CHALLENGES TO DEVELOPMENTAL PARADIGMS:

Implications for Theory, Assessment and Treatment

Edited by

PHILIP R. ZELAZO
RONALD G. BARR, M.D.

Montreal Children's Hospital
McGill University

LEA LAWRENCE ERLBAUM ASSOCIATES, PUBLISHERS
1989 Hillsdale, New Jersey Hove and London

5 Infants' Perception of Faces and Speech Sounds: Challenges to Developmental Theory

Andrew N. Meltzoff
Patricia K. Kuhl
University of Washington

It is no accident that some of the strongest challenges to theories of early development have come from research concerning infants' knowledge about faces and speech. Why is that so? Might faces and speech be privileged signals in the world of the infant?

For many years developmentalists presented infants with visual targets composed of two-dimensional, static, inanimate patterns—like checkerboards—and theories of psychological development were advanced on this basis. Similarly, infants were presented with simple auditory stimuli like clicks and tones. A motivation behind these studies, or at least an implicit assumption, was that infants' reactions to simple stimuli should be assessed first. Once we understood these responses, we would enrich things by presenting gradually more complex signals. Our scientific theories would develop in small steps—first cataloging simple sensory reactions, then progressing to the understanding of higher-order psychological processes—much like the infants themselves were thought to develop from sensory respondents to cognitive beings.

The research discussed in this chapter takes a complementary but very different approach. We conduct experiments with natural faces and voices with all their complexities. The results of this research show that young infants are capable of being more than simple sensory respondents. Even the youngest infants perceive, categorize, and act within a psychological framework that includes extensive knowledge about faces, speech sounds, and certain correspondences between the two. We believe that these complex signals have a privileged status in the perceptual world of the infant.

There are three good reasons for this privileged status. First, evolution may have prepared infants with finely articulated and innate reactions to these universal and communicatively important stimuli. Second, infants' caretakers, while not commonly presenting babies with checkerboards or pure tones, will constantly present them with faces and speech sounds. Experience with caretakers may influence infants' reactions. Third, and less often contemplated by theorists, infants themselves have faces and voices. The experience they have with their own bodies and the perceptual consequences of their own actions may contribute to a special knowledge about face/voice dynamics.

We have investigated infants' reactions to faces and voices by presenting them with tasks involving imitation, cross-modal perception, and the categorization of speech sounds. Each of these provide a different, but as we will show, complementary perspective on exploring the infant's world.

Imitation provides a window into the infant's mind, for it necessitates that the infant perceive an adult's production, "translate" the perceived act into an analogous one of his own, and execute a motor plan so that the output corresponds to the model. Furthermore, if infants can imitate targets after delays, this requires them to act on the basis of some memory or internal representation of the absent event (Meltzoff, 1985a, 1988a,b,c). In this chapter the phenomenon of facial imitation is discussed. The data show that infants in the first hours and days of life can duplicate the facial gestures of an adult experimenter, with substantial implications for our theories of man's "initial state," the starting point of human psychological development.

The ability to recognize equivalences in information from different senses is termed cross-modal perception, which is the topic of our second line of research. Without some capacity to integrate information across modalities we would not experience one unitary world, but five separate ones, one for each of our different sensory modalities. Cross-modal perception is not only highly adaptive, it forms the basis of some of our most pleasurable aesthetic experiences. Ballet is neither exclusively visual nor auditory, but relies, in Keats' words, on the fact that "eyes/And ears act with that unison of sense/Which marries sweet sound with the grace of form." Here, we discuss recent experiments examining cross-modal aspects of speech perception, highlighting a primitive capacity to "lip-read" that can be found in young infants.

Our third line of research concerns categorization, the proclivity to group similar things together, which is so basic to thought and language. Infants' ability to recognize similarities among discriminably different inputs establishes order out of what might otherwise be sensory chaos. Our concern is with infants' ability to organize speech into linguistically appropriate categories. We discuss research showing that infants can indeed "sort" sounds into categories, and furthermore that their groupings correspond to linguistic (phonetic) categories. The perception of phonetic categories in infancy is critical to the development of speech and language.

Each of these three topics in infancy—imitation, cross-modal perception, and speech-sound categorization—presents challenges to current theory and suggests new hypothèses for characterizing early perceptual-cognitive development. We discuss the theoretical background and findings in each of these domains and also delineate areas of convergence between them.

THE IMITATION OF FACIAL GESTURES

The imitation of facial gestures by infants has long been singled out as a special phenomenon by developmental theorists. It has classically been discussed as a developmental milestone that was first passed at about 1 year of age. The late onset of facial imitation was embodied not only in virtually all introductory texts written before 1977 but also in scales of infant development (e.g., Uzgiris & Hunt, 1975) that we in turn used in cross-cultural studies, with handicapped populations, and to predict language development and other significant cognitive achievements.

This broad consensus about the late emergence of facial imitation rested on three sources. These were to be found in the sensory literature, the Piagetian literature, and the learning literature—an uneasy alliance to be sure, but nonetheless all supported the classic view that facial imitation was a late-developing skill.

On the sensory side, the initial laboratory studies in the 1960s emphasized certain limitations in infants' visual behavior. One oft-quoted constraint was that infants in the first few months seemed to process information about the outline of a visual figure, but not its internal features (Milewski, 1976; Salapatek, 1975). This was dubbed the "externality effect." Because the imitation of facial gestures requires infants to perceive the interior features of a face, it was thought that visual constraints alone might make such behavior impossible in young infants.

However, the initial sensory research of the 1960s and early 1970s was conducted using static geometric figures. When an infant is presented with a repetitive mouth-opening or tongue-protrusion display, such as used by Meltzoff and Moore (1977, 1983a), the interior facial features are moving and of potential visual interest in themselves. Considered this way, there may not be incompatibility between the imitation results and the original sensory research. Indeed, since the 1977 demonstration of facial imitation, three studies of sensory development can be cited in support of this perspective. Both Bushnell (1979) and Girton (1979) showed that infants under 2-months-old fixated an internal element when it moved. Gannon and Swartz (1980) found that a sufficiently complex and attractive internal pattern was visually detected by infants as young as 1-month-old. It appears, then, that the presumed constraint of the externality effect might not actually be a constraint on facial imitation. This newer research has brought the work on sensory functioning and early imitation more in line with one another.

A second basis for thinking that young infants could not imitate facial gestures is derived from Piagetian theory. According to Piaget (1962), infants in the first months are capable of vocal and manual imitation but only slowly progress to facial imitation by 1-year-old-age. This prediction is essentially in line with common sense. The theory states that manual and vocal imitation are the initial forms of imitation, because infants can directly compare the adults' target and their own response within the same sensory modality. For example, if an adult opens and closes his hand, infants can relate this to actions they see themselves perform. A direct visual-visual comparison is possible between the adult model and the infants' response. Similarly, vocal imitation involves direct comparisons between the infant's vocalizations and the adult's sounds as monitored through the auditory modality.

Facial imitation is regarded as a more difficult task because both visual and auditory guidance is impossible. The infant can see the adult's target gesture, but he cannot see his own face. According to standard developmental theory, facial imitation ought to be more difficult than manual or vocal imitation, because the infants have no direct way to compare their own actions and the adults'. Facial imitation has often been called "invisible imitation" to denote that it involves a coordination of the seen and the unseen (Piaget, 1962). Piaget drew intriguing parallels between the development of facial imitation and object permanence. Both involve coordinating the visible and invisible, and it was an impressive prediction of his theory that these seemingly disparate achievements should both appear during the same developmental stage, at about 1-year-of-age, during sensorimotor stage 4. Albeit in a variety of vernaculars, numerous textbooks and journal papers have restated Piaget's basic position that facial imitation is impossible before about 1 year because an infant cannot see his or her own face.

The thesis that facial imitation appears in late infancy also gathered support from a third source, from classic learning theory. Like Piaget, learning theorists sought to explain imitation in infancy without recourse to an innate capacity for mimicry. Infants learned to pair a parent's gesture with a matched response in the same way that they learned to pair any other arbitrary stimulus (a buzzer) with a motor movement (head turn to the right). The similarity between the stimulus and response was not thought to facilitate this learning in any way (Skinner, 1953). Extending this model to early development, Gewirtz and Stingle (1968) made the prediction that "The first imitative responses must occur by chance, through direct physical assistance, or through direct training" (p. 379). In other words, the infant began with no capacity to imitate adults, and then gradually acquired particular imitative routines through adult shaping.

In 1977 Meltzoff and Moore conducted an experiment to test the notion that facial imitation was a late-emerging skill. The report pointed out the need for several methodological safeguards that had not been incorporated in previous studies of infant imitation. The experimental paradigm used by Meltzoff and Moore is described in detail elsewhere (Meltzoff & Moore, 1983b), but two controls are particularly noteworthy. First, they highlighted the need to distinguish imitation from a general arousal response. For example, they noted that imitation of a tongue protrusion could not be inferred simply by comparing the baseline frequency of infant tonguing to the rate of tonguing when a tongue protrusion gesture was demonstrated by the adult. Even if there were a significant increase in response to the adult gesture, this would not be sufficient to establish true imitation. An alternative, more parsimonious interpretation is that young infants become aroused when seeing a moving adult face and that increased tonguing is a by-product of this general arousal.

To distinguish imitation from general arousal, Meltzoff and Moore utilized a "cross-target comparison." They showed infants a variety of gestures in a repeated-measures design, with each gesture being presented by the same adult, at the same distance from the infant, and in the same temporal pattern. For example, infants were shown a tongue-protrusion gesture and then a mouth-opening gesture. The imitation hypothesis predicts significantly more infant tongue protrusion to the adult tongue-protrusion display then to the adult mouth-opening display. Conversely, it predicts significantly more infant mouth opening to the adult mouth-opening display than to the adult tongue-protrusion display. If this pattern of differential responding obtains, it cannot be accounted for by a general arousal response because, in both cases, the same face at the same distance is moving at the same rate.

A second methodological point concerns the scoring of the infant responses. Earlier studies of infant imitation attempted to score the infants' responses live, using the same experimenters both to present the gestural displays and classify the infants' responses. In Meltzoff and Moore's work infants were recorded using two video systems, one to record the infant's behavior and one the adult's. The scorers were shown only the videotaped record of the infant and remained naive to the adult display during scoring.

Using these and other controls, Meltzoff and Moore (1977) reported that infants in the first 2- to 3-weeks-of-life could imitate adult facial movements. In Study 1, four gestures were tested in a repeated-measures design: lip protrusion, mouth opening, tongue protrusion, and sequential finger movement. The infants' responses were videotaped and subsequently coded by observers who were able to judge at greater than chance accuracy the gesture the infant was mimicking. In this study the infant was free to respond while the adult presented the gestures. Study 2 extended this work by testing whether the infants could imitate if they were prevented from responding immediately. A design was needed in which infants were shown the target gestures, but not allowed to respond until after the experimenter was no longer modeling the gestures. The approach taken was to use pacifiers. A pacifier was put in the infant's mouth during the demonstration of the gesture. After the display stopped the pacifier was removed and the experimenter maintained a passive-face pose for a 2.5-minute response period. The pacifier was then reinserted and the experimenter followed the identical procedure with a second gesture.

Videotaped records showed that the infants sucked actively on the pacifiers during the stimulus-presentation period. Apparently the sucking reflex took precedence over the imitative response. The design was thus successful in prompting infants to delay their facial imitation. Even under these conditions there was evidence for imitation of the two displays used, mouth opening and tongue protrusion. The data showed significantly more infant mouth openings in response to the mouth-opening display than during a baseline period or to the tongue-protrusion display ($p < .05$ in both cases). Similarly, there was more infant tongue protrusion to the adult tongue-protrusion display than during baseline or after the mouth-opening display ($p < .005$ in both cases). The results of Study 2 extended those of Study 1 by showing that infants could imitate facial gestures even if they did not engage in imitative co-action during the display itself.

The initial prompt for using 2- to 3-week-old infants was that theory held facial imitation to be absent in this age group. However, given advances in our understanding of the intricacies of early mother-infant interaction (Brazelton & Tronick, 1980; Trevarthen & Marwick, 1986), staunch learning theorists could suppose that such behavior had been shaped during these episodes. If this were so, the apparent challenge to theory presented by early imitation would vanish. This debate, focusing as it does on the role of experience in the attainment of a particular skill, is not unfamiliar in psychology.

If postnatal interaction with parents is a necessary condition for imitation, then newborn infants in the first hours of life should fail on the task. To test this, Meltzoff and Moore (1983a) conducted a study of imitation using newborns. The study involved 40 infants with a mean age of 32 hours old. The youngest subject was just 42 minutes old at the time of test. The infants were shown both the mouth-opening and the tongue-protrusion gestures in a repeated-measures design. Pilot work revealed that infants were not maximally responsive if the experimenter repeatedly presented the display. The design therefore entailed an alternation between 20 seconds of gesturing, then 20 seconds of a passive face, and so on for a total of 4 minutes. After the first 4-minute period with one gesture, say mouth opening, the experimenter would switch to the opposite display, tongue protrusion, and follow the identical procedure for a second 4-minute period. The experimental periods were electronically timed and there were no breaks or pauses anywhere in the test.

The results provided strong support for the existence of imitation in newborns. There were significantly more infant mouth openings in response to the adult mouth display ($M = 7.1$) than to the adult tongue display ($M = 5.4$) ($p < .05$). Conversely, there were significantly more infant tongue protrusions in response to the adult tongue display ($M = 9.9$) than to the adult mouth display ($M = 6.5$) ($p < .001$).

We can conclude that infants have the capacity to imitate facial gestures well before the 1 year age period predicted by Piagetian theory. Despite the initial surprise of these findings and debate about the meaning of the phenomenon (see Meltzoff & Moore, 1983b, 1985 for a review), the basic fact that young infants can match adult facial displays has now been reported by researchers in at least six independent labs since the 1977 report (Abravanel & Sigafoos, 1984; Dunkeld, 1978; Field, Woodson, Greenberg, & Cohen, 1982; Fontaine, 1984; Heimann & Schaller, 1985; Jacobson, 1979). The basic finding that certain gestures elicit matching responses now seems secure. What does it imply about the psychology of young infants?

While recognizing that other interpretations are possible, the interpretation favored by Meltzoff and Moore was that infants are able to appreciate equivalences in body transformations they see and ones they feel themselves make. The infants are thought to engage in an active matching process in which information about the adult's gesture is registered by vision and information about the self is picked up through proprioception (Meltzoff & Moore, 1977, 1983a, 1983b, 1985). This implies that young infants have the ability to recognize cross-modal equivalences for form or movements across the different perceptual systems. Data from a number of laboratories, including our own, demonstrate that this is indeed the case (Bower, 1979, 1982; Meltzoff & Borton, 1979). We turn now to a specific example of this work, illustrating cross-modal functioning involving facial movements in a situation quite different from imitation.

CROSS-MODAL SPEECH PERCEPTION AND VOCAL IMITATION

Speech perception is classically considered a purely auditory phenomenon. Speech information is carried by sound waves. It is transduced by the ear and auditory system. Experiments are typically conducted using stimuli that are purely auditory in nature, so that the potential contribution that vision could make to the perception of speech has been largely unaddressed by developmental theorists.

Yet clinicians concerned with the hearing-impaired patient were aware that adults could "hear with the eye." Two lines of recent experimental work have now brought "lipreading" phenomena to the attention of theorists. The first is a series of studies showing that normal adult listeners in ordinary listening environments are strongly influenced not only by what they hear but by what they see a person's lips doing (Green & Kuhl, 1989; Green, Kuhl, & Meltzoff, 1988; Kuhl, Green, & Meltzoff, 1988; McGurk & MacDonald, 1976; Summerfield, 1979). The second finding involves infants rather than adults (Kuhl & Meltzoff, 1982; MacKain, Studdert-Kennedy, Spieker, & Stern, 1983). This work demonstrates that young infants can detect equivalences between the auditory pattern of speech and the visual pattern provided by articulatory movements. In a sense, young infants display a rudimentary form of "lipreading." These findings altered our opinions concerning the purely auditory nature of infant speech perception—we began to think of it as a cross-modal phenomenon (Kuhl & Meltzoff, 1982, 1984, 1988).

The fact that speech information can be conveyed by eye as well as by ear is manifest at cocktail parties and in other situations when we watch the lip movements of the speaker to improve our understanding of what the speaker is saying. Laboratory work with adults documents the substantial contribution visual information can make in noisy environments or when speech is degraded by filtering. Sumby and Pollack (1954) demonstrated that seeing the face of a talker whose speech was presented in noise was equivalent to providing a 15–20 decibel boost in the auditory signal. Similarly, Grant, Ardell, Kuhl, and Sparks (1985) showed that an otherwise unintelligible auditory signal became intelligible if the subject watched the face of the talker. In the Grant et al. study, the auditory signal was a simple pure tone that followed the pitch contour of a speaker's voice. In isolation the tone provided no linguistic information, but when subjects watched the face and listened to the tone, speech was perceived with 80% accuracy. Evidently vision can play a substantial role in speech perception, at least in adults.

The major goal of our infant research was to test whether young infants could recognize the correspondence between the visual and auditory manifestations of a speech act (Kuhl & Meltzoff, 1982). We wanted to test whether infants recognized that an /a/ vowel sound (as in "pop") corresponded to one articulatory gesture and that an /i/ sound (as in "peep") corresponded to another articulatory gesture. In order to test this we placed infants within a three-sided enclosure (Fig. 5.1). A film of two faces articulating the vowels was projected onto the front wall of the enclosure. One face was articulating the /a/ vowel, and the other the /i/ vowel. The two faces were life-sized and in color. They were specially filmed and edited so that they would articulate in perfect temporal synchrony with one another. Each vowel sound was approximately 1 sec in duration, with a silent interval of approximately 2 sec between sounds. The auditory vowel sounds, either the /a/ or the /i/, were presented from a loudspeaker placed midway between the two faces.

This experimental set-up allowed us to rule out two possible bases that infants might use to detect an auditory-visual match: (a) The central placement of the loudspeaker ruled out any spatial cues, because the sound source was equidistant from each face. (b) The precise temporal synchrony between the two faces ruled out temporal cues. When the /a/ vowel was presented it was equally synchronized with the opening and closing of the visual /a/ articulation as with the visual /i/ articulation. The only way infants could solve this problem was by knowledge that the auditory /a/ corresponded with the articulation involving a wide-open mouth and the auditory /i/ with the articulation involving narrowed lips with the corners pulled back.

A total of 32 infants between 18 and 20 weeks old were tested. The procedure entailed a short familiarization period, followed by the test period. Familiarization consisted of two sequential 10-sec periods in which each visual face was presented without sound. The infant's attention was then brought back to the midline by flashing a small light between the faces. Finally, one soundtrack (/a/ or /i/) was activated and the films of the two faces were allowed to play for a 2-min test period. The infants' fixations to the two faces were videorecorded for subsequent blind scoring.

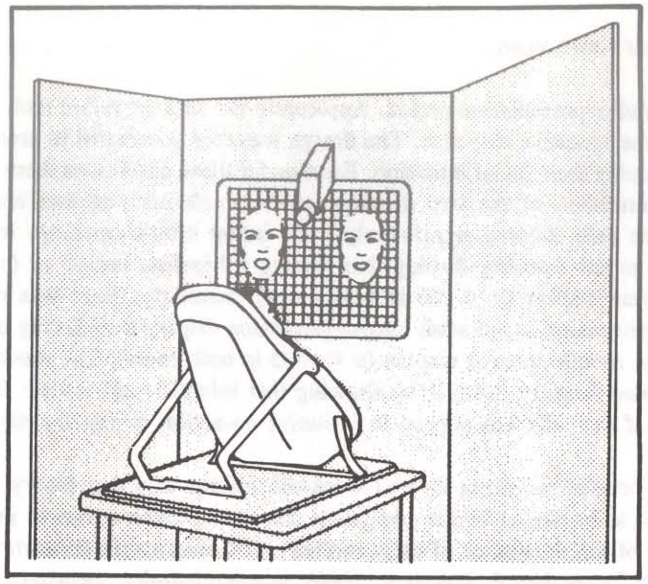

FIG. 5.1. Technique used to test infants' cross-modal perception of speech. The infants watched two faces while listening to a single sound emanating from a loudspeaker centered midway between the two faces. From Kuhl & Meltzoff (1982).

We hypothesized that the auditory stimulus would influence the infants' visual fixations. If infants could detect correspondence between "auditory speech" and "visual speech," they should look longer at the face that produced movements appropriate to the sound they heard. The hypothesis was supported. Of the total fixation time, 73.6% was devoted to the face that "matched" the soundtrack, which is significantly greater than the 50% chance level, $t(31) = 4.67$, $p < .001$. At the level of individual subjects, 24 of the 32 infants fixated the matching face longer than the mismatching face ($p < .01$, binomial test). There were no overall preferences for one visual display or the other, no right vs. left side biases, and no other significant interactions.

What is the invariant that allows these face-voice matches? What information links the visual and auditory signals? One hypothesis was that infants' ability to detect auditory-visual correspondences for speech is based on an intermodal representation of the phonetic unit (Kuhl & Meltzoff, 1982, 1984, 1988). But before this line of reasoning could be pursued we needed to assess a different, competing hypothesis—one stating that the invariant had nothing to do with the representation of speech but instead involved the detection of some other property, wholly independent of the phonetic content of the utterance. One such hypothesis concerned timing information.

The timing hypothesis could be a problem under the following conditions. Suppose the auditorially-presented /a/ vowels happened to be longer in duration than the /i/ vowels, and the /a/ articulatory acts were similarly longer. If this were the case, infants could have succeeded on the task simply by using cross-modal timing information. Therefore, the recognition of "other than phonetic" cross-modal invariants could explain the results. These concerns are real enough because infants' ability to recognize temporal patterns cross-modally has been demonstrated (Bahrick, 1983; Dodd, 1979; Spelke, 1979; Walker, 1982). Of course, we had designed an experiment to rule out the use of temporal information by arranging the two visual stimuli so they were articulating in synchrony with each other. Nonetheless a direct empirical test of the temporal hypothesis seemed worthwhile.

To test this, we altered the original speech signals by removing the frequency information that specified the identity of the vowel while retaining the entire temporal/amplitude envelope of the sounds. Each of the original vowel utterances was replaced by a pure-tone stimulus of 200 Hz that precisely matched its durational characteristics, its amplitude envelope over time, and its alignment in relation to the visual stimuli. The tone began when the lips opened. It became louder as the lips widened. And it became softer and eventually terminated as the lips went back to their closed resting position. If the infants in Study 1 had based their solutions on temporal invariants, they should also succeed on this task. Our hypothesis was that in the absence of spectral information specifying the identity of the vowel, the infants would fail at the task.

These altered stimuli were used with another group of 32 infants of the same age as the original sample. The test procedure was identical to that of Study 1. As predicted, the results fell to chance. It was not that the infants were inattentive to the faces in the presence of these altered stimuli. Infants spent an average of 93.1% of the test time staring at one or the other of the articulatory gestures, which did not differ from that spent in Study 1 (90.3%). However, the direction of their visual fixations were not driven by these altered auditory signals. Of the 32 infants, only 17 looked longer at the "matched" face. The mean percentage of total fixation time devoted to the matched face was 54.6%, $t(31) = 0.78$, $p > .50$.

Taken together, the results of the two experiments show that the temporal information contained in the auditory stimuli was not sufficient to produce the cross-modal speech effect. The results suggest that spectral (formant frequency)

information is critical to the detection of these face-voice correspondences. Work in our laboratory is now being directed at isolating the nature of the spectral information that is supporting the cross-modal effect, and we have made progress in specifying the effective stimulus (Kuhl & Meltzoff, 1984, 1988). Regardless of the outcome of this ongoing research, the fact that infants recognize that particular sounds correspond to lips moving in particular ways supports the notion that speech itself may be intermodally represented as early as 4 months of age.

Vocal imitation. Additional support for our hypothesis that speech is represented as an intermodal event for infants came from the infants' vocal behavior. Not only did the subjects direct their visual attention, they also produced systematic vocalizations. The infants appeared to imitate vocally both the prosodic characteristics of the two vowels and the vowel sounds themselves.

The prosodic characteristics of the adult's utterances—their intonation contour and duration—were duplicated by the infants, as reported in Kuhl and Meltzoff (1982). Figure 5.2 displays a computer analysis of the adult's auditory signal and one infant's vocal responses. The figure shows the pitch (fundamental frequency) of both the adult model and the infant's response as a function of time. The adult produces a "declarative" intonation contour, a slight rise followed by a gradual fall in the pitch of the voice. The infant's production mimics the rise-fall contour. Notice that it is the overall shape or form of the contour that is matched, not its absolute frequency. The infant's productions are higher in absolute pitch than the adult's because infants cannot produce adult pitches with their short vocal cords; there is an anatomical constraint on the possible frequencies that can be achieved by such young infants. Nonetheless the pitch *pattern* is abstracted and matched. Note also that overall durations of the productions are very similar. The target speech sounds were of about 1 sec duration, and the infants' responses match this aspect of the signal as well. Infant vocalizations of this type are not high baseline events in this age group; the literature on early babbling and cooing confirms that it is not common for 4-month-old infants to

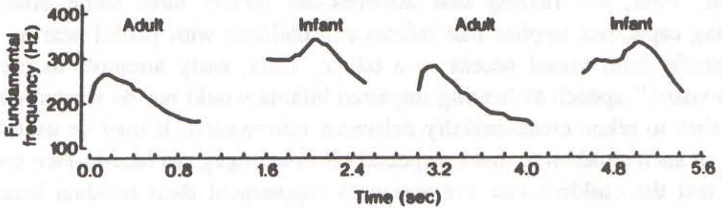

FIG. 5.2. Vocal imitation of pitch contour by 4-month-old infants. The fundamental frequencies (pitches) of the adult's and infant's vowels are plotted as a function of time. From Kuhl & Meltzoff (1982).

produce sustained vowels of this duration (Kent & Murray, 1982; Netsell, 1981; Oller, 1980). Future studies that systematically vary the pitch contour and duration seem worthwhile.

In addition to imitation of prosody, we examined the data for evidence that infants were mimicking the phonetic segments of speech, that is, the particular vowels they heard. Recall that half the infants were randomly assigned to hear the /a/ vowel and half the /i/ vowel. To test for imitation we computer analyzed all of the infants' vowel-like responses. Infants of this age cannot produce perfect /a/ and /i/ vowels. Yet one can analyze the formant frequencies of the infants' vocalizations that test whether those hearing /a/ produced sounds whose formants were more "/a/-like," while those hearing /i/ produced sounds whose formants were more "/i/-like."

According to "distinctive feature theory" (Jakobson, Fant, & Halle, 1969), the formant frequencies of /i/ are spread widely apart (exhibiting the distinctive feature "diffuse"), while /a/'s formants are close together in frequency (exhibiting the distinctive feature "compact"). The formants of the infants' sounds were measured by computer, and the values of the diffuse-compact feature were calculated. The results supported the imitation hypothesis. Infants hearing /i/ produced significantly more "/i/-like" sounds, ones that were more diffuse while infants hearing /a/ produced sounds that were significantly more compact (Kuhl & Meltzoff, 1988).

We suggest that both vocal imitation and the recognition of auditory-visual correspondences for speech manifest a knowledge of the relationship between audition and articulation. Moreover, we believe that these abilities have a common origin—the infant's intermodal representation of speech (Kuhl & Meltzoff, 1988). Although infant speech perception is commonly treated as an auditory phenomenon, it may be profitable to investigate infant speech as an intermodal event. Knowledge about faces and voices is closely coordinated even by 4-months-of-age.

This theoretical perspective has at least two implications for more clinical concerns. First, the finding that 20-week-old infants have some primitive lipreading capacities implies that infants and children with partial hearing loss might profit from visual access to a talker. Thus, early attempts to present "audio-visual" speech to hearing-impaired infants would not be wasted due to an inability to relate cross-modally delivered information. It may be useful for parents of such children to make a special effort to engage in face-to-face speaking so that the children can use vision to supplement their residual hearing. Research with adults shows that visual access to a talker is equivalent to boosting the auditory signal almost 20 dB (Sumby & Pollack, 1954); there is thus reason to believe that *en face* speaking could be useful in keeping partially deaf infants tuned into the verbal channel.

Second, we note that research with adults has shown that some speech information can be delivered through another sensory modality, namely the skin, via a

variety of prosthetic devices that encode auditory frequency information spatially on the skin using either vibratory or electrocutaneous stimulation. The results of this work show that adults can effectively combine auditory, visual, and tactile information into one coherent speech signal (Sparks, Kuhl, Edmonds, & Gray, 1978). On the basis of our research and the theory discussed earlier, it may be suggested that such devices, which are effective with adult hard-of-hearing populations, might also be effectively adapted for use with younger-age populations.

CATEGORIZATION OF SPEECH SOUNDS

The mastery of language requires that infants both draw distinctions and perceive similarities in the sound stream. Regarding the first, language acquisition requires infants to distinguish among the smallest entities that convey meaning—the phonetic units of the language. Infants must have the requisite auditory acuity for frequency, intensity, and duration to perceive differences between words such as "pat" and "bat," which differ by a single phonetic unit. Regarding the second, infants must be able to group sounds together that are auditorially discriminable but phonetically equivalent. In order to speak and understand language, a phonetic unit such as the consonant /d/ must be perceived as the same invariant unit across wide changes in the acoustic "cues" that indicate the unit's identity (Kuhl, 1985a, 1987).

Research on speech has shown that the natural variation found in spoken language has a profound effect on the acoustic determinants of phonemes. The acoustic cues that identify a unit change dramatically when it is spoken by different talkers, appears in different positions in a syllable, when its neighboring phonetic units change, or when speech is spoken at a faster rate (Klatt, 1975; Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Miller & Liberman, 1979; Peterson & Barney, 1952). This is why the acoustic cues that identify speech sounds are considered highly context dependent, and speech sounds are often described as exhibiting a "lack of acoustic invariance." There is obviously "invariance" at the *perceptual* level that we register; but the interesting problem is that there is no clear invariant at the *physical*, i.e., acoustic, level that can be specified.

This extreme variation in the acoustic realizations of phonetic units is the main problem preventing the recognition of speech by a computer. For example, when a phonetic unit is produced by a new talker, or when it is produced in different contexts, computers fail to categorize the unit correctly because the criterial acoustic cues defining the identity of the unit vary. Yet, adults perceive phonetic units despite these variations. The question is: Can infants organize speech sounds into "categories," perceiving different acoustic realizations of a given phonetic type as "the same"?

In order to do so, the infant has to do more than perform a simple discrimination task (Kuhl, 1985a). Consider the categorization of simple vowel sounds. Suppose a man, woman, and child have each produced two vowels, /a/ (as in "pot") and /æ/ (as in "pat"). What is required in order to "sort" the six sounds into two phonetic categories?

There are two processes involved: (a) The infant has to focus on the critical properties that separate the two categories /a/ and /æ/, that is, their formant patterns, and (b) they have to ignore the salient differences *within* each of the categories, such as the differences between the voices of the different talkers, and differences in loudness and duration, that are irrelevant to the categorization of /a/ and /æ/. This is similar to the classic cases of object categorization studied by cognitive psychologists, in which subjects must sort exemplars differing in perceptually salient and overlapping ways into groups. For example, a common case of categorization would be the sorting of squares and rectangles when they differ in size, color, and orientation. Subjects have to focus on the critical properties defining squares and rectangles while ignoring the irrelevant ones, such as color and size.

The research we discuss here shows that human infants in the first half year of life can succeed on these phonetic categorization problems. In particular, work has focused on the problem of "talker normalization"—recognizing equivalence for the same speech sound when it is spoken by talkers of different age and sex (Kuhl, 1979, 1983, 1985a). Talker normalization is of special theoretical interest because speech varies radically when spoken by different talkers. This is because individuals' vocal tracts have different overall dimensions and the absolute formant frequencies that result are quite different, even when the same vowel is produced (Peterson & Barney, 1952). The largest acoustic differences occur between males and females and between adults and children (Fant, 1973). Thus, recognizing the equivalence for vowels spoken by different talkers provides an interesting test of infants' categorization ability. Adults sort these acoustically diverse entities into linguistically relevant categories. Do infants?

The technique used to test infants' ability to categorize speech sounds across talkers involved a conditioning procedure (Kuhl, 1985b). Figure 5.3 shows the experimental situation used in several of Kuhl's studies. An infant sits on the mother's lap and is engaged by an assistant who slowly rotates objects as a way of engaging the infant's attention. To the left of the infant is a loudspeaker that repeatedly emits a background speech sound, for instance the vowel /a/. Periodically this stimulus is changed to a new vowel, for example an /i/ vowel. The infant is trained to turn her head to the loudspeaker whenever the sound is changed. A correct head turn causes a box on top of the loudspeaker suddenly to light up revealing a dancing toy bear (Fig. 5.3, bottom panel). A glimpse at the dancing bear seems to engage infants sufficiently that they rapidly learned to turn their heads whenever they hear a change in the speech sound. They seem to enjoy the game.

The categorization problem was posed to infants in the following way. During

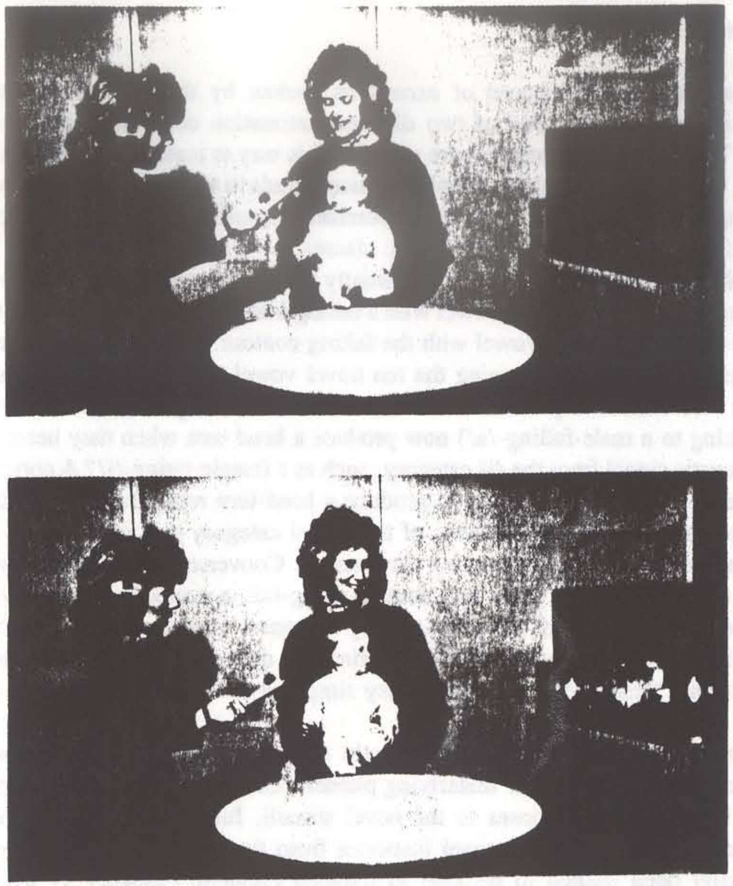

FIG. 5.3. The head-turn technique used to study infants' categorization of speech. Infants sit on a parent's lap and watch the toys held by an assistant, while listening to sounds repeated over a loudspeaker located to their left (top). They quickly learn to turn toward the loudspeaker when the sound changes from one phoneme (/a/) to another (/i/) because it signals the opportunity to see an animated dancing bear (bottom). From Kuhl (1987).

training infants learned to produce a head-turn response when the background vowel (/a/) was changed to another vowel (/i/) spoken by the same talker using the same intonation contour. The subsequent test of categorization then involved presenting infants with novel instances of the /a/ and /i/ vowels. A total of twelve stimuli were used in the tests, six /a/ vowels and six /i/ vowels. Each vowel category was composed of exemplars spoken by three talkers (male, female, and child) using one of two different intonation contours (rising and falling). The intonation contours were varied in this way to make it an even more difficult, but realistic, problem. People produce sounds in varying tones of voice; infants have to be able to ignore this prominent acoustic difference between vowels if they are to respond to phonetic classes.

In Kuhl (1979) half the infants were initially trained to make a head turn to a male talker pronouncing an /i/ vowel with a falling intonation contour and half to the male producing an /a/ vowel with the falling contour. Each infant was then given the categorization test using the ten novel vowel stimuli. Would infants who had been trained to produce a head turn to a male-falling-/i/ (and to inhibit head turning to a male-falling-/a/) now produce a head turn when they heard a novel acoustic signal from the /i/ category, such as a female-rising-/i/? A correct categorization response would be to produce a head-turn response because the novel exemplar was still an instance of the vowel category /i/, even though it varied in the talker and pitch contour dimensions. Conversely when the infants heard a novel /a/ vowel, such as a female-rising-/a/, a correct categorization would be the inhibition of the head-turning response, because it fell into the nonreinforced phonetic category. The experimental question was the following: Would infants "categorize" or would they simply produce head-turns to all of the new stimuli?

The results showed that infants correctly sorted novel instances of vowel productions according to their underlying phonetic category. Figure 5.4 depicts infants' head-turning responses to the novel stimuli. Infants consistently produced head-turn responses to novel instances from the phonetic category they had initially been trained to respond to (labeled Phonetic Category 1) while consistently inhibiting head-turn responses to novel instances from the phonetic category they had not been trained to respond to (labeled Phonetic Category 2). Novelty itself, therefore, did not prompt a head-turn response. Infants produced the response only for exemplars that belonged to the appropriate phonetic category. Infants treated all /a/'s (and all /i/'s) as equivalent, despite the fact that they could clearly hear the differences among the /a/'s and among the /i/'s that were produced by the variation in talker and intonation contour.

Recently these findings were extended to a different vowel contrast (Kuhl, 1983). These vowels were chosen to be maximally confusable in that they are adjacent in "vowel space." The stimuli were /a/ (as in "cot") and /ɔ/ (as in "caught"). Data of Peterson and Barney (1952) on the formant frequencies of /a/ and /ɔ/ show that these two vowels are similar, even when produced by a single talker. A test of the /a/ and /ɔ/ contrast was therefore conducted because it constitutes one of the most difficult vowel contrasts in English. Should infants demonstrate equivalence classification for these vowel categories, one might reasonably suppose they could do so for all vowel categories in English (Kuhl, 1983).

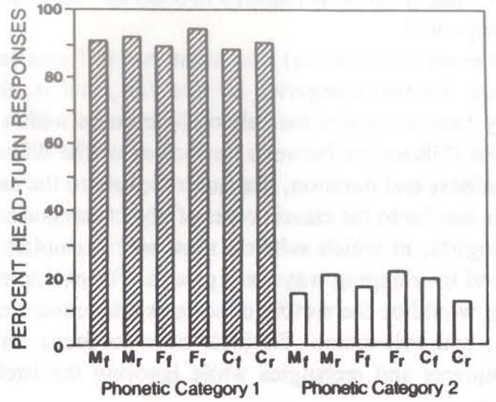

FIG. 5.4. Results from the speech categorization experiment involving talker normalization. Infants were trained to produce a head-turn response to a single vowel from one of the phonetic categories (designated as Phonetic Category 1). The vowel was always produced by a male talker using a falling pitch contour (symbolized as Mf). In this period infants were also trained to refrain from turning to the male-falling production of the opposite vowel (designated Phonetic Category 2). After training, novel vowels from both phonetic categories were introduced in order to test infant categorization. These were produced by female (F) and child (C) talkers with either a rising (r) or falling (f) pitch contour. The correct categorization response is to perform a head-turn for all stimuli belonging to Phonetic Category 1 and not to stimuli belonging to Category 2. From Kuhl (1987).

The same age group, 6-month-olds, was tested using the same head-turn technique. Once again the test involved recognizing the vowels across changes in three talkers (male, female, and child) and two intonation contours (rising and falling). The results of this experiment clearly showed that infants can correctly categorize these vowels. As expected, the contrast was slightly more difficult, but the results were highly significant. The mean percent correct scores were 76.3%, with the performance of individual infants ranging from 51.4% to 89.8%. Categorization tests have also been successfully extended to consonant classes (Kuhl, 1980) and ones involving pitch (Kuhl, 1985a).

This work clearly shows that infants are capable of sorting speech sounds into categories, even though it is difficult to characterize the sorting rule they use. However, even the exact nature of the information used by adult listeners in the categorization of vowels produced by different talkers cannot be identified. Although something about the locations of the formant frequencies is involved, the information cannot be frequency-specific (because this varies across talkers).

Rather, listeners must be using more abstract aspects of the signal such as spectral "patterns" or "shapes" (Fant, 1973). More research on adults and infants will be necessary before the basis of this categorization skill is isolated. Suffice it to say that it is not based on a simple, frequency-specific property of the sounds. There is real categorization across acoustic diversity.

These studies complement and add substantially to existing information concerning infants' perceptual organization of speech. Most experiments on speech perception in early infancy have focused on the infant's ability to detect a *difference* between speech sounds on a continuum, with all other variation controlled (e.g., studies on "categorical perception"; Eimas, Siqueland, Jusczyk, & Vigorito, 1971). The main focus of Kuhl's experiments is the perception of *similarity* among stimuli that varied in perceptually prominent ways but still belonged to the same underlying phonetic category. Thus, these experiments examine categorization—the ability to render perceptually distinguishable events equivalent—and provide the first strong evidence that 6-month-old infants are capable of recognizing phonetic equivalence across salient perceptual changes. By 6-months-of-age infants perceptually organize linguistically relevant speech-sound categories. Their ability to correctly categorize phonetic units across talkers and intonation variation far outstrips the capacity of any computer so far built. This speech recognition problem remains one of the most intractable ones in the field of artificial intelligence. It cannot be solved by computers, but is solved with facility by babies, who seem to be "natural categorizers."

CONCLUSIONS

The research discussed here shows that infants have substantial knowledge about speech sounds, faces, and correspondences between the two. What are the implications of these findings for theory?

The research on speech-sound categorization demonstrates that young infants can sort sounds into phonetic categories. The infant's skill at imposing this kind of organization on diverse acoustic stimuli has direct implications for the development of speech and language (Kuhl, 1985a). It is, in fact, an essential prerequisite for the acquisition of language, because language development could not proceed normally if the same word, when spoken by different people, were not perceived to be the same. Moreover, categorization is critical to the infant's own ability to reproduce the sound patterns he or she hears. Due to anatomical constraints, the infant cannot match the speech of adults by reproducing identical absolute formant frequencies. To imitate an adult, the infant must produce a sound in his own frequency range that, although acoustically different, is nonetheless perceived as the equivalent unit. As the experiments show, infants can indeed detect the requisite auditory equivalences.

The research also shows that infants are born with a capacity to duplicate

certain facial movements they see with ones of their own. The subjects in the tests were not 1-year-olds, nor 1-month-olds, but literally 1-day-olds. What psychological mechanism underlies this behavior? Meltzoff and Moore proposed the AIM hypothesis. In this view early imitation reflects a process of active intermodal mapping (AIM) in which infants use the equivalences between visually and proprioceptively perceived body transformations as a basis for organizing their responses. This hypothesis also implies that early imitation is mediated by an internal representation of the adult's act. If so, the newborn could not be using a mental image of the adult's display in the sense of a visual picture or iconic copy, because then we would be left with the problem of how infants could ever link up the visual image of the adult's act with the motor image of their own movements. Meltzoff and Moore therefore suggested that neonates can pick up supramodal information about the adult's movement pattern that is used directly as the basis for the infant's own motor plans. This abstract representation constitutes the model that directs the infant's actions and against which he matches his motor performance (Meltzoff, 1985b; Meltzoff & Moore, 1977, 1983a, 1983b).

All of this imputes fairly complicated processing to young infants. The AIM hypothesis would gain force if converging evidence showed that young infants are capable of complex intermodal representations, especially ones involving facial movements. It is therefore of special interest that another phenomenon discussed here, the cross-modal speech effect, requires that infants recognize a complex mapping between audition and oral movements. While we can say with certainty that the visual-motor mappings involved in imitation are functional at birth, the same claim cannot yet be made about the auditory-visual mappings.

In fact, there remain three developmental alternatives for the cross-modal speech phenomenon. First, the infants may simply have learned which articulatory gestures go with which sounds by watching and listening to adults. This would reduce to pure associative learning. Second, it may be that these auditory-articulatory mappings are innately specified. If so, then a follow-up newborn study, as was performed on the gestural imitation case, may yield positive results. However, there is also an intriguing third alternative—namely, that the infants' own babbling experience may play an important role (Kuhl & Meltzoff, 1984). This interpretation is interesting because it ties together several of the phenomena discussed in this essay.

In the babbling account, infants are conceived of as carefully monitoring their own vocal play during cooing and babbling. They "feel" their articulatory movements through proprioception and can perceive the consequences of these articulatory efforts through audition. Thought of in this manner, babbling is a way of practicing the basic act of speaking, practicing the production of phonetic units at will.

How could this cooing and babbling experience help infants in the cross-modal situation? It could help only if infants can relate the speech acts they see the adult perform in the experiment to the auditory-articulatory events they themselves produced during babbling. The research from our laboratories indicates that it is very likely. Infants' ability to imitate visual gestures demonstrates that they can relate mouth movements they see to their own mouth movements. The mouth-opening movement in Meltzoff and Moore's imitation experiments is similar to the mouth opening used to produce /a/ in Kuhl and Meltzoff's cross-modal speech case. There is thus a foothold on the articulatory side—infants may relate their own speech mouth movements to those they see the adult perform. There is also a similar foothold on the auditory side. Kuhl's speech categorization work indicates that infants can recognize the equivalence between the vowels uttered across talkers, including those produced by children and adults. It therefore is reasonable to suppose that the infants in the cross-modal experiment can recognize equivalences between the vowels they hear in the experiment and their own previous productions during babbling.

In short, infants have the requisite tools, as manifest by the imitation of visual mouth movements and the cross-talker categorization of vowels, to use cooing and babbling experience to help solve the cross-modal task. Babbling provides infants with an auditory-articulatory event in which /a/ sounds are systematically produced by /a/ articulations. The cross-modal experiment now re-poses that question for another's body, not one's own: Which articulatory act is linked with an /a/ sound? The knowledge gained during their own babbling may contribute to infants' ability to recognize cross-modal equivalences for speech in others. We are investigating the impact of babbling on our cross-modal speech test by testing infants prior to the age at which cooing and babbling begins.

We have thus come full circle. Infants' ability to imitate, to recognize cross-modal speech equivalences, and to sort sounds into speech categories may be linked. At a minimum their facility at facial imitation and speech categorization may contribute to their success on auditory-visual speech tasks. But even if intermodal speech tasks were solved by newborns, and therefore babbling experience played no role, it would not entirely divorce the phenomena we have discussed. All of these phenomena suggest that infants are capable of, indeed quite engaged by, complex equivalence mappings we might have thought beyond their powers. Table 5.1 summarizes the findings.

The speech categorization case is described as an auditory-auditory mapping in the table, because infants must group acoustically diverse events as similar. Whatever makes /a/'s the same across variations in talker and intonation pattern is not a simple, frequency-specific cue. Infants may categorize on the basis of some higher-order information, perhaps the "shape" or "form" of the spectral information. Alternatively, they may have a phonetic prototype of an /a/ in mind against which similarity is judged (Grieser & Kuhl, 1989).

The cross-modal speech case takes this one step further. Rather than connecting one acoustic event to another within the same modality, the mapping must extend across perceptual systems. This leads to the suggestion that phonetic units

TABLE 5.1
Four Phenomena in Early Infancy and the Types of Mapping Functions They Suggest

Phenomenon	*Type of Mapping*
Categorization of speech	Auditory → Auditory
Cross-modal speech effect	Auditory → Visual
Vocal imitation	Auditory → Motor
Gestural imitation	Visual → Motor

may be specified in a form that unites both auditory and articulatory specifications. The phonetic level of language may be intermodally represented (Kuhl & Meltzoff, 1982, 1984, 1988). This research also shows that young infants can vocally imitate, duplicating both the prosodic characteristics of the display as well as the phonetic information. Hearing the speech sounds drives motor production; hence it illustrates auditory-motor mapping.

Finally, facial imitation also entails a complex equivalence mapping. As in the case of vocal imitation, infants must do more than simply recognize a match that is presented to them. They visually pick up information about facial movement and must use this as a basis for action. Successful gestural imitation involves a mapping from vision to motor performance. Human gestural acts, as well as speech products, may be amenable to a supramodal description that unites within one common framework self-production and the perception of acts by others.

That young infants can succeed on all the tasks depicted in Table 5.1 suggests they are not simply sensory respondents. Evidently the course of psychological development cannot be characterized as having an early stage in which infants first respond to "basic sense data" in the form of retinal images or raw acoustic energy, only later coordinating these sensations into organized perceptions, and then finally progressing to the coordination of perception and action. Instead, infants, even newborns, are attuned to perceiving certain higher-order categories and events in the world. Their categories unite sensorily disparate events and seem to operate not only within single receptor systems, but across perceptual modalities as well. Some of the stitching together and construction that we used to think took place in later infancy seems already to be done (Bower, 1982).

This theoretical perspective has implications for clinical practice. The notion that young infants can handle multimodal stimulation and abstract higher-order commonalities across different input modalities suggests possible treatment strategies in cases of sensory handicap. If infants have an auditory handicap, there is hope that vision may help complement the residual hearing that remains, as in the lipreading example we discussed earlier. Similarly it has been suggested that infants with visual handicaps will profit from substituting auditory information for vision (Bower, 1979, 1982). To some extent treatment strategies based on this type of theory of infant development have been successful (Aitken & Bower, 1982).

In conclusion, one cannot help but be struck by the fact that young infants do not respect our traditional disciplinary lines. As adult scientists we often specialize in vision research, or hearing, or motor organization. Infants are not specialists in that sense. They are interdisciplinary in nature. Twenty years ago psychologists studying children migrated toward infancy research in hopes of breaking things down into simpler component parts. We are now catching glimpses of these "psychological primitives," and the new findings suggest that they are anything but simple. The challenge we now face is to outline a theory of psychological development in which imitation, cross-modal perception, and speech categorization are conceived of as initial state settings from which development proceeds, rather than as an endpoint of infancy to which it is proceeding. Because the young human infant has stubbornly refused to become simple, we must make our psychological theories and our attendant treatment strategies more complex.

ACKNOWLEDGMENTS

We gratefully acknowledge the support of grants from the National Institutes of Health (HD-22514 and HD-18286) and from the John D. and Catherine T. MacArthur Foundation. We also wish to thank Keith Moore, Calle Fisher, Craig Harris, and Karen Wolak for help on all phases of the research.

REFERENCES

Abravanel, E., & Sigafoos, A. D. (1984). Exploring the presence of imitation during early infancy. *Child Development, 55,* 381–392.

Aitken, S., & Bower, T. G. R. (1982). The use of the Sonicguide in infancy. *Journal of Visual Impairment and Blindness,* March, 91–100.

Bahrick, L. E. (1983). Infants' perception of substance and temporal synchrony in multimodal events. *Infant Behavior and Development, 6,* 429–451.

Bower, T. G. R. (1979). *Human development.* San Francisco: W. H. Freeman.

Bower, T. G. R. (1982). *Development in infancy* (2nd ed.). San Francisco: W. H. Freeman.

Brazelton, T. B., & Tronick, E. (1980). Preverbal communication between mothers and infants. In D. R. Olson (Ed.), *The social foundations of language and thought* (pp. 299–315). New York: W. W. Norton.

Bushnell, I. W. R. (1979). Modification of the externality effect in young infants. *Journal of Experimental Child Psychology, 28,* 211–229.

Dodd, B. (1979). Lip reading in infants: Attention to speech presented in- and out-of-synchrony. *Cognitive Psychology, 11,* 478–484.

Dunkeld, J. (1978). *The function of imitation in infancy.* Unpublished doctoral dissertation, University of Edinburgh, Scotland.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science, 171,* 303–306.

Fant, G. (1973). *Speech sounds and features.* Cambridge, MA: MIT Press.

Field, T. M., Woodson, R., Greenberg, R., & Cohen, D. (1982). Discrimination and imitation of facial expressions by neonates. *Science, 218,* 179–181.

Fontaine, R. (1984). Imitative skills between birth and six months. *Infant Behavior and Development, 7,* 323–333.

Gannon, E. C., & Swartz, K. B. (1980). Perception of internal elements of compound figures by one-month-old infants. *Journal of Experimental Child Psychology, 30,* 159–170.

Gewirtz, J. L., & Stingle, K. G. (1968). Learning of generalized imitation as the basis for identification. *Psychological Review, 75,* 374–397.

Girton, M. R. (1979). Infants' attention to intrastimulus motion. *Journal of Experimental Child Psychology, 28,* 416–423.

Grant, K. W., Ardell, L. H., Kuhl, P. K., & Sparks, D. W. (1985). The contribution of fundamental frequency, amplitude envelope, and voicing duration cues to speechreading in normal-hearing subjects. *Journal of the Acoustical Society of America, 77,* 671–677.

Green, K. P., & Kuhl, P. K. (1989). The role of visual information in the processing of place and manner features in speech perception. *Perception and Psychophysics, 45,* 34–42.

Green, K. P., Kuhl, P. K., & Meltzoff, A. N. (1988). Factors affecting the integration of auditory and visual information in speech: The effect of vowel environment. *Journal of the Acoustical Society of America, 84,* Supplement 1, S155.

Grieser, D. L. & Kuhl, P. K. (1989). Categorization of speech by infants: Support for speech-sound prototypes. *Developmental Psychology, 25.*

Heimann, M., & Schaller, J. (1985). Imitative reactions among 14–21 day olds. *Infant Mental Health Journal, 6,* 31–39.

Jacobson, S. W. (1979). Matching behavior in the young infant. *Child Development, 50,* 425–430.

Jakobson, R., Fant, C. G. M., & Halle, M. (1969). *Preliminaries to speech analysis: The distinctive features and their correlates.* Cambridge, MA: MIT Press.

Kent, R. D., & Murray, A. D. (1982). Acoustic features of infant vocalic utterances at 3, 6, and 9 months. *Journal of the Acoustical Society of America, 72,* 353–365.

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

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. Kavanagh, & C. A. Ferguson (Eds.), *Child Phonology: Perception* (Vol. 2, 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.

Kuhl, P. K. (1985a). Categorization of speech by infants. In J. Mehler & R. Fox (Eds.), *Neonate cognition: Beyond the blooming, buzzing confusion* (pp. 231–262). Hillsdale, NJ: Lawrence Erlbaum Associates.

Kuhl, P. K. (1985b). Methods in the study of infant speech perception. In G. Gottlieb & N. A. Krasnegor (Eds.), *Measurement of audition and vision in the first year of postnatal life: A methodological overview* (pp. 223–251). Norwood, NJ: Ablex.

Kuhl, P. K. (1986). Infants' perception of speech: Constraints on characterizations of the initial state. In B. Lindblom & R. Zetterström (Eds.), *Precursors of early speech* (pp. 219–244). New York: Stockton Press.

Kuhl, P. K. (1987). Perception of speech and sound in early infancy. In P. Salapatek & L. B. Cohen (Eds.), *Handbook of infant perception:* From perception to cognition (Vol. 2, pp. 275–382). New York: Academic Press.

Kuhl, P. K., Green, K. P. & Meltzoff, A. N. (1988). Factors affecting the integration of auditory and visual information in speech: The level effect. *Journal of the Acoustical Society of America, 83,* Supplement 1, S86.

Kuhl, P. K., & Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. *Science, 218,* 1138–1141.

Kuhl, P. K., & Meltzoff, A. N. (1984). The intermodal representation of speech in infants. *Infant Behavior & Development, 7,* 361–381.

Kuhl, P. K., & Meltzoff, A. N. (1988). Speech as an intermodal object of perception. In A. Yonas (Ed.), *Perceptual development in infancy: Minnesota symposia on child psychology* (Vol. 20, pp. 235–266). Hillsdale, NJ: Lawrence Erlbaum Associates.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74,* 431–461.

MacKain, K., Studdert-Kennedy, M., Spieker, S., & Stern, D. (1983). Infant intermodal speech perception is a left hemisphere function. *Science, 219,* 1347–1349.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264,* 746–748.

Meltzoff, A. N. (1985a). Immediate and deferred imitation in fourteen- and twenty-four-month-old infants. *Child Development, 56,* 62–72.

Meltzoff, A. N. (1985b). The roots of social and cognitive development: Models of man's original nature. In T. M. Field & N. Fox (Eds.), *Social perception in infants* (pp. 1–30). Norwood, NJ: Ablex.

Meltzoff, A. N. (1988a). Infant imitation and memory: Nine-month-olds in immediate and deferred tests. *Child Development, 59,* 217–225.

Meltzoff, A. N. (1988b). Infant imitation after a 1-week delay: Long-term memory for novel acts and multiple stimuli. *Developmental Psychology, 24,* 470–476.

Meltzoff, A. N. (1988c). Imitation, objects, tools, and the rudiments of language in human ontogeny. *Human Evolution, 3,* 45–64.

Meltzoff, A. N., & Borton, R. W. (1979). Intermodal matching by human neonates. *Nature, 282,* 403–404.

Meltzoff, A. N., & Moore, M. K. (1977). Imitation of facial and manual gestures by human neonates. *Science, 198,* 75–78.

Meltzoff, A. N., & Moore, M. K. (1983a). Newborn infants imitate adult facial gestures. *Child Development, 54,* 702–709.

Meltzoff, A. N., & Moore, M. K. (1983b). The origins of imitation in infancy: Paradigm, phenomena, and theories. In L. P. Lipsitt (Ed.), *Advances in infancy research* (Vol. 2, pp. 265–301). Norwood, NJ: Ablex.

Meltzoff, A. N., & Moore, M. K. (1985). Cognitive foundations and social functions of imitation and intermodal representation in infancy. In J. Mehler & R. Fox (Eds.), *Neonate cognition: Beyond the blooming, buzzing confusion* (pp. 139–156). Hillsdale, NJ: Lawrence Erlbaum Associates.

Milewski, A. E. (1976). Infants' discrimination of internal and external pattern elements. *Journal of Experimental Child Psychology, 22,* 229–246.

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

Netsell, R. (1981). The acquisition of speech motor control: A perspective with directions for research. In R. E. Stark (Ed.), *Language behavior in infancy and early childhood* (pp. 127–156). New York: Elsevier North Holland.

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

Peterson, G., & Barney, H. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America, 24,* 175–184.

Piaget, J. (1962). *Play, dreams and imitation in childhood.* New York: W. W. Norton.

Salapatek, P. (1975). Pattern perception in early infancy. In L. B. Cohen & P. Salapatek (Eds.), *Infant perception: From sensation to cognition* (Vol. 1, 133–248). New York: Academic Press.

Skinner, B. F. (1953). *Science and human behavior.* New York: Macmillan.

Sparks, D. W., Kuhl, P. K., Edmonds, A. E., & Gray, G. P. (1978). Investigating the MESA (Multipoint Electrotactile Speech Aid): The transmission of segmental features of speech. *Journal of the Acoustical Society of America, 63,* 246–257.

Spelke, E. S. (1979). Perceiving bimodally specified events in infancy. *Developmental Psychology, 15,* 626–636.

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America, 26,* 212–215.

Summerfield, O. (1979). Use of visual information for phonetic perception. *Phonetica, 36,* 314–331.

Trevarthen, C., & Marwick, H. (1986). Signs of motivation for speech in infants, and the nature of a mother's support for development of language. In B. Lindblom & R. Zetterström (Eds.), *Precursors of early speech* (pp. 279–308). New York: Stockton Press.

Uzgiris, I. C., & Hunt, J.McV. (1975). *Assessment in infancy: Ordinal scales of psychological development.* Urbana: University of Illinois Press.

Walker, A. S. (1982). Intermodal perception of expressive behaviors by human infants. *Journal of Experimental Child Psychology, 33,* 514–535.