceived the unified phonetic information underlying the diversity in acoustic information.

The cause of the age-related perceptual differences in the magnitude of the trading relation is not directly revealed by this study, and warrants further exploration. One possible reason for the difference might be a lowered sensitivity to frequency differences among formant transition onsets in children vs adults (Eguchi, 1976); however, the lack of an age effect in the "ay-day" test eliminates the possibility of an absolute age difference in frequency sensitivity for F_1 onset values in our stimuli. Children at this age are apparently equal to adults in their perceptual use of a 230- vs 430-Hz F_1 onset difference to signal a difference in degree of alveolar stop closure; that is, they do not differ from adults in their use of that acoustic information as a *primary* cue to a phonetic distinction. They deviate from adults only in their use of the same acoustic information as a *secondary* cue to a multiply-cued phonetic contrast. This would suggest that the age difference is more likely related to developmental changes in selective attention to perceptual information than it is to changes in basic auditory sensitivity. Converging support for this hypothesis comes from Bernstein's (1979) report that children

are less consistent than adults in using F_0 as a secondary cue to stop contrasts.

A second possibility is that the age difference in perception of multiple acoustic cues to a phonetic contrast might also relate in some way to child vs adult production differences. Children 6 years of age produce shorter VOTs (Kent, 1981), and they show less of a VOT distinction (Kent, 1976) for stop consonants in syllable-initial position, relative to adults' productions. Furthermore, children's VOT for stops in /s/ + stop clusters is about 12 msec, averaging across three places of articulation (see Fig. 2 in Bond & Wilson, 1980), whereas in adult production the average VOT is 23 msec, again averaging across three places of articulation (see Table 1 in Klatt, 1975). Since children produce both word-initial voiceless stops and those following initial /s/ with a shorter VOT than adults, this means that they start phonation earlier after the release of the constriction. In turn, this would imply a lower F_1 onset frequency in children's voiceless stops than in adults', at least for those following /s/, and that the F_1 onset frequency differences would therefore be smaller for children's voiced-voiceless distinctions in production. The obtained smaller trading relation in the children, for our /s/ + stop cluster, would seem to imply lowered *perceptual* and *productive* use of the F_1 onset distinction, relative to adults. This hypothesized relation between children's smaller perceptual trading relation and their production of smaller voicing category distinctions could be tested by examining children's gap durations and F_1 onsets in "say-stay" production relative to their perceptual equivalence tests for "say-stay." A relationship between perception and production abilities in 3-year-olds, for example, has been reported for the contrasts /w/, /r/, and /l/ (Strange & Broen, 1981), and has also been indicated by the research of Bailey and Haggard (1980) on voicing distinctions.

Perception of running speech in the natural environment depends upon a listener's ability to integrate multiple acoustic cues, which may interact in complex ways to specify phonetic category information. Yet developmental research on perceptual integration of multiple acoustic cues specifying phonetic content has been sorely lacking. As the results of the present study indicate, examining children's and adults' perception of simple one-cue word-initial differences provides little information about developmental changes in listeners' abilities to integrate and utilize these cues for phonetic perception in multiple-cue contexts, which more closely approximate the diverse information available to a listener in natural speech. In order to better understand developmental changes in the perception of speech it is important that we begin to examine perceptual abilities that more closely approximate those necessary for the perception of speech in the natural environment.

REFERENCES

- Abramson, A. & Lisker, L. Voice onset time in stop consonants: Acoustic analysis and synthesis. In D. E. Cummins (Ed.), *Proceedings of the 5th Congress of International Acoustics*. Liege: Thone, 1965.
- Bailey, P. J., & Haggard, M. P. Perception-production relationships in the voicing contrast for initial stops in 3-year-olds. *Phonetica*, 1980, 37, 377-396.
- Bailey, P. J., & Summerfield, Q. Information in speech: Observations on the perception of [s]+stop clusters. *Journal of Experimental Psychology: Human Perception and Performance*, 1980, 6, 536-563.
- Bernstein, L. E. Developmental differences in labeling VOT continua with varied fundamental frequency. *Journal of the Acoustical Society of America*, 1979, 65(Suppl.), S112 (Abstract).
- Best, C. T., Morrongoello, B. A., & Robson, R. C. Perceptual equivalence of acoustic cues in speech and nonspeech perception. *Perception and Psychophysics*, 1981, 29, 191-211.
- Bond, Z. S., & Wilson, H. F. /s/ plus stop clusters in children's speech. *Phonetica*, 1980, 37, 149-158.
- Dorman, M. F., Raphael, L. J., & Isenberg, D. Acoustic cues for a fricative-affricate contrast in word-final position. *Journal of Phonetics*, 1980, 8, 397-405.
- Dorman, M. F., Raphael, L. J., & Liberman, A. M. Some experiments on the sound of silence in phonetic perception. *Journal of the Acoustical Society of America*, 1979, 65, 1518-1532.
- Dorman, M. P., Studdert-Kennedy, M., & Raphael, L. J. Stop consonant recognition: Release bursts and formant transitions as functionally equivalent, context-dependent cues. *Perception and Psychophysics*, 1977, 22, 109-122.
- Eimas, P., Siqueland, E. R., Jusczyk, P., & Vigorito, J. Speech perception in early infancy. *Science*, 1971, 171, 304-306.
- Eguchi, S. Difference limens for the formant frequencies: Normal adult values and their development in children. *Journal of the American Audiological Society*, 1976, 1, 145-149.
- Fitch, H. L., Halwes, T., Erickson, D. M., & Liberman, A. M. Perceptual equivalence of two acoustic cues for stop consonant manner. *Perception and Psychophysics*, 1980, 27, 343-350.
- Haggard, M. P., Ambler, S., & Callow, M. Pitch as a voicing cue. *Journal of the Acoustical Society of America*, 1970, 47, 613-617.
- Harris, K. S., Hoffman, H. S., Liberman, A. M., Delattre, P. C., & Cooper, F. S. Effect of third-formant transitions on the perception of the voiced stop consonants. *Journal of the Acoustical Society of America*, 1958, 30, 122-126.
- Kent, R. D. Anatomical and neuromuscular maturation of the speech mechanism: Evidence from acoustic studies. *Journal of Speech and Hearing Research*, 1976, 19, 421-447.
- Kent, R. D. Sensorimotor aspects of speech development. In R. Aslin, J. Alberts, & M. Peterson (Eds.), *Development of Perception: Psychobiological Perspectives* (Vol. 1). New York: Academic Press, 1981.
- Kewley-Port, D., & Preston, M. Early apical stop production: A voice onset time analysis. *Journal of Phonetics*, 1974, 2, 195-210.
- Klatt, D. Voice onset time, frication, and aspiration of word-initial consonant clusters. *Journal of Speech and Hearing Research*, 1975, 18, 686-705.
- Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. Perception of the speech code. *Psychological Review*, 1967, 74, 430-460.
- Lisker, L. Is it VOT or a first-formant transition detector? *Journal of the Acoustical Society of America*, 1975, 57, 1547-1551.
- Lisker, L., Liberman, A. M., Erickson, D. M., Dechovitz, D., & Mandler, R. On pushing the voice-onset-time (VOT) boundary about. *Language and Speech*, 1977, 20, 209-216.

- Miller, J. L., & Liberman, A. M. Some effects of later-occurring information on the perception of stop consonant and semivowel. *Perception and Psychophysics*, 1979, 25, 457-465.
- Repp, B. H. Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. *Psychological Bulletin*, 1982, 92, 81-110.
- Repp, B. H. Relative amplitude of aspiration noise as a voicing cue for syllable-initial stop consonants. *Language and Speech*, 1979, 22, 173-189.
- Repp, B. H., Liberman, A. M., Eccardt, T., & Pesetsky, D. Perceptual integration of acoustic cues for stop, fricative and affricate manner. *Journal of Experimental Psychology: Human Perception and Performance*, 1978, 4, 621-637.
- Simon, C., & Fourcin, A. J. Cross-language study of speech-pattern learning. *Journal of the Acoustical Society of America*, 1978, 63, 925-935.
- Strange, W., & Broen, P. The relationship between perception and production of /w/, /r/, and /l/ by three-year-old children. *Journal of Experimental Child Psychology*, 1981, 31, 81-102.
- Summerfield, Q., & Haggard, M. On the dissociation of spectral and temporal cues to the voicing distinction in initial stop consonants. *Journal of the Acoustical Society of America*, 1977, 62, 436-448.
- Wolf, C. The perception of stop consonants by children. *Journal of Experimental Child Psychology*, 1973, 16, 318-331.
- Zlatin, M. A., & Koenigsnecht, R. A. Development of the voicing contrast: Perception of stop consonants. *Journal of Speech and Hearing Research*, 1975, 18, 541-553.
- Zlatin, M. A., & Koenigsnecht, R. A. Development of the voicing contrast: A comparison of voice onset time in stop perception and production. *Journal of Speech and Hearing Research*, 1976, 19, 78-92.

RECEIVED: November 15, 1982; REVISED: March 11, 1983.