THE COMPARATIVE PSYCHOLOGY OF AUDITION:
Perceiving Complex Sounds

Edited by

ROBERT J. DOOLING
The University of Maryland

STEWART H. HULSE
Johns Hopkins University

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

13 On Babies, Birds, Modules, and Mechanisms: A Comparative Approach to the Acquisition of Vocal Communication

Patricia K. Kuhl
University of Washington

INTRODUCTION

The vocal repertoires of animals constitute one of the most noticeable differences among them. The song of a bird, the croak of a frog, the sonar signals of a bat, the sounds that crickets make, and the calls of whales are strikingly different. Hearing any one of these signals allows us to name the animal that produced it. Birds do not croak like frogs, bats do not call like whales, and crickets do not chirp like birds. Vocalizations are a species' signature; they identify animals as members of one group rather than another.

Like birds, bats, and frogs, babies will produce a species-specific signal at a very young age—one that identifies them as members of the human species. The signal is "canonical babbling," repetitive consonant-vowel syllables such as /mamamama/ or /dadadada/. My 8-month-old baby has just begun this type of babbling, and in doing so she has produced the first observable sign that she will acquire language. Although I had heard infant babbling many times in the course of my research, I was awestruck at the regularity in its form and the precision of its timing when it occurred in my own child, and was somewhat surprised by its failure to occur early in response to my ever-constant modeling. I was reminded that this milestone of human speech occurs at the appointed time regardless of the language in which the infant is being reared, the educational background and socioeconomic status of the infant's parents, the infant's general motor and intellectual capabilities (within normal limits), and (apparently) parent prompting.

Acquisition of conspecific signals does not appear to be left to chance. All male swamp sparrows sing, all male humpbacks call, and all normal babies babble. Such obligatory behavior must be well protected from the nuances of individual development. Nature seems to have put a premium on ensuring that all appropriate members of the species produce the right signal at the right time.

What ensures this? How does nature guarantee that infants of a species will learn to produce one set of sounds as opposed to another? Are we protected from incorrect learning by restrictions on what we can hear? or on what our vocal tracts can produce? Certainly production constraints narrow the options somewhat: birds cannot croak, given their vocal apparatus, and whales cannot produce birdsong, given theirs. Simple sensory limitations account for others: Crickets probably cannot hear speech, and frogs probably cannot hear the sonar signals produced by bats. In some instances, then, the selection of signals for reproduction by the young is limited by fairly peripheral constraints, either motor or sensory.

But what of the remaining cases? Swamp sparrows produce signals that are quite different from their not-so-distant relatives, the song sparrow (see Marler, 1973, for review). The humpback whale's vocal signals differ greatly from those produced by other whales. Vocal tract constraints cannot explain these differences; nor can peripheral auditory constraints. And what of human infants? Why do they mimic speech rather than other reproducible signals? Peripheral constraints alone cannot explain the vocal repertoires of babies and birds. Experience has been shown to play a critical role. Birds who are socially isolated or deafened will not produce conspecific song (Marler, 1973), and babies who are born deaf will not produce canonical babbling (Oller, 1986) nor learn to talk normally. But even here there is a further complication. The required auditory experience must be of the right kind. Hearing just any species' vocalizations is not sufficient to promote learning. Song sparrows will not produce their own song when they hear only the swamp sparrow song; nor will song sparrows learn to produce the swamp sparrow's song. Similarly, their is evidence suggesting that young infants do not mimic nonspeech auditory events (Kuhl & Meltzoff, 1988).

What accounts for selective vocal learning and selective imitation? How do infants of a species know which signals are the right kind?

Here, ethologists interested in birds and speech scientists interested in babies find themselves in the same box. We both have to explain what it is that allows the infant of the species to learn selectively. Ethologists have approached the problem by manipulating what the bird hears and observing what the bird eventually produces. But because speech scientists cannot experimentally manipulate what babies hear, our approach in studying human infants has been to conduct experiments involving auditory perception. We examine what babies can perceive, and use these data to test theories concerning the nature of the mechanisms underlying speech perception in infants.

These experiments reveal that human infants exhibit a remarkable sensitivity to human speech (see Kuhl, 1987a, for recent review). They appear to have an innate ability to perceive the universal set of phonetic[1] distinctions that are

[1]The term *phonetic* is used to signify any difference in any language that is sufficient to distinguish two words. The term *phonemic* is used to specify distinctions that are used in a particular language.

appropriate for speech in any language. Moreover, the work discussed in this chapter shows that infants recognize complex equivalences between phonetically equal events—between two phonetically equal but discriminably different auditory events (Kuhl, 1979a, 1983, 1985a), between phonetically equal events delivered to two different modalities, as when an auditorially presented speech sound is related to the sight of a person producing that sound (Kuhl & Meltzoff, 1982, 1984a), and when an auditory event, produced by someone else, is related to the motor movements necessary to reproduce it oneself (Kuhl & Meltzoff, 1982, 1988). How are these complex equivalences detected? Is it simply infants' general sensory and cognitive processes that account for these remarkable abilities—or is there something more?

The production side poses similar questions. Normal speech will not occur unless the infant can hear, but given this, and given exposure to a specific language, influences of the mother tongue will appear quite early. By the end of the first year, young children already sound like infants being reared in a particular language environment; they have an *accent*. The American infant will sound distinctly different from a Russian, French, African, or Chinese infant (e.g., de Boysson-Bardies, Sagart, & Durand, 1984). Thus, vocal learning is affected by the infant's linguistic environment. But this learning is somehow restricted to sounds that are speech. How does the mechanism specify the signal to be learned, so as to restrict vocal learning to speech as opposed to other sounds in the environment? As in birds, the constraints on learning are not peripheral ones; neither motor limitations nor perceptual limitations account for infants' selectivity in learning. Their vocal learning must be guided by something more.

The *something more* that we subscribe to in explaining both innate perceptual abilities and selective vocal learning is the notion that for each species in which vocal communication plays a critical role, specialized perceptual mechanisms exist. These special mechanisms are as unique to the species as the sounds that are produced by that species. Thus, we say it is the bird's *auditory template* and the baby's *speech module* that explain why birds sing and babies babble, and why both display early perceptual recognition of their conspecific signals.

Accepting that infants demonstrate both innate perceptual abilities and early restrictions on vocal learning, the challenge to theory is to describe the mechanisms that underlie these abilities. What kinds of mechanisms are they? What information do they specify? Are the mechanisms dedicated to the processing of conspecific signals? Do they involve perceptual systems, production systems, or both?

The purpose of this chapter is to review the current status of our answers to these questions regarding the human infant's acquisition of speech. Important phenomena regarding infants' perception and production of speech are reviewed and two theories that have been advanced to explain these abilities are described. The comparative approach is shown to have made two important contributions to the study of special mechanisms in human infants. Experiments on animals' perception of speech have provided critical data on the question of specialized

mechanisms. More generally, theory building has benefited from the interchange between developmental ethologists and developmental speech scientists who have examined each other's methods, data, and arguments.

I. MODULES AND MECHANISMS

There are two very different characterizations of infants' "initial state" regarding speech (Kuhl, 1986a, 1987a). One account argues that, from the start, the perceptual mechanisms underlying speech in infants include a phonetic-level representation of speech. On this view infants are born with an "innate phonetics"—linguistic-unit representations, either segments or features, preexist in the child in some form. It is this representation, one necessarily involving mechanisms that evolved especially for speech, that explains infants' abilities to detect complex equivalences.

The second account is quite different. On this view, there is no preexisting phonetic-level representation of speech; no formal description of phonetic units exists innately. By this account phonetic-level representations are formed later, perhaps as infants begin to map various acoustic forms onto objects and events in the world. According to this model, infants' initial speech perception abilities are attributable to their more general auditory and cognitive abilities.

By this description, the key points for the first account are (a) phonetic-level representation, and (b) specialized mechanisms. The key points for the second account are (a) no phonetic-level representation, and (b) mechanisms that are general.

Before going further, we need to decide what to call the two accounts. There are three dichotomous terms that have been used historically to characterize the two positions: *phonetic* versus *auditory*, *special* versus *general*, and *motor* versus *sensory*. The terms phonetic, special, and motor were used to characterize the first account; the terms auditory, general, and sensory, were used to characterize the second account.

The problem is that the terms associated with each position, while loosely associated, are not mutually exclusive. The term phonetic has been strongly associated with the *Motor Theory*, which argues that the phonetic-level representation can be specified only in motor terms (Liberman & Mattingly, 1985). However, others have not assumed that a phonetic-level representation must be specified in motor terms; instead, it has been argued that such a representation could be specified in auditory terms (Diehl & Kluender, in press), or in an abstract form not specific to either modality (i.e., in an amodal form) (Kuhl & Meltzoff, 1982, 1984; Studdert-Kennedy, 1986). Similarly, describing the alternative account as auditory restricts it to explaining behavior that is exclusively auditory in nature; the main postulate of the theory is that the underlying mechanisms are general rather than specialized. When referring to the two opposing views in this chapter, I will call the first account the Special Mechanism account (hereafter SMA). SMA argues for phonetic-level representations, and consequently for specialized mechanisms, but does not specify their exact form. I term the second account the "General Mechanism account" (hereafter GMA). GMA postulates general mechanisms and no phonetic-level representation.

Finally, it is worth noting that this debate about the nature of the mechanisms underlying the perception of complex signals is not restricted to speech. Ethologists have long favored the notion that complex perception, especially species-typical behavior, is accomplished by specialized neural mechanisms. Moreover, specialization has been advanced as a general theory of the perceptual processing of complex stimuli; it is a theory that now pervades all of the psychology of perception. The theory, advanced by Fodor (1983), centers on the concept of the *module*—the highly specialized neural architecture that does the computational work required to perceive eccentric stimuli. Modules do things like perceive speech, recognize objects, localize sound, track things that move in space, and detect color. Modules have a particular set of properties. They are first and foremost *modularized;* that is, they are separate from other modules, and they use specialized, rather than general-purpose mechanisms, to do their work. They operate only on stimuli of a particular kind (domain specific), their computations depend only on resources that are internal to the module (informationally encapsulated), and they are not accessible by higher-order mechanisms (cognitively impenetrable). Their operations are rapid and mandatory. Most pertinent to this discussion, they are *innate*.

It is not difficult to portray speech as a canonical case of an eccentric stimulus in need of a module. The problems inherent in the nature of speech, the extremely complex mapping between acoustic events and phonetic percepts, and the problem of segmenting the continuous stream of speech into linguistically appropriate units such as phonetic segments or features, appear to be intractable problems for computers (Klatt, 1986). Moreover, the complex equivalences detected by babies—between acoustic events that are physically different but phonetically similar, between the sight of a face and the sound of a voice when they both indicate the same phonetic unit, and between sounds articulated by someone else and then reproduced with our own mouths—are difficult to explain without reference to some kind of specialized mechanisms (Kuhl, 1986b).

A speech module for babies would indeed present a solution, but recent data suggest that the alternative be considered. Experiments on animals' perception of speech shows that they also demonstrate perceptual phenomena such as categorical perception, and that their categorical boundaries also move when the context is changed (Kuhl, 1987b). It is these phenomena that have been used as evidence for an innate speech module in humans. Moreover, recent data on infants' cognitive abilities suggest that they detect complex equivalences outside of the realm of speech, between sensory information presented to different

modalities and even between sensory events and their motor equivalents. Thus, when the signal is speech, certain of the perceptual abilities thought to be species-specific are not; animals display them as well. And when the perceivers are human, complex perceptual abilities are not restricted to speech; stimuli in other domains evoke them as well.

The question is: What do we want to impute to the baby? Do infants come into the world equipped with special mechanisms that provide both a means for detecting phonetic equivalence and a means for segmenting the stream of speech into its component parts (SMA)? Or is there no phonetic-level representation of speech, in which case infants' abialities are attributed to their more general sensory and cognitive abilities(GMA)?

II. FOCUS OF THE DEBATE: THE BASIS OF INFANTS' ABILITIES

The two accounts just described do not take different positions regarding infants' capabilities. Both positions agree that infants' perception of speech shows remarkable sophistication. Instead, the differences lie in how they view the nature of the mechanisms that underlie infants' abilities. Thus the debate centers on the *basis* of behavior. The question is: Are infants' abilities based on special mechanisms (SMA) or more general ones (GMA)?

Traditionally, the basis question has been approached in two different ways. We consider these two ways and add a third.

Consider first the two traditional approaches. The first compares the perception of speech sounds with that of nonspeech sounds that are designed to mimic speech acoustically without being perceived as speech. The second compares the perception of speech by human and nonhuman listeners. I have argued elsewhere (Kuhl, 1986b, 1987a) that while both these traditional approaches address the SMA vs. GMA debate, they do not answer the same question. The distinction made here is a simple point of logic, which is offered to explain a point of view about the contributions of nonspeech tests.

I have argued that studies using nonspeech ask whether the mechanisms underlying speech perception are speech specific. No one disagrees with this point. But having determined the answer to this question we have to decide what we can conclude. Consider the easy case first. If speech and nonspeech findings completely diverge, as they did in early tests of speech perception (e.g., Mattingly, Liberman, Syrdal, & Halwes, 1971), it is sensible to conclude that the mechanisms underlying the phenomenon are speech-specific, and, thus, evolved especially for speech.

A problem emerges, however, if the opposite result obtains, with speech and nonspeech showing complete convergence (Miller et al., 1976; Pisoni, 1977; Pisoni, Carrell, & Gans, 1983). It is logical to conclude from such results that the

mechanisms are *not* speech-specific. However, it cannot be argued unambiguously that such mechanisms did not evolve especially for speech. This argument cannot be made because the terms *speech-specific* and *especially evolved for speech* are not synonymous. Mechanisms could, in principle, have evolved especially for speech without being speech-specific.

Consider the following interpretation: Nonspeech sounds carefully designed to mimic the speech signal are processed as speech because they *fool* the special speech mechanism. Thus, the mechanisms that evolved especially for speech did so in such a way that they did not exclude nonspeech signals (Kuhl, 1978, 1986b). What this leaves us with is a situation in which results showing complete speech-nonspeech convergence can be explained by either alternative. Using the SMA, special mechanisms for speech have evolved but are fooled by nonspeech signals that mimic speech. The GMA argues that speech and nonspeech are processed similarly because there are no special mechanisms and both signals are handled by more general ones.

Therefore, studies involving nonspeech signals are most easily interpreted when the outcomes of the studies show a complete dissociation between speech and nonspeech, as they did in the early studies. Complete divergence of speech and nonspeech is easily interpreted as strong support for the theory that speech requires *special* mechanisms, ones different from those used in the processing of nonspeech signals. When studies on nonspeech demonstrate the opposite, that is, complete convergence between speech and nonspeech, the opposing claim cannot be unambiguously advanced.

Animal studies contribute to the debate in a different way. Tests of speech perception in animals answer a simple question: Can the perceptual phenomenon exist in the absence of mechanisms that evolved especially for speech? If animals replicate speech effects we can assert without ambiguity that special mechanisms are not *necessary* to account for the phenomenon. Animal replications do not prove that special mechanisms are not at work in humans, but the results eliminate the need for positing them to explain specific phenomena.

Both kinds of experiments help build theories. For example, if speech–nonspeech convergence is found with the same stimuli that animals succeed on, we have no reason to impute special processing of these stimuli. Furthermore, if both speech–nonspeech comparisons and animal tests fail at the same *level* of complexity the theoretical implications are strong. It would be at this level that evidence for special speech mechanisms would have been obtained. Taken together, the two approaches provide valuable complementary evidence for theory construction.

The speech phenomena demonstrated by infants will be discussed in the next section and studies done to investigate the basis of infants' abilities will also be described. In most instances studies on the basis of infants' abilities use nonspeech or animals to address the question. To these two traditional methods of testing the underlying basis of the effects in infants I add a third. This new

approach involves asking whether or not similar phenomena have been observed with other than auditory stimuli. In other words, the broader issue of domain specificity will be addressed. If the detection of complex equivalences, such as those involving cross-modal or imitative abilities, are exclusive to speech, then this supports SMA. But if such abilities are demonstrated more generally in infants, and appear to be part of their native cognitive endowment, then there may be no reason to claim that the abilities are part of a specialized subsystem for speech.

III. INFANTS' DETECTION OF COMPLEX EQUIVALENCES

The most striking thing about infants' perception of speech is not their ability to detect fine differences between sounds that will eventually convey meaning, though they do that quite well. The most striking thing is their ability to detect similarity—equivalence—between stimuli that are phonetically equal but physically different. The detection of phonetic equivalence is in fact a critical problem for theory.

It is precisely this problem that causes computers to fail at speech recognition. Speech segments are coarticulated; this means that the acoustic cues for an individual unit vary dramatically depending on the context in which the unit appears. The phonetic unit /b/ in *bat* is not physically identical to the /b/ in *beet, bit, boat,* and *boot*. Yet, as adults we are good at recognizing its equivalence across these contexts—so good that it is difficult to view the detection of equivalence as a problem for theory. Similarly, when we perceive a phonetic unit as being the same regardless of whether the talker is a male, female, or young child, our recognition of the an equivalence is automatic. Yet neither of these perceptual feats can be performed by the most sophisticated computer (Klatt, 1986). The major point for this discussion is that infants are also good at equivalence detection. Solved by babies, but not by machines, the detection of equivalence is a central problem in speech perception; if we understood how it was done it would be a major breakthrough.

There are three classes of phenomena, each involving the detection of complex equivalences for phonetically similar but physically different stimuli, that have been demonstrated in infants. The three classes of phenomena include

(1). Auditory equivalence, the detection of equivalences between two auditory stimuli, as represented by the phenomena of categorical perception (Eimas, Siqueland, Jusczyk, & Vigorito, 1971) and of equivalence classification (Kuhl, 1979a, 1983, 1985a),

(2). Auditory-Visual equivalence, the detection of a correspondence between auditory and visual representations of the same phonetic unit (Kuhl & Meltzoff, 1982, 1984; MacKain, Studdert-Kennedy, Spieker, & Stern, 1983), and

(3). Auditory-Motor equivalence, as demonstrated by vocal imitation, wherein an auditory stimulus produced by someone else evokes the motor movements necessary to reproduce that signal oneself (Kuhl & Meltzoff, 1982, 1988; Lieberman, 1984).

Auditory Equivalence

There are two examples of auditory equivalence demonstrated by infants that theories regarding the initial state of perceptual mechanisms will have to explain. One is the classic phenomenon of *categorical perception* (CP), in which listeners are shown to be more sensitive to changes in a speech stimulus at the boundary between phonetic categories than they are in the middle of the category. The other, *equivalence classification,* is more like what cognitive psychologists classically refer to as categorization. It involves tests of infants' abilities to recognize equivalence between two auditory stimuli that they can easily discriminate.

Categorical Perception (CP)

One of the most significant early findings in favor of innate special mechanisms came from the discovery of categorical perception in infants (Eimas et al., 1971). In adults the phenomenon involved the following demonstration. A continuum of sounds was generated by computer along which an acoustic dimension was altered in small physically equal steps. Tests showed that while the acoustic dimension changed continuously along the continuum in a stepwise fashion, perception was discontinuous. The stimuli were heard as a series of stimuli (e.g., /ra/'s) that changed abruptly to a new series of stimuli (e.g., /la/'s) at some point on the continuum (Fig. 13.1, top). Moreover, the ability to discriminate between

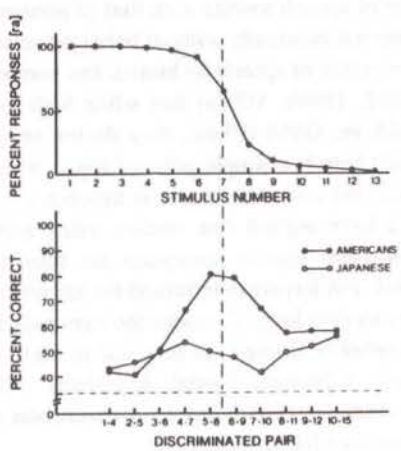

FIG. 13.1. Categorical perception in American and Japanese adults. From Miyawaki et al. (1975).

sounds taken from the series was constrained. Adults could discriminate quite easily between sounds that fell in different categories, but discrimination between sounds in the same category was quite difficult (Fig. 13.1, bottom) (Miyawaki et al., 1975).

Perception was shown to be categorical only for contrasts that were phonemic (made a difference between words) in the adult's language. Japanese adults, for whom the /ra-la/ distinction is not phonemic, did not produce the characteristic peak in the discrimination function for /ra-la/stimuli (Miyawaki et al., 1975). Their ability to discriminate the stimuli hovered near chance (Fig. 13.1, bottom).

This phenomenon immediately raised a question about development: Do infants demonstrate CP initially, or only after experience with a specific language? The question was answered by Eimas et al.'s (1971) work on 1-month-old infants' perception of speech. Eimas showed that infants could discriminate computer-generated sounds that straddled the adult-defined phonetic boundary but failed to discriminate within-category stimuli (Eimas et al., 1971; Eimas, 1974, 1975). Infants' ability to do this in the absence of a protracted period of experience in producing or listening to speech suggested that the phenomenon was not learned. Infants appeared to partition the stimulus continuum just as adults did, right from the start. Further evidence against the learning account came from the finding that infants demonstrated the effect for all phonemic contrasts, whether native to the language environment in which they were raised or not (Aslin, Pisoni, Hennessey, & Perey, 1981; Streeter, 1976). These findings were extremely important because they were the first to suggest that infants might be innately endowed with mechanisms specialized for speech.

Other related phenomena such as *context effects* and *trading relations*, were also investigated with infants. These studies stemmed from early work focused on the acoustic analysis of speech. This early work showed that the acoustic cues underlying phonetic perception were context-dependent (Liberman et al., 1967). The phonetic units surrounding a target unit, the specific talker who produced it, the rate at which it was spoken, and its position in a syllable were shown to alter the acoustic cues that specified a particular phonetic unit. That adult listeners were sensitive to these contextual differences had been shown in studies of CP in adults (Best, Morrongiello, & Robson, 1981; Miller & Liberman, 1979; Summerfield & Haggard, 1977). These studies confirmed the perceptual effect of context by showing that the exact location of the phonetic boundary on a continuum was altered by the context in which the phonetic unit appeared.

An example of a context effect is that provided by a change in the rate of speech. Studies suggest that adult listeners may take rate-of-articulation information into account when making decisions about the phonetic identity of a particular phonetic unit. Miller and Liberman (1979) demonstrated that the location of the boundary between the consonants /b/ and /w/ changed as a function of the duration (and thus, to an adult, the perceived rate) of the syllable. In their first experiment, a /ba-wa/ syllable continuum was lengthened to indicate slower

speech by increasing the duration of the vowel. For this *long* syllable continuum, the boundary was located at a longer transition duration than it was for the *short* syllable continuum. In a second experiment, the syllable was lengthened in a different way, one not associated with a slower rate of articulation. The syllable was lengthened by adding formant transitions to the end of the original vowel, which created the perception of a final consonant on the syllables (/bad-wad/) but did not signal a slowed rate of speaking. Here the effect was reversed; the perceptual boundary moved toward shorter transition durations.

Eimas and Miller (1980) showed that 2- to 3-month-old infants demonstrate one part of this effect. Using the same stimuli used by Miller and Liberman (1979), these authors selected syllables from the long and short /ba-wa/ continua. Syllables were chosen to create four stimulus pairs, including both within-category pairs and between-category pairs from each continuum. Infants were tested using the HAS (high amplitude sucking) technique. The results demonstrated that infants discriminated only the between-category pairs on both continua, thus suggesting that infants are sensitive to contextual information.

Trading relations effects are similar to context effects, but take the argument one step further. In these cases, the cues that are necessary to achieve a particular phonetic percept not only change with the context, but in a specific compensatory way. The value along one acoustic dimension determines the value that is required along a second dimension. The effects of the two dimensions appear to be additive, such that an increased value on the first dimension must be accompanied by a decreased value on the second dimension.

There are two examples of trading relations in infants. Both support the claim that infants are sensitive to compensatory effects. The first case involves a trading relation between the duration of the first formant and the VOT required to perceive a voiceless stop (Summerfield & Haggard, 1977). Miller and Eimas (1983) tested this trading relation on 2- to 3-month-olds using the HAS technique. They created two continua varying in VOT from 5 ms to 55 ms. This variation in VOT is sufficient to change an adult's percept from /ba/ to /pa/. One continuum was constructed with short (25 ms) transitions and the other with long (85 ms) transitions. For adults, the boundary value between voiced and voiceless stops on the short continuum occurred at about 25 ms, while on the long continuum the boundary occurred at about 45 ms. The infants were tested with four pairs of stimuli: 5 ms vs. 35 ms short (perceived as /b/ and /p/ respectively by adults); 35 ms and 55 ms short (both perceived as /p/ by adults); 5 ms vs. 35 ms long (both perceived as /b/ by adults); 35 ms vs. 55 ms long (perceived as /b/ and /p/ respectively by adults). The infants provided evidence of discriminating only the 5 ms vs. 35 ms *short* pair and the 35 ms vs. 55 ms *long* pair. Thus, the data suggest that the location of enhanced discriminability on these two continua occurs at different places for infants, as well as adults, thus providing evidence of trading relations in infants.

A second example of trading relations that has been tested with infants in-

volved the contrast *say* vs. *stay*. Adult studies show that inserting a silent gap between the /s/ and the vowel in the word *say* induces the perception of a voiceless stop, so as to create the word *stay*. More importantly, the length of the silent gap inserted in these situations interacts with the spectral aspects of the remainder of the syllable. Best et al. (1981) synthesized two continua ranging from say to stay. One continuum was synthesized with formant transitions appropriate for /t/ and one was synthesized without these transitions. Best et al. showed that the silent duration required to perceive *stay* varied for the two continua. When the spectral information for "t" was more complete, less silence was required to produce *stay* than when the spectral information for "t" was less well specified. Best et al. argued that this perceptual trading relation was due to the listener's knowledge of the association of these two cues in the production of the sound.

Recently Eimas (1985) provided evidence for this trading relation in infants. He tested 2- to 4-month-olds using stimuli from the two continua. Pairs of stimuli were drawn such that adults perceived the stimulus pair as containing two *say* stimuli, two *stay* stimuli, or one of each. In all cases, discrimination was evidenced to be similar to an adult's; that is, infants failed to detect the difference between two syllables heard by adults to be equivalent (two versions of say, or two of stay), but always provided evidence of discriminating two syllables heard as different by adults.

In summary, the data that are available on context effects and trading relations in infants provide support for the notion that infants are sensitive to these effects. These effects are important to theory because they show that the perceptual boundary between phonetic categories moves. The fact that the boundary is not fixed makes it difficult to attribute these effects to a simple mechanism, and suggests the possibility that the perceived equivalence between acoustic events derives from their common articulatory origin. But infants have not yet produced these sounds, and therefore such motor knowledge has to be argued to be built-in. An alternative view (the GMA) is that these perceptual effects are the result of the functional characteristics of the auditory system. That is to say, it is possible that these perceptual effects derive from the complex way in which the auditory system combines acoustic information in perception, irrespective of its status as speech (Kuhl & Padden, 1983).

The Basis of CP

Tests on Nonspeech Signals. Studies on adults have shown that CP can be replicated using stimuli that mimic speech sounds varying in VOT (Miller et al., 1976; Pisoni, 1977) and for nonspeech analogs mimicking the /ba-wa/ rate effect (Pisoni et al., 1983). The agreement between the speech and nonspeech data for adult listeners naturally led to a strong interest in the performance of infants in discrimination tasks involving nonspeech stimuli. Two of the studies cited above

have been examined in young infants, one involving the nonspeech correlate of VOT, tone-onset time (TOT), and the other, the nonspeech correlate of /ba/ and /wa/ in tests of the context effect of rate. Jusczyk, Pisoni, Walley, and Murray (1980) tested the discrimination of sounds varying in tone-onset time (TOT). The stimuli were synthesized to duplicate those used by Pisoni (1977) on adults. Jusczyk et al. predicted that infants, like adults, would discriminate only those stimuli that straddled the −20 ms or +20 ms TOT boundaries. They tested a number of stimulus pairs: −70 ms vs. −40 ms; −40 ms vs. −10 ms; −30 ms vs. 0 ms; −20 ms vs. +10 ms; −10 ms vs. +20 ms; 0 ms vs. +30 ms; +10 ms vs. +40 ms; and +40 ms vs. +70 ms.

Contrary to their prediction, the sucking recovery scores indicated that only the −70 ms vs. −40 ms and +40 ms vs. +70 ms contrasts were discriminated. The data provided support for the notion that infants perceive three categories on the TOT continuum, but the data suggested that the boundaries for these categories were located in different places for infants than for adults, and that in infants the TOT nonspeech boundary does not coincide with the VOT speech boundary. The fact that discrimination was symmetrical, that on both sides of the continuum discrimination was not evidenced until the tones were temporally offset by at least 40 ms, suggests that infants may require a longer interval between the onsets of two tones before perceiving them as nonsimultaneous. The result is an important one, because it shows a dissociation between speech and nonspeech in infants that is not present in adults.

In a recent report, Jusczyk, Pisoni, Reed, Fernald, and Myers (1983) replicated with infants the context effect involving rate using the nonspeech analogs of /ba/ and /wa/. The results demonstrated that, just as when listening to speech, infants needed a shorter transition duration to detect a change in *short* stimuli, and a longer transition duration for detecting a change in *long* stimuli. This means that infants are sensitive to overall duration in nonspeech as well as in speech, and that transition duration is processed relationally for both signals.

Thus, the results for infants' perception of nonspeech are mixed, with the study of the context effect of rate (Jusczyk et al., 1983) supporting a close agreement between speech and nonspeech, and the TOT study (Jusczyk et al., 1980) failing to do so. A strong conclusion about the agreement between speech and nonspeech data by infants is not possible at this time; further experiments addressing this issue are needed.

Tests on CP in Animals. The research completed on animals' perception of speech is at present quite extensive (Kuhl, 1986b); three examples are cited here to illustrate the findings. The first data are from the first study that examined an animal's ability to categorize sounds from a speech-sound continuum (Kuhl & Miller, 1975). This test focused on the characteristic labeling functions obtained in speech. The question for animals was whether the boundary between the two categories on the continuum coincided with the phonetic one, or appeared some-

place else. The second data are more recent and focus directly on tests of the "phoneme boundary effect" (Kuhl & Padden, 1982, 1983). The question here is whether or not in the absence of any experience in *labeling* the stimuli on the continuum, an animal will demonstrate enhanced discriminability at the boundaries between phonetic categories, like human infants do. The third is our most recent result (Stevens, Kuhl, & Padden, 1988) and it concerns the context effect of rate demonstrated with the syllables /ba/ and /wa/.

The Kuhl and Miller (1975) study resembled an adult categorization experiment, only with animals. The question was: Where would an animal place the boundary on a phonetic continuum? Chinchillas were trained to distinguish computer-synthesized versions of the two endpoint stimuli on a /da-ta/ continuum, 0 msec VOT and +80 msec VOT. During training, they were not given any exposure to the rest of the test continuum. When performance on these endpoint stimuli was near perfect, a generalization paradigm was used to test the intermediate stimuli, those between /da/ and /ta/ on the continuum (+10 msec VOT to +70 msec VOT, in 10-msec steps). The design of the experiment was that during generalization testing, half of the trials would involve the endpoint stimuli. On these trials, all of the appropriate feedback was given, just as it had been during the training phase. On the other half of the trials, the intermediate stimuli were presented.

The intermediate trials were the ones most critical for theory. On trials involving intermediate stimuli, the feedback was arranged to indicate that the animal was always correct, no matter what the response. There was no training on these stimuli, and thus no clue was provided to the animal telling him how to respond and thus where to place the boundary on the continuum.

The data are shown in Fig. 13.2 (top). The mean percentage of /da/ responses to each stimulus on the continuum are plotted for chinchillas and human adults. The curves were generated by the same least-squares method. The resulting phonetic boundaries, located at 35.2 msec VOT for humans and 33.3 msec VOT for animals, did not differ significantly. A subsequent study using a totally different procedure and monkeys rather than chinchillas demonstrated that the location of the boundary on a /da-ta/ continuum was located at +28 msec, in good agreement with the chinchilla data (Waters & Wilson, 1976).

Kuhl and Miller (1978) extended these tests to continua involving other voiced-voiceless pairs, namely bilabial (/ba-pa/) and velar (/ga-ka/) contrasts. These stimuli were of interest because human listeners' boundaries differ with the place of articulation specified by the particular voiced-voiceless pair. The new tests involving the bilabial and velar stimuli were run exactly as the previous ones. The endpoint VOT values were 0 and +80 msec. The intermediate stimuli (+10 to +70 msec in 10-msec steps) were presented with feedback indicating that the animal was correct regardless of his performance. Thus, no training occurred on these stimuli. The results again demonstrated excellent agreement between the human and animal categorization data. The boundary values for the

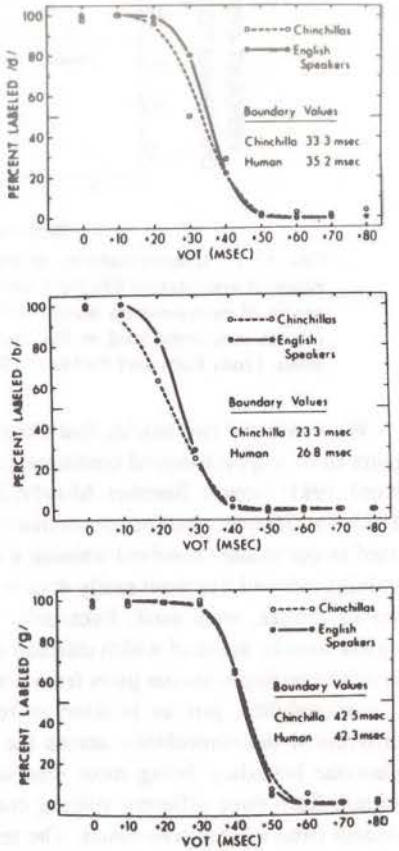

FIG. 13.2. Discrimination of sounds from speech continua by animals and human adults tested on /d-t/ (top), /b-p/ (middle), and /g-k/ (bottom) stimuli. The locations of the boundaries for the two groups did not differ significantly. From Kuhl and Miller (1978).

bilabial stimuli were 26.8 msec VOT for humans and 23.3 msec VOT for animals (Fig. 13.2, middle), which were not significantly different. The boundary values for the velar stimuli were 42.3 msec VOT for humans and 42.5 msec VOT for animals (Fig. 13.2, bottom). Again, the values did not differ significantly.

Taken together, the data suggested that animals' natural boundaries coincided with humans' phonetic ones, but to this point no studies had been done on animals' discrimination of specific pairs of stimuli from the continuum. Since it is the enhanced discriminability between categories—the phoneme boundary effect—that sets speech apart from other phenomena in psychophysics and in cognitive psychology, and since infants appear to demonstrate this effect without learning to experience (Eimas et al., 1971) discrimination tests were considered important.

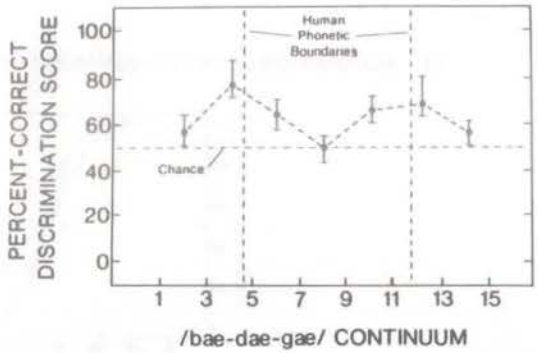

FIG. 13.3. Discriminability of pairs of speech stimuli taken from a place-of-articulation (/b-d-g/) continuum. The bars indicate the entire *range* of performance across animals for each stimulus pair. Performance was enhanced at the locations of humans' phonetic boundaries. From Kuhl and Padden (1983).

We conducted two studies that directly addressed discriminability of stimulus pairs from a speech-sound continuum (Kuhl & Padden, 1982, 1983). (See also Kuhl, 1981; Sinnott, Beecher, Moody, & Stebbins, 1976; and Morse & Snowdon, 1975, for different types of speech discrimination tests on animals). The technique used in our studies involved training a monkey on a same-different task. During training, stimuli that were easily discriminable like a tone versus a noise or a click versus a buzz, were used. Eventually the monkey had to discriminate various vowel sounds, some of which differed only in intensity or pitch. The experiment involved testing stimulus pairs from a speech sound continuum to examine their discriminability, just as is done in tests on infants. The CP model predicts differential discriminability across the continuum, with pairs that straddle the phonetic boundary being most discriminable. Kuhl and Padden (1982) used stimuli from three different voicing continua and Kuhl and Padden (1983) used stimuli from a place continuum. The results of the two experiments were identical—in both cases animals demonstrated the discrimination typical of CP.

The results from the /bæ-dæ-gæ/ experiment are shown in Fig. 13.3. The average percentage correct discrimination score is given for each pair. The locations of the human phonetic boundaries are marked by dashed vertical lines. As shown, the best performance occurred on stimulus pairs 3 vs. 5, 9 vs. 11, and 11 vs. 13. These are the only pairs that differ significantly from chance, and involve stimuli from different phonetic categories for humans. Thus, while stimulus pairs were always separated by an equal physical distance on the continuum, their perceived differences were not equivalent. Discriminability was poor when the stimuli involved pairs taken from the same involved pairs taken from different phonetic categories.

Most recently, our tests on animals have been extended to the phenomenon of

"context effects." The particular example that we have tested involved the /ba-wa/ distinction and its dependence on rate (Stevens, Kuhl, & Padden, 1988). Recall that when the two syllables /ba/ and /wa/ are produced at *fast* as opposed to *slow* rates of speech, the boundary between them is located at two different places on the respective continua. On the fast continuum, shorter transition durations are required to change the percept from /ba/ to /wa/; on the slow continuum, longer transition durations are required to change the percept from /ba/ to /wa/. The problem for a theory of speech perception is to explain how the underlying mechanisms specify a different boundary for fast versus slow speech.

Our tests on macaques used the same stimuli used to test infants by Eimas and Miller (1980). The stimulus pairs had been chosen by these authors so as to include pairs that straddled the adult-defined boundaries on both the fast and slow /ba-wa/ continua, and pairs that fell within a single phonetic category. These same pairs were used to test macaques. Our results mirrored those found with infants. Macaques discriminated only the pairs that straddled adult human boundaries, while failing to show discrimination of pairs of stimuli that fell within a single phonetic category (Stevens, Kuhl, & Padden, 1988). Thus, it appears that macaques also show context effects for speech; their boundaries on speech continua move when the context is changed.

Taken as a whole, these data lend support to the notion that enhanced discrimination near phonetic boundaries can be demonstrated in mammals other than man. I have argued elsewhere that this finding supports two conclusions, one about evolution and the other about infant performance (Kuhl, 1986b; Kuhl & Padden, 1983). First, in the evolution of language, the choice of the particular phonetic units used in communication was strongly influenced by the extent to which the units were ideally suited to the auditory system (Kuhl, 1988; Stevens, in press). It has been argued (Kuhl, 1979b, 1981; Kuhl & Padden, 1983; Stevens, 1972, 1981) that the perception of certain auditory properties, such as spectral shape, detection of rapid formant change, and temporal order, served as a set of constraints on the acoustics of language. The second conclusion is that since animals demonstrate these speech phenomena, the fact that infants do so is not sufficient evidence *by itself* to support the notion that the mechanisms underlying the effects in infants are ones that evolved especially for speech. Animals demonstrate these effects in the absence of special mechanisms; infants may do so as well.

Equivalence Classification

A second kind of auditory equivalence is demonstrated by infants, one more like what cognitive psychologists call "categorization"—the ability to "render discriminably different things equivalent" (Bruner, Goodnow, & Austin, 1956).

Categorization is a phenomenon that characterizes all of perception. As stimuli typically vary along many dimensions, categorization requires that we recog-

nize similarities in the presence of considerable variance. Often the exact criteria used to categorize are not obvious. Consider the categories *cat* and *dog*. Describing what distinguishes them, and thus what uniquely categorizes them, is not simple. They both have two eyes, two ears, four legs, fur, a tail, and so on. Configurational properties of the face probably distinguish them, but trying to describe these features is difficult. Yet we would not expect an adult to mistakenly identify a cat as a dog, or vice versa.

In speech, a similar categorization problem exists. Take a simple example, such as the vowel categories /a/ as in "cot" and /æ/ as in "cat." The differences between the two vowels are not subtle to the human ear; they are clearly different. But trying to program a computer to identify these vowels correctly when they are spoken by different individuals demonstrates it to be a very difficult problem. When different talkers produce these two vowels, there is overlap in the physical cues, the formant frequencies, that represent the two categories. The explanation for this has to do with the fact that people with different-sized vocal tracts (like males, females and children) produce different resonance frequencies when they create the same mouth shape. Thus far, no one has successfully described an algorithm that correctly recovers which of the vowels a speaker produced when acoustic information (the formant frequency values) is the only thing provided. In humans, various attempts to explain the processes by which we normalize the speech produced by different talkers have been offered; most of them involve computation of some kind (Lieberman, 1984, for review).

The critical question for the current discussion is whether infants recognize equivalence when the same vowel is produced by different talkers. Are all /a/'s the same to the baby, regardless of the talker who produced them? It is of no small import to the child that such an ability exists early in life. Vocal-tract normalization is critical to the infant's acquisition of speech. Their vocal tracts cannot produce the frequencies produced by the adult's vocal tract, so infants must normalize speech to imitate it.

How can the question be posed to the infant? We want to know if they can sort vowel sounds into two categories. When we ask whether infants can categorize, we want evidence that they can perceptually group a variety of instances into Type A and Type B events, even though the various A's (or B's) are clearly differentiable. To perform such a task requires that infants recognize similarity among discriminably different instances representing the A (and B) category, while ignoring the irrelevant differences between the various A's (and B's). Thus, categorization requires a process in which the perceiver perceptually establishes two groups of stimuli; in each category equivalence must be detected while irrelevant variations (though discriminable) must be ignored.

Two main features distinguish tests of "equivalence classification" from the tests of CP. First, the stimuli representing the categories are discriminably different. In tests of CP, the stimuli differ on a single acoustic parameter, and thus evidence of categorization is taken from infants' failure to evidence discrimination. In tests of equivalence classification, the stimuli vary along a number of dimensions and are clearly discriminable. Thus, categorization, if it occurs, cannot be attributed to a failure to discriminate the stimuli. In the first speech experiment on equivalence classification with infants, Kuhl (1979a) examined infants' discrimination of two vowel categories, /a/ (as in *pop*) and /i/ (as in *peep*). The stimuli used in the experiment varied along three dimensions, phonetic identity (/a/ versus /i/), pitch contour (rising versus falling), and talker identity (male, female, or child). Stimuli belonging to the same category (all /a/'s for example) were shown to be easily discriminable from one another by infants. The question was: Can infants perceptually group all of the /a/'s and all of the /i/'s?

Second, the categorization approach requires the infant to produce an equivalent response to stimuli that are perceived to be equivalent, rather than to produce a response based on the detection of a difference between two stimuli. In order to explore infants' abilities to categorize, a technique had to be developed that required the infant to report the perception of *similarity* rather than to report the perception of a *difference*. Because every member of the category is perceptually different, having infants' responses depend on their perception of a difference would not allow one to address the categorization question. Instead, we wanted infants to signal that they heard a similarity between a novel stimulus and a stimulus they heard previously. Moreover, we wanted this perception of *sameness* not to be based on a failure to discriminate the two stimuli.

A technique was developed by Kuhl (1985b) that achieves this goal. It uses a simple conditioning procedure that is shown in Fig. 13.4. The infant sits on a parent's lap and is visually engaged by an assistant, who manipulates toys silently. A speech sound, such as the vowel sound /a/, plays repeatedly from the loudspeaker at the infant's left. The infant quickly learned that when the sound changes from the vowel /a/ to the vowel /i/ a bear playing a drum inside a black box on top of the loudspeaker is turned on for a short period of time. Eventually the infant anticipates the occurrence of the bear and produces a head-turn in the direction of the box when the sound /i/ is played.

Once trained, the infant produces head-turning responses only when /i/ vowels occur, and does not turn during presentations of the vowel /a/. We want to know what infants will do when they are presented with new instances of /a/ and /i/vowels, instances clearly different from the /a/ and /i/ stimuli heard during training. To find out, infants were initially trained to make a response when an /a/ vowel, produced by a male voice with a falling pitch contour, was changed to an /i/ vowel. The two stimuli were acoustically matched in every other detail. After this initial training, infants were tested with novel stimuli representing the two categories, ones produced by female and child talkers, with either rising or falling pitch contours. All of the novel stimuli differed perceptually from the two initial training stimuli, and infants were shown to be capable of discriminating all of the /a/'s from one another. The hypothesis was that the infant's initial training to respond to a single /i/ sound would generalize to all members of the category; that is, we argued that if infants perceived all /i/'s as

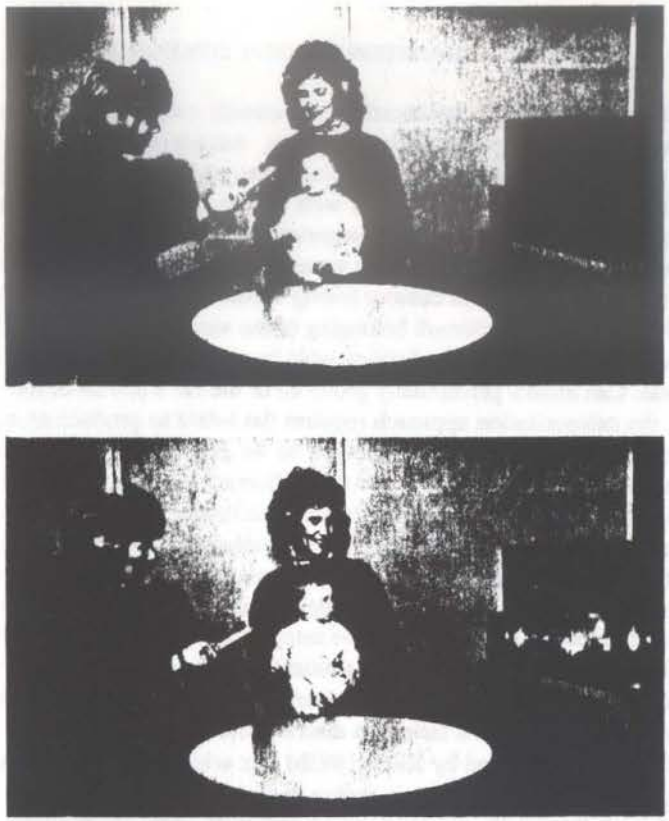

FIG. 13.4. A head-turn technique used to test infants' categorization of speech sounds. Infants are trained to produce a head turn toward the loudspeaker on the infant's left when a speech sound from one phonetic category is changed to a sound from a second phonetic category. If infants do so at the appropriate time, a visual reinforcer is presented. Once training is complete, novel stimuli from both categories are presented to test infants' abilities to categorize them. From Kuhl (1986b).

perceptually equivalent, then if the infant had been trained to produce a head-turn response to the male's /i/ vowel, but not to his /a/ vowel, the infant would produce that response to all novel /i/'s (ones produced by females or children), but not to equally novel /a/'s.

The results show that this hypothesis was correct (Kuhl, 1979a). Infants responded correctly to the novel vowels. If the infant had been trained to turn to the male's /a/, then all novel /a/'s evoked the response, while none of the novel /i/'s did. The same was true if infants were trained to turn to the male's /i/—all novel /i/'s evoked the response, but not the novel /a/'s. Figure 13.5 shows the

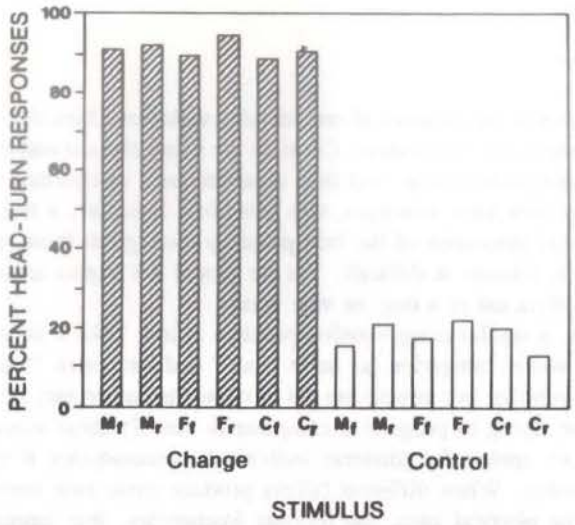

FIG. 13.5. Group data for the /a-i/ vowel categorization experiment. Per cent head-turn responses to each novel stimulus that belonged to the phonetic category that was initially reinforced (Change stimuli) and the phonetic category that was not initially reinforced (Control stimuli). The training stimuli were the far left "Change" stimulus and the far left "Control" stimulus, each produced by a male (M) voice with a falling (f) pitch contour. After training, novel vowels produced by female (F) and child (C) talkers, with either rising (r) or falling pitch contours, were introduced. The data show excellent generalization from the training stimuli to novel stimuli representing the same phonetic category. From Kuhl (1987).

per cent head-turn responses to all of the stimuli introduced in the experiment. As shown, infants produced head-turn responses only to the stimuli that were members of the phonetic category that they were initially trained to respond to. They failed to produce head-turn responses to equally novel stimuli that were not members of the phonetic category that they were trained to respond to. More surprisingly, an analysis of infants' first-trial responses showed that infants performed correctly on the very first trial. These results suggest that six-month-old infants categorize all /a/'s (and all /i/'s) as the same.

Kuhl (1983) extended these results to vowel categories that are much more similar from an acoustic standpoint and therefore much more difficult to categorize. The vowels were synthesized versions of /a/ (as in *cot*) and /ɔ/ (as in *caught*). In naturally produced words containing these vowels, the overlap in the first two formant frequencies is so extensive that the two categories cannot be separated on this acoustic dimension (Peterson & Barney, 1952). Moreover, in most dialects used in the United States, talkers do not distinguish between the two vowels. The experiment was run just as before. Infants were trained on the

/a/ and /ɔ/ vowels spoken by a male talker. Then, novel vowels spoken by female and child talkers, with additional random changes in the pitch contours of these vowels, were introduced. Results of the /a–ɔ/ study showed that most of the infants performed as well as infants in the /a–i/ experiments had performed. For these infants correct performance to the novel vowels occurred on the first trial (Kuhl, 1983).

In a study just completed (Kuhl, Wolak, & Green, in preparation) we examined another difficult vowel contrast, /a/ as in cot and /æ/, as in cat. To make the test more similar to the situation typically experienced by infants growing up, all of the vowels were spoken naturally, as opposed to our previous tests, in which the vowels were computer-synthesized. The number of talkers was also increased from three (previous studies) to twelve. The twelve talkers included 5 males, 5 females, and 2 children. The variation in voices was quite astounding, even to the adult ear, and this made it all the more difficult to attend to the essential differences between the vowels /a/ and /æ/. Adults who were tested in this task told us that they had to pay attention in order to perform perfectly. How did infants fare? Six-month-olds tested on this task performed at about 76% correct, significantly above the 50% chance level ($p < 0.01$).

Thus, studies of vowel categories show that by six months of age infants recognize equivalence classes that conform to the vowel categories of English. Given the demonstration that infants categorize variants for an easily discriminable contrast (/a-i/), as well as for difficult contrasts (/a-ɔ/ and /a-æ/), infants probably demonstrate this vowel constancy for all vowel categories in English.

How do infants recognize speech categories? We have obtained preliminary evidence that infants' vowel categories may be organized around a central "good" stimulus—a prototype of the category (Grieser & Kuhl, 1989). Our evidence consists of the results of studies that show that infants' spontaneous tendencies to generalize to new instances of a vowel category are affected by the perceived goodness (to an adult ear) of the stimulus that infants were initially trained to respond to.

The experiment was conducted in the following way. We synthesized a variety of stimuli that conformed to points in a two-formant coordinate vowel space within the /i/ vowel category. We asked adults to rate the *goodness* of these vowels on a scale of 1 to 7. Based on these ratings, we chose one stimulus that was rated as a good version of the /i/ vowel and another rated as a poor version of /i/. Even the poor one was always classified as an /i/ rather than as some other vowel. We then synthesized 32 variants around the good and poor /i/ vowel stimuli. These variants formed "rings" around the stimuli. Each ring was a specified distance in mels[2] from these points (30, 60, 90 and 120 mels). On each ring, eight stimuli were synthesized, for a total of 32. Once again, adults were

[2]The mel scale equates for perceived changes in pitch at a variety of different frequencies. Using a mel scale to specify the vowel's formants was an attempt to make all stimuli on each of the rings equally different from the target vowel.

asked to rate the goodness of each of the variants (N = 64) around both the good and poor versions of /i/.

The tests on infants were designed to examine whether generalization around a good stimulus differed from generalization around a poor stimulus. The head-turn technique was used to examine infants' generalization to variants around the two points. The results showed that generalization around the good stimulus was significantly broader than generalization around the poor stimulus. When trained on a good variant of /i/, infants responded to many more stimuli in vowel space than they did when trained on a poor variant of /i/. In fact, the same /i/ stimuli were responded to differently depending upon whether the infant had been trained to respond to a good as opposed to a poor /i/ stimulus.

These studies suggest that some points in vowel space are ideal candidates for category centers, because they are associated with perceptual stability over a broad array of category variants. Other points in vowel space are poor candidates, as perception is not stable and generalization to novel exemplars is weak. These data support the notion first expressed by Stevens (1972), who argued that vowel categories were organized so as to take advantage of the quantal nature of perception. This phenomenon is consistent with prototype theory (Medin & Barsalow, 1987; Rosch, 1975), and is the first data that we are aware of that suggest that infants' speech categories demonstrate internal structure and organization by 6-months-of-age—and, in particular, that speech may be represented by prototypes in infants.

We turn now to studies of equivalence classification on consonant classes. They are of great interest theoretically. Because consonants cannot occur in isolation, and have to be coarticulated with vowels, the acoustic cues to a particular consonant vary a great deal depending on the vowel that precedes or follows it (Öhman, 1966, but see Stevens & Blumstein, 1981, for a description of cues that may be invariant across context). Thus, while we hear the /b/ in *bat, but, boot,* and even those in *tab, tub,* and *tube* as the same segment /b/, there is no theory that explains this, and computers are unable to identify a given segment as the same across different contexts. We therefore want to know whether infants perceive the similarity between consonants across vowel contexts.

Studies of equivalence classification for consonant classes have been undertaken in our lab (Hillenbrand, 1983, 1984; Kuhl, 1980). These experiments used the same basic design as the tests on vowels just described. Infants were trained to differentiate two CV syllables, whose initial consonants differed. In Kuhl (1980), we reported experiments on fricatives in which syllables beginning with /s/, as in *sell,* were contrasted with syllables beginning with /ʃ/, as in *shell.* During training the consonants were spoken by a female talker and they appeared in an /a/ vowel context (/sa/ vs. /ʃa/). Once training on these syllables was complete, novel /s/ and /ʃ/ syllables were introduced. These new syllables differed both in the talker who produced them (2 male and 2 female talkers) and the vowel context in which the consonants appeared (/a/, /i/, and /u/). Thus, infants not only had to recognize the consonant regardless of the talker, they also

had to recognize the consonant regardless of the vowel context in which it appeared.

In all, 22 novel syllables were introduced. The infant's task was to categorize the novel syllables by producing a head-turn response to those beginning with one of the consonants (either /s/ or /ʃ/) and to inhibit the head-turn response to the opposite category. The experiments were also run with the fricatives /f/ and /θ/, and in the initial as well as in the final positions of syllables. (See Kuhl, 1980, for full details.)

It was a challenging task, more difficult than any of the vowel categorization tests in which the within-category variation was not as extreme. Yet infants performed well in the task, with some of them producing a near-errorless performance (Kuhl, 1980). It was clear that infants were capable of recognizing that the /s/'s in /si/, /sa/, and /su/ were the same and that all were distinct from the /ʃ/'s in /ʃi/, /ʃa/, and /ʃu/, which were themselves the same.

A complete discussion of the issues raised by the data is beyond the scope of this chapter. The main issue posed concerns segmentation: Do these results suggest that infants segment the syllables into two component parts, consisting of a consonant (/C/) and a vowel (/V/), and that the basis of category recognition is the common consonantal segment contained in each? If so, it would provide strong evidence in support of a phonetic level of representation for infants. Such an explanation is consistent with their performance, but we cannot go this far in explaining infants' performance on these tests (Kuhl, 1985a, 1986c). The conservative posture claims only that infants hear similarities between the initial (or final, since we tested both) "portions of the syllables" (Kuhl, 1985a). We do not know that infants hear two segmental events in these syllables, and that one of them, the consonant, is the same. It will take further experiments to decide this. For now, however, we can say that there is some evidence that infants can cope with the extreme variations in the acoustic cues underlying consonants, and this is impressive.

Experiments on equivalence classification have thus demonstrated infant's abilities to perceive a constancy of sorts for speech sounds. Infants recognize auditory equivalence for vowels spoken by different people, and for consonants in different contexts. At least in the case of vowels, we have evidence to suggest that infants' speech categories are organized around a *good* stimulus. These abilities are remarkable, and must be of enormous help in the infant's acquisition of phonology.

The Basis of Equivalence Classification

There are no nonspeech studies analogous to equivalence classification. There are, however, some data on animals' abilities to perform in tasks of this kind. In addition, there are data suggesting that equivalence classification is not restricted to the domain of speech, but occurs for other classes of stimuli as well.

Early studies on animals (Baru, 1975; Burdick & Miller, 1975; Kuhl & Miller, 1975) involved tests that were similar to "equivalence classification." In these tests, animals were trained on a subset of sounds from two different categories and then tested for generalization to novel members of both categories. Kuhl and Miller used chinchillas to test categories of CV syllables differing in voicing (/t/ vs. /d/), with the talker and vowel context varying. Burdick and Miller and Baru reported on the perception of the vowels /a/ and /i/, with talker varying. These studies were the first to demonstrate that animals could learn to respond correctly to discriminably different instances representing a phonetic category, including novel ones.

Recent studies have provided additional support for the idea that non-human animals perceptually group speech sounds from a given phonetic category together. Kluender and Diehl (1987), for example, have shown that Japanese quail learn to categorize natural consonant-vowel syllables beginning with /d/, as opposed to /b/ or /g/, and that this learning generalizes to syllables having different vowels.

It is also worth noting that the ability to detect equivalences between discriminably different members of a category by infants has been shown for stimuli outside the domain of speech. Cohen and Strauss (1979) showed that 7-month-olds could form a category of a specific female, regardless of her orientation, or of female faces in general. Infants can form categories based on stimulus configuration (Milewski, 1979), the general characteristics of human faces (Fagan, 1976), and possibly even number (Starkey, Spelke, & Gelman, 1983). In other words, infants' abilities to recognize equivalence among discriminably different stimuli that belong to a category are not specific to speech; they are illustrated in many different perceptual domains.

Auditory-Visual Equivalence

Thus far infants' detection of equivalence for diverse auditory events has been discussed. Now we extend the discussion to the detection of cross-modal equivalence for speech, wherein categorization abilities go beyond those involving auditory perception. Recent studies on adults from our own lab (Green & Kuhl, in press; Kuhl, Green, & Meltzoff, 1988; Grant, Ardell, Kuhl, & Sparks, 1985) and others (McGurk & MacDonald, 1976; Massaro & Cohen, 1983; Green & Miller, 1985; Summerfield, 1979) show that the perception of speech is strongly influenced by information gleaned from watching the face of a talker. This raises profound problems for a theory of speech perception because it means that visual information, such as watching a talker's lips come together to produce the consonant /b/, is somehow equated in perception to acoustic information that auditorially signals the consonant /b/. (See Kuhl & Meltzoff, 1988, for discussion.) One important question about such complex "cross-modal" equivalences is how information as different as the sight of a person producing speech, and the auditory speech event that is the result of production, come to be related in development.

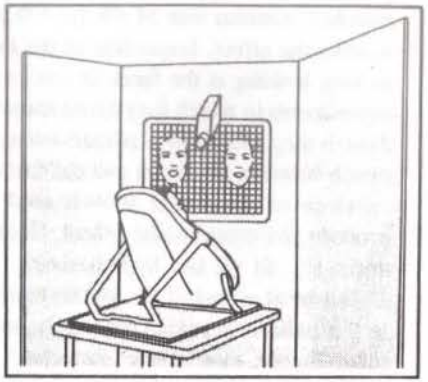

FIG. 13.6. Experimental set-up used to test the cross-modal perception of speech in infants. Infants view two faces producing the vowels /a/ and /i/ while a single sound (either /a/ or /i/) is presented from a loudspeaker located midway between the two facial images. (From Kuhl and Meltzoff, 1982.)

We designed an experiment to pose this problem to infants. We asked whether infants could relate the sight of a person producing a speech sound to the auditory concomitant of that event (Kuhl & Meltzoff, 1982). Infants were shown two filmed faces, side by side, of a woman articulating two different vowel sounds (Fig. 13.6). One face displayed productions of the vowel /a/, the other of the vowel /i/. While viewing the two faces, a single sound, either /a/ or /i/, was presented from a loudspeaker located midway between the two facial images. This eliminated any spatial cues as to which of the two faces produced the sound. The two facial images articulating the sounds moved in perfect synchrony with one another; the lips opened and closed at the exact same time, thus eliminating any temporal cues. The only way an infant could solve the problem was by recognizing a correspondence between the sound and the mouth shape that normally caused that sound. In other words, infants had to perceive a cross-modal match between the auditory and visual representations of speech.

Thirty-two infants ranging in age from 18 to 20 weeks were tested. They were placed in an infant seat facing a three-sided cubicle (Fig. 13.6). The experiment had two phases, a familiarization phase and a test phase. During familiarization, infants saw each of the two faces for 10 sec in the absence of sound. Following this, both faces were presented side by side, and the sound was turned on. Infants were video- and audio-recorded. An observer who was uninformed about the stimulus conditions scored the videotaped infants' visual fixations to the right or left stimulus.

The hypothesis was that infants would prefer to look at the face that *matched* the sound. The results confirmed this prediction; infants looked longer at the face that matched the vowel they heard. Infants presented with the auditory /a/ looked longer at the face articulating /a/; those who heard /i/ looked longer at the face articulating /i/. The effect was strong—of the total looking time, 73% was spent on the matched face ($p < 0.001$) and 24 of the 32 infants demonstrated the effect ($p < 0.01$). There were no other significant effects—no preference for the face

located on the infant's right as opposed to the infant's left side, or for the /a/ face as opposed to the /i/ face. There was no significant difference in the strength of the effect when the matching stimulus was located on the infant's right as opposed to the infant's left. (See Kuhl & Meltzoff, 1984a, for full details.)

We then replicated the findings with 32 additional infants and a new research team (Kuhl & Meltzoff, 1984b). All other details of the experiment were identical. The results again showed that infants looked longer at the face that matched the sound they heard. Of the total fixation time, infants spent 62.8% fixating the matched face ($p < 0.05$), and 23 of the 32 infants demonstrated the effect ($p < 0.01$). Recently another team of investigators has also replicated this cross-modal matching effect for speech using disyllables such as *mama* versus *lulu* and *baby* versus *zuzi* in a design similar to ours (MacKain et al., 1983).

Most recently we have extended the tests to another vowel pair (/i-u/), thus including the third "point" vowel in the set of vowels tested. The point vowels are maximally distinct, both acoustically and articulatorily, and occur at the three endpoints of the triangle which defines "vowel space" (Peterson & Barney, 1952). The test was conducted just as it had been previously, only this time infants watched faces producing the vowels /i/ and /u/, and listened to either /i/ or /u/ vowels. The results showed that the effect could be extended to a new vowel pair. The mean percentage of fixation time to the matched face was 63.8% ($p < 0.05$), and 21 of the 32 infants looked longer at the matched face ($p < 0.05$) (Kuhl & Meltzoff, 1984b).

Thus, 4-month-olds perceive auditory-visual equivalents for speech. They recognize that /a/ sounds "go with" wide-open mouths, /i/sounds with retracted lips, and /u/ sounds with pursed lips. What accounts for infants' cross-modal speech perception abilities? Have infants learned to associate an open mouth with the sound pattern /a/ and retracted lips with /i/ simply by watching talkers speak? Does some other kind of experience play a role in this ability? Our tests are now being conducted on younger infants to examine the learning account; we are specifically interested in whether or not experience in babbling plays a role in the effect (Kuhl & Meltzoff, 1984a).

Presuming for the moment that these effects can be demonstrated quite early, thus reducing the possibility that learning explains them, two theoretical possibilities were suggested (Kuhl & Meltzoff, 1984a). One is that the effect derives from a phonetic representation of speech such as that suggested by the SMA, the other that the effect is independent of speech, and based on infants' more general cognitive abilities (GMA). The main postulate of SMA is that the perceived match between the auditory and visual information is based on mediation by a representation of the phonetic unit, in this case, the vowels /a/ and /i/ in a form that specified both their auditory and visual instantiations. These representations would account for the detection of equivalence in the phonetic information perceived through the two modalities; the representation links the two stimuli.

Our second account, similar to the GMA, was very different. We argued that

it was possible that the auditory and visual speech information was related by some other property, one that directly tied information such as the formant frequencies of the sound /a/ to the sight of a wide-open mouth. This account held that mediation at the phonetic level was not necessary to perceive a match between the two stimuli; it might be done on the basis of simple physics (Kuhl & Meltzoff, 1984b). A series of experiments aimed at separating the two explanations was designed.

The Basis of Auditory-Visual Equivalence

Our first question about the effect was whether or not a nonspeech sound that mimicked certain features of the auditory stimulus could replace it in the cross-modal test. We argued that the use of nonspeech stimuli helped identify what aspect of the auditory signal was necessary and sufficient to evoke the matching response. Was it necessary that the auditory signal contain enough information to identify the vowel or was a single feature of the vowel, presented in a nonspeech context, sufficient?

The nonspeech tests were conducted in two steps. The first was to verify that the cross-modal matching effect depended upon the spectral information in the vowels rather than temporal information. Vowels are defined primarily in terms of spectral information (formant frequencies) rather than in terms of temporal or amplitude information, so it was important to test whether the spectral information in the vowels was essential. Because the auditory and visual vowel stimuli had been matched on all temporal and amplitude parameters, we assumed that infants' matches must be based on the spectral differences between the /a/ and /i/ vowels. We thus hypothesized that if we altered this spectral information, taking the formant frequencies out of the sounds, infants could no longer succeed on the cross-modal task.

To test this directly, the /a/ and /i/ vowels used in our original study were altered to remove their formant frequencies while leaving whatever temporal and amplitude information remained. Using computer analysis techniques, we extracted the amplitude envelopes of the vowels and their precise durations. Then we computer synthesized pure-tone stimuli with a frequency of 200 Hz (the average value of the female talker's fundamental frequency), one for each of the 40 original /a/ and /i/ vowels. Each pure-tone stimulus exactly followed the amplitude envelope of its speech-stimulus original.

These pure-tone stimuli could not be identified as /a/ or /i/, yet when they were played while looking at the faces, the resulting display was quite engaging. Because the temporal properties of the tones matched the original vowels, the tones became louder as the mouths grew wider and softer as the mouths drew to a close. Thus, if infants in our task could discover a match between auditory and visual stimuli on time-intensity cues alone, they should succeed. If, however, the spectral properties of the vowels were necessary, the results should drop to

chance. Arguing that the temporal-envelope properties of the stimuli were insufficient for success in our original experiment, we favored the spectral hypothesis.

The results were in support of the spectral hypothesis; infants' cross-modal performance dropped to chance. The mean percentage fixation time to the matched stimulus was 54.6% ($p > 0.50$), with only 17 of the 32 infants demonstrating the effect. Inspection of the looking data revealed that infants spent just as long looking at the faces in this experiment as they had in the previous three experiments in which they heard speech sounds rather than tones, so it was not as though they found these stimuli uninteresting. However, they could not detect a match between the tones and the faces. We had shown, then, that the temporal envelope of the vowel stimuli used in our experiment was not sufficient to produce the cross-modal effect. Some aspect of the spectral information was necessary, as we had hypothesized.

But what aspect of the spectral information was needed? Did the information in the auditory stimulus have to be sufficient to identify it as an /a/ or an /i/ in order for the match to be detected? Or would a simpler spectral property be sufficient? As a second step, we undertook a variety of tests involving nonspeech stimuli that captured spectral features of the /a/ and /i/ vowels. Additional tests using pure tones were conducted to see whether representing just the "grave-acute" distinctive feature (i.e., the low versus high tonal quality of the vowels) would be sufficient. Our tests had shown that pure tones can reliably be related to auditorially or visually presented vowels by adults; in both cases, they associate low tones with /a/ and high tones with /i/ (Kuhl & Meltzoff, 1988). We used nine different pure tones ranging from 250 Hz to 4000 Hz. We also used three-formant analog stimuli. These were made up of three pure tones whose frequencies matched the formant frequencies; thus, the 3-tone analogs more closely resembled the spectral properties of speech than did the simple pure tones. The results of these tests showed that neither simple distinctive features of the vowels (as represented by pure tones), nor our three-tone analog representations of the vowels were sufficient for infants. They could not detect a match between a nonspeech auditory stimulus and a face mouthing speech (Kuhl & Meltzoff, 1988). Only when the full signal was presented did infants relate the auditory and visual concomitants of speech.

Thus our studies on the basis of the effect suggest that nonspeech analogs do not work. Infants do not detect matches between auditory nonspeech events and faces that make speech movements; they may already know that mouths produce speech, rather than tones or chords. The fact that we have identified a dissociation between speech and nonspeech is intriguing, in light of the fact that nonspeech experiments on categorical perception have replicated the results of speech experiments.

Finally, it is clear that cross-modal perception in infants is not unique to speech. Meltzoff and Borton (1979) conducted experiments on cross-modal perception in 4-week-olds, showing that they could detect equivalences between

information delivered tactually and visually. In this study infants were given one of two pacifiers, either one with nubs on it or a smooth one. The pacifier was then removed and two visual stimuli were presented, a sphere with nubs on it and a smooth one. The results showed that infants who sucked on the smooth pacifier looked at the smooth sphere while those who had sucked on the nubby pacifier fixated the nubby sphere. We can not claim, then, that cross-modal perception in infants is speech-specific.

Auditory-Motor Equivalence

Thus far in discussing the infant's detection of equivalences in speech we have focused on the perception of speech through different sensory modalities—auditory and visual. We now turn our attention to speech production to examine another aspect of equivalence that infants detect for speech.

As adults, we can produce a specific auditory target, such as a vowel, on the first try; it is not a trial-and-error process. Auditory signals are directly related to the motor commands necessary to produce them because adults have rules that dictate the mapping between articulation and audition. This mapping is quite sophisticated. Experiments show that if an adult speaker is suddenly thwarted in the act of producing a given sound by the introduction of a sudden load imposed on his lip or jaw, compensation is essentially immediate (Abbs & Gracco, 1984). The adjustment can occur on the very first laryngeal vibration, prior to the time the adult has heard anything. Such rapid motor adjustments suggest a highly sophisticated and flexible set of rules relating articulatory movements to sound.

How do auditory-articulatory mapping rules develop? Evidence suggests that at least one important mechanism for learning them is vocal imitation.

Among mammals, humans are the only animals who give evidence of vocal learning, that is, learning the species vocal repertoire by hearing it and mimicking it. We share this ability with a few select species of passerine birds (Marler, 1973). Presumably it is the mechanism of imitation that guides vocal learning. The power of its effects can be seen in the fact that early auditory exposure to a specific language pattern puts an indelible marker on one's speech patterns. Foreigners try to rid themselves of their language-specific phonetic errors and their foreign accents, but it is notoriously difficult to do so.

We presume then that at some point young infants must mimic the speech patterns that they hear others produce. But how early are infants capable of doing this? Some relevant data can be adduced from the earliest age at which infants from different language environments produce phonetic units that are unique to their own native language. The data on infant babbling show that infants produce sounds that are not specific to any one language (Oller, 1986; Stark, 1980). But by the time first words emerge, infants will begin to produce sounds that are typical of their language, but are rare in other languages. Moreover, these infants will have an accent. They will have adopted the prosodic features of the language—its cadence, rhythm and tempo, as well as its characteristic intensity and

intonation contours. There is some research suggesting that as early as 12-months-of-age these differences are discernible (de Boysson-Bardies et al., 1984). The data thus suggest that at some point prior to the onset of speech and perhaps as early as 12-months-of-age, infants have acquired enough information about the phonetic units and prosody of their native language to produce it in a way that is characteristic of their native tongue. Thus, by this time, evidence of vocal learning exists.

A more direct approach to the question is to examine vocal imitation experimentally. From Piaget on, reports have appeared that are highly suggestive of vocal imitation of at least one prosodic aspect of speech, its pitch (Kessen, Levine, & Wendrich, 1979; Lieberman, 1984; Papousek & Papousek, 1981; Piaget, 1962); however, all but one of these studies (Kessen et al., 1979) involved natural interactions between adults and infants, and as such are subject to methodological problems (Kuhl & Meltzoff, 1988). Natural observations of mothers and their infants are usually subject to the question "who is imitating whom?" The Kessen et al. study tested infants in multiple sessions over several months, giving them repeated practice and feedback, so the issue of training is unresolved in the study.

With these issues in mind we sought evidence of vocal imitation in our own experiments on infants' cross-modal perception of speech (Kuhl & Meltzoff, 1982; 1988). The cross-modal studies provided a controlled setting in which to study vocal imitation. Recall our experimental set-up. Infants sat in an infant seat facing a three-sided cubicle. They viewed a film of a female talker producing vowel sounds. Half of the infants were presented with one auditory stimulus while the other half were presented with a different auditory stimulus. The stimuli were totally controlled, both visually and auditorially. There were no human interactions with the infant during the test, and thus no chance for spuriously shaping and/or conditioning of a response. The room was a soundproof chamber and a studio-quality microphone was suspended above the infant to obtain clear recordings that could be perceptually or instrumentally analyzed. Finally, the stimulus on film being presented to the infant occurred once every 3 sec, with an interstimulus interval of about 2 sec. This was ideal for encouraging turn-taking on the part of the infant. We found that infants in this setting were calm and highly engaged by the face-voice stimuli. They often listened for a while, smiled at the faces, and then started "talking back." Our question was: Do infants' speech vocalizations match those they hear?

In our initial report we described data that were highly suggestive of infants' imitation of the prosodic characteristics of the signal (Kuhl & Meltzoff, 1982). We observed infant matching of the pitch contour of the adult model's vowels. Both the adult's and infant's responses are shown in Figure 13.7. Instrumental analysis showed that the infant produced an almost perfect match to the adult female's rise-fall pattern of intonation. While the infant has shorter vocal folds and therefore produces a higher fundamental frequency the pitch pattern of a rapid rise in frequency followed by a more gradual fall in frequency duplicates

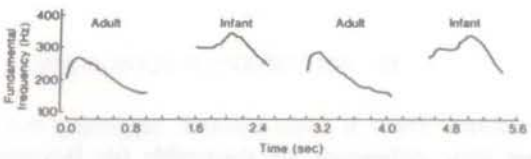

FIG. 13.7. Pitch matching in tests of vocal imitation in infants. From Kuhl and Meltzoff (1982).

that of the adult. The two contours were perceptually very similar. The infant's response also matched the adult's in duration. Because vocalizations with this rise-fall pattern and of this long duration are not common in the utterances of four-month-olds, it was highly suggestive of vocal imitation. But because we had not varied the pitch pattern of the vowel in the experiment it was not possible to conclude definitively that infants could differentially match the pitch contour of vowels.

A more rigorous test of the young infant's ability to imitate relates to their matching of the phonetic segments of speech. Half of the infants in our experiments had heard /a/ vowels while the other half had heard /i/ vowels. This allowed a good test of the differential imitation of speech sounds. All of the vowel-like vocalizations produced by the infants in the /a-i/ studies were analyzed. Vowel-like sounds were defined on the basis of acoustic and articulatory characteristics typical of vowels. The sounds had to be produced with an open mouth, rather than one that was closed. They had to have a minimum duration of 500 msec. They had to be *voiced,* that is, vocalized with normal laryngeal vibration, and could not be aspirated or voiceless sounds. They could not be produced on an inhalatory breath. Vocalizations that occurred while the infant's hand was in his mouth could not be reliably scored and were excluded. Consonant-like vocalizations were also scored, but they occurred rarely and were always accompanied by vowel-like sounds.

Once identified, the sounds were submitted to analysis. Perceptual scoring was done by having a trained phonetician listen to each infant's productions and judge whether, on the whole, they were more "/i/-like" or /a/-like." Infants at this age cannot produce perfect /i/ vowels, due to anatomical restrictions. They can, however, produce other high front vowels such as /ɪ/ or /ɛ/. Similarly, a perfect /a/ is rare in the vocalizations of the 4-month-old, but similar central vowels, such as /æ/ and /ʌ/ are producible by infants at this age. Thus, the judgment made by the observer was a forced-choice one concerning whether an infant's vocalizations were more /a/-like or more /i/-like.

We then asked judges to predict whether infants had been exposed to /a/ as opposed to /i/, based on the infant's vocalizations. If judges can do so with greater than chance (50%) accuracy, then there is evidence for vocal imitation. The results confirmed this prediction. Infants produced /a/-like vowels when listening to /a/ and /i/-like vowels when listening to /i/, allowing the judges to

predict accurately in 90% of the instances the vowel heard by the infant. These results were highly significant ($p < 0.01$) (Kuhl & Meltzoff, 1988).

We are now involved in the instrumental analysis of the sounds. Using distinctive feature theory to guide our instrumental analyses, we measured the graveness and compactness of the infants' vowel productions. The results demonstrated that infants' vocal responses to /a/ were significantly more grave, that is, they had a lower center of gravity, than their responses to /i/. Similarly, their responses to /a/ were significantly more compact, that is, they had formants spaced more closely together, than their responses to /i/. Taken together, the two analyses provide some evidence that 4-month-old infants are engaged in vocal imitation of the phonetic segments of speech.

The Basis of Auditory-Motor Equivalence

Our first question was again related to the effectiveness of nonspeech sounds. Could the auditory stimulus in vocal imitation studies be replaced by a nonspeech stimulus? The specific questions we were interested in were these: What happens when infants listen to speech as opposed to nonspeech sounds? Do vocalizations occur as frequently as they do when infants listen to speech? And if they occur, do these vocalizations sound like those given in response to the speech stimulus mimicked by the nonspeech analog?

Recall that in our cross-modal studies involving speech infants heard one of the three point vowels, /a/, /i/, or /u/. In two other studies infants heard nonspeech stimuli consisting of either pure tones or a three-tone analog stimulus. In the pure-tone study (Kuhl, Wolak, & Meltzoff, in preparation), nine pure tones were used, varying from 125 Hz to 4000 Hz. In the three-tone analog study, the tones matched the formant frequencies of the vowels. In neither of these nonspeech studies could any of the sounds be identified as speech.

Our original study included a nonspeech test in which a single pure tone was presented (Kuhl & Meltzoff, 1982). The results of the study suggested that nonspeech sounds were not effective elicitors of vocalization. We reported that the infants tested in the speech condition versus those tested in the tone condition produced a differential amount of vocalization. Infants who heard speech produced cooing sounds typical of speech. The infants who were presented with the nonspeech tone did not produce speech-like vocalizations. They had watched the same faces, heard sounds of the same duration and intensity, and were given just as long to reply. But they did not produce speech. In the 1982 paper we reported that 10 of the 32 infants hearing speech produced speech-like vocalizations whereas only a single infant hearing nonspeech produced speech-like vocalizations, and this difference was significant ($p < 0.01$).

We can now extend these results to a much larger sample. To date we have analyzed the vocalizations of all of the infants who participated in the two /a-i/ studies for a total of 64 infants. In addition, we have analyzed the vocalizations of the first half (72 infants) of the 144 infants tested in the pure-tone study (Kuhl

& Meltzoff, 1988). The results strongly show the superiority of human speech in eliciting infant vocalizations. Infants listening to speech produce speech, while infants listening to tones do not. Fully 40 of the 64 infants listening to speech in our sample produce vocalizations that are typical of speech, while only 5 of the 72 infants hearing nonspeech produce sounds of this type ($p < 0.001$). Infants listening to nonspeech do not tend to produce speech-like vocalizations; instead, they squeal, gurgle, grunt, or produce raspberries. Apparently, infants talk only to faces that are talking to them. Thus we see a dissociation between speech and nonspeech in our studies of vocal imitation.

Having these data on vocal imitation in hand, we can now ask whether the tendency to mimic human acts is unique to speech? Once again, the answer from cognitive development is clear. Meltzoff's work on facial and gestural imitation has shown quite convincingly that very young infants (in some instances newborns) imitate adult facial and manual gestures such as tongue protrusion, mouth opening, and the opening and closing of the hand (Meltzoff & Moore, 1977, 1983). Thus, we cannot claim that infants' imitative capacity regarding vocalization is a specialization that is speech specific.

IV. A RETURN TO THEORY

We began this chapter by noting the fact that for both animal and human species, the notion of "special mechanisms" has been offered to explain infants' early responsiveness to species-specific signals (SMA). For the case of human infants acquiring speech, however, a second account was described that is a viable alternative. The second account holds that general mechanisms may be sufficient to account for infants' abilities (GMA). Both accounts attempt to explain infants' detection of complex equivalences in speech: between auditory events that are physically different and easily discriminable, between speech stimuli presented to different modalities, as in auditory-visual speech perception, and between an auditory speech stimulus and its motor equivalent.

How do our two models account for the equivalence data? Recall that SMA argues that in each of the cases a phonetic representation of speech units mediates perception. This is its key point; without higher-order representations, these events cannot be equated in perception. They are not linked to each other in any other way. Auditory equivalence is perceived because two different speech events (such as the vowel /a/ spoken by two different people) are tied to the same phonetic representation. Thus, even though their surface acoustic properties are not the same, their common representation renders them equal. When auditory and visual versions of /a/ are detected, or when sounds are equated to the motor movements used to produce them, this account holds that it is because the two stimuli have a common underlying phonetic representation (Kuhl & Meltzoff, 1984a). The auditory and visual instantiations of speech are not themselves directly tied. Nor are sounds and the motor movements that produce them. They

are linked up by virtue of the fact that they are both independently tied to the higher-order representation of the phonetic segment. Without higher-order representations, these events cannot be linked, and equivalence would not be detected (Kuhl & Meltzoff, 1984a).

What of GMA? How does it explain the data on infants' detection of equivalence? The mainstay of its argument is that infants' detection of equivalence for speech does not depend on a representation of phonetic units. On this view, perception of equivalence is not mediated by pre-existing representations of phonetic units because innately stored representations of phonetic units do not exist. Infants' capabilities are explained by their general auditory and cognitive abilities.

Regarding infants' detection of auditory equivalence, the GMA holds that this is due purely to the perception of auditory similarity. This is true both for two vowels spoken by different people, and for cues that "trade" in perception. On this view, these stimuli can be perceived to be auditorially equivalent in the absence of any other speech-specific processing. Critical data for this position are provided by the animal studies reviewed earlier; they show that animals detect auditory equivalences for speech.

According to the GMA higher-order equivalences involving cross-modal and imitative abilities are also not dependent upon preset phonetic representations. Critical data here are those provided by studies of infants' general cognitive abilities. Research on infants' cognitive development clearly demonstrates that these abilities exist outside the domain of speech. Thus, the key point argued by the GMA is that infants do not need special mechanisms to accomplish cross-modal and imitative tasks for speech; such mechanisms already exist for the perception of objects and faces.

Having summarized each account's approach to the data on equivalence detection, we address the evidence presented from tests on the basis of the effect. The SMA is most forcefully supported as an explanation for an effect when nonspeech tests fail, when animals fail, and when no other domain but speech gives evidence of the effect. The GMA is supported for effects in which nonspeech tests succeed, animals succeed, and other domains provide evidence of similar effects. What pattern of results was obtained? Did clear support for one or the other account emerge?

Consider first the pattern of results with nonspeech. There is evidence that CP effects in infants can be replicated with nonspeech, although the difference between VOT results and TOT results remains puzzling. Context effects and trading relations have also been demonstrated using nonspeech analogs with infants. The only effects that have clearly failed using nonspeech are complex tasks like cross-modal speech perception and vocal imitation. These effects appear to require the whole stimulus. We might therefore draw a line between the detection of auditory equivalence and the detection of equivalence for higher-order intermodal relations. Perhaps the detection of intermodal equivalences is indeed based on more specialized mechanisms.

We look for confirmation of this hypothesis in tests on animals. CP effects can be replicated in animals. Moreover, context effects have now been replicated, though only one example has been tested. Tests of equivalence classification also show that animals are capable of perceiving speech categories. Tests of cross-modal perception and imitation have not been completed, but a reasonable guess would be that these tests would fail. Animals are not known to be proficient on these tasks, particularly on imitation (Meltzoff, 1988). If we imagine, for the sake of argument, that animals will fail on these tasks, then a similar pattern of results with nonspeech and animals would have emerged, with both suggesting that *auditory* equivalence is less likely to require special mechanisms than are more complex *intermodal* equivalences.

Lastly, we look at the evidence for domain specificity. Are any of these equivalences detected by infants unique to speech? Here we have to conclude that speech is not unique. Equivalence classification, cross-modal perception, and imitation are cognitive abilities that appear to be quite robust in infants. One might have thought that evidence of such sophisticated talent would be rare in infants. It is not. Is speech a special case of these more complex skills? It may turn out to be, but one need not posit this, given infants' apparent cognitive capacity for the detection of higher-order equivalences.

V. SUMMARY AND CONCLUSIONS

There are two distinct characterizations of infants' initial state for speech processing. Both conceed that infants demonstrate speech phenomena that are extremely sophisticated. Infants' detection of complex equivalences—between discriminably different auditory events, between speech information delivered auditorially and visually, and between the auditory and motor instantiations of a speech event—suggest an initial organization of speech that is highly conducive to the acquisition of an intermodally represented speech system. The Specialized Mechanism Account explains this by imputing phonetic-level representations of speech to the infant at birth. On this view, infants' detection of equivalence is due to the mediating effects of a phonetic-level representation. The General Mechanism Account claims that phonetic-level representations do not exist at birth and that infants' capabilities are due to their more general sensory and cognitive abilities. This account holds that phonetic-level representations are built up only later as the child acquires language.

Experiments directed towards identifying the basis of these effects were reviewed. These experiments include tests on nonspeech signals, tests on animals' perception of speech, and tests on equivalence detection in domains other than speech. These experiments show that both nonspeech and animal tests replicate auditory equivalence effects. Importantly, though, nonspeech signals fail to reproduce the auditory-visual cross-modal effect and fail to induce vocal imitation. It is tempting to conclude, then, that these higher-order equivalences require

special mechanisms. Yet, the detection of higher-order equivalences by infants is not restricted to speech; they are demonstrated in other domains as well. Thus, even complex behaviors such as these may not be due to a domain-specific speech module. It appears, then, that even if speech is intermodally represented in infants, it may not require "special mechanisms" to be organized in that way. Rather, speech may draw upon a natural proclivity to represent information intermodally. At present, no clear evidence in favor of a phonetic-level representation of speech has been presented. Until further tests have been conducted, claims about infants' phonetic representation of speech are most wisely offered and debated, but not yet acclaimed as definitely proven.

ACKNOWLEDGMENTS

The author and the work described here were supported by grants from the National Science Foundation (BNS 8316318) and from the National Institutes of Health (HD-18286 and HD 22514). I am indebted to Karen Wolak, Craig Harris, and Kerry Green for assistance in the experiments, Karen Wolak for helpful comments on the manuscript, Andy Meltzoff for discussions of the issues raised here, and KKM for inspiration.

REFERENCES

Abbs, J. H., & Gracco, V. L. (1984). Control of complex motor gestures: Orofacial muscle responses to lead perturbations of the lip during speech. *Journal of Neurophysiology, 51,* 705–723.

Aslin, R. N., Pisoni, D. B., Hennessey, B. L., & Perey, A. J. (1981). Discrimination of voice onset time by human infants: New findings and implications for the effects of early experience. *Child Development, 52,* 1135–1145.

Baru, A. V. (1975). Discrimination of synthesized vowels [a] and [i] with varying parameters (fundamental frequency, intensity, duration, and number of formants) in dog. In G. Fant & M. A. A. Tatham (eds.), *Auditory analysis and perception of speech* (pp. 91–101). New York: Academic Press.

Best, C. T., Morrongiello, B., & Robson, R. (1981). Perceptual equivalence of acoustic cues in speech and nonspeech perception. *Perception and Psychophysics, 29,* 191–211.

Bruner, J. S., Goodnow, J. J., & Austin, G. A. (1956). *A study of thinking.* New York: Wiley.

Burdick, C. K., & Miller, J. D. (1975). Speech perception by the chinchilla: Discrimination of sustained [a] and [i]. *Journal of the Acoustical Society of America, 58,* 415–427.

Cohen, L. B., & Strauss, M. S. (1979). Concept acquisition in the human infant. *Child Development, 50,* 419–424.

de Boysson-Bardies, B., Sagart, L., & Durand, C. (1984). Discernible differences in the babbling of infants according to target language. *Journal of Child Language, 11,* 1–15.

Diehl, R. L., & Kluender, K. R. (in press). On the objects of speech perception. *Ecological Psychology.*

Eimas, P. D. (1974). Auditory and linguistic processing of cues for place of articulation by infants. *Perception and Psychophysics, 16,* 513–521.

Eimas, P. D. (1975). Auditory and phonetic coding of the cues for speech: Discrimination of the [r-l] distinction by young infants. *Perception and Psychophysics, 18,* 341–347.

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

Eimas, P. D., & Miller, J. L. (1980). Contextual effects in infant speech perception. *Science, 209,* 1140–1141.

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

Fagan, J. F., III. (1976). Infants' recognition of invariant features of faces. *Child Development, 47,* 627–638.

Fodor, J. A. (1983). *The modularity of mind.* Cambridge, MA: MIT Press.

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. (in press). The role of visual information in the processing of place and manner features in speech. *Perception and Psychophysics.*

Green, K., & Miller, J. L. (1985). On the role of visual rate information in phonetic perception. *Perception and Psychophysics, 38,* 269–276.

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

Hillenbrand, J. (1983). Perceptual organization of speech sounds by infants. *Journal of Speech and Hearing Research, 26,* 268–282.

Hillenbrand, J. (1984). Speech perception by infants: Categorization based on nasal consonant place of articulation. *Journal of the Acoustical Society of America, 75,* 1613–1622.

Jusczyk, P. W., Pisoni, D. B., Reed, M. A., Fernald, A., & Myers, M. (1983). Infants' discrimination of the duration of a rapid spectrum change in nonspeech signals. *Science, 222,* 175–177.

Jusczyk, P. W., Pisoni, D. B., Walley, A., & Murray, J. (1980). Discrimination of relative onset time of two-component tones by infants. *Journal of the Acoustical Society of America, 67,* 262–270.

Kessen, W., Levine, J., & Wendrich, K. A. (1979). The imitation of pitch in infants. *Infant Behavior and Development, 2,* 93–99.

Klatt, D. (1986). Problem of variability in speech recognition and in models of speech perception. In J. S. Perkell & D. H. Klatt (Eds.), *Invariance and variability in speech processes* (pp. 300–324). Hillsdale, NJ: Lawrence Erlbaum Associates.

Kluender, K. R., Diehl, R. L., & Killeen, P. R. (1987). Japanese quail can learn phonetic categories. *Science, 237,* 1195–1197.

Kuhl, P. K. (1978). Predispositions for the perception of speech-sound categories: A species-specific phenomenon? In F. D. Minifie & L. L. Lloyd (Eds.), *Communicative and cognitive abilities—Early behavioral assessment* (pp. 229–255). Baltimore: University Park Press.

Kuhl, P. K. (1979a). 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. (1979b). Models and mechanisms in speech perception: Species comparisons provide further contributions. *Brain, Behavior and Evolution, 16,* 374–408.

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, Vol. 2, Perception* (pp. 41–66). New York: Academic Press.

Kuhl, P. K. (1981). Discrimination of speech by nonhuman animals: Basic auditory sensitivities conducive to the perception of speech-sound categories. *Journal of the Acoustical Society of America, 70,* 340–349.

Kuhl, P. K. (1983). Perception of auditory equivalence classes for speech in early infancy. *Infant Behavior and 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. (1986a). Infants' perception of speech: Constraints on characterizations of the initial state. In B. Lindblom & R. Zetterstrom (Eds.), *Precursors of early speech* (pp. 219–244). New York: Stockton Press.

Kuhl, P. K. (1986b). Theoretical contributions of tests on animals to the special-mechanisms debate in speech. *Experimental Biology, 45,* 233–265.

Kuhl, P. K. (1986c). Reflections on infants' perception and representation of speech. In J. S. Perkell & D. H. Klatt (Eds.), *Invariance and variability of speech processes* (pp. 19–30). Hillsdale, NJ: Lawrence Erlbaum Associates.

Kuhl, P. K. (1987a). 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). Orlando, Fl.: Academic Press.

Kuhl, P. K. (1987b). The special-mechanisms debate in speech research: Categorization tests on animals and infants. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 355–386). Cambridge, England: Cambridge University Press.

Kuhl, P. K. (1988). Auditory perception and the evolution of speech. *Human Evolution, 3,* 19–43.

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,* Suppl. 1, S86(A).

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

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

Kuhl, P. K., & Meltzoff, A. N. (1984b). *Imitation, representation, and cross-modal perception in infants.* International Conference on Infant Studies, New York.

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

Kuhl, P. K., & Miller, J. D. (1975). Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants. *Science, 190,* 69–72.

Kuhl, P. K., & Miller, J. D. (1978). Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli. *Journal of the Acoustical Society of America, 63,* 905–917.

Kuhl, P. K., & Padden, D. M. (1982). Enhanced discriminability at the phonetic boundaries for the voicing feature in macaques. *Perception and Psychophysics, 32,* 542–550.

Kuhl, P. K., & Padden, D. M. (1983). Enhanced discriminability at the phonetic boundaries for the place feature in macaques. *Journal of the Acoustical Society of America, 73,* 1003–1010.

Kuhl, P. K., Wolak, K. M., & Green, K. P. (in preparation). Infants' detection of auditory equivalences in speech: Vowel categories.

Kuhl, P. K., Wolak, K. M., & Meltzoff, A. N. (in preparation). Infants' cross-modal perception of speech: Studies on the basis of the effect.

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

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition, 21,* 1–36.

Lieberman, P. (1984). *The biology and evolution of language.* Cambridge, MA: Harvard University Press.

MacKain, K., Studdert-Kennedy, M., Spieker, S., & Stern, D. (1983). Infant intermodal speech perception is a left-hemisphere function. *Science, 219,* 1347–1348.

Marler, P. (1974). Constraints on learning: Development of bird song. In N. F. White (Ed.), *Ethology and Psychiatry: The Clarence M. Hincks Memorial Lectures for 1970.* (pp. 69–83). Toronto: University of Toronto Press.

Massaro, D. W., & Cohen, M. M. (1983). Evaluation and integration of visual and auditory information in speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 9,* 753–771.

Mattingly, I. G., Liberman, A. M., Syrdal, A. K., & Halwes, T. (1971). Discrimination in speech and nonspeech modes. *Cognitive Psychology, 2,* 131–157.

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

Medin, D. L., & Barsalou, L. W. (1987). Categorization processes and categorical perception. In S. Harnad (Ed.), *Categorical perception:* The ground work of cognition (pp. 455–490). Cambridge, England: Cambridge University Press.

Meltzoff, A. N. (1988). The human infant as *homo imitans*. In T. R. Zentall & B. G. Galef (Eds.), *Social learning: Psychological and biological perspectives* (pp. 319–341). Hillsdale, NJ: Lawrence Erlbaum Associates.

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. (1983). Newborn infants imitate adult facial gestures. *Child Development, 54,* 702–709.

Milewski, A. E. (1979). Visual discrimination and detection of configurational invariance in 3-month infants. *Developmental Psychology, 15,* 357–363.

Miller, J. L., & Eimas, P. D. (1983). Studies on the categorization of speech by infants. *Cognition, 13,* 135–165.

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

Miller, J. D., Wier, C. C., Pastore, R. E., Kelly, W. J., & Dooling, R. J. (1976). Discrimination and labeling of noise-buzz sequences with varying noise-lead times: An example of categorical perception. *Journal of the Acoustical Society of America, 60,* 410–417.

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A. M., Jenkins, J. J., & Fujimura, O. (1975). An effect of linguistic experience: The discrimination of /r/ and /l/ by native speakers of Japanese and English. *Perception and Psychophysics, 18,* 331–340.

Morse, P. A., & Snowdon, C. T. (1975). An investigation of categorical speech discrimination by rhesus monkeys. *Perception and Psychophysics, 17,* 9–16.

Öhman, S. E. G. (1966). Coarticulation of VCV utterances: Spectrographic measurements. *Journal of the Acoustical Society of America, 39,* 151–168.

Oller, D. K. (1986). Metaphonology and infant vocalizations. In B. Lindblom & R. Zetterström (Eds.), *Precursors of early speech* (pp. 21–35). New York: Stockton Press.

Papousek, M., & Papousek, H. (1981). Musical elements in the infant's vocalization: Their significance for communication, cognition, and creativity. In L. P. Lipsitt & C. K. Rovee-Collier (Eds.), *Advances in infancy research* (Vol. 1, pp. 164–224). Norwood, NJ: Ablex.

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

Piaget, J. (1951). *Play, dreams and imitation in childhood.* New York: W. W. Norton.

Pisoni, D. B. (1977). Identification and discrimination of the relative onset time of two component tones: Implications for voicing perception in steps. *Journal of the Acoustical Society of America, 61,* 1352–1361.

Pisoni, D. B., Carrell, T. D., & Gans, S. J. (1983). Perception of the duration of rapid spectrum changes in speech and nonspeech signals. *Perception and Psychophysics, 34,* 314–322.

Rosch, E. (1975). Cognitive reference points. *Cognitive Psychology, 7,* 532–547.

Sinnott, J. M., Beecher, M. D., Moody, D. B., & Stebbins, W. C. (1976). Speech sound discrimination by monkeys and humans. *Journal of the Acoustical Society of America, 60,* 687–695.

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

Starkey, P., Spelke, E. S., & Gelman, R. (1983). Detection of intermodal numerical correspondences by human infants. *Science, 222,* 179–181.

Stevens, E., Kuhl, P. K., & Padden, D. (1988). Macaques show context effects in speech perception. *Journal of the Acoustical Society of America, 84,* Suppl. 1, 577(A).

Stevens, K. N. (1972). The quantal nature of speech: Evidence from articulatory-acoustic data. In E. E. David, Jr. & P. B. Denes (Eds.), *Human communication: A unified view* (pp. 51–66). New York: McGraw-Hill.

Stevens, K. N. (1981). Constraints imposed by the auditory system on the properties used to classify speech sounds. Evidence from phonology, acoustics, and psychoacoustics. In T. Myers, J. Laver, & J. Anderson (Eds.), *Advances in psychology: The cognitive representation of speech.* Amsterdam: North-Holland.

Stevens, K. N. (in press). On the quantal nature of speech. *Journal of Phonetics.*

Stevens, K. N., & Blumstein, S. E. (1981). The search for invariant acoustic correlates of phonetic features. In P. Eimas & J. Miller (Eds.), *Perspectives on the study of speech.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Streeter, L. A. (1976). Language perception of 2-month-old infants shows effects of both innate mechanisms and experience. *Nature, 259,* 39–41.

Studdert-Kennedy, M. (1986). Development of the speech perceptuomotor system. In B. Lindblom & R. Zetterström (Eds.), *Precursors of early speech* (pp. 205–217). New York: Stockton Press.

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

Summerfield, Q., & Haggard, M. (1977). On the dissociation of spectral and temporal cues to the voicing distinction in initial stop consonants. *Journal of the Acoustical Society of America, 62,* 435–448.

Waters, R. S., & Wilson, W. A., Jr. (1976). Speech perception by rhesus monkeys: The voicing distinction in synthesized labial and velar stop consonants. *Perception and Psychophysics, 19,* 285–289.