Perception & Psychophysics
1990, 48 (6), 559-570

Young infants' perception of liquid coarticulatory influences on following stop consonants

727

CAROL A. FOWLER
Dartmouth College, Hanover, New Hampshire
and Haskins Laboratories, New Haven, Connecticut

CATHERINE T. BEST
Wesleyan University, Middletown, Connecticut
and Haskins Laboratories, New Haven, Connecticut

and

GERALD W. McROBERTS
Haskins Laboratories, New Haven, Connecticut

Phonetic segments are coarticulated in speech. Accordingly, the articulatory and acoustic properties of the speech signal during the time frame traditionally identified with a given phoneme are highly context-sensitive. For example, due to carryover coarticulation, the front tongue-tip position for /l/ results in more fronted tongue-body contact for a /g/ preceded by /l/ than for a /g/ preceded by /r/. Perception by mature listeners shows a complementary sensitivity—when a synthetic /da/-/ga/ continuum is preceded by either /al/ or /ar/, adults hear more /g/s following /l/ rather than /r/. That is, some of the fronting information in the temporal domain of the stop is perceptually attributed to /l/ (Mann, 1980). We replicated this finding and extended it to a signal-detection test of discrimination with adults, using triads of disyllables. Three equidistant items from a /da/-/ga/ continuum were used preceded by /al/ and /ar/. In the identification test, adults had identified item ga5 as "ga," and da1 as "da," following both /al/ and /ar/, whereas they identified the crucial item d/ga3 predominantly as "ga" after /al/ but as "da" after /ar/. In the discrimination test, they discriminated d/ga3 from da1 preceded by /al/ but not /ar/; compatibly, they discriminated d/ga3 readily from ga5 preceded by /ar/ but poorly preceded by /al/. We obtained similar results with 4-month-old infants. Following habituation to either ald/ga3 or ard/ga3, infants heard either the corresponding ga5 or da1 disyllable. As predicted, the infants discriminated d/ga3 from da1 following /al/ but not /ar/; conversely, they discriminated d/ga3 from ga5 following /ar/ but not /al/. The results suggest that prelinguistic infants disentangle consonant-consonant coarticulatory influences in speech in an adult-like fashion.

The mappings are complex between the phonetic structure of a spoken message and the acoustic structure in the speech signal that conveys the message to a listener. So too, therefore, is the reverse mapping between acoustic signal and phonetic message. Of course, mature listeners recover phonetic properties despite the complexity of these mappings. Adults have extensive experience hearing and producing the sounds of speech, as well as an active knowledge of the lexicon and syntax of their language, all of which potentially aid recovery of a speech message. Yet what of very young infants, who have much more limited speech listening experience, even less experience producing speechlike sounds, and no comprehension of words or syntactic rules? What structure do they recover from the acoustic speech signal? Certainly, the acquisition of language entails recovery of phonetic structure from the acoustic signal. But when does the capability to recover phonetic structure emerge? Previous findings indicate that certain relevant achievements, such as perceptual constancy, perceptual equivalence and trading of phonetically equivalent acoustic properties, and use of context in speech perception, are present long before the infant utters or understands its first meaningful word, and even before it begins to produce syllable-like babbles (Bertoncini, Bijeljac-Babic, Jusczyk, Kennedy, & Mehler, 1988; Eimas, 1985; Eimas & Miller, 1980a, 1980b; Grieser & Kuhl, 1989; Kuhl, 1979, 1980, 1983; Morse, Eilers, & Gavin, 1982).

This research was supported by NIH Grant DC00403 to Catherine T. Best. We wish to thank the following people for their contributions to completion of the project: Virginia Mann for a helpful discussion of a possible auditory account of the results; Michael Donaghu for help collecting and scoring the data of the adult listeners in Experiments 1 and 2; and Glendessa Insabella, Stephen Luke, Peter Kim, Laura Klatt, Meredith Russell, Jean Silver, Pam Speigel, and Jane Womer for help collecting, scoring, and analyzing the infant data in Experiment 3. We also thank our adult subjects, and we are particularly grateful to the parents of the infant subjects for their interest in the project and their willingness to permit their children's participation. Gerald W. McRoberts is currently in the Department of Psychology, Stanford University, Stanford, CA 94305. Reprint requests should be sent to Carol A. Fowler or Catherine T. Best at Haskins Laboratories, 270 Crown St., New Haven, CT 06511.

None of those reports, however, has focused on infants' perception of the particular complex mappings between acoustic and phonetic structure that arise from coarticulation. Coarticulation is of particular interest because of the ways in which it complicates the acoustic consequences of phonetic-segment production. The language-learning child must disentangle those complications in order to come to recognize the segmental structure of speech.

Talkers coarticulate phonetic segments—that is, they implement the phonetic properties of neighboring consonants and vowels in overlapping time frames. The effects work in both directions in time. As an example of anticipatory coarticulation, vowels followed by nasal consonants are nasalized (e.g., Kent, Carney, & Severeid, 1974); as an example of carryover, or perseverative, coarticulation, /g/ preceded by /l/ is fronted (Mann, 1980). The consequence of such coarticulatory overlap is that coarticulating phonetic segments have converging effects on common acoustic dimensions of a speech signal within a given time frame (see, e.g., Fant & Lindblom, 1961). Accordingly, one must ask how even mature listeners deal with the converging effects of diverse segmental properties on common acoustic dimensions. Research shows that adults deal remarkably successfully with the convergences, behaving as though they have disentangled the converging influences on the acoustic signal. Listeners treat acoustic information for a segment x, occurring in the temporal domain of segment y, as information for x. This holds, for example, for anticipatory vowel information that appears in the domain of a preceding fricative (Whalen, 1983) or in the domain of an earlier transconsonantal vowel (Fowler & Smith, 1986; Martin & Bunnell, 1981); it also holds for anticipatory information about a nasal consonant that appears in the temporal domain of a preceding vowel (Krakow, Beddor, Goldstein, & Fowler, 1988), and for the carryover effects of one consonant occurring in the domain of another (Mann, 1980). In the last-cited research, the high front (alveolar) position of tongue-tip contact for an /l/ pulls the tongue-body forward, whereas /r/ does not exert a fronting effect. As a result, the velar contact for a /g/ is pulled forward in the mouth (i.e., $F3$ onset frequency is raised in the direction of the $F3$ onset frequency for /da/) when it is preceded by an /l/ but not when preceded by an /r/. Compatible with this, if a synthetic continuum for /da/ to /ga/ is preceded by either /al/ or /ar/, adults hear more /g/s following /l/ than /r/, indicating that some of the tongue-fronting information that occurs in the temporal domain of the stop consonant is perceptually attributed to the preceding /l/ (Mann, 1980).

In addition to the classic coarticulatory effects just described, prosodic and nonlinguistic properties of an utterance are coproduced with phonetic segments, and they converge with the segmental influences on the acoustic signal. For example, prosody affects the durational properties and fundamental frequency ($F0$) of an utterance, both of which also reflect systematic variation due to the consonants and vowels on which the prosody is realized (see, e.g., Klatt, 1976; Silverman, 1987). Rate variation illustrates nonlinguistic influences. In speaking, changes in rate have durational effects that may converge with phonetic variation (for example, durational differences related to vowel height), phonological-segmental variation (e.g., differences in phonological length), and prosodic variation (e.g., durational differences related to stress patterns). As in cases of segmental coarticulatory influences, listeners apparently disentangle the prosodic and nonlinguistic influences on the signal. For example, they judge intonational accents as if the effects of vowel height on $F0$ had been eliminated (Silverman, 1987), while, for its part, the contribution of vowel height to the $F0$ contour is used as information for vowel height (Reinholt-Peterson, 1986). In addition, the effects of speech-rate variations are effectively eliminated from the phonetic sources of variation in formant-transition duration that distinguish /b/ from /w/ (e.g., Miller & Liberman, 1979).

The question arises whether the ability to perceive phonetic segments with these converging influences disentangled requires experience producing coarticulated speech. That is, must the speaker/hearer learn to associate the intended phonetic segments with their complex and temporally overlapping acoustic consequences? The prebabbling infant under about 7 months of age lacks this kind of experience because it is not yet producing syllabic combinations of consonant-like and vowel-like sounds. The relevant articulatory experience might be acquired, then, during the last half of the first year, as the infant begins to produce reduplicated and nonreduplicated babbling (see, e.g., Oller, 1980; Stark, 1980). Alternatively, the relevant factor may not be articulatory experience per se, but rather the development of a sizable lexicon beyond 50 or so words, which may enable the child to recognize the efficiency of using a phonological system for lexical organization. We suspected, however, that adult-like perceptual disentangling of coarticulatory influences in the speech signal might be evident even earlier in development than either of these possibilities. Our prediction was derived from an account of speech perception that posits articulatory gestures as the primitives of both speech perception and speech production (Best, in press; Fowler & Rosenblum, in press; see also Liberman & Mattingly, 1985). The specific reasoning that led to the studies reported here was that young infants should show perceptual sensitivity to coarticulatory influences as a consequence of a basic perceptual tendency to recover information in stimulation about the source event that produced the signal (e.g., Gibson, 1966, 1979). To test our hypothesis, in the present study we examined how very young, prebabbling infants handle coarticulatory influences when perceiving speech. Findings on this issue are also relevant to accounts that focus on basic auditory processes (e.g., Diehl & Kluender, 1989); we address two such accounts in our General Discussion.

Infants do show evidence, in other domains, of adult-like perception of the acoustic speech signal. For exam-

ple, they exhibit perceptual equivalence of temporal and spectral information for a stop consonant in a *say-stay* context (Eimas, 1985; see also Morse et al., 1982; cf. Eilers & Oller, 1989).[1] This pattern replicates earlier findings with adults by Best, Morrongiello, and Robson (1981; see also Fitch, Halwes, Erickson, & Liberman, 1980; review by Repp, 1982). Infants also show shifts in boundaries between voicing categories along a voice-onset time (VOT) continuum as the starting frequency of *F*1 is varied, demonstrating a trading relation between temporal and spectral information about stop voicing (Miller & Eimas, 1983), again in keeping with adult findings (Summerfield & Haggard, 1977). Finally, as Carden, Levitt, Jusczyk, and Walley (1981) had found earlier in a study of context effects in adult speech perception, infants fail to distinguish fricationless /fa/ and /θa/, but do distinguish them when the same frication noise is placed before the truncated syllables (Levitt, Jusczyk, Murray, & Carden, 1989).

Specifically regarding infants' handling of the convergence of multiple aspects of linguistic structure on a single acoustic dimension, however, less is known. They do show adult-like normalization for the influence of a nonlinguistic factor—speech-rate variations—when discriminating /b/-/w/ syllables that vary in formant-transition duration (Miller & Eimas, 1983; cf. Jusczyk, Pisoni, Reed, Fernald, & Myers, 1983). To our knowledge, however, no one has looked at infants' perception of convergences caused by concurrent production of multiple linguistic properties of an utterance—in particular, by coarticulation of segmental properties. As we suggested earlier, perceptual disentangling of the acoustic effects of multiple gestural influences on the speech signal are important to the child's discovery of the segmental organization of its native language.

Therefore, in the present study, we examined prelinguistic infants' ability to separate coarticulatory influences on a speech signal, before the age at which infants begin to produce syllabic babbling themselves. We chose to use Mann's (1980) stimuli,[2] because experience producing /r/ and /l/, and consonant-consonant (CC) sequences in general, typically emerges rather late in language development, during the preschool years; those properties are not evident in the vocalizations of 4- to 5-month-olds, and are rare even in the babbling of much older infants. The first two experiments with adult listeners were designed to verify earlier findings of perceptual "normalization" of coarticulatory influences between adjacent consonants, and to extend those findings to performance under conditions similar to those used in infant discrimination testing procedures. These first two studies also served to identify the appropriate stimulus pairings for use in the final experiment with 4- to 5-month-old infant listeners. We predicted that, even prior to producing syllable-like babbling, infants would show the same pattern of perceptual sensitivity to coarticulatory influences as adults.

EXPERIMENT 1

In the first experiment, we replicated a portion of Mann's (1980) Experiment 1, using a subset of her stimuli. In Mann's research on adult listeners, the boundary along a synthetic /da/-/ga/ continuum was shifted by a preceding naturally produced /al/ syllable as compared to a preceding /ar/ or no preceding syllable at all. Specifically, /ga/ responses increased in the context of /al/. Mann interpreted the findings as suggestive evidence that perception takes into account the carryover coarticulatory fronting effects of /l/ on a following velar consonant when identifying a following consonant as having a velar or alveolar place of articulation. Our primary purpose in this study was to determine whether we could identify the critical stimulus items needed for the infant test (Experiment 3) and for an adult test under conditions approximating those of the infant discrimination procedure (Experiment 2). Specifically, the latter two procedures required that we obtain three equidistant items along the /da/-/ga/ continuum, one of which adults identify consistently as /da/ in both the /al/ and the /ar/ context, one consistently identified as /ga/ in both contexts, and a crucial item midway between these two which is identified predominantly as /ga/ following /al/ but as /da/ following /ar/.

Method

Subjects. The subjects were 9 undergraduate students and 1 graduate student. All were native speakers of English who reported normal hearing, and all were naive to the purposes of the experiment. The undergraduates received course credit for their participation.[3]

Materials. We used a subset of Mann's (1980) stimuli. They consisted of "hybrid" disyllables of which the first syllable was naturally produced and the second was synthesized. Use of natural initial syllables ensures that natural coarticulatory information for a following stop consonant is available to the listeners; use of synthetic final consonant-vowel (CV) syllables permits sensitive detection of shifts in identification of the synthetic consonant along a continuum according to coarticulatory context.

The first syllables of each disyllabic nonsense word were stressed /al/ or /ar/ produced by a male speaker of English in the context of following /da/ or /ga/. Durations of each of the four precursor syllables were as follows: "al(da)," 261 msec; "al(ga)," 262 msec; "ar(da)," 248 msec; and "ar(ga)," 242 msec. As Mann's (1980) measurements indicate, major differences between /al/ and /ar/ syllables are that /ar/ has a higher *F*2 and a lower *F*3 than /al/. For the four syllables we used, estimates of the offset frequencies of *F*2 and *F*3 were, respectively, 1012 and 2720 Hz for "al(d)"; 1060 and 2720 Hz for "al(g)"; 1566 and 1824 Hz for "ar(d)"; and 1402 and 2018 Hz for "ar(g)." In the isolated /ar/ and /al/, the place of articulation of the stop consonant following the /r/ or /l/ in the original disyllabic productions was identifiable due to anticipatory coarticulation. Each /al/ and /ar/ syllable was spliced onto each member of a seven-item /da/-/ga/ synthetic speech continuum to create four distinct VCCV continua. Stimuli in the CV synthetic continuum differed in the onset of *F*3, which ranged from 2690 to 2104 Hz in approximately even steps. Onsets of *F*1 and *F*2 were 310 and 1588 Hz. Steady states for *F*1, *F*2, and *F*3 were 649, 1131, and 2448 Hz. Transitions were 100 msec in duration. While these are

rather long transitions for stop consonants, we chose to retain Mann's original stimuli; in any case, they were clearly stops rather than glides. Total CV durations were 230 msec, including a 50-msec closure interval following the /al/ or /ar/ precursor.

Pairing of each natural VC syllable with each continuum member gave 28 distinct disyllables. A test order was created consisting of 10 tokens of each of the 28 disyllables in random order with 3.5 sec between trials in the test and a 7-sec pause after each block of 28 stimuli.

Procedure. The subjects listened to tape-recorded stimulus presentations over headphones in a sound-attenuated room. They were tested in groups of 1-3 students. They were instructed to identify the second consonant in each disyllable as "d" or "g" (by writing the appropriate letter on an answer sheet), guessing if necessary.

Results and Discussion

Figure 1 displays the percentage of "g" responses to synthetic CV continuum members separately for the four continua. The top display in the figure compares the outcome when precursor syllables were "al(d)" and "ar(d)"; the bottom display presents the results when precursors were "al(g)" and "ar(g)." In an analysis of variance with the factors continuum (Items 1-7), precursor syllable (/al/ or /ar/), and stop context of the precursor as originally produced (/d/ or /g/), all main effects and interactions reached significance. The main effect of continuum [$F(6,54) = 144.76, p < .0001$], which accounted for most of the variance in the analysis (72%), reflected the increase in "g" responses with a decrease in onset $F3$ in the synthetic continuum. The main effect of precursor [$F(1,9) = 29.33, p = .0005$] reflected the effect of interest, a lower percentage of "g" responses associated with the precursor /ar/ as compared to /al/. The main effect of contextual stop [$F(1,9) = 6.43, p = .03$] reflected a lower percentage of "g" responses for precursors originally produced in the context of following /d/ than /g/. Interactions involving the factor continuum appeared largely to reflect the smaller magnitude of main effects and interactions at the endpoints of the continuum where "g" responses were at floor or ceiling. The interaction of precursor syllable × context stop consonant [$F(1,9) = 14.04, p = .0046$] was significant because the effect of context consonant was present only for the /ar/ precursor, and, on the other side, because the effect of precursor syllable was present only for the "al(d)"-"ar(d)" precursor pair. Mann (1980) obtained this interaction as well (see her Figure 3); however, her effect of precursor syllable was reduced, rather than eliminated, for the "al(g)"-"ar(g)" precursors.

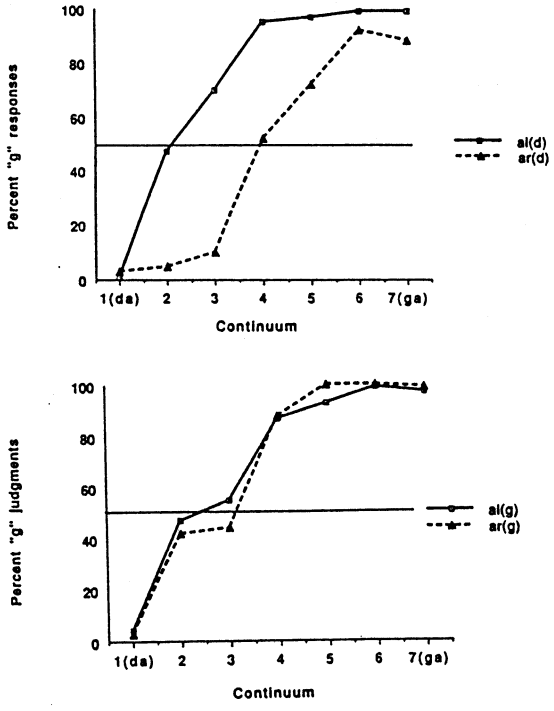

Figure 1. Identification functions averaged across 10 adult listeners, for synthetic /da/-/ga/ continuum preceded by stressed "al(d)" and "ar(d)" (top) in Experiment 1. Bottom: data on "al(g)" and "ar(g)" continua.

Just one of the two possible pairs of continua that we might use with infants provided an outcome meeting our requirements. With precursors "al(d)" and "ar(d)," as depicted in Figure 1 (top), the fifth CV along the continuum (henceforth ga5) was identified predominantly as "ga" preceded by both precursor syllables (97% of the time after /al/ and 72% after /ar/), while the first (da1) was identified predominantly as "da" in both contexts (93% after /al/ and 98% after /ar/). The crucial third CV (henceforth d/ga3) was identified predominantly as "ga" after /al/ (70%), but as "da" after /ar/ (90%). Pairing these CVs with /al/ and /ar/ allowed us to test two between-category discriminations in Experiments 2 and 3, one for each preceding context (ald/ga3 vs. alda1 and ard/ga3 vs. arga5) and two within-category discriminations (ald/ga3 vs. alga5 and ard/ga3 vs. arda1), with the acoustic differences matched among between- and within-category pairs. Thus, the within- and between-category pairs pattern oppositely between the /al/ context and the /ar/ context.

In the other possible pair of continua (with "al(g)" and "ar(g)" precursors; Figure 1 bottom), while continuum members 5 and 1 were convincingly "ga" and "da," respectively (with percent identification >92% in each response category), and while responses to the third continuum member was predominantly "ga" with the "al" precursor (55%) and "da" with the "ar" precursor (56%), the 11% separation in response rates to the third continuum member was small and unreliable [$t(9) = 1.03$]. Possibly the precursor effect diminishes (Mann, 1980) or, here, is eliminated, in the context of following /g/ because information for /g/ in "ar(g)" promotes "ga" identifications more so than does /g/ information in "al(g)."[4] This effect of anticipatory coarticulation on "g" identifications in the "ar(g)" context balances the complementary effect of carryover coarticulation on listeners' tendency to report more "g"s following "al" than "ar" in the "ar(g)-al(g)" continua. As for reasons why effects of the precursor syllables originally followed by /g/ were present in Mann's findings and not in our own, the most

likely reason is that we used just one of her six (three stressed and three unstressed) tokens of each precursor syllable. Rather than pursue this issue, however, which was not a primary focus of our study, we dropped the "al(g)" and "ar(g)" precursors and performed the remaining experiments with "al(d)" and "ar(d)" precursors.

Testing the foregoing between- and within-category discriminations using "al(d)" and "ar(d)" precursors with prelinguistic infant listeners may help to determine whether prebabbling infants show an adult-like effect of precursor syllable on their responses to continuum members. Before testing infants, however, we ran a further study with adults. Experiment 2 was designed to ensure that adult discrimination performance, under conditions similar to the infant discrimination procedure used in Experiment 3, would reflect the categorizations suggested by the identification data collected in Experiment 1.

EXPERIMENT 2

For Experiment 2, we chose a signal-detection discrimination procedure for adults. This was necessary to verify that the stimulus pairs we had chosen on the basis of the results of Experiment 1 would maintain their category memberships when presented under listening conditions that approximated the discrimination task we planned to use with our infant listeners. Accordingly, adults listened to sequences of varying numbers of identical (background) disyllables (either of the critical stimuli ald/ga3 or ard/ga3), in which a new disyllable (/al/ or /ar/ followed by either da1 or ga5) was presented at an unpredictable point near the end of the sequence. They hit a response key whenever they detected a change from the background disyllables. We performed a signal-detection analysis on the data.

Method

Subjects. The subjects were 12 undergraduates who participated for course credit. All were native speakers of English who reported normal hearing. All were naive with respect to the experimental hypotheses.

Materials. The test consisted of 48 sequences evenly divided among the four conditions of the experiment (background disyllable ald/ga3 changing either to alda1 or alga5, and analogous sequences using ard/ga3 changing either to arda1 or arga5). Across sequences, the change or target disyllable occurred after as few as 10 repetitions of the background disyllable or as many as 33 repetitions. The target disyllable was presented one time in each sequence, and it was followed by two repetitions of the background disyllable before the sequence ended. Distance of the target disyllable from the beginning of the sequence was balanced across lists. There was a 1,500-msec interval (offset to onset) between disyllables in a sequence. On the second channel of the tape, a tone pulse marked the onset of each disyllable. That pulse, input to a computer, enabled association of keypress responses signaling detection of a target disyllable with each disyllable in a sequence.

Procedure. Listeners were tested individually. The stimuli were presented over a loudspeaker (as in the infant experiment) in a quiet listening room. The subjects were instructed to hit a key on a computer terminal keyboard whenever they heard a change from the background disyllable, however subtle the change might be. They were not told that there was just one target disyllable per sequence; accordingly, they were allowed to hit the key as many times as they chose on each trial of the experiment. They were told, however, that the change would never occur before the 11th disyllable of a given trial; this would allow them to get used to the background disyllable's sound before listening for a change.

Measures were hits, misses, false alarms, and correct rejections, converted to d' measures.

Results

Figure 2 displays the d's for the four conditions. As the figure shows, d' measures were considerably higher for the two between-category discriminations than for their corresponding within-category discriminations. In an analysis of variance with the repeated measures factors precursor syllable (/al/ or /ar/) and direction of shift (to ga5 or da1), neither main effect was significant (both Fs < 1), but the interaction was highly significant [$F(1,11) = 57.77$, $p < .0001$]. The interaction reflects two significant outcomes: (1) poor discrimination ($d' = .07$) of d/ga3 from da1 in the context of /ar/, but good discrimination of the same shift in the context of /al/ ($d' = 2.38$), and (2) poor discrimination of d/ga3 from ga5 in the context of /al/ ($d' = .57$), but good discrimination in the context of /ar/ ($d' = 2.40$). Pairwise comparisons (Scheffé tests) verified that d's for the between-category discriminations were significantly larger than those for within-category discriminations [/al/, $F(1,11) = 7.32$, $p = .006$; /ar/, $F(1,11) = 12.14$, $p = .0009$]. Pairwise comparisons of the two between-category discriminations and of the two within-category discriminations were nonsignificant (both Fs < 1). Finally, excepting the d' values for the within-category discrimination with /ar/ as the precursor syllable, all conditions showed significantly positive d's, indicating significant evidence of discrimination [for the within-category discrimination involving /al/, $t(11) = 2.97$, $p = .01$]; there were no negative d's in the two between-category conditions.

On the basis of these findings, we considered our stimulus pairings appropriate for testing with prelinguistic infants.

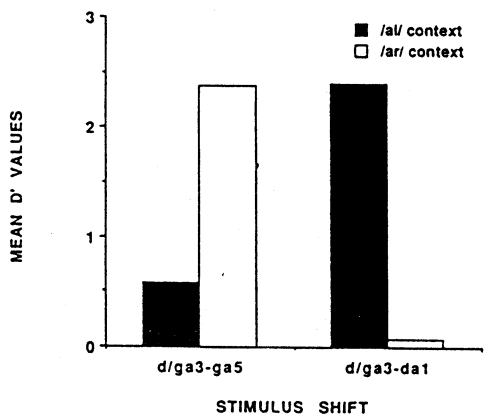

Figure 2. Average d' values of 12 adult listeners in the signal detection test for discrimination of d/ga3 from ga5 and from da1 preceded by /al/ and by /ar/ (Experiment 2).

EXPERIMENT 3

In the final experiment, we examined 4- to 5-month-olds to determine whether or not they disentangle coarticulatory influences as adults do. We predicted that our prebabbling infants would discriminate the stimulus pairs determined to be between-category in the adult tests, but would fail to discriminate the pairs that were within-category for adults. That is, the infants should show the same context-dependent reversal in performance levels as had the adults in Experiment 2 when discriminating d/ga3 from ga5 and from da1, suggesting perceptual sensitivity to the converging influences of multiple phonetic segments on a single acoustic dimension.

The infants participated in a habituation procedure comparable to the signal-detection task of the adults in Experiment 2. Following habituation to either the ald/ga3 or ard/ga3 disyllable, infants received one of two stimulus shifts: to the corresponding ga5 disyllable or to the corresponding da1 disyllable. Fixation time before and after the shift was examined for evidence of dishabituation to the novel stimuli.

Method

Subjects. The subjects were 48 infants from the communities surrounding Wesleyan University, between 4 and 6 months of age (M = 4 months, 17 days; range = 4 months to 5 months, 29 days). Twelve infants were tested in each of the four test conditions (see Procedure), with males and females approximately equally distributed across conditions. Data from an additional 16 infants were excluded because of crying/fussing (3), inattention to the visual stimulus (6), performance scores greater than 2 SD beyond the mean for the infant's test condition (1), equipment problems (2), and experimental error (4). Thus, the success rate was 75%. The dropout rate was approximately evenly distributed across the experimental conditions.

The subjects were solicited via mailings and follow-up phone calls to parents listed in the birth announcements of newspapers for Middletown, CT, and neighboring towns. This recruitment procedure yields an approximate 25%-30% acceptance rate.

Materials. There were four 30-min stimulus tapes, one for each test condition. The stimuli were recorded in synchrony on two channels of a four-track tape, with tone pulses recorded on a third track, 15 msec preceding the onsets of each pair of items on the stimulus channels. There were 1,500-msec interstimulus intervals between disyllables on the stimulus channels, as in Experiment 2. The tone pulses were used to signal a computer as to when stimulus presentations could be initiated, terminated, or switched between channels (see Procedure). The d/ga3 stimulus preceded by the precursor syllable for the appropriate condition (/al/ or /ar/) was recorded on one channel of the tape, while synchronized repetitions of the appropriate ga5 or da1 disyllable were recorded on the other channel.

Procedure. Each subject was tested on one of four test comparisons: (1) ald/ga3 – alga5; (2) ald/ga3 – alda1; (3) ard/ga3 – arga5; and (4) ard/ga3 – arda1. Conditions 1 and 4 presented within-category comparisons according to the adult findings, whereas Conditions 2 and 3 presented between-category comparisons.

We employed the infant-controlled visual fixation discrimination procedure described by Miller (1983). In this procedure, the infant is operantly conditioned to fixate a rear-projected slide of a brightly colored checkerboard in order to receive audio presentations of speech stimuli. The stimuli were presented at a comfortable listening level (70 dB) over a loudspeaker (Jamo) hidden a few feet above the target slide. A computer (Atari-800) initiated and terminated the stimulus presentations from a continuously playing tape deck (Otari 5050 MXB), and determined which channel of the tape was presented over the loudspeaker, on the basis of keypress input from a trained observer. The observer viewed a video monitor conveying input from a camera focused on the infant's face (under control of a cameraperson) in order to detect the infant's fixations of the target slide. The observer was separated from the infant and loudspeaker by a sound-treated wall. To further assure that (s)he was "deaf" to the stimuli that the infant heard, the observer wore headphones and listened to music throughout the session. In addition, the observer was unaware of when during the test the stimulus shift trials actually occurred, because the number of habituation trials varied from infant to infant, depending on their fixation patterns. The observer's lack of awareness about the course of the test session was underscored by the fact that the cameraperson invariably had to let them know when the test had ended.

The infant's fixation behavior determined the division of the test session into individual trials. Whenever the infant gazed away from the target slide for more than 2 sec, the slide was automatically shut off for 1 sec and then redisplayed to begin a new trial. Once the infant habituated to the familiarization stimulus during the habituation phase of the test, the speech presentations were shifted to the novel stimulus on the second audio channel during the test phase. The habituation criterion was a decline in the infant's fixations on two consecutive trials to a level below 50% of the mean of the two highest preceding trials. Stimulus presentations were shifted to the test channel on the next trial following that on which the habituation criterion was met. The exact details of the procedure and experimental set-up are described in Best, McRoberts, and Sithole (1988).

To assess the interjudge reliability of observations of the infants' visual fixations, the videotapes of 29 test sessions were rescored by members of the research team (60% of the sessions). Included were all sessions for which there was any question about the infants' fixation pattern and/or behavioral state (e.g., fussing), as well as an equal number of unquestioned sessions. Interobserver correlations were quite high, ranging between .95 and .99, with one exception at .78 (the latter session was retained because the single test trial on which the observers disagreed was not one of the critical trials surrounding the stimulus shift).

Results and Discussion

We computed the mean looking times for the two trials immediately preceding the stimulus shift (habituation level) and for the first two postshift trials beginning when the infant heard at least one test stimulus presentation (dishabituation). Some infants failed to look at the slide during the first trial or so after the shift because they had habituated to 0 during the first part of the test, and hence they failed to hear any postshift stimuli during those first postshift trials. Because at least one postshift stimulus was needed for the infant to have an opportunity to discriminate between preshift and postshift stimuli, then, we did not include in the dishabituation mean any non-looking trial(s) immediately following the shift. Once the infant looked even briefly enough to hear one postshift stimulus, the true dishabituation trials began (see Best et al., 1988). The summary data are shown in Figure 3 for the four conditions of the experiment. Qualitatively, the response pattern in Figure 3 is very similar to that of the adult listeners shown in Figure 2. As predicted, t tests (one-tailed) revealed significant recovery after the stimulus shift in the two conditions predicted to provide

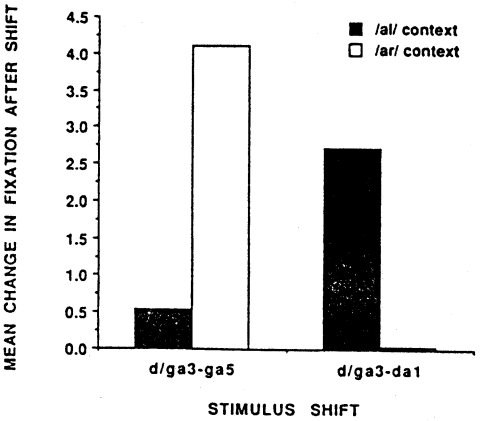

Figure 3. Infants' response recoveries (in seconds) following the stimulus change in each condition (12 subjects per condition) of the infant-controlled visual fixation habituation procedure; results indicate extent of infant discrimination of d/ga3 from ga5 and from da1 preceded by /al/ or /ar/ (Experiment 3).

between-category comparisons [ald/ga3 to alda1, $t(11) = 2.74$, $p = .01$, and ard/ga3 to arga5, $t(11) = 2.04$, $p = .03$], and no significant recovery in the remaining conditions. Compatibly, an analysis of variance on pre- and posthabituation looking times with the factors precursor syllable (/al/ or /ar/) and direction of shift (to ga5 or to da1) yielded no main effects (both Fs < 1) but did yield a significant interaction [$F(1,44) = 4.57$, $p = .038$]. The interaction is significant because the relative recovery magnitudes in the two shift directions (ga5, da1) pattern oppositely, depending on the preceding context.

Accordingly, for prelinguistic infants as for adults, a stop consonant that is ambiguous between /d/ and /g/ is heard as less "d"-like in the context of /l/ than in the context of /r/. That is, both mature listeners and prelinguistic infants effectively remove the coarticulatory fronting influence that /l/ has on a following velar consonant.

GENERAL DISCUSSION

That prelinguistic infants show the same interaction in the two syllable contexts as adults do demonstrates conclusively that, in this instance at least, neither experience producing coarticulated speech, nor acquisition of language-specific lexical items is required for perceptual elimination of coarticulatory influences on acoustic information for a phonetic segment. Another finding in the literature is relevant to an interpretation of the outcome. Mann (1986) tested Japanese listeners on the disyllables used in Mann (1980) and in the present experiments. This language group is of interest because the Japanese language does not make a phonemic /l/-/r/ distinction. Mann identified two groups of Japanese listeners on the basis of their ability to label stimuli consistently as "l" and "r." In one group, listeners were at chance on the average (58% correct, $p > .1$) in identifying the final consonants of /al/ and /ar/. In another, they were near perfect (98% correct). Remarkably, both groups of listeners showed shifts in the /da/-/ga/ boundary in the context of preceding /l/ as compared to /r/. Moreover, the magnitude of the shift was the same in the two groups of Japanese listeners as in a third group of native English listeners. Apparently, a listener need not be able to classify consonants into distinct phonological categories in order to extract their different coarticulatory influences on neighboring consonants. How, then, is the extraction to be explained?

If both mature listeners who cannot reliably classify /l/s and /r/s into different phonemic categories and prelinguistic infants show the same perceptual response patterns as do mature listeners who command the phonemic distinction, presumably an explanation for the response patterns must derive from something that all three groups have in common. One possibility is the auditory systems of these listeners.

Mann (1986) considers and rejects one such account of the Japanese listeners' performance patterns. It is that auditory nerve fibers are known to exhibit forward masking by one acoustic signal that precedes another by 50-100 msec. The masking effect is such that the response of the auditory nerve is depressed to stimuli in the same frequency range as that of the preceding masking stimulus (Delgutte & Kiang, 1984; Harris & Dallos, 1979; Smith, 1977). Psychophysical tests of human listeners reveal compatible response patterns (Elliot, 1971; Moore, 1978).

In Mann's stimuli, /al/ but not /ar/ has an $F3$ offset frequency close to the onset frequency of $F3$ for stimuli at the /da/ end of the /da/-/ga/ continuum. Accordingly, preceding /al/ should selectively depress auditory-nerve sensitivity to stimuli at that end of the continuum, giving rise to the observed increase in "ga" responses.

For several reasons, we reject this account of our findings and of Mann's (1980, 1986). First, as Mann (1986) points out, the auditory masking interpretation is weakened by findings of Mann and Liberman (1983). They employed the same stimuli as those under test here; however, the critical $F3$ transitions for /da/ or /ga/ were presented to one ear, and the remainder (base) of the disyllable was presented to the other ear. This manner of presenting speech stimuli gives rise to a "duplex" percept in which the $F3$ transition is apparently heard in two ways at once. It is integrated with the information in the opposite ear, giving rise, in that location, to a /da/ or /ga/ percept for the second syllable of the disyllable; it is simultaneously heard as a pitch glide in the ear receiving the transition. Under these conditions, Mann and Liberman obtained two findings that are important for the present purposes. First, context effects of /l/ on "d" and "g" classifications were present, eliminating the auditory nerve (or in fact any other peripheral influence) as a source of the context effects. Second, context effects were absent in the classifications of the pitch glides, weakening any account of the context effects that ascribed them to masking originating in higher level (central) auditory-system processing per se.

A final reason to reject an auditory masking account is that the offset frequency of *F*3 of /l/ (2711 Hz averaged across the multiple natural /al/ tokens in Mann's stimuli) is closest to the endpoint /da/'s *F*3 onset frequency (2690 Hz) and becomes progressively farther from the other continuum members' *F*3 onsets as we approach ga7 (2104 Hz). Since, in the auditory masking literature, effects are largest for stimuli closest in frequency to the context stimulus, auditory effects should be largest on the /da/ endpoint and progressively smaller thereafter (Mann, personal communication, February 1, 1990). However, this is opposite to the pattern of context effects found in Mann (1980, 1986), Mann and Liberman (1983), and the present study. Furthermore, masking should be absent outside the critical band surrounding 2711 Hz (approximately 400 Hz), but the first continuum member outside that band is d/ga3, the stimulus on which the largest context effects were obtained.

If the perceptual elimination of coarticulatory influences is not to be explained by appeal to masking, how is it to be explained? Possibly the findings of Mann and Liberman (1983) permit a further inference about the domain in which an explanation for the context effects should be sought. Mann and Liberman found that only formant transitions that are experienced in the same spatial location (ear) as the rest of the disyllable and that are experienced as part of the disyllable are subject to context effects. The dichotic shift in perceived location of the transition must be associated with a perceptual "parsing" of the acoustic signal, in which the transitions and the remainder of the disyllable serve as joint acoustic consequences of a single coherent sound-producing event. If so, then context effects may arise only when the context counts perceptually as part of the same sound-producing event that gave rise to the transitions. Yet parsing into distinct segmental influences on a single sound-producing event must be based on relevant information in the acoustic signal. If so, perhaps there is also an informational basis in the signal for the context effects, rather than a basis in the auditory mechanisms of the listener.

Consider one implication of an inference that the context effects are information-based. The information in an acoustic speech signal is about its gestural source in the vocal tract. That is, the structure in a speech signal is directly caused by the actions of the moving vocal tract; accordingly, to the extent that different actions of the vocal tract pattern the air pressure changes differently, structure in the acoustic signal provides information about its articulatory gestural source. It need not follow from this, of course, that listeners use acoustic structure in that way. However, there is reason to suppose that they do.

Across perceptual modalities, perceiving is the only means by which organisms can come to know the environment in which they participate as actors. But perception can be the means by which the environment is known only if stimulation at the sense organs—structured energy patterns in the air and light, for example—serves not as something to be perceived and experienced in itself, but rather as information about the causal sources of its structure in the environment (see, e.g., Gibson, 1966, 1979). As visual perceivers, we *see* environmental sources of structure via reflected light; we do not see the structure in the light itself, even though it is the light and not the environment that stimulates the retina. We use the structure in reflected light to recover its environmental causes. Compatibly, as haptic perceivers, we experience manipulable objects in the environment, not the skin and joint-angle deformations they cause. Accordingly, as auditory perceivers, we should hear environmental sources of structure in acoustic signals, not the acoustic signals themselves, which should serve, instead, as information bearers. In speech, the sources of acoustic structure are linguistically significant actions of the vocal tract (see Best, 1984, in press; Browman & Goldstein, 1986; Fowler, 1986, 1989; Fowler & Rosenblum, in press; see also Liberman & Mattingly, 1985).

Setting aside for the moment the possible influence of perceptual learning, information in the acoustic signal about its origin in a sound-producing event in the environment—including vocal tract actions—is available to any organism with an auditory system able to register the relevant acoustic structure. This includes prelinguistic infants, adult speakers from any language community, and even nonhuman animals with appropriate auditory systems.

How, then, is perceptual elimination of coarticulatory influences of /l/ on following /g/ to be explained from this perspective? The /l/ in /alga/ is produced in part by creating a constriction between the tip of the tongue and the alveolar ridge of the palate. A /g/ is produced by creating a constriction between the back of the tongue and the soft palate. The forward constriction of the /l/ pulls the whole tongue forward, however. When production of the two phonetic segments overlaps, the constriction location for the following /g/ is fronted along the soft palate. The alveolar constriction, the soft-palate constriction, and the causal effects of the former on the latter all have acoustic consequences. To the extent that the consequences are specific to those actions, the acoustic signal can specify those actions to a sensitive perceiver who then can ascribe the fronting to its source, the alveolar constriction. This information, if it is there at all, is as available to a prelinguistic infant as it is to a mature listener of any language community and even to a variety of nonhuman animals.[5]

As for the effect of learning a specific language on recovery of phonetic properties from an acoustic speech signal, our interpretation is similar to Mann's (1986). We have argued that listeners can recover information about vocal tract actions from acoustic speech signals. Mann refers to this as a "universal" level of perception, to contrast it with a distinct, language-specific phonological level in which the linguistic significance of perceived gestures

is appreciated. We will refer to the distinction in terms of attunement of attention, rather than perceptual levels. There is a mode of attending to acoustic speech signals that is available to listeners who participate in a particular language community and who have, therefore, discovered the linguistic significance, if any, of phonetic-gestural distinctions conveyed by an acoustic speech signal. This mode of attending is available to mature language users, but not to prelinguistic infants or to nonhuman animals (cf. Note 5). Although this linguistically informed mode of attending to the signal is essential to linguistic interpretation of an utterance in the listener's native language (e.g., Best et al., 1981; Best, Studdert-Kennedy, Manuel, & Rubin-Spitz, 1989), it may hinder explicit classification according to phonetic differences that are not phonologically distinctive in the native language (e.g., Werker & Logan, 1985). In making "l"-"r" classifications, Japanese listeners are impaired by their difficult-to-overcome tendency to ignore phonetic distinctions that are phonologically nondistinctive in their language. In contrast, all listeners can recover phonetic gestures of the vocal tract from the acoustic signal and can disentangle coarticulatory interactions among gestures, at least those that involve carryover, insofar as the acoustic signal specifies them. We suggest that prelinguistic infants eliminate coarticulatory influences of /l/ on /g/ precisely because the signal does specify the distinct articulatory correlates of /l/ and /g/ when the two segments are coarticulated.

Before concluding in this way, however, we will consider an alternative, auditory, account of the findings of the present research that is also consistent with the inference that the context effects observed in this research are information-based. Mann considered this interpretation in her original article (1980), but not in her later one (1986), perhaps for a reason that we will outline shortly; two reviewers of the present manuscript requested that we consider the interpretation. We will do so and explain why we consider it untenable.

The account ascribes the context effects of /l/ and /r/ on /d/-/g/ perception to auditory contrast. Contrast effects are widely observed in research obtaining perceptual judgments from subjects (see Warren, 1985 for a review), and on that basis alone, contrast might be considered a plausible or even likely cause of the present findings. In this instance, the high *F*3 of /al/ as compared to /ar/ may have a contrastive effect on judgments of the *F*3 transition of the following synthetic CV, leading listeners to judge it lower in frequency and hence more characteristic of /g/ than /d/. While the duplex perception experiment of Mann and Liberman (1983), cited earlier, rules out a locus for such an effect in the auditory system periphery, some contrast effects are thought to be more central in origin. In an example cited by one reviewer, Johnson (1944) found that immediately prior experience hefting weights gave rise to contrast effects on weight judgments; however, he observed informally that an interpolated weight that subjects considered extraneous to the experimental setting—in particular, a book or chair that subjects might have moved during a rest break in the experimental proceedings—was "without apparent effect upon their scales of value based upon lifting the stimulus weights" (p. 436). If these informal observations are accurate and general, then perhaps the findings of Mann and Liberman (1983), and hence of the present investigation, can be explained in terms of contrast effects at a cognitive level. In particular, possibly in the research of Mann and Liberman (1983), the presence of context effects on the second syllable of the disyllables, but not on the isolated pitch glides, occurred because, as we suggested earlier, the pitch glides but not the disyllables' CVs were judged perceptually to constitute distinct objects from the influencing VCs.

An account in terms of auditory contrast makes qualitatively the same predictions concerning effects of spectral consequences of coarticulatory overlap on perception as does our proposed articulatory account. Acoustic effects of coarticulation are generally assimilatory, and contrastive effects of the coarticulating segment's acoustic consequences will always work to neutralize the perceptual effects of the assimilations. Qualitatively, this will also be the effect if listeners, as we suggest, ascribe coarticulatory influences to the coarticulating, rather than the influenced (target), phonetic segment.

Even so, for two reasons, we discount the explanation of perception of coarticulatory context effects in terms of auditory contrast. The first reason concerns Mann's (1986) findings with Japanese listeners who were at chance in identifying /l/ and /r/, but who nonetheless exhibited context effects indistinguishable from those of English listeners and of Japanese listeners able to make the identifications. While findings of Mann and Liberman (1983) exclude a peripheral locus for any contrast effects just as they eliminate a peripheral locus for masking, findings of Mann (1986) with the first-mentioned group of Japanese listeners exclude a late, cognitive, locus—the locus at which Johnson's (1944) subjects would have excluded books and chairs from having a contrastive effect on weight judgments. Those listeners exhibited differential effects of context on phonetic segments that they could not label differentially. Accordingly, the contrast effects cannot arise early and they cannot arise late. There remains the possibility, of course, that contrast effects occur at some intermediate level of processing, less peripheral than the level at which duplex effects arise and more peripheral than that at which phonemic classifications occur. However, the articulatory account does not require such proliferation of processing levels, because it ascribes the effects to the relation between articulation and the acoustic signal, of which listeners are presumed to make use in perception. In articulation, phonetic segments are not discrete along the time axis; accordingly,

listeners perceive a phonetic segment's domain to include its entire articulatory extent, insofar as it is specified acoustically and detectable auditorily.

A second reason to discount an explanation of the present findings in terms of auditory contrast is that the account does not explain the broader array of earlier findings concerning listeners' perception of coarticulated speech. It falls short in two domains, one relating still to spectral consequences of segment-to-segment coarticulatory overlap (classical coarticulatory effects) and the other to the acoustic consequences of other kinds of articulatory overlap.

In the literature, there are two complementary findings concerning listeners' perceptions as guided by spectral consequences of segment-to-segment coarticulatory overlap. One finding is exemplified by the present research. Listeners appear to eliminate effects of coarticulatory assimilations in their judgments of coarticulated segments, so that phonetic segments that are subject to coarticulatory overlap are both identified and discriminated as if the acoustic consequences of coarticulation were eliminated. Other research shows, however, that the acoustic effects of coarticulation are nonetheless perceptually effective as information for the coarticulating segment itself (e.g., Fowler, 1984; Fowler & Smith, 1986; Martin & Bunnell, 1981; Whalen, 1984). Indeed, in the research of Fowler (1984; Fowler & Smith, 1986), both findings are obtained using the same stimuli. That is, effects of coarticulatory assimilations appear to have been eliminated in discriminations of influenced segments, but nonetheless they serve as information for the coarticulating segment itself. Contrast effects can explain elimination of the effects of coarticulatory assimilations on perception of a target segment influenced by a coarticulating segment, but it is not obvious how they could put the effects back in elsewhere. Our account of perception,in fact, motivated the research of Fowler cited above, and predicted the obtained outcomes.

The second research domain in which the contrast account fails, in our view, has to do with listeners' perceptual handling of other kinds of articulatory overlap, including prosodic and nonlinguistic variables that yield converging effects on fundamental frequency as reviewed in our introduction. The perceptual results are analogous to those in the literature just reviewed. That is, listeners judge intonation contours as if effects on the fundamental frequency contour of declination (Pierrehumbert, 1979; Silverman, 1987) and of segmental perturbations such as vowel height (Silverman, 1987) had been eliminated. Moreover, as in the literature on classic coarticulation effects, the "eliminated" effects are not eliminated in perception generally; they are eliminated only from listeners' judgments of the pitch melody of an utterance. Phonetic segmental perturbations of the fundamental frequency contour of an utterance, including those due to variation in vowel height and consonant voicing, serve as information for their causes, namely vowel height (Reinholt-Peterson, 1986) and consonant voicing (Silverman, 1986), respectively. It is not obvious that a contrast account would handle even the elimination of the other than intonational convergences on fundamental frequency from perception of the pitch melody, because articulatory overlap does not cause acoustic assimilation in these cases. Nor, analogous to the difficulties for the contrast account that we outlined relating to classic coarticulatory effects, does the contrast account appear to explain why the convergences, eliminated from one set of judgments (here, relating to intonational melody), do contribute to another set (that is, to judgments of vowel height or consonant voicing). An explanation that invokes recovery of the origins of the acoustic pattern in vocal tract actions, however, does provide a unified account of the whole set of findings.

For these reasons, among others, we conclude that perception of coarticulated speech by adults and infants indexes their recovery of talkers' linguistically significant vocal tract actions; it does not index auditory contrast.

REFERENCES

BERTONCINI, J., BIJELJAC-BABIC, R., JUSCZYK, P. W., KENNEDY, L. J., & MEHLER, J. (1988). An investigation of young infants' perceptual representations of speech sounds. *Journal of Experimental Psychology: General*, **117**, 21-33.

BEST, C. T. (1984). Discovering messages in the medium: Speech and the prelinguistic infant. In H. E. Fitzgerald, B. Lester, & M. Yogman (Eds.), *Advances in pediatric psychology* (Vol. 2, pp. 97-145). New York: Plenum.

BEST, C. T. (in press). The emergence of language-specific phonemic influences in infant speech perception. In H. Nusbaum & J. Goodman (Eds.), *The transition from speech sounds to spoken words: Development of speech perception*. Cambridge, MA: MIT Press.

BEST, C. T., MCROBERTS, G. W., & SITHOLE, N. N. (1988). The phonological basis of perceptual loss for non-native contrasts: Maintenance of discrimination among Zulu clicks by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception & Performance*, **14**, 345-360.

BEST, C. T., MORRONGIELLO, B., & ROBSON, R. (1981). Perceptual equivalence of acoustic cues in speech and nonspeech perception. *Perception & Psychophysics*, **29**, 191-211.

BEST, C. T., STUDDERT-KENNEDY, M., MANUEL, S., & RUBIN-SPITZ, J. (1989). Discovering phonetic coherence in acoustic patterns. *Perception & Psychophysics*, **45**, 237-250.

BROWMAN, C., & GOLDSTEIN, L. (1986). Towards an articulatory phonology. *Phonology*, 3, 219-252.

CARDEN, G., LEVITT, A., JUSCZYK, P. W., & WALLEY, A. (1981). Evidence for phonetic processing of cues to place of articulation: Perceived manner affects perceived place. *Perception & Psychophysics*, **29**, 26-36.

DIEHL, R. L., & KLUENDER, K. (1989). On the objects of speech perception. *Ecological Psychology*, **1**, 121-144.

DELGUTTE, B., & KIANG, N. Y. (1984). Speech coding in the auditory nerve: IV. Sounds with consonant-like dynamic characteristics. *Journal of the Acoustical Society of America*, **75**, 897-907.

EILERS, R. E., & OLLER, D. K. (1989). Conflicting and cooperating cues to final stop consonant voicing by infants and adults. *Journal of Speech & Hearing Research*, **32**, 307-316.

EIMAS, P. D. (1985). The equivalence of cues in the perception of speech by infants. *Infant Behavior & Development*, **8**, 125-138.

EIMAS, P. D., & MILLER, J. L. (1980a). Contextual effects in infant speech perception. *Science*, **209**, 1140-1141.

EIMAS, P. D., & MILLER, J. L. (1980b). Organization in the perception of information for manner of articulation. *Infant Behavior & Development*, **3**, 367-375.

ELLIOT, L. L. (1971). Backward and forward masking. *Audiology*, **10**, 65-76.

FANT, G., & LINDBLOM, B. (1961). Studies of minimal speech and sound units. *Speech Transmission Laboratory: Quarterly Progress Report*, **2/1961**, 1-11.

FITCH, H., HALWES, T., ERICKSON, D. M., & LIBERMAN, A. M. (1980). Perceptual equivalence of two acoustic cues for stop-consonant manner. *Perception & Psychophysics*, **27**, 343-350.

FOWLER, C. A. (1984). Segmentation of coarticulated speech in perception. *Perception & Psychophysics*, **36**, 359-368.

FOWLER, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics*, **14**, 3-28.

FOWLER, C. A. (1989). Real objects of speech perception. *Ecological Psychology*, **1**, 145-160.

FOWLER, C. A., & ROSENBLUM, L. D. (in press). The perception of phonetic gestures. In I. G. Mattingly & M. Studdert-Kennedy (Eds.), *Modularity and the motor theory of speech perception*. Hillsdale, NJ: Erlbaum.

FOWLER, C. A., & SMITH, M. R. (1986). Speech perception as "vector analysis": An approach to the problems of segmentation and invariance. In J. Perkell & D. Klatt (Eds.), *Invariance and variability of speech processes* (pp. 123-139). Hillsdale, NJ: Erlbaum.

GIBSON, J. J. (1966). *The senses considered as perceptual systems*. Boston, MA: Houghton-Mifflin.

GIBSON, J. J. (1979). *The ecological approach to visual perception*. Boston, MA: Houghton-Mifflin.

GRIESER, D. A., & KUHL, P. K. (1989). Categorization of speech by infants: Support for speech-sound prototypes. *Developmental Psychology*, **25**, 577-588.

HARRIS, D., & DALLOS, P. (1979). Forward masking of speech by the auditory nerve system. *Journal of Neurophysiology*, **42**, 1083-1107.

JOHNSON, D. (1944). Generalization of a scale of values by the averaging of practice effects. *Journal of Experimental Psychology*, **34**, 425-436.

JUSCZYK, P. W., PISONI, D. B., REED, M., FERNALD, A., & MYERS, M. (1983). Infants' discrimination of a rapid spectrum change in nonspeech signals. *Science*, **222**, 175-177.

KENT, R. D., CARNEY, P. J., & SEVEREID, L. R. (1974). Velar movement and timing: Evaluation of a model for binary control. *Journal of Speech & Hearing Research*, **17**, 470-488.

KLATT, D. (1976). Linguistic uses of segment duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America*, **59**, 1208-1221.

KLUENDER, K., DIEHL, R., & KILLEEN, P. (1987). Japanese quail can learn phonetic categories. *Science*, **237**, 1195-1197.

KRAKOW, R., BEDDOR, P., GOLDSTEIN, L., & FOWLER, C. (1988). Coarticulatory influences on the perceived height of nasal vowels. *Journal of the Acoustical Society of America*, **83**, 1146-1158.

KUHL, P. K. (1979). Speech perception in early infancy: Perceptual constancy for spectrally dissimilar vowel categories. *Journal of the Acoustical Society of America*, **66**, 1668-1679.

KUHL, P. K. (1980). Perceptual constancy for speech-sound categories in early infancy. In G. H. Yeni-Komshian, J. F. Kavanaugh, & C. A. Ferguson (Eds.), *Child phonology: Vol. 2. Perception* (pp. 41-66). New York: Academic Press.

KUHL, P. K. (1983). Perception of auditory equivalence classes for speech in early infancy. *Infant Behavior & Development*, **6**, 263-285.

LEVITT, A., JUSCZYK, P. W., MURRAY, J., & CARDEN, G. (1989). Context effects in two-month-old infants' perception of labiodental/interdental fricative contrasts. *Journal of Experimental Psychology: Human Perception & Performance*, **14**, 361-368.

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

MANN, V. A. (1980). Influence of preceding liquid on stop-consonant perception. *Perception & Psychophysics*, **28**, 407-412.

MANN, V. A. (1986). Distinguishing universal and language-dependent levels of speech perception: Evidence from Japanese listeners' perception of "l" and "r." *Cognition*, **24**, 169-196.

MANN, V. A., & LIBERMAN, A. M. (1983). Some differences between phonetic and auditory modes of perception. *Cognition*, **14**, 211-235.

MARTIN, J. G., & BUNNELL, H. T. (1981). Perception of anticipatory coarticulation effects in /stri, stru/ sequences. *Journal of the Acoustical Society of America*, **69**, S92. (Abstract)

MILLER, C. (1983). Developmental changes in male-female voice classification by infants. *Infant Behavior & Development*, **6**, 313-330.

MILLER, J. L., & EIMAS, P. D. (1983). Studies on the categorization of speech by infants. *Cognition*, **13**, 135-165.

MILLER, J. L., & LIBERMAN, A. M. (1979). Some effects of later-occurring information on the perception of stop consonant and semivowel. *Perception & Psychophysics*, **25**, 457-465.

MOORE, B. C. J. (1978). Psychophysical tuning curves measured in simultaneous and forward masking. *Journal of the Acoustical Society of America*, **63**, 524-532.

MORSE, P. A., EILERS, R. E., & GAVIN, W. J. (1982). The perception of the sound of silence in early infancy. *Child Development*, **53**, 189-195.

OLLER, K. D. (1980). The emergence of the sounds of speech in infancy. In G. H. Yeni-Komshian, J. F. Kavanaugh, & C. A. Ferguson (Eds.), *Child phonology: Vol. 1. Production* (pp. 93-112). New York: Academic Press.

PIERREHUMBERT, J. (1979). The perception of fundamental frequency declination. *Journal of the Acoustical Society of America*, **66**, 363-369.

REINHOLT-PETERSON, N. (1986). Perceptual compensation for segmentally-conditioned fundamental-frequency perturbations. *Phonetica*, **43**, 31-42.

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

SILVERMAN, K. (1986). F0 segmental cues depend on intonation: The case of the rise after voiced stops. *Phonetica*, **43**, 76-91.

SILVERMAN, K. (1987). *The structure and processing of fundamental frequency contours*. Unpublished doctoral dissertation, Cambridge University.

SMITH, R. L. (1977). Short-term adaptation in single auditory-nerve fibers: Some post-stimulatory effects. *Journal of Neurophysiology*, **40**, 1098-1112.

STARK, R. E. (1980). Stages of speech development in the first year of life. In G. H. Yeni-Komshian, J. F. Kavanaugh, & C. A. Ferguson (Eds.), *Child phonology: Vol. 1. Production* (pp. 73-92). New York: Academic Press.

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

WARREN, R. M. (1985). Criterion shift rule and perceptual homeostasis. *Psychological Review*, **92**, 574-584.

WERKER, J. F., & LOGAN, J. S. (1985). Cross-language evidence for three factors in speech perception. *Perception & Psychophysics*, **37**, 35-44.

WHALEN, D. H. (1983). Vowel information in postvocalic fricative noises. *Language & Speech*, **26**, 91-100.

WHALEN, D. H. (1984). Subcategorical mismatches slow phonetic judgments. *Perception & Psychophysics*, **35**, 49-64.

NOTES

1. The latter authors reported a case of failure of perceptual equivalence, in both infants and adults, for the contributions of release burst and vowel length to perceived voicing of a final stop. However, these two acoustic cues do not result from a unitary phonetic gesture, and so would not be expected to show perceptual equivalence.

2. We thank Virginia Mann for loaning us her stimuli.

3. The authors also completed the test, but their data were not included in the final analyses.

4. We do not know why such an asymmetry should occur. However, perhaps if /l/ pulls /g/ forward, /g/ does not correspondingly pull /l/ back very far due to /l/'s fixed, and /g/'s sliding, place of articulation along the palate.

5. In this respect, we disagree with Kluender, Diehl, and Killeen (1987), who conclude that the Japanese quail's ability to categorize novel CV syllables on the basis of the initial consonant is not attributable to perceived articulation. ("On what basis do these quail correctly categorize new tokens? The possibility that their categorizations are based on a knowledge of articulatory commonalities can be excluded."—p. 1196) We would not be surprised to find that quail could categorize novel instances of active humans into the classes "walking" or bipedal "hopping"; moreover, if they could, we would presume that the categorizations were based on the perceived distal events of people either walking or hopping as those events are conveyed by information in reflected light to the eye. It seems to us no less plausible to suppose that quail can categorize novel instances of utterances into classes /d/-initial and /b/-initial, on the basis of the perceived distal events of vocal-tract-like systems producing those consonants as those articulations are conveyed by information in acoustic speech signals.

(Manuscript received March 19, 1990;
revision accepted for publication July 24, 1990.)