3

Sources of Variability in
Early Speech Development

Michael Studdert-Kennedy
*Queens College and Graduate Center, City University of New York,
and Haskins Laboratories*

This paper considers the origins of differences among children, and within a child from time to time, in the early development of speech. The paper is biased toward viewing these differences as special cases of general variability in animal behavior and its development. Some variability among children is surely genetic in origin (Lieberman, Chapter 4); this is the stuff of natural selection. Other variability is precisely what we expect in a system growing from an open genetic program (Mayr, 1974), that depends on loosely invariant properties of the environment to specify the course of development (for elaboration, see below, and for an excellent brief discussion, see Lenneberg, 1967, Ch.1). Finally, variability within a child is a precondition of the adaptive biological process that we term "learning" (Fowler & Turvey, 1978). However, I will come to all these matters only in the last section of the paper.

My first concern, and the topic of the early parts of the paper, is the apparent differences between capacities of infants and older children. Ferguson (Chapter 2) notes two main areas of research in child phonology: speech perception in infants and the sound systems of individual children aged 2–4 years, as shown by their speech productions. The relation between these two bodies of work is, indeed, "problematic," as Ferguson remarks. For, on the one hand, we have an infant apparently capable not only of discriminating virtually every adult segmental contrast with which it is presented, but also of discriminating speech sound categories across speakers and perhaps even across intrinsic allophonic variants (for a comprehensive review, see Aslin, Pisoni & Jusczyk, 1983). On the other hand, we have an older child producing a bewildering variety of sounds in its attempts to reproduce a particular adult word.

The discrepancy is not simply between perception and production. For we also find the older child, even up to the age of 5 or 6 years, making substantial numbers of perceptual errors on consonant contrasts (voicing, nasality, place of articulation) that would, seemingly, have caused no difficulty at all when it was an infant (see Barton [1980] for a review). Of course, these are cross-sectional comparisons. But the data are well established, and would usually be taken to reflect the child's course of development rather than sampling error.

How then are we to resolve the paradox? The first step is to acknowledge that different tasks place different demands on infant and older child: to detect the difference between two patterns of sound (discrimination) is not necessarily to recognize each pattern as an instance of a category (identification) (Barton, 1980, p. 106). Moreover, even when the tasks assigned to infant and older child are the same (i.e. discrimination), different behavioral measures may give different results: Recovery from habituation to a nonsense syllable upon presentation of a new syllable, as measured by high amplitude sucking or by heart rate, may not draw on the same capacities as choosing which of two nonsense words refers to a particular wooden block (Garnica, 1971). If we assume, as seems reasonable, that the older child has not lost capacities for discriminating between sounds of the surrounding language that it possessed as an infant, we must conclude that those capacities are not sufficient for more explicitly communicative tasks (Oller & Eilers, 1983; Oller & MacNeilage, 1983).

Yet the origin of the paradox is more than methodological. It also arises because infant speech research has "... generally taken for granted a phonological unit corresponding to the 'segment' [or, we may add, feature] of contemporary phonological theories, even though researchers have sometimes been familiar with the problems of relating such abstract units to the processes of speech perception" (Ferguson, Chapter 2). Ferguson has a different and, I believe, more fruitful approach. For rather than viewing the child as "acquiring" its phonology from the adult, Ferguson sees the adult's phonology as growing out of the child's (cf. Locke, 1983; Menyuk & Menn, 1979). Moreover, like Moskowitz (1973), and in accord with sound biological principle (Waddington, 1966), Ferguson sees this growth as a process of differentiating smaller structures from larger. The child does not build words with phonemes: phonemes emerge from words. In short, Ferguson shuns the preformationist view (long banished from embryology, but still thriving in psychology) that attributes adult properties to the child; he seeks rather to trace the epigenetic course from child to adult.

In the next few sections, I will sketch a view of infant speech development over roughly the first year of life that attempts to resolve the problematic relation between the apparent capacities of infant and older child. Broadly, my view is that two wrong turns have led into the impasse. First, a too

narrow notion of development has encouraged undue concentration on the infant's initial state. For the biologist, development begins with the first division of the fertilized egg and ends with death. At each moment, the organism is sufficient for adaptive response to current internal and external conditions. Birth is certainly an occasion of abrupt discontinuity and of radical changes in conditions, but prenatal and postnatal development do not differ in principle: The infant's state at birth is simply the first state that psychologists can conveniently study.

Of course, we may treat the whole process teleologically, seeing the end in the beginning. That, in my view, is the second wrong turn. For the habit of describing infants' presumed percepts (and articulations) in linguistic terms has diverted attention from the central problem of early speech development, namely, imitation. We have been easily diverted because it seems natural that, if an adult speaks a word or grasps the air with a hand, a young child can repeat the word or imitate the hand movements. But how, in fact, does the child do this? What information in the acoustic or optic array specifies the executed movements? How is the information transduced into muscular controls? We are far from even imagining an answer to the last question. But we may gain leverage on the former (the very question to which the infant, learning to speak, must itself find an answer), if we couch our descriptions in auditory and motoric, rather than in linguistic, terms. We begin then with a brief summary of what is known about speech perceptuomotor processes in adults.

CEREBRAL ASYMMETRY FOR LANGUAGE IN ADULTS

Brain lateralization offers a window through which we may view the early stages of imitative processes essential to language development. To justify this claim my first assumption is that the association between lateralizations for language and manual praxis in more than 90% of the human population (Levy, 1974) is not mere coincidence. Second, I assume that lateralization of hand control evolved in higher primates to facilitate bimanual coordination by assigning unilateral control to a bilaterally innervated system (Mac-Neilage, Studdert-Kennedy & Lindblom, ms.). Third, I assume that speech and language exploited the already existing neural organization of the left hemisphere to develop a characteristic structure, analogous in certain key respects to the structure of coordinated hand movements.

I have no space to develop the analogy here (for elaboration, see Mac-Neilage [1983] and MacNeilage, Studdert-Kennedy & Lindblom [1984]). In any case, for present purposes, the necessary assumption is simply that language evolved in the left hemisphere for reasons of motor control. The assumption is consistent with studies of aphasics (Milner, 1974), of split-brain

patients (Zaidel, 1978) and of the effects of sodium amytal injection (Borch-grevink, 1983; Milner, Branch & Rasmussen, 1964), showing that in most right-handed individuals the right hemisphere is essentially mute: The bilat-erally innervated speech apparatus is controlled from the left side.

My final assumption is that a capacity to perceive speech—more exactly, to break its patterns into components matched to the motor components of articulation—evolved alongside the motor system in the left hemisphere. The assumption is consistent with numerous studies of dichotic listening (Kimura, 1961, 1967; Studdert-Kennedy & Shankweiler, 1970), and has drawn further support from studies of split-brain patients. Levy (1974) showed that only the left hemisphere of these patients can carry out the phonological analysis needed to recognize written rhymes; Zaidel (1976, 1978) showed that, while the right hemisphere may have a sizeable auditory and visual lexicon, only the left hemisphere can carry out the auditory-phonetic analysis necessary to identify synthetic nonsense syllables or the phonological analysis necessary to read new words.

In short, the stated assumptions and their supporting evidence justify the claim that the speech perceptuomotor system is vested in the left hemisphere of most normal right-handed individuals. Let us turn now to the development of this system over the first year of life.

CEREBRAL ASYMMETRY FOR SPEECH IN INFANTS

Perception

A number of perception studies has demonstrated dissociation of the left and right sides of the brain for perceiving speech and nonspeech sounds at, or very shortly after, birth. For example, Molfese, Freeman, and Palermo (1975) measured auditory evoked responses, over left and right temporal lobes, of ten infants, ranging in age from one week to 10 months. Their stimuli were four naturally spoken monosyllables, a C major piano chord and a 250–4000 Hz burst of noise. Each stimulus lasted 500 msec and was presented about 100 times, at randomly varying intervals. Median amplitude of response was higher over the left hemisphere for all four syllables in nine out of ten infants, higher over the right hemisphere for the chord and the noise in all ten infants; the one child who responded to speech with higher right hemisphere amplitude had a left-handed mother. Molfese (1977) has reported similar asymmetries for syllables and pure tones in neonates.

Segalowitz and Chapman (1980) studied 153 premature infants with a mean gestational age at testing of 36 weeks. They measured reduction of limb tremor over a 24-hour period, at the end of a daily regimen of exposure to 5-minute spells of speech (the mother reading nursery rhymes) or music

(Brahms' "Lullaby"), presented six times a day at 2-hour intervals. Tremor in the right arm (but not in the right leg, nor in the left arm or leg) was significantly more reduced by speech than by music or by silence (control group). The mechanism of the effect is not understood, nor whether it is due to cortical or subcortical asymmetries.

Finally, Best, Hoffman and Glanville (1982) tested forty-eight 2-, 3- and 4-month old infants for ear differences in a memory-based dichotic task. They used a cardiac orienting response to measure recovery from habituation to synthetic stop-vowel syllables and to Minimoog simulations of concert A (440 Hz) played on different instruments. In the speech task, a single dichotic habituation pair (either [ba-da] or [pa-ta]) was presented nine times, at randomly varying intervals. On the tenth presentation, one ear again received its habituation syllable, while the other received a test syllable (either [ga] or [ka]), differing in place of articulation from both habituation syllables. An analogous procedure was followed in the musical note task. The results showed significantly greater recovery of cardiac response for right ear test syllables in the 3- and 4-month olds, and for left ear musical notes in all age groups. The authors suggest that right-hemisphere memory for musical sounds develops before left-hemisphere memory for speech sounds, and that the latter begins to develop between the second and third months of life.

Neither these nor any of the several other studies with similar findings (see Best et al., 1982 for a brief review) indicate what properties of the signal mark it as speech. We may note, however, that those properties are evidently present in isolated syllables, natural or synthetic, and do not depend on the melody or rhythm of fluent speech. Moreover, the results of Best et al. (1982) invite the inference that infant speech sound discrimination, attested by numerous studies, engages left-hemisphere mechanisms no less than does adult speech sound discrimination.

Production

Evidence for early development of the production side of the perceptuomotor link is tenuous, but suggestive. Kuhl and Meltzoff (1982) showed that 4- to 5-month old infants looked longer at the video-displayed face of a woman articulating the vowel they were hearing (either [i] or [a]) than at the same face articulating the other vowel *in synchrony*. The preference disappeared when the signals were pure tones matched in amplitude and duration to the vowels, so that infant preference was evidently for a match between mouth shape and spectral structure. Similarly, MacKain, Studdert-Kennedy, Spieker, and Stern (1983) showed that 5- to 6-month old infants preferred to look at the face of a woman repeating the disyllable they were hearing (e.g. [zuzi]) than at the synchronized face of the same woman repeating another disyllable (e.g. [vava]). In both of these studies, infant preferences were for natural

structural correspondences between acoustic and optic information. Since these two sources of information have a common origin in the articulations of the speaker, we may reasonably infer that the infant is sensitive to information that specifies articulation. (For related work on adult "lip-reading," see Campbell & Dodd, 1979; Crowder, 1983; McGurk & MacDonald, 1976; Summerfield, 1979).

Two more items complete the circle. First, Meltzoff and Moore (1977) showed that 12- to 21-day old infants could imitate both arbitrary mouth movements, such as tongue protrusion and mouth opening, and (of interest for the development of manual sign languages) arbitrary hand movements, such as opening and closing the hand by serially moving the fingers. Here mouth opening was elicited without vocalization; but had vocalization occurred, its structure would necessarily have reflected the shape of the mouth. Kuhl and Meltzoff (1982) do, in fact, report as an incidental finding of their study that 10 of their 32 infants "... produced sounds that resembled the adult female's vowels. They seemed to be imitating the female talker, 'taking turns' by alternating their vocalizations with hers" [p. 1140]. Of course, we have no indication that this incipient capacity, demonstrated under conditions of controlled attention in the laboratory, is actively used by 5-month old infants in the more variable conditions of daily life.

The second item of evidence is a curious aspect of the study by MacKain et al. (1983), cited earlier: Infant preferences for a match between the facial movements they were watching and the speech sounds they were hearing were statistically significant only when they were looking to their right sides. Fourteen of the eighteen infants in the study preferred more matches on their right sides than on their left. Moreover, in a follow-up investigation of familial handedness, MacKain and her colleagues learned that six of the infants had left-handed first- or second-order relatives. Of these six, four were the infants who displayed more left-side than right-side matches.

These results can be interpreted in the light of work by Kinsbourne and his colleagues (e.g. Kinsbourne, 1972; Lempert & Kinsbourne, 1982). This work suggests that attention to one side of the body may facilitate processes for which the contralateral hemisphere is specialized. If this is so, we may infer that infants with a preference for matches on their right side were revealing a left hemisphere sensitivity to articulations specified by acoustic and optic information. Thus, we have preliminary evidence that 5- to 6-month old infants, close to the onset of babbling, already display the beginnings of a speech perceptuomotor link in the left hemisphere.

Here we should strike a note of caution. The evidence reviewed up to this point does not demonstrate that specialized phonetic processes are occurring in the infant. In fact, whatever mechanisms for imitating articulation may be developing in these early months seem to be no different, in principle, than corresponding specialized mechanisms for imitating movements of hand, face,

and body. What distinguishes the speech perceptuomotor link at this stage of development is first, its locus in the brain, and second, its modality. The capacity to imitate vocalizations seems to be peculiar to certain birds, certain marine mammals, and man.

SPEECH PERCEPTION IN INFANTS

0-6 months[1]

As is well-known, infants in the first 6 months of life discriminate almost any adult segmental contrast on which they are tested. Particularly striking, in the early years of this work (initiated by Eimas and his colleagues [Eimas, Siqueland, Jusczyk & Vigorito, 1971]) was 1- and 4-month old infants' discrimination of synthetic syllables along a stop consonant voice-onset time continuum. Discrimination was measured by recovery (or no recovery) of high-amplitude sucking on a nonnutritive nipple, in response to a change in sound (or no change for a control group), after habituation to repeated presentation of another sound. Like adults, infants readily discriminated between acoustically different items belonging to different (English) phonetic categories, but not between acoustically different items belonging to the same category. This finding, fortified by similar results on continua of stop consonant place of articulation (Eimas, 1974), consonant manner (Eimas and Miller, 1980 a, b), and the [r]–[l] distinction (Eimas, 1975) for example, encouraged the hypothesis that "... these early categories serve as the basis for future phonetic categories" (Eimas, 1982, p. 342).

However, there is a confusion here between two different types of category. On the one hand, we have categories comprising more-or-less random variations in the precise acoustic properties of a single syllable, spoken repeatedly with identical stress and at an identical rate by the same speaker: These are the patterns mimicked by a synthetic series, varied along a single acoustic dimension. On the other hand, we have the categories of natural speech, comprising intrinsic allophonic variants, formed by the execution of a particular phoneme in a range of phonetic contexts, spoken with varying stress, at different rates, and by different speakers. The latter are presumably the "future phonetic categories" to which Eimas refers, while the former are auditory categories to which infants, chinchillas (for VOT, Kuhl & Miller, 1978), and macaques (for place of articulation, Kuhl & Padden, 1983) have

---

[1] The periods used here are not fixed stages of development. They are simply convenient headings that correspond roughly to a period before babbling (0–6 months) on which much infant perceptual researcher has focused, and a period of babbling (7–12 months) on which there has been little perceptual research.

been shown to be sensitive in synthetic speech studies (see also Kuhl, 1981). The proper interpretation of these studies would seem to be that infants (and an open set of other animals) can discriminate the several contrasts tested, if they are presented in an invariant acoustic context.

Evidence for phonetic categories from studies of contrasts across varying acoustic contexts differs depending on the nature of the variation. Talker variations, at least on the few contrasts that have been tested, seem to cause little difficulty for infant (Hillenbrand, 1983; Kuhl, 1979a), dog (Baru, 1975), cat (Dewson, 1964), or chinchilla (Burdick & Miller 1975). Cross-talker categories, then, seem to be auditory rather than phonetic. (We may note, in passing, that such findings present a puzzle for accounts of speaker normalization that rest on the listener's presumed knowledge of the speaker's phonetic space [Gerstman, 1968; Ladefoged & Broadbent, 1957]).

Studies of contrasts across variations in phonetic context have given less consistent results. Warfield, Ruben, and Glackin (1966) trained cats to discriminate between the words *cat* and *bat*, but found no transfer of training to other minimal pairs beginning with the same segments. Holmberg, Morgan, and Kuhl (1977) studied fricative perception in 6-month old infants. They used an operant head-turning paradigm, in which the infant was conditioned to turn its head for visual reinforcement when repeating sounds from one category were changed to repeating sounds from another. They found that infants discriminated [f]/[θ] and [s]/[š] across variations in vowel context (e.g. [fɑ], [fɪ], [fu]) and syllable position (e.g. [fɑ], [ɑf]). Kuhl (1980) reports similar results for an infant trained to discriminate [d]/[g].

Katz and Jusczyk (1980), cited in Jusczyk (1982), reasoned that a more stringent test of infant phonetic categorization would be to show that infants more readily learn to discriminate between (that is, to generalize within) phonetically-based groupings than arbitrary groupings of the same syllables. In a head-turning study of 6-month old infants, they found that most infants learned to discriminate between sets of syllables, paired for consonant onset, but differing in vowel (e.g. [bi] and [bɛ] versus [di] and [dɛ]), but not between sets, arbitrarily paired, differing in both consonant and vowel (e.g. [bɛ] and [di] versus [bi] and [dɛ]). However, none of the infants learned to discriminate either phonetic or arbitrary groupings of [b] and [d] followed by four vowels ([i, ɛ, o, ʊ]). Jusczyk (1982) interprets the results as providing some "... weak support for ... perceptual constancy for stop consonant segments occurring in different contexts" [p. 378].

Before commenting on this study, let us compare its results with those of Miller and Eimas (1979), who used a similar set of stimulus materials to ask a different experimental question: Are infants sensitive to the structure of syllables? That is to say, do infants perceive syllables holistically, as seamless, undifferentiated patterns, or do they perceive the structure of syllables, analyzing them into their component segments (consonants and

vowels)? Miller and Eimas used a high-amplitude sucking paradigm to test
2-, 3- and 4-month old infants. One group of infants successfully discrimi-
nated between sets of syllables, paired for consonant onsets, but differing in
vowel ([b a] and [b æ] versus [d a] and [d æ]), as did the infants of Katz and
Jusczyk. However, another group also discriminated between sets arbitrarily
paired, differing in both consonant and vowel ([b a] and [d æ] versus [b æ] and
[d a]), which the infants of Katz and Jusczyk did not do. Miller and Eimas
interpreted their positive outcome as evidence that infants are sensitive to
the segmental structure of syllables.

A similar conflict in results emerges at a "feature" level when we com-
pare a study by Hillenbrand (1983) with the second and third experiments of
Miller and Eimas (1979). Hillenbrand used a head-turning paradigm to test
the capacity of 6-month old infants to discriminate between sets of syllables
differing on a single feature (oral–nasal, as in [b a] and [d a] versus [m a] and
[n a]) and sets of syllables differing on arbitrary combinations of two features
(oral–nasal and place of articulation, as in [b a] and [ŋ a] versus [n a] and [g a]).
He found that infants were significantly better at discriminating the single
feature phonetic groups than the arbitrary double feature groups. He con-
cluded that infants were sensitive to the auditory correlates of consonantal
features. Miller and Eimas (1979), on the other hand, tested 2-, 3- and 4-
month old infants with a high amplitude sucking procedure on single-feature
phonetic groups analogous to those of Hillenbrand (voicing versus place of
articulation; oral–nasal versus place of articulation), and on the correspond-
ing double-feature sets where the two features were arbitrarily combined.
Pooling data from the two experiments, they found that infants assigned to
experimental conditions displayed significantly more recovery from habitu-
ation than control infants, and that there was no significant difference in
recovery for the two types of syllable set. They concluded from the lack of
reduction in performance across set types that infants were sensitive to the
structure of consonantal segments, that is, to their particular combinations
of features.

We have then a conflict in data from the three studies: 2- to 4-month old
infants, tested with high amplitude sucking, discriminate between arbitrary
sound classes that are indiscriminable for 6-month old infants, tested with
operant head-turning. If the results are valid, and not mere sampling error,
we have a paradox similar to that for infants and older children with which
we began. We may resolve the paradox on the same two fronts. Method-
ologically, we must acknowledge a commonplace of psychophysical testing
for many years (e.g. Woodworth, 1938, Ch. 17): Different behavioral mea-
sures may give different results, even in the same individual, at roughly the
same time. Moreover, since demonstrating a capacity takes precedence over
demonstrating its absence, and since 6-month old infants are unlikely to have
lost capacities for discriminating among the sounds of the surrounding lan-

guage that they possessed at 3 months, we must conclude that high-amplitude sucking is a more sensitive measure of infant discriminative capacity than operant head-turning. Thus, the two head-turning studies failed to reveal infant conditioning to arbitrary groupings of syllables because task difficulty and behavioral measure interacted—a possibility raised by Jusczyk (1982, p. 379).[2] The attempt to develop a more stringent test of infant consonant categorization across vowel contexts than that used by Holmberg, et al. (1977) for fricatives was therefore not successful.

Beyond the methodological issue lies the matter of interpretation. Consider, first, the conclusion from Miller and Eimas (1979) that infants are sensitive to the segmental structure of syllables and the featural structure of segments. Unfortunately, the conclusion is not forced by the data, since, as Aslin et al. (1983) point out, an infant discriminating, say [ba] and [na] from [da] and [ma], has simply to detect that one (or both) of the syllables in the second set is different from either of the syllables in the first set. In other words, the infant can discriminate the patterns holistically without analysis. Miller and Eimas (1979) recognize this fact ("... we know of no way to make this distinction [holistic/analytic] experimentally with infant subjects"), but justify their preference for the analytic interpretation, because "There is ... rather extensive behavioral as well as neurophysiological evidence for an analysis into components or features in human and non-human pattern perception" [both quotations from p. 355, footnote 2]. I do not doubt this evidence, but it does not justify our attributing analytic capacities to the 3-month old—particularly when, by doing so, we create a paradoxical discrepancy between the capacities of infant and older child.

Consider, next, the evidence that infants can form "phonetic" categories across a variety of acoustic contexts. Here again the data are overinterpreted. Since every phonetic contrast is marked by an acoustic contrast (if it were not, how would the infant learn to talk?), phonetic and auditory perception cannot be dissociated in the infant (though they can be in the adult: Best, Morrongiello & Robson, 1981; Best & Studdert-Kennedy, 1983; Liberman, Isenberg & Rakerd, 1981; Mann & Liberman, 1983; Schwab, 1981). This fact is recognized by Miller & Eimas (1979, p. 365), and by Aslin, et al., (1983). What we are left with then is evidence that infants, in their first 6 months of life, can detect auditory similarities across certain adult phonetic categories. Incidentally, apart from the study of cats mentioned above (Warfield, et al.,

---

[2] This interpretation assumes that arbitrary groups were, in fact, more difficult to discriminate than "phonetic" groups. Perhaps it is easier to detect a difference between groups, if all members of one group differ from all members of another group on the same dimension ("phonetic") than if each member of one group differs from each member of another on a different dimension (arbitrary). The difference in task difficulty might then be great enough to show up, if the criterial response is itself relatively difficult (head turning), but not if the response is relatively easy (high amplitude sucking).

1966), we have no evidence, so far as I know, that other animals cannot do the same. Of course, proving the null hypothesis for animals is a thankless task.

Finally, we may ask what role categories, whether auditory or phonetic, are presumed to play in the infant's learning to speak. Eimas (1982) argues that "... the acquisition of the complex rule systems of linguistics requires that the young child treat all instantiations of a phonetic category as members of a single equivalence class" [p. 346]. He adds in a footnote, "... if the child treats each possible member of the two voicing categories of English as separate entities and not as perceptually identical events or at least as members of the same equivalence class, then acquisition of the rule for pluralization will necessarily be painfully slow, if ever learned" (p. 346, footnote 5). Eimas goes on to justify the search for perceptual constancy in infants on grounds of parsimony, because "... it would effectively eliminate explanations based on receptive experience" (p. 346).

There are several things wrong here. First is the implication that accounts of biological development calling for experience to direct its course are somehow not parsimonious, perhaps even not "biological." In fact, just the reverse is true. Precisely because full genetic specification is costly, even the lowliest behaviors of non-human animals may depend on broadly invariant external conditions to guide development (see Immelmann, Barlow, Petrinovich & Main, 1981; Lenneberg, 1967, Ch. 1; Mayr, 1974; and the brief discussion below). Second, the notion of rule is prescriptive, as though speakers applied rules much as they do in a game of chess. In fact, a phonological rule is simply a description of regularities in speech; the processes by which these regularities arise are completely unknown (for discussions, see Menn, 1980; Menyuk & Menn, 1979). Finally, once again, the outcome of development (the formation of phonological structures that control adult speaking) is posited to be already in place at a time when development has scarcely begun. I do not doubt that infants can form auditory categories, but there is no evidence that this capacity is either needed for or brought to bear on early speaking.[3] If it were, we would be hard put to explain the word-by-word development of adult phones that Ferguson (Chapter 2) describes, or the relatively slow accumulation of the first 50 (or so) words. We may indeed suspect that the emergence of auditory-motoric categories, around the beginning of the third year, is a factor in triggering the explosive growth of the child's vocabulary

---

[3] Jusczyk (1982) makes the same point, proposing the "...possibility [that]...recognition of phonetic identities is not achieved until the child is engaged in learning how to read" (p. 365, footnote 3). If "recognition" here means "metalinguistic awareness," Jusczyk may be right. But functional categories surely predate the alphabet, both ontogenetically and historically. The alphabet (like dance notation) can only succeed because its units correspond to functional units of perceptuomotor control. The task for the child, learning to read, is to discover these units in its own behavior.

(at an average rate of perhaps 5–10 words a day) over the next 4 or 5 years (Miller, 1977, pp. 150 ff.).

In short, we can resolve the paradoxical discrepancy between the capacities of infants and older children if we refrain from regarding precursors of a behavior as instances of the behavior itself. No doubt, infant kicking (and stepping when the infant is held erect) are precursors of walking and, with normal growth in an appropriate environment, will develop into walking (Thelen, 1983). But infant kicks and steps are not strides.

7–12 months

None of the foregoing should be interpreted as claiming that phonetically relevant development of the infant's perceptual system is not going forward during the first 6 months of life. However, the first (and still sparse) behavioral evidence of such development comes from older infants.

Eimas (1975) showed that 4–6 month old English infants discriminated between English [r] and [l]. On the assumption that Japanese infants would have done the same, and given the well-known fact that native Japanese speakers, who know no English, do not make this discrimination (Miyawaki, Strange, Verbrugge, Liberman, Jenkins & Fujimura, 1975), Eimas suggested that learning the sound system of a language may entail loss of the capacity to discriminate contrasts not used in the language. Similar suggestions have been made by Aslin and Pisoni (1980), and Locke (1983).

Werker, Gilbert, Humphrey & Tees (1981) have traced the onset of perceptual loss to the second 6 months of life, a period when the infant is perhaps first attending to individual words and the situations in which they occur (Jusczyk, 1982; MacKain, 1982). Their initial finding was that 7-month-old Canadian English infants, tested in a head-turning paradigm, could discriminate between naturally spoken contrasts in Hindi as English-speaking adults could not. Werker (1982) followed this up by tracking the decline of discriminative capacity in cross-sectional and longitudinal studies. She used a conditioned head-turning paradigm to test three groups of infants on two non-English sound contrasts: Hindi voiceless, unaspirated retroflex versus dental stops (cf. Locke, 1983, pp. 90–92), and Thompson (Interior Salish, an American Indian Language) voiced, glottalized velar versus uvular stops. On the Hindi contrast, the number of infants successfully discriminating were: 11/12 at 6–8 months, 8/12 at 8–10 months, 2/10 at 10–12 months; for the Thompson contrast the results were essentially the same. (An infant was classified as having failed to discriminate only if it had successfully discriminated an English contrast both before and after failure on a non-English contrast). Finally, Werker (1982) reports longitudinal data for six Canadian English infants on the same two non-English contrasts. All six discriminated both contrasts at 6-8 months, but at 10-12 months none of them made the

discrimination. By contrast, the one Thompson and two Hindi infants so far
tested at 10-12 months could all make the called for discrimination in their
own language.

Perceptual loss is not permanent, since capacity can be recovered by adults
learning a new language (e.g. MacKain, Best & Strange, 1981). Nor can the
effect be general, since sufficiently salient foreign contrasts can presumably
be discriminated even by adults. We may suspect then that loss is focused on
relatively fine auditory contrasts, specifying slight differences in the space-
time coordinates of a single articulator's movements, and that it arises as
a side effect of the infant's developing "attention" to closely related con-
trasts in its own language. This is not to suggest that the younger infant
is not "attending" to speech during its early months. Rather, its search for
meaning and communicative function (Trevarthen, 1979) may initially be
guided by the rhythm and melody of speech (Mehler, Barrière and Jasik-
Gerschenfeld, 1976). Only when these larger patterns have begun to take
form (Menn, 1978a) are the infant's capacities for segmental discrimination,
readily demonstrated in the laboratory, brought to bear on the speech it
hears at home.

SPEECH PRODUCTION IN THE INFANT

The infant, by definition, does not speak (Latin: *infans*, not speaking).
But there is now ample evidence that the discontinuity between babble and
speech, posited by Jakobson (1968), is not real. Oller (1980) provides a
taxonomy of the emerging stages from phonation (0–1 month) to variegated
babbling (11–12 months). Oller, Wieman, Doyle & Ross (1975) describe
similarities between patterns of babbling and early speech (cf. MacNeilage,
Hutchinson & Lasater, 1981). Vihman, Macken, Miller, Simmons, & Miller
(in press) demonstrate parallels in the distribution and organization of sounds
in speech and babble during the period (roughly 9–15 months) when they
overlap.

What is the origin of this continuity? The first possibility is that the sound
distributions of babble and early speech are similar because the infant begins
to learn the sounds of the language around it and to practice them during its
second 6 months of life. Locke (1983, Ch. 1) has marshalled evidence against
this view. First, he has collated data on the babbling of 9- to 12-month old
infants growing up in 14 different language environments, distributed across
some half dozen language families (Locke 1983, Table 1.3, p. 10). These in-
fants were certainly old enough to have begun to discover the sound patterns
of their languages and, indeed, if the data on perceptual loss reviewed above
have any generality, perceptual discovery had already begun. Yet of the 143

consonantal sounds entered in Locke's table, over 85% correspond to one of the twelve most frequent sounds in the babbling of English children, a strikingly homogeneous distribution. Second, Locke has reviewed some dozen studies that have looked for drift in the sounds of infant babbling, during the second 6 months of life, toward the sounds of the surrounding language. Most of the studies either found no evidence of drift or were inconclusive. Finally, Locke has reviewed available studies on the babbling of deaf infants and infants with Down's syndrome. Despite the common belief that deaf babbling fades before the end of the first year, several studies agree that it may continue well into early childhood (5 6 years). But what is remarkable is that the developmental course of babbling up to 12 months is similar in deaf and hearing infants, and, incidentally, in Down's syndrome infants. For example, the relative proportions of labial, alveolar, and velar consonants follow essentially the same course: Only after the 12th month does the expected preponderance of labial movements in deaf children begin. The three strands of evidence converge on a process of articulatory development, independent of the surrounding language and common to all human infants.

We are left, then, with the second possible account of the continuity between babble and speech, namely that, as Locke proposes, the phonetic proclivities of adults and infants are similar. Both are largely determined by anatomical and physiological constraints on the signaling apparatus. What these constraints may be has only recently come under scrutiny (Kent, 1980; Lindblom, 1983a; Ohala, 1983b).

Of course, this hypothesis immediately raises the question of language change: If all adult speakers develop from a common infant base, why do languages differ? The question is too large, and my competence too small, for adequate treatment here. However, I note several points. First, as Locke (1983) has shown, many infant biases (e.g. for open rather than closed syllables, for stops over fricatives, for singleton consonants over clusters, and so on) are indeed preserved by many groups of adult speakers (i.e., languages); it is this fact that the continuity of babble and speech reflects. At the same time, infant preferences are not rigid, because, as Darwin taught, no animal structure specifies a unique function. A structure (e.g. the vocal apparatus) permits an unspecifiable, though presumably limited, range of functions, and the natural variability of behavior offers this range for selection. Second, infant articulatory capacities are a subset of the capacities of mature speakers. As skill develops, the range of response, available for selection by a variety of sociocultural forces, widens. Certainly, the exact course of historical change will never be fully specified for language, any more than for, say, clothing, cuisine, or social organization. Nonetheless, there would seem to be no reason, in principle, why we should not develop a cultural–evolutionary account of language diversity (Lindblom, 1984), compatible with relatively fixed infant articulatory proclivities.

In short, perceptual and motor development of speech over the first year of life, as manifested in infant behavior, may justly be seen as parallel, independent processes. No doubt, physiological changes in the perceptual and motor centers of the left hemisphere are taking place to prepare for the ultimate connection between the two systems. These processes may be analogous to those in songbirds, such as the marsh wren, in which the perceptual template of its species' song is laid down many months before it begins to sing (Kroodsma, 1981). But behavioral evidence of the perceptuomotor link appears only with that song, just as behavioral evidence of the link appears in the infant only with its first imitation of an adult sound.

FROM BABBLING TO SPEECH

The transition from babbling to speech is a murky period. At this stage we see the first clear evidence of a perceptuomotor link, but know little about what the child perceives. Even when the perceptual data come in, it will be a delicate task to determine their relevance. For, as we have noted, a capacity demonstrated in the laboratory does not tell us how, or even if, that capacity is put to use in learning to speak. Consequently, we may have to place as much weight on shaky inference from the child's productions as on firm evidence from perceptual studies.

At this stage, we also find it increasingly difficult to refrain from describing the child's productions by means of phonetic transcriptions. Of course, we do not want to refrain: Transcription is our readiest mode of description, because children have vocal tracts very like adults' and make sounds like adults' sounds. Yet transcription is a double-edged blade. For it is precisely in order to understand the apparently segmented structure of speech (and the resulting adult capacity to transcribe) that we are studying its ontogeny. As is well known, phonetic segments are not readily specified either in articulation or in the signal, so that their functional reality has had to be inferred, in the first instance, from adult behaviors, such as errors of perception (Browman, 1978) and production (Shattuck-Hufnagel, 1983), backward talking (Cowan, Leavitt, Massaro & Kent, 1982), aphasic deficits (Blumstein, 1981) and, not least, use of the alphabet. By relying on a descriptive apparatus that derives from characteristics of mature speakers, we put ourselves in danger of attributing to the child properties it does not yet possess.

Despite these difficulties, headway has been made, and a view of the child as something other than a preformed adult is beginning to emerge (see especially Menn, 1978a, b, 1980, 1983; Menyuk and Menn, 1979). A striking aspect of this view, though not a surprising one, is the lavish variability of the child's productions. In these last few paragraphs, I will briefly consider how we might approach this variability.

Variability within a child

Ferguson (Chapter 2) presents compelling arguments for regarding the word as the unit of contrast in early speech; he defines a word as "... any apparently conventionalized sound-meaning pair." The definition is important, because it draws attention to the fact that a word is not simply a pattern of sound, but a pattern of sound appropriate to a particular situation (Menyuk and Menn, 1979). To discriminate one word from another, to recognize a word, and to use it correctly, entails discriminating and recognizing various nonlinguistic properties of a situation. Thus, a child's failure to discriminate or recognize a word in a perceptual test may reflect nonlinguistic as much as linguistic factors. Moreover, many of the child's spoken variations may reflect variability in the situations in which the child has heard the word and in the varying salience of its phonetic properties in those situations. The same adult word may then be a different word to the child in different situations.

Nonetheless, highly variable productions of a given word do occur within essentially the same situation. Ferguson (Chapter 2; Ferguson & Farwell, 1975, p. 423, footnote 8) lists ten different attempts by a child (K at approximately 1 year, 3 months) to say *pen* within one half-hour session. Ferguson comments, "She seemed to be trying to sort out features of nasality, bilabial closure, alveolar closure, and voicelessness." Waterson (1971) describes numerous such instances for her child, P, in similar phonetic terms, noting as a common occurrence that "features" lose their order and become recombined into patterns quite unlike the adult model.

Perhaps, however, we would do well to avoid featural terminology. We might attempt a more direct articulatory description as do Menyuk and Menn (1979), describing one of Menn's (1978a) subjects Jacob's protowords: "... Jacob was varying the timing of front-back articulations against the timing of lowering and raising the tongue" [p. 61]. Of course, this is little more than a gloss on phonetic transcriptions. Yet, in the absence of cineradiographic or even acoustic records, the gloss may "... help us see more clearly what it is the child needs to learn and to look at it in a way less coloured by our knowledge of mature linguistic behavior" (Menyuk & Menn, 1979, p. 61; cf. Kent, 1984). For we then see the speaking of a word not as a bundling of features into concatenated segments, but as a distribution of interleaved movements of articulators over time (Browman & Goldstein, ms.). In the adult, repeated coordination of particular movements in recurrent patterns has crystallized into structures that form the phonological elements of the language. For the child the movements have yet to be organized.

Three points deserve emphasis. First, despite the variability of a child's productions, they also display surprising accuracy. The phone classes of Ferguson and Farwell (1975) show much variability in voicing and manner—due perhaps to unskilled timing of closure and release—yet remarkable homo-

geneity in place of articulation. Also, K's attempts at *pen* did not include, for example, [gʌk]: Almost every attempt included some recognizable property of the adult word. This means that the acoustic structure of adult words specifies for the child at least some rough pattern of configurations of the vocal tract, necessarily the product of a specialized perceptuomotor link. Yet, second, the link is not precisely predetermined: It must develop. Not only the movements, but their relative timing and sequencing must develop. These are complex processes that almost certainly require active movement for their neural control structures to take form. Perhaps it is the normal function of babbling to promote growth of these structures in the left hemisphere. In any event, we are now led to see, and this is my third point, that genetically programmed variability is a condition of the child's learning to speak. In general, the longer the life span of an animal, the longer the period of parental care, and the more complex the mature behavior, the more likely is the behavior to develop through an open genetic program (Mayr, 1974; though, for an exception, see below). Such a program relies on experience to select and, if necessary, shape the needed behavior from a reservoir of variable responses (Fowler & Turvey, 1978).

Variability among Children

As earlier noted, some individual differences in the course of development are genetic or congenital in origin. MacKain (in press) describes several extreme cases of children born without a tongue who approach a surprisingly normal phonetic repertoire by an idiosyncratic path of development. Yet other differences arise from the plasticity of an open system, sensitive to environmental contingencies and equipped with a variable repertoire of responses. Adaptive response to some particular, short-term aspect of the environment may lead an individual down an idiosyncratic path, because the precise order in which the parts of the system assemble themselves is not preordained. Here we may draw a useful analogy with the self-stabilizing processes in embryological development termed "canalization" (Waddington, 1966, p. 48). Waddington describes how various regions of an embryo differentiate into eyes, arms, legs, and so on. Each region has many possible paths to the same end. The exact path is determined, in part, by chance factors in the embryonic environment; equifinality is assured by fixed constraints inside and outside the developing region. Similarly, we may suppose, no single path is prescribed for the development of a phonological system. Many paths, determined by partially fixed, partially variable perceptual, motoric, and social conditions lead to the same end (Lindblom, MacNeilage, & Studdert-Kennedy, 1983).

Certainly, there may be a "normal" path, the product of articulatory proclivity or "ease" (Locke, 1983) and perceptual salience. But a child can

readily be diverted from the path by accidents of the speech it hears or of its physical structure and growth.

For example, if final fricatives become salient for a particular child, due to adult lexicon in some recurrent situation, the child may try them and be successful, yet be unable (through lack of consonant harmony in the target word or other "output constraints" [Menn, 1978b]) to execute the initial consonants of the words. A vowel-fricative routine is then established which the child can bring to bear on words that most children would attempt with the standard stop-vowel sequence, followed by a "deleted" fricative (e.g. Waterson, 1971, p. 185). Yet the deviant child will ultimately come upon the same phonological system as its peers.

Here we should note that even quite simple behaviors in nonhuman animals may develop through an open genetic program. The filial and sexual imprinting of mallard ducklings or domestic chicks on slow-moving objects (such as a walking human, or even a red plastic cube revolving on the arm of a rotary motor [Vidal, 1976]) is well known. The effect is possible because genetic "instructions" are loose; they do not specify the form and color of the mother bird, but only her typical rate of movement. Evolution can afford such imprecision because the normal environment provides the duckling with only one slow-moving object, its mother. If the combination of gross genetic "instructions" and a more or less invariant environment permits essential functions (here, protection from predators and species identification) to develop, there will be no selective pressure for more exact genetic specification.

For the imprinting of precocial birds, the behavior is roughly fixed, while eliciting conditions are only loosely specified. For the development of language, both the behavior and the eliciting conditions are loosely specified.[4] Presumably, the infant has certain minimal, perhaps quite general, capacities (its "initial state"), including sensitivity to the contingencies of its own behavior, the basis perhaps of social responsiveness (Watson, 1972, 1981), while the social environment normally offers the infant certain more-or-less invariant invitations to interact. So, within weeks of birth we find the infant watching intently its mother's eyes, face, and hands, as she talks and plays, and we detect certain inchoate communication patterns in postures of the infant's head, face, and limbs, and in prespeech movements of tongue and lips (Trevarthen, 1979). But at this stage, not even the modality of language is fixed. For if the infant is born deaf, it will learn to sign no less readily than its hearing peer learns to speak. Thus, the neural substrate is also shaped by environmental contingencies; and the left hemisphere, despite its predis-

---

[4] I am not proposing that language can take any arbitrary form. On the contrary, its general form, that is, its two-leveled hierarchical structure of phonology and syntax, emerges necessarily from its function. Innumerable details of form within these levels must result from more-or-less invariant perceptuomotor, cognitive and pragmatic constraints, of which we know, at present, very little.

position for speech, is then usurped by sign (Neville, 1980; Neville, Kutas & Schmidt, 1982; Studdert-Kennedy, 1983, pp. 175 ff. and pp. 219ff.). In fact, recent studies of aphasia in native American Sign Language signers show remarkable parallels in forms of breakdown between signers and speakers with similar left hemisphere lesions (Bellugi, Poizner & Klima, 1983).

The differences between deaf and hearing individuals are certainly gross. Yet every child grows in its peculiar niche with its peculiar anatomical and physiological biases, and must therefore discover its own "strategy" for fulfilling the human communicative function. (The term "strategy" should be stripped of its cognitive, not to say military, connotations in this context, as it is in standard ethological usage.) Indeed, language, as a sociobiological system, exploits the potential for diverse strategies to mark social groups by channeling speakers into distinctive linguistic styles and dialects—to which, of course, children are highly sensitive (e.g. Local, 1983). Thus, individual differences and individual adaptive response make language a force for social cohesion and differentiation. (For examples of stable diversity within species of bee, treefrog, anemonefish, ruff, and other animals, see Krebs & Davies, 1981, Ch. 8).

Finally, individual differences offer an opening for research. Presumably, there are limits on possible strategies. But what these limits may be we do not know. As data from longitudinal studies of individual children accumulate, strategies may cluster, until it is possible to sketch their limits. Such work may lead toward clearer notions of "perceptual salience" and "ease of articulation." Thus, we come back to the constraints on individuals by which phonological elements emerge and phonological systems organize themselves (Lindblom, MacNeilage & Studdert-Kennedy, 1983).

ACKNOWLEDGMENTS