470

Theory and Research in Behavioral Pediatrics

Volume 2

Edited by

Hiram E. Fitzgerald, Ph.D.
Professor of Psychology
Michigan State University
East Lansing, Michigan

Barry M. Lester, Ph.D.
Director of Developmental Research
Child Development Unit
Children's Hospital Medical Center
Assistant Professor of Pediatrics
Harvard Medical School
Boston, Massachusetts

and

Michael W. Yogman, M.D.
Associate Chief, Division of Child Development
Children's Hospital Medical Center
Assistant Professor of Pediatrics
Harvard Medical School
Boston, Massachusetts

PLENUM PRESS • NEW YORK AND LONDON

1984

Vrensen, G., & Cardozo, J. N. Changes in size and shape of synaptic connections after visual training: An ultrastructural approach of synaptic plasticity. *Brain Research*, 1981, *218*, 79–97.

Wesa, J. M., Chang, F. F., Greenough, W. T., & West, R. W. Synaptic contact curvature: Effects of differential rearing on rat occipital cortex. *Developmental Brain Research*, 1982, *4*, 253–257.

Wiesel, T. N., & Hubel, D. H. Receptive fields of cells in striate cortex of very young, visually inexperienced kittens. *Journal of Neurophysiology*, 1963, *26*, 994–1002. (a)

Wiesel, T. N., & Hubel, D. H. Simple cell responses in striate cortex of kittens deprived of vision in one eye. *Journal of Neurophysiology*, 1963, *26*, 1003–1017. (b)

Wiesel, T. N., & Hubel, D. H. Comparison of the effects of unilateral and bilateral eye closure on cortical unit responses in kittens. *Journal of Neurophysiology*, 1965, *28*, 1029–1040.

Will, B. E., Rosenzweig, M. R., Bennett, E. L., Herbert, M., & Morimoto, H. Relatively brief environmental enrichment aids recovery of learning capacity and alters brain measures after postweaning brain lesions in rats. *Journal of Comparative and Physiological Psychology*, 1977, *91*, 33–50.

Williams, J. T., & Zieglgansberger, W. Neurons in the frontal cortex of the rat carry multiple opiate receptors. *Brain Research*, 1981, *226*, 304–308.

Wise, S. P., & Herkenham, M. Opiate receptor distribution in the cerebral cortex of the rhesus monkey. *Science*, 1982, *218*, 387–389.

Yang, H. Y., Hong, J. S., & Costa, E. Regional distribution of len and met enkephalin in rat brain. *Neuropharmacology*, 1977, *16*, 303–307.

Discovering Messages in the Medium

Speech Perception and the Prelinguistic Infant

CATHERINE T. BEST

To parents and those who work with infants, one of the most remarkable developments during the first year of life is the rapid growth of vocal communication skills prior to language. Initially, the infant communicates by cries, but during the second half-year, the speechlike sounds of babbling emerge. By twelve months, babbling not only conveys feelings and needs to others but may also express infants' observations of regularities in the events and objects of their world.

The sounds that infants make are but one facet of their progress toward verbal communication. More hidden from our view, yet also important, is their *perceptual* grasp of the speech around them. In this chapter on infant speech perception, speech will be considered as the medium through which language is expressed vocally, much like the sounding of musical instruments is the medium through which a symphony is expressed. Both language and symphony are structurally complex systems, with many levels of concurrent organization, which are reflected in the organization of the medium. Speech carries the multiple messages that can be conveyed verbally, and hence carries information about the complex structural organization of vocal communication, which includes not only the structure of words and sentences but also broader aspects, such as stress, conversational rhythm, and voice

---

Catherine T. Best • Department of Human Development, Cognition, and Learning, Teachers College/Columbia University, New York, New York 10027. Work on the chapter was supported in part by a Spencer Young Faculty grant and Biomedical Research grant 42-USC 241 awarded through Teachers College.

characteristics (e.g., speaker gender, age, identity, and emotional state). It also carries information about finer grained structures such as consonants, vowels, and syllables, and the vocal gestures that produce them.

A listener's knowledge about the organization of vocal communication sets limits on which messages can be recognized in speech. That is to say, one must know or learn which structures to listen for within the wealth of information that the medium reflects about the events forming a vocalization (stated in the spirit of E. J. Gibson, 1977, and J. J. Gibson, 1966). Although infants may recognize some aspects of nonlinguistic structures in vocal communication, they are limited in their recognition of language structure as such. To use language, then, they must still discover the existence and meaning of many of the messages in speech, particularly those of words and phrases. But where do they start and how do they proceed in their discoveries during the first year? For infants who have not yet discovered the word, what messages are perceptually available to them in speech that might expedite that discovery?

The central concern of this chapter is with the nature of messages that prelinguistic infants may hear in human speech, at the level of the finer grained structures that we know as consonants and vowels. This issue has been addressed in the last twelve years of research on infant speech perception. Two basic themes about the information that infants perceive in speech have emerged. Both themes presume that some innate mechanism(s) of the auditory perceptual system fully account for infant speech perception, thus implying a mechanistic view that the young perceiver's role in seeking information in speech is rather passive. According to the first theme, infants possess species-specialized perceptual mechanisms that are tuned to linguistic contrasts among phonemes, those individual consonants and vowels we adults often associate with letters or letter combinations in words. The other theme proposes that infant speech perception is shaped by the auditory system's response to acoustic components of the stimulus; that is, the perceptual process is stimulus-bound and intrinsically neutral with respect to speech versus other sounds. These neutral acoustic attributes include the bits of noise and frequency changes, interspersed with silent gaps and humming or buzzing, which comprise a physical description of the speech signal.

It will be argued here that neither theme adequately explains how infants perceive the speech they normally hear during development. In their stead, the features of a third perspective based on ecological considerations (see also Fowler, Rubin, Remez, & Turvey, 1980; Summerfield, 1978) will be outlined, which posits a more active, information-seeking role for the perceiver. This alternative view is that infants actively attend for information in the speech medium about the natural forces that structured it, particularly how it was shaped by the human vocal tract. For the sake of simplicity, this perceptual focus on how speech is structured by its vocal source will be referred to as *speech source perception*. This term is offered rather than *articulatory perception* (see Studdert-Kennedy, 1981a; Summerfield, 1978), in order to encompass not only the articulatory gestures of the mouth and tongue but also the anatomical structure of the human vocal tract and its variations according to speaker characteristics (e.g., sex, age, and emotional state). As will be argued, a theory focused on the vocal tract sources of the speech medium's acoustic structure has greater potential than the other two themes for explaining how and why infants might begin to develop language based on the speech they hear. In short, it would provide the infant a more direct avenue by which to both discover and produce words.

But how is this vocal source information conveyed in speech? Simply stated, for now, the acoustic properties of sounds are determined by the structure and movements of the sound-making object, including the human vocal tract (Fant, 1960; Flanagan, 1973). Thus, speech carries information about vocal configurations and gestures (Cooper, 1981; Dudley, 1940; Paget, 1930); the *speech source* view proposes that this vocal tract information is available to perception. This view will be explained in more detail, with support from recent research with adults and infants. It will also be suggested that the specialization of the human left cerebral hemisphere for language reflects an attunement to detect information in speech about the articulatory gestures of the speaker's vocal tract. The chapter concludes by noting that speech source perception is only one contribution to the infant's development of language. In order to discover words and develop language, infants must also learn about the broader aspects of language from the natural context in which speech occurs.

1. SETTING THE CONTEXT

1.1. Language and the Prelinguistic Infant

Prelinguistic infants, by definition, do not yet produce true words; that is, their vocalizations apparently do not refer to objects and events in the way that the words of the adult language community do. It should be noted that it is not possible to draw a sharp chronological division between prelinguistic and linguistic periods in the development of either speech production or speech perception. Generally, however, the first year of life is considered to be prelinguistic.

Two complementary and interdependent questions provide a guide for understanding the prelinguistic antecedents of language development: What is vocal language that a prelinguistic infant may come to know it? and What is the prelinguistic infant that she or he may come to know vocal language? (adapted from McCulloch, 1965). The next few sections will focus on the former question, to frame the subsequent discussion of infant speech perception research. They will describe the basic characteristics of speech that are important for understanding the task facing a prelinguistic perceiver. Once the stage has been set, the three theoretical views about the way infants process speech will be described in greater depth.

1.2. What Is Vocal Language?

What type of information or messages does speech carry that prelinguistic infants might perceive? Prelinguistic infants do not yet produce words, nor do most infants under 9–10 months yet comprehend spoken words (e.g., Lenneberg, 1967). Thus, we should not expect younger infants to perceive any information that is defined by word meanings. Nevertheless, some coherent information in human speech must be available to prelinguistic infants, for they do eventually discover words.

To discover words in the speech directed toward them, presumably infants would, in part, have to (a) disembed from continuous speech the recurring subpatterns that become familiar words; (b) recognize the invariance in the pattern of a word, across the variations in acoustic detail that occur when it is produced by different speakers or in different contexts; and (c) recognize the relevant differences that do specify meaningfully different patterns. And in order to produce words, they would also have to recognize how to imitate or approximate subpatterns from a language-user's speech, even though the acoustic output of their own smaller and differently proportioned vocal tracts differs substantially from that provided by their older models (Goldstein, 1979; Lieberman, Harris, Wolff, & Russell, 1971).

In language research with adults, phonemes (consonants and vowels) have often been considered to be the building blocks of words. According to that perspective, the achievement of the perceptual tasks previously listed would seemingly be founded on perception at the phonemic level of speech. To date, most infant speech perception research has focused on phonemes or their combinations in syllables, on the apparent assumption that perception of these subword units must be precursory to the perception of words.

Language users easily recognize words and phonemes when listening to conversational speech. However, these recognitions are no small feat for prelinguistic listeners, who lack a language system that could help them solve some apparent puzzles in adult speech perception. The source of these puzzles, which lies in the acoustic characteristics of speech, is discussed next. (For more extensive discussions of adult speech perception, see Fowler *et al.*, 1980; Liberman, 1982; Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Pisoni, 1978).

1.3. The Puzzles in the Acoustic Shape of Speech

To imagine the infant's difficulty, recall listening to a stretch of conversation in an unfamiliar language. Foreign speech typically seems like a relatively continuous flow of sounds, in which the quality of some phonemes may be unfamiliar (e.g., the /r/ of French or Spanish), and the boundaries of individual words often may be indecipherable. This can make it difficult for a listener to recognize a foreign word uttered in different sentences and by different people. The infant's problem is compounded because, in contrast to a language user, infants presumably do not know what words are, so this concept cannot guide their discovery of word boundaries in the flow of conversational speech. Similarly, a language user may have difficulty recognizing the precise qualities of an unfamiliar foreign phoneme when it is uttered in different words and by different people. Yet, relative to mature language users, to infants all of the phonemes occurring even in their native language environment would be comparatively unfamiliar. Infants also presumably lack certain concepts about the linguistic role of phonemes that may guide the language user's recognition of individual phonemes in conversational speech.

1.3.1. Acoustic Continuities and Discontinuities in Running Speech

One reason a sentence spoken in an unfamiliar language sounds indivisible is that utterances in natural conversation are a fairly continuous stream of sound. This is partly attributable to the cohesive intonation, or pitch contour of the voice. But it results also from the vocal-tract movement trajectories that interconnect the adjacent words in sentences or phrases. Speakers in conversation rarely pause between words, instead usually moving in connected fashion from one to the next, just as a runner usually adjusts to changes in terrain or direction without pausing between step cycles. Sometimes neighboring words in informal speech even become contracted (e.g., "what are you . . ." becomes "wadaya . . ." or "whatcha . . ."). Thus, the raw acoustic properties of conversational speech do not always reveal clear boundaries between words.

Conversely, the vocal tract can make other relatively rapid adjustments that do cause obvious acoustic discontinuities. These breaks can occur within as well as between words, however. For instance, in the word "so" there is a rather sudden change from the lack of vocal-cord vibration during the voiceless /s/ to the onset of vibration for the voiced sound /o/. This causes an acoustic break between the noiselike, aperiodic hiss of the /s/ and the voiced acoustic periodicity of the vowel. The paradox is that a knowledgeable listener perceives these discontinuities as an integral part of a word, where appropriate, rather than as breaks between words.

The speech properties just discussed can be seen in Figure 1. A spectrogram is one way of visualizing the acoustic components of speech. As indicated earlier, these acoustic characteristics are determined by the structure and movements of the vocal tract (Fant, 1960; Flanagan, 1973).

The spectrographic analysis shows the relative acoustic intensity (darkness level) of the frequency components in the speech signal (ordinate) as they change over time (abscissa). The wide, horizontally varying bars of increased density within the dark vertical striations are called *formants,* the lowest being referred to as the first format (F1), and correspond to the time-varying resonant frequencies of various relatively hollow spaces or chambers in the vocal tract. In the vowel "ee," for example, a small resonating chamber is formed at the front of the mouth between the edges of the tongue blade pressed against the upper teeth and the close approach of the soft palate to the base of the tongue, while a relatively large resonating chamber forms at the back of the mouth behind the base of the tongue. This results in a low-frequency F1 and a high-frequency second format (F2).

The vertical striations in which the formats appear represent the individual energy pulses emitted by each vocal-fold vibration. The more closely packed the striations are, the briefer the periods between pulses and hence the higher the pitch of the voice. In Figure 1, the man raised his voice pitch substantially for the word "saying" and then dropped it for "(to) me little," raising it again toward the end of "girl." This degree of pitch modulation is more exaggerated than normal, and often occurs when parents talk playfully to their babies (Kaye, 1980). The dappled, nonstriated patches represent aperiodic acoustic noise produced by air turbulence at some point in the vocal tract as with the tongue-tip constriction near the upper front teeth for the /s/ of "saying."

1.3.2. Continuities and Discontinuities among and within Phonemes

Since infant speech research has focused primarily on phonemes, the sentence in Figure 1 is printed beneath the spectrogram for reference. The letters for the vowels and consonants are roughly lined up under the midpoint of their portion of the acoustic signal. The match between phoneme and acoustic information is only approximate, however, because of inherent difficulties in determining the acoustic span of a phoneme. At times, discontinuities appear to fall within rather than between phonemes, as was true at the level of words in a sentence. For instance, the /t/ in "to" encompasses

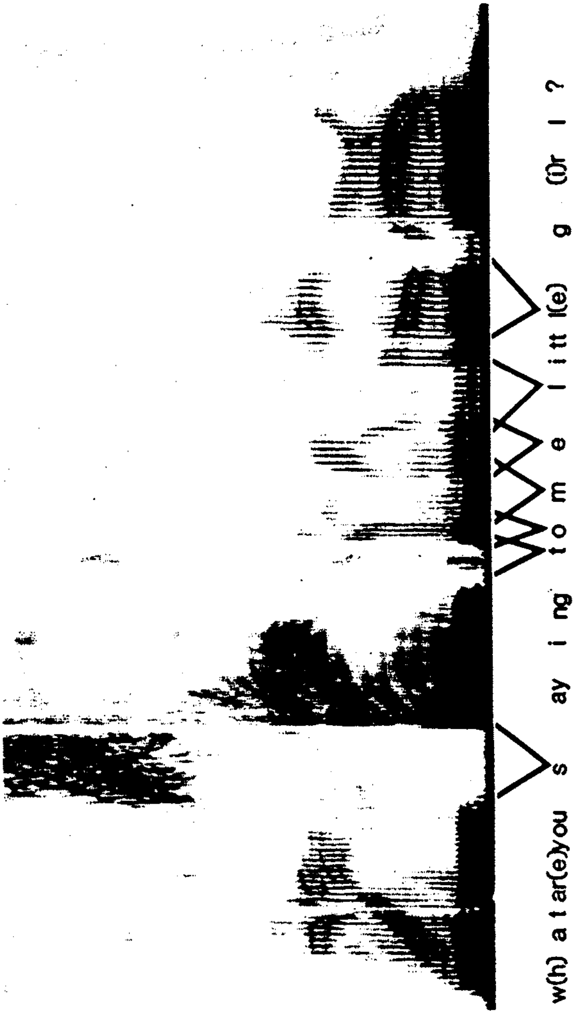

FIGURE 1. Spectrogram of the sentence "What are you saying to me, little girl?" spoken by a man to a young baby. Reduced 65% for reproduction.

an aperiodic noise burst, the following brief period of breathy aspiration, and the subsequent rapid transitions of the format frequencies (i.e., the upglide at onset of the lowest formant and downglides in the pale higher formants). Again, the knowledgeable listener may paradoxically perceive acoustic discontinuities as integral to a unitary phoneme.

Another difficulty with matching phonemes to acoustic segments derives from the temporal overlap of adjacent phonemes. Speakers do not simply finish one phoneme and, at precisely that time, begin the next. Interconnected trajectories characterize the neighborly relations not only of words, but also of vowels and consonants, for example, the trajectory in "me" from the lips being closed for /m/ to the lips being open and tongue blade high for "ee." The structure of the vocal tract does not permit an instantaneous change from one configuration to another, just as a runner's leg cannot instantaneously change from flexed to extended position. In the sample sentence, not only is a discrete boundary missing between the words "to" and "me," but there is none within "to" to define the end of /t/ and the beginning of the vowel. Yet a perceiver familiar with the language can identify individual phonemes as well as individual words.

The trajectories between target vocal-tract configurations, however, account only partially for the acoustic interconnection between phonemes. Often, vocal-tract adjustments for a phoneme begin one to several segments ahead of it, or persist one to several segments beyond. In other words, there is some *coarticulation* among nearby phonemes (e.g., Bell-Berti & Harris, 1979; Fowler, 1980). While pronouncing the /t/ in "too," a speaker usually is already rounding his lips appropriate for the "oo." This lip-rounding is not a standard property of /t/—for the /t/ in "tee," the corners of the lips are instead pulled back slightly for the following "ee."

1.3.3. Phonemic Context Effects and Acoustic Variability

Coarticulation among phonemes causes the acoustic characteristics of any item to be assimilated to its neighbors. The articulatory difference between the two /t/'s results in "too" beginning with a somewhat lower frequency noise burst than "tee." The paradox or puzzle is that, although the perceiver recognizes an invariant identity for a vowel or consonant across various phonemic contexts, there is no clearcut invariance in its raw acoustic properties.

Movement trajectories also contribute to this acoustic variability problem. Their shapes are determined by the vocal configurations they interconnect and there are rarely definable boundaries in conversational speech between a static configuration and a trajectory into or out of it. Figure 1 indicates that, because of the differences in the surrounding phonemes, the first and the last /l/ in "little" are acoustically different, both in the flanking format trajectories and in the exact frequencies of the flatter formants midway through the "segment."

1.3.4. Vocal Tract Variations and Perceptual Normalization

Not illustrated in the figure is a broader problem of acoustic variability: the acoustic contextual variation caused by different speakers. Of importance are the differences found between males and females, or between children and adults. On the average, female vocal tracts are smaller than those of males, which biases the acoustics of female speech toward higher frequencies in voice pitch and in formant frequencies. More important, though, are the age and gender differences in proportional relations among vocal-tract areas. The ratio between the distance from the vocal cords to the base of the tongue, versus the distance from the lips to the base of the tongue, is greatest for adult males and smallest for young infants (Goldstein, 1979; Lieberman, Harris, Wolff, & Russell, 1971).

Because formant frequencies are determined by the sizes of the vocal resonating chambers, these vocal-tract ratio differences cause age and gender differences in the proportional relations among formant frequencies for a given vowel. It has been impossible thus far to derive a simple mathematical formula for the formant frequency relations of a vowel produced by proportionally differing vocal tracts. In other words, there is no invariant acoustic description of formant frequency relations across men's, women's, and children's utterances of the vowel (Bernstein, 1981; Broad, 1981; Kent & Forner, 1979). The puzzle is that listeners, at least those familiar with the language, immediately hear the vowel's identity across a variety of vocal tracts differing in size and proportion. This perception of constancy in the face of speaker-specific acoustic variations has been referred to as the vocal-tract *normalization* problem.

1.3.5. Summary

These acoustic properties thus pose a number of difficulties for the perceptual capture of words or phonemes from conversational speech, even for adults listening to their native language. These difficulties can cause a sentence spoken in an unfamiliar language to sound like a rather undivided flow, when the listener is not prepared to handle them. However, when adults listen to their native language, they can identify discrete phonemes and words. Most likely, this is because they already know the phonemes and many of the words in their language, as well as the permissible ways by which items of either type can combine. Infants, on the other hand, do not have this knowledge of language. Yet they must be able to "solve these puzzles" in perceiving speech in order to ultimately discover words, since the words directed to infants are usually embedded in phrases or sentences (Kaye, 1980) and are presented to them by a variety of people in different speech contexts. What might infants perceive in speech, at the level of the phoneme, that could help them to recognize discrete words within the flow of sound? For consideration of this question, the discussion now turns to the infant speech perception literature.

2. THE FOUNDATIONS OF INFANT SPEECH PERCEPTION RESEARCH

Research on infant speech perception has largely been guided by two theoretical approaches to one overriding issue and its underlying assumptions, as indicated earlier. Although the questions and discussion presented thus far have been oriented around the eventual discovery of words by infants who initially have very limited knowledge about vocal communication, this has not been the major issue in research and theory on infant speech perception. Instead, the primary theoretical focus has been on how infants solve the acoustic puzzles of phoneme perception (e.g., see Jusczyk, 1981a,b). Its main underlying assumptions have been that (a) the basic perceptual unit in speech *is* phoneme-sized, (b) the speech percept derives from intraperceiver transformation(s) of the acoustic properties of the stimulus, and (c) the source of the transformation is an innate mechanism of the auditory system.

Much of the research has been generated by a controversy over the nature of the transformation(s) and supporting mechanism that can be traced to a similar controversy in experimental work on adult speech perception. On one side is the *phonetic* interpretation of infant speech perception, which posits that the perceptual mechanism is uniquely human by nature and differs qualitatively from the means for perceiving other sounds. Proposals about the exact properties of the specialized phonetic mechanism have ranged from a comparator that matches incoming speech sounds to the neuromotor commands for producing them (the motor theory of speech perception: Liberman *et al.*, 1967) to innate categories of linguistic features of phonemes (e.g., Eimas, Siqueland, Jusczyk, & Vigorito, 1971) that may be mediated by innately tuned neural feature-detectors (e.g., Cutting & Eimas, 1975).

On the other side of the controversy is the *psychoacoustic* approach, which presumes the machinery for the speech-to-percept transformation to be neither uniquely human nor limited to the perception of speech. In the psychoacoustic view, the general organization of the mammalian (or primate) auditory system yields an invariant stimulus-bound response whenever a given acoustic property occurs, regardless of the class of sound (e.g., speech vs. nonspeech) to which the individual is listening.

In the following summary and interpretation of the literature, it will be argued that the psychoacoustic-phonetic controversy in infant speech perception research is misguided. Both views are inadequate because they fail to consider the relation between infant and language. Following that review and discussion, a promising alternative theoretical perspective on infant speech perception will be described: the ecologically motivated (e.g., J. J. Gibson, 1966; Summerfield, 1978) *speech source perception* view outlined earlier. But at this point, the issue that has guided existing infant speech perception research, the psychoacoustic-phonetic controversy, must be placed in proper historical perspective.

2.1. The Empirical Beginnings

Research on infant perception of phonemes began with two reports in 1971, both of which gave a phonetic interpretation to the underlying processes. Each study employed a variant of the habituation paradigm to map the limits of infants' phoneme categories via their discriminations of syllable pairs differing in initial consonant. One of the studies found that 5 to 6-month-olds can discriminate natural utterances of /ba/ and /ga/ (Moffitt, 1971). The consonants in the tested syllables are both voiced stop consonants (along with /d/; the voiceless stops are /p/, /t/, and /k/), but they differ in place of articulation. Since the infants discriminated between consonants that differed solely in place of articulation, the author concluded that "linguistic-perceptual capacities are present during early life" (p. 717).

The other study (Eimas *et al.*, 1971) has received the preponderant attention in subsequent research. In this study, 1- and 4-month-olds were presented with computer-synthesized versions of /ba/ and /pa/, which differ in the articulatory property called *voice onset time* (VOT). That is, in the voiceless /p/ of /pa/, the vocal cords begin vibrating later with respect to the lip-opening gesture than is the case with the voiced /b/ of /ba/. Therefore, /p/ has a longer VOT than /b/. The acoustic consequences of articulatory differences in VOT are many (Lisker, 1978), but research attention has primarily focused on the time difference between the consonantal noise burst and the onset of periodic voicing, referred to here as *acoustic* VOT. It is usually confounded with other acoustic differences between a voiced-voiceless consonant pair.

Computer synthesis was used in the Eimas *et al.* study to produce a systematic series or continuum of syllables, which varied in equal-sized steps along the acoustic VOT dimension. Such acoustically controlled continua usually cannot be produced by a human speaker because of mechanical constraints on possible vocal tract movements (although in the case of stop voicing, human speakers can produce a range of different acoustic VOT values). Adult listeners typically fail to hear the gradual steps of acoustic change along synthetic continua between two contrasting consonants. Instead, they identify all stimuli as exemplars of one or the other phoneme category, and a sharp boundary on the continuum separates the two perceptual categories. Adult listeners also discriminate between acoustically different pairs of synthetic stimuli much better when the members are from different phoneme categories than from the same category. This pattern of identification and discrimination results has been termed *categorical perception* (see Figure 2). It was originally taken as evidence for a specialized phonetic mode of perception, since the nonspeech continua that had been similarly studied were perceived continuously (i.e., no clear labelling boundary or no clear performance peak in discrimination ability) rather than categorically (e.g., Liberman *et al.*, 1967; Repp, 1982).

To determine whether infants also perceive speech categorically, and by presumption phonetically, Eimas *et al* (1971) tested their discrimination of synthetic syllables that differed in acoustic VOT, but either did or did

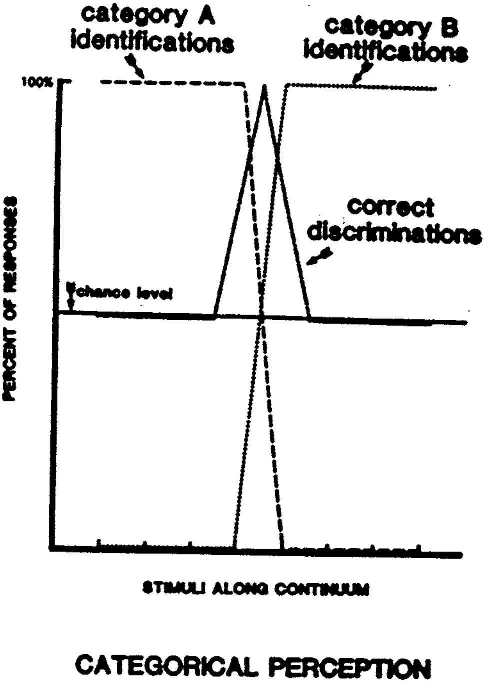

FIGURE 2. A schematic diagram of ideal results of categorical perception tests.

not differ according to American English /pa/ versus /ba/. The within-category pairs differed by the same magnitude of acoustic VOT as did the between-category pairs. The infants discriminated the between-category difference much better than the within-category difference, in agreement with the adult phoneme boundary. This finding led the authors to conclude that infants have an innate capacity to perceive speech linguistically, that is, in terms of adult phoneme categories.

These two early studies were scientifically intriguing, and encouraged a profusion of infant speech research that still continues. Some researchers essentially accepted the premise that the means by which acoustic properties are translated to a percept is speech-specialized, that is, specifically phonetic in nature. They went on to explore the range and nature of phoneme contrasts that young infants discriminate. As the psychoacoustic-phonetic controversy indicates, however, other researchers took issue with the phonetic perspective. The studies of this latter group have attempted to show that categorical perception is due to general psychoacoustic mechanisms, which respond to particular acoustic attributes whether they appear in speech or nonspeech sounds, but which occur most frequently in speech (e.g., Blumstein, 1980; Stevens & Blumstein, 1978).

2.2. Infant Perception of Phoneme Contrasts

Many of the early studies employed synthetic syllable continua, since explaining categorical perception has been a key interest in the psychoacoustic-phonetic controversy. According to the definition of categorical perception, percentage-correct discrimination of stimuli from a continuum must show a performance peak at the position of the category boundary for labeling of those stimuli (see Figure 2) (Studdert-Kennedy, Liberman, Harris, & Cooper, 1970). Therefore, both identification and discrimination data are needed to assess categorical perception. However, because there is no currently acceptable test for infants' identification of sounds, the conclusions of the infant research are based on discrimination data only (see Jusczyk, 1981a). Since the infant discrimination data cannot be compared to infant identification responses, it may be better to refer to the peaks and troughs in their discrimination of synthetic speech continua by some other term, such as perceptual *boundary effect* (Kuhl, 1981b; Wood, 1976).

A number of other infant studies from the phonetic perspective have employed natural rather than synthetic speech stimuli, thus focusing on discrimination of contrastive spoken examplars rather than on categorical discrimination. In both types of study, the discrimination data have been derived from either habituation tests or tests of generalization of conditioned operant responses, and the subjects have usually been between 1 and 7 months of age (prior to the reported onset of word comprehension). The consonant contrasts that were tested have usually occurred in syllable-initial position, which makes phoneme discrimination in natural speech easier for young children than do medial or final positions (Schvachkin, 1973).

The categorical discrimination studies have found that infants show a boundary effect for a variety of synthetic consonant contrasts, with their discrimination peak usually occurring at or near the adult American English identification boundary. These contrasts include stop consonant voicing (specifically, /p/ vs. /b/, and /d/ vs. /t/) as cued by acoustic VOT (Eimas, 1975a; Eimas *et al.*, 1971; Streeter, 1976) or by another naturally occurring cue, the extent of frequency change in the first formant (F1) transition (J. L. Miller & Eimas, 1981).

Infants also show a boundary effect for place of articulation differences between stop consonants (/b/ vs. /d/ vs. /g/). (C. L. Miller & Morse, 1976; C. L. Miller, Morse, & Dorman, 1977; Morse, 1972; Williams & Bush, 1978), even when the consonants occur in the middle of vowel-consonant-vowel (VCV) syllables (Jusczyk & Thompson, 1978) or in the final position of VC or CVC syllables (Jusczyk, 1977). Infant boundary effects have also been found for place of articulation distinctions between liquid consonants (/r/ vs. /l/, Eimas, 1975b). There is an infant boundary effect for place of articulation differences among fricatives (voiceless: /f/ vs. "th" as in "thanks"; voiced: /v/ vs. "th" as in "that"), although the infants' discriminations may differ in some respects from adults' (Jusczyk, Murray, & Bayly, 1979). Infants also show a boundary effect for manner of articulation differences between /b/–/w/ (same place of articulation, but stop vs. semivowel manner: Eimas & Miller, 1980a; Hillenbrand, Minifie & Edwards, 1979) and between /b/ –/m/, although they do not discriminate the latter contrast as categorically as adults, that is, they show moderate discrimination of the within-category pairs (Eimas & Miller, 1980b).

Adults perceive synthetic continua between vowel contrasts in a more continuous or less categorical fashion than consonants, unless the vowels are severely shortened in duration (e.g., Crowder, 1973; Liberman *et al.*, 1967; Pisoni, 1973, 1975). Infants show similar effects in perception of the "ee-ih" vowel distinction (Swoboda, Morse, & Leavitt; Swoboda *et al.*, 1978).

Consistent with the boundary effect findings, infants discriminate a fairly wide array of the natural consonant and vowel contrasts in spoken English. They discriminate natural stop voicing distinctions (Trehub & Rabinovitch, 1972), as well as some contrasts that children often do not produce correctly until late in phonological development (3–5 years), and which often cause persistent articulatory difficulties: certain fricative place (/s/–"sh") and voicing (/s/–/v/) contrasts (Eilers & Minifie, 1975; Eilers, Wilson, & Moore, 1977), the place contrast between the liquid consonants

/w/–/r/ (Eilers, Oller, & Gavin, 1978), and the consonant clusters /sl/–/spl/ (Morse, Eilers, & Gavin, 1982). Infants also discriminate the naturally produced vowels "ee" versus "ih" (Eilers *et al.*, 1978), as well as the "ee-ah-oo" triad, whether they occur in isolation or in CV syllables (Trehub, 1973).

To summarize, young prelinguistic infants discriminate many consonant contrasts and often show a boundary effect akin to the adult category boundary. They even discriminate some consonant contrasts that children produce late and often misarticulate. Infants also discriminate vowel contrasts, and do so less categorically than consonants, again like adults. (For more extensive discussion of the theoretical particulars, see Aslin & Pisoni, 1980; Cutting & Eimas, 1975; Eilers, 1980; Eilers & Gavin, 1981; Eimas, 1974a, 1975a; Jusczyk, 1981a, 1981b; Kuhl, 1978, 1980, 1981a; Mehler & Bertoncini, 1978; Morse, 1978; Trehub, Bull, & Schneider, 1981; and Walley, Pisoni, & Aslin, 1981.)

The performance pattern does not, however, indicate whether infants discriminate by psychoacoustic or phonetic means. Initially, researchers who took a phonetic view assumed that the boundary effect was evidence for a speech-specialized perceptual process. But that assumption could be questioned, and was submitted to test by researchers on both sides of the theoretical dichotomy. The alternative posed by the psychoacoustic perspective was, of course, that the perceptual boundary might be an attribute of the auditory system's response to the acoustic properties, rather than the phonemic identities, of the speech sounds.

2.3. Is the Boundary Effect "Phonetic" or "Psychoacoustic"?

A direct test of this question was to see whether infants show a discrimination boundary effect for some phoneme-differentiating acoustic property even when it occurs outside a speech context. Both Morse (1972) and Eimas (1974b, 1975b) isolated the major acoustic cue that had been manipulated to produce a place of articulation continuum, by stripping away the other acoustic properties that were shared by the contrasting phonemes, and presented young infants with the isolated cue continuum to discriminate. Morse (1972) tested infans with the isolated F2 for the /ba/–/ga/ contrast, whereas Eimas tested the isolated F2 cue for /da/–/ga/ (1974b), and the isolated F3 cue for /ra/–/la/ (1975b). Although in each case the isolated formants were the sole acoustic property that distinguished the phoneme contrast, and hence were crucial to adult categorical perception of that contrast, outside of their natural context they sounded like nonspeech "bleats."

The argument was that if the infant boundary effect reflects a uniquely speech-related perceptual specialization, it should occur in the perception of a particular acoustic cue only when that cue actually specifies a phoneme distinction. The infants in the Morse and Eimas control studies did show a boundary effect for the full syllables but failed to show one for the isolated formants, and in fact discriminated the latter very poorly. The authors interpreted their findings as support for the phonetically specialized nature of infant boundary effects.

The psychoacoustic perspective offered an alternate interpretation, however. The crucial stimulus attribute for psychoacoustically based boundaries could be the *interrelation* between the distinctive acoustic cue and the nondistinctive information provided in the other formants (e.g., Jusczyk, 1981a). Such a relational attribute would be destroyed by presenting the distinctive cue in isolation. Therefore, the Morse and Eimas studies could not definitively answer the controversy.

A more appropriate nonspeech control should maintain the interrelations among the acoustic features involved in a phoneme distinction. One such nonspeech distinction is the difference in risetime, or time from onset of a sound until it reaches its maximum intensity, between plucked versus bowed violinlike sounds, which is analogous to the "sha-cha" distinction in speech. Adults had been reported to perceive the pluck-bow distinction categorically, even though it is nonspeech (Cutting & Rosner, 1974). Jusczyk, Rosner, Cutting, Foard, and Smith (1977) extended that finding to infants and presented it as evidence that infant perceptual boundary effects have a psychoacoustic basis. However, subsequent replications of the adult study uncovered a stimulus problem. The acoustic differences among the original pluck-bow stimuli were not of equal magnitude throughout the continuum; when this source of acoustic discontinuity was removed, adults no longer perceived pluck-bow categorically (Rosen & Howell, 1981). Since Jusczyk *et al.* employed the original pluck-bow continuum, the infant findings must be questioned (Jusczyk, 1981a,b).

A second infant nonspeech study was subsequently run, using *tone onset time* (TOT) differences between the individual tones of a two-tone chord, which is an analogue for the acoustic VOT distinction in speech (Jusczyk, Pisoni, Walley, & Murray, 1980). Adults show a sharp TOT boundary in line with their boundary for acoustic VOT (Pisoni, 1977). Thus, the phonetic uniqueness of adult categorical perception has been called into question by the TOT results, along with similar reports on other nonspeech contrasts (e.g., J. D. Miller, Wier, Pastore, Kelly, & Dooling, 1976). In the Jusczyk *et al.* (1980) study, the infants discriminated TOT differences nearly categorically, leading the authors to conclude that earlier reported VOT boundary effects may not be unique to speech perception by infants either.

Nonetheless, the TOT findings also fail to offer a definitive choice between the phonetic and the psychoacoustic interpretations of infant perceptual boundaries. In contrast to adults, the infants failed to discriminate

TOT as categorically as acoustic VOT, and the position of their TOT boundary differed significantly from their acoustic VOT boundary. Jusczyk *et al.* (1980) claim that this does not damage the general psychoacoustic stand, since TOT only partly captures the *articulatory* VOT distinction (i.e., the psychoacoustic key could be some other, untested acoustic attribute of articulatory VOT). However, the developmental data are at odds with this logic. The TOT and acoustic VOT boundaries *do* match for adults, indicating that a perceptual change must occur between infancy and adulthood. Yet the infant and adult VOT boundaries match. Therefore, it is the TOT boundary that changes developmentally, and not the VOT boundary, in contradiction to the claim that infants come to perceive speech distinctions via psychoacoustic means (Jusczyk, 1981b).

In any event, the evidence from the infant nonspeech perception research is weak. The currently available nonspeech control studies present a theoretical stalemate between the phonetic and the psychoacoustic claims about the nature of infant speech perception. However, the proponents of the controversy have argued that the answer may lie elsewhere in the premises of the psychoacoustic-phonetic distinction. It might be settled by exploring whether phonetic perception is uniquely human, a claim made by the phonetic perspective that is rejected by the psychoacoustic perspective.

2.4. Is the Boundary Effect Uniquely Human?

Empirical attacks on the claim that humans are the sole possessors of categorical phoneme perception have involved assessing whether other mammals or primates show abrupt shifts in perceptual sensitivity around the phoneme boundaries. It was reasoned that, if animals showed a boundary effect, general psychoacoustic factors must then account for categoricity in speech perception, since by definition infrahumans cannot perceive in a humanly specialized manner. The relevance of this research to infant speech perception is that infants and animals are nonusers of language, but only infants are human and have the capacity to develop human language.

Researchers of animal speech perception have reported boundary effects, similar to those found with infants, for discrimination of the stop consonant voicing distinction by chinchillas (South American rodent) (Kuhl, 1981b; Kuhl & Miller, 1975; J. D. Miller, Henderson, Sullivan, & Rigden, 1978), and by rhesus monkeys (Waters & Wilson, 1976), and for the stop consonant place of articulation distinction by monkeys (Morse & Snowden, 1975; Sinnott, Beecher, Moody, & Stebbins, 1976). Chinchillas also discriminate the vowels "ah" and "ee" (Burdick & Miller, 1975); as do dogs (Baru, 1975), and exhibit boundary effects in go- no-go categorizations of voicing among stop consonants, with their boundaries falling at the position of human adult boundaries (Kuhl & Miller, 1978). Thus, the psychoacoustic interpretation is that the perceiver's knowledge of language is not a necessary precondition for the boundary effect, and apparently neither is membership in the human species.

Researchers who support the psychoacoustic view of speech perception have used three animal findings to propose the following picture of the evolution of speech perception and production: the mammalian auditory system has specialized notches in sensitivity for certain regions of certain acoustic dimensions. These psychoacoustic specializations placed selective pressures on the choice of phoneme contrasts by human languages. The productions of the phonemes chosen must have capitalized on just the acoustic domains that are most neatly suited to mammalian psychoacoustic specializations (Kuhl, 1981b; Kuhl & Miller, 1975, 1978; J. D. Miller, 1977; Stevens, 1972). By extension, human infants possess those same psychoacoustic sensitivities (Aslin & Pisoni, 1980; Jusczyk, 1981a,b; Kuhl, 1978; Walley, Pisoni, & Aslin, 1981), and their attention is thus captured by the acoustic attributes of the language in their environment.

However, the claims of this psychoacoustic proposal belie the clarity of the infrahuman data. The animal data fail in several ways to match those of human adults. Recall the earlier argument that a determination of categorical perception depends on data from labeling and from discrimination tests; the animal research has necessarily relied only on discrimination data. Indeed, animal discrimination boundaries are less sharply defined than human adult phoneme boundaries (Kuhl & Miller, 1978). In other words, animals are noticeably better than human adults at discriminating within-category acoustic differences in a place of articulation continuum, and worse at discriminating between-category differences, indicating lowered categoricity (Morse & Snowden, 1975). The between-species categoricity difference is consistent with the greater sensitivity of human adults to formant-onset frequency changes, by a factor of about two (Sinnot *et al.*, 1976). In addition, only human adults show reaction time increases, large ones in fact, when making within-category discriminations (Sinnott *et al.*, 1976). Finally, the absolute position of the human adult category boundary is more stable than the monkey boundary, in the face of variations in the acoustic range covered by a synthetic phoneme continuum (Waters & Wilson, 1976). Humans obviously do show speech-relevant perceptual specializations beyond the limits of the other mammals tested.

The animal and nonspeech research may suggest that the boundary effect is not absolutely speech-specific and species-specific. However, there are clear and unexplained differences remaining between the human adults' perception of speech and the control studies on animal speech perception and infant nonspeech perception. Thus, the basic theoretical choice between the phonetic and psychoacoustic explanations of speech perception, especially in infants, is still open. If studies of the boundary effect have failed to solve the psychoacoustic-phonetic quandary, then possibly a more abstract characteristic of speech perception would (such as the perceptual constancy of phoneme identity across phonemically irrelevant acoustic variations).

2.5. Phonemic Perceptual Constancy in Infants

Recall the earlier discussion of some puzzles in the fit between speech acoustics and perception, particularly the lack of satisfactory acoustic descriptions for the invariant identity of a phoneme across different contexts of surrounding phonemes (the acoustic variability problem) or as uttered by differently proportioned vocal tracts (the normalization problem). In spite of these puzzles, adults perceive the identity of a phoneme spoken in widely different words and by different vocal tracts with seeming immediacy and effortlessness.

To see whether infants show a similar perceptual constancy, Fodor and colleagues (Fodor, Garrett, & Shapero, 1970; Fodor, Garrett, & Brill, 1975) trained them to respond operantly to a pair of vowel-differing syllables that either began with the same consonant (e.g., "pee"–"poo")or began with different consonants (e.g., "pee"–"kah"). In both conditions, the consonants differed acoustically because of the change in vowel context. The authors wanted to assess whether, despite the acoustic variability, the infants learned the operant response more easily for the consonantal match. The infants were later tested on a new syllable (e.g., "pah"), to determine whether they generalized the learned operant response more consistently from the consonant-matched pair to a new syllable beginning with the same consonant. The infants did learn and generalize more consistently for consonantal matches than consonantal mismatches. If they had learned to associate syllable pairs simply by remembering the pairing of their dissimilar acoustic properties, they should have responded to the mismatched-consonant pairs as consistently as to the consonant-matched pairs. The authors concluded that these prelinguistic infants had maintained perceptual constancy for consonantal identity in the face of concurrent acoustic variations, and that this ability must depend "on innately determined phonological identities (p. 180)."

There are two problems with this claim. Their task was difficult for infants to learn, and several attempted replications or extensions have failed. In addition, the psychoacoustic perspective offered an alternative interpretation: The perceptual constancy might reflect a response to some (yet uncovered) higher order acoustic invariant shared by the varying instances of a given consonant (Kuhl, 1980, 1981b). Using a different technique than Fodor *et al.*, Kuhl and her colleagues found perceptual constancy in young infants for the vowels "ah" versus "ee" (Kuhl, 1979) and for the fricatives /f/ versus /s/ (Holmberg, Morgan, & Kuhl, 1977) across different neighboring phoneme contexts and different speakers. Thus, consistent with the Fodor *et al.* report, the infants solved the acoustic invariance problem. Since perceptual constancy was maintained across different speakers the infants also solved the normalization problem. But whereas Fodor *et al.* favored the phonetic viewpoint, Kuhl and others favor the psychoacoustic viewpoint (e.g., Jusczyk, 1981b; Walley *et al.*, 1981), in part because chinchillas and dogs show perceptual constancy for "ah" versus "ee" across speakers and pitch contours (Baru, 1975; Burdick & Miller, 1975). Chinchillas also show such constancy for /t/–/d/ even in different vowel contexts (Kuhl & Miller, 1975).

The perceptual constancy findings thus indicate that infants can somehow solve two seemingly knotty acoustic puzzles to reach an important perceptual aspect of phoneme identity. Once again, the findings apparently do not allow a theoretical choice between the phonetic and psychoacoustic explanations of infant speech perception. Also, as will be argued in the next section, neither is the theoretical choice decided by considerations about the innateness of infant phoneme perception.

2.6. Innateness of Infant Phoneme Perception Effects

A pervasive notion on both sides of the psychoacoustic-phonetic controversy has been that a boundary effect or perceptual constancy effect is innate if infants show it "at the earliest age tested" (e.g., Aslin & Pisoni, 1980; Eimas, 1975a; Eimas *et al.*, 1971; Jusczyk, 1981a,b; Kuhl, 1978). Curiously, the empirical foundation for this belief includes almost no data before 1 month, a handful of studies on 2-month-olds, and many studies that have collapsed data across 1–4 months, 4–6 months, 6–8 months, or 10–12 months. Few have compared different age groups (cf. Best, Hoffman, & Glanville, 1982; Eilers, Wilson, & Moore, 1977; Werker, 1983; Werker & Tees, 1982).

This view apparently assumes that "an infant is an infant is an infant," across at least the first 6 months of life. In studies that averaged over several months of age, the "earliest age" cannot be trusted since it refers to the youngest infant they tested, even though group data were reported and appropriate age analyses were almost never run (but see 1- versus 4-month-age differences in Eimas *et al.*, 1971). Thus, it is nearly impossible to assess which, if any, perceptual boundaries are innate, or presumably biologically determined and inborn. All we can note are the ages below which a given perceptual effect has *not* been shown; for this review, the conservative assumption will be that the average age tested is the correct "earliest age" to show the reported effect.

The literature offers the following observations: First, infant boundary effects matching the adult findings have been shown only by a mean of 2½ months (e.g., Eimas, 1974b, 1975b; Morse, 1972). Conversely, newborns and 1-month-olds do not discriminate VOT absolutely categorically (Eimas *et al.*, 1971; Molfese & Molfese, 1979). Two-month-olds do not perceptually integrate two types of fricative acoustic information as adults do (Jusczyk *et al.*, 1979) nor do they discriminate phoneme contrasts under demands on short-term memory (Best *et al.*, 1982; Morse, 1978). These data hint at a perceptual change sometime between 1 and 2½ months, which is consistent with widespread biobehavioral and social changes around 6–10 weeks (e.g., Clifton, Morrongiello, Kulig, & Dowd, 1981; Emde & Robinson, 1979; Haith, 1979), which includes vocal behavior (Oller, 1980; Stark, 1980), suggesting that early changes in speech perception should be further explored (see Werker, 1983, for an example of important age changes in speech perception by older infants).

When infant and adult categorical discrimination has been directly compared, 3-month-olds' (Eimas & Miller, 1980b) and even 7- to 8-month-olds' performance differs significantly from that of adults (Aslin, Pisoni, Hennessy, & Percy, 1981; Eilers, Wilson, & Moore, 1976). Differences from adults, in fact, persist until at least 5–6 years of age (L. E. Bernstein, 1979; Garnica, 1973; Robson, Morrongiello, Best & Clifton, 1982; Schvachkin, 1973; Simon & Fourcin, 1978; Werker & Tees, 1981; Zlatin & Koenigsknecht, 1976). Therefore, boundary effects cannot be considered innate in some absolute sense.

Second, perceptual constancy for vowels is only certain as early as a mean of 2½ months (Kuhl & Miller, 1975). Perceptual constancy for consonants has not been reported earlier than 4 months (Fodor *et al.*, 1975) or 6 months (Holmberg *et al.*, 1977; Kuhl, 1980). Thus, arguments for the innateness of perceptual constancy effects (e.g., Jusczyk, 1981b) should also be held in check.

The exact timing, causes, and nature (i.e., phonetic vs. psychoacoustic) of changes in speech perception cannot be inferred from this literature, however. Even if age had been systematically studied, conclusions would still be limited by the near-exclusive reliance on discrimination measures. The coincidence of an adult phoneme boundary and a peak in infant discrimination is not sufficient evidence to claim "adultlike" perception of the contrast. As argued earlier, assessment of categorical perception requires both labeling and discrimination tests. Simple discrimination cannot reveal whether the distinction was perceived *as a phoneme contrast* (see also Jenkins, 1980), a concern that is equally relevant to the animal research. Discrimination indicates only that *some* difference was detected. Since phoneme discrimination is dissociable from phoneme category identification in aphasics (Blumstein, Cooper, Zurif, & Caramazza, 1977; Riedel, 1981), it is equivocally involved in aspects of perception that are closer to the meaning of language than mere acoustic contrast detection. Because language-dependent perceptual qualities are more likely to change developmentally, discrimination is inadequate as the sole measure of development in speech perception.

It is uncertain which, if any, speech perception effects are innate, and how they might change developmentally prior to the discovery of words. More crucial for the auditory-phonetic controversy, the following questions remain unanswered: Even if perception of a phoneme contrast is innate, is that innately possessed quality phonetic or psychoacoustic in nature? And if perceptual change does occur during prelinguistic infancy, what is the nature of the change? Each side of the controversy has provided answers (see Aslin & Pisoni, 1980; Trehub, Bull, Schneider, 1981; Walley, Pisoni, & Aslin, 1982; cf. Eilers, 1980; Eilers, Gavin, & Wilson, 1979; Eimas, 1975a), and the auditory-phonetic choice remains unclear. One potential motive force for early perceptual changes can be eliminated, however: if they do occur, they could not have a linguistic motivation from the infant's perspective. This fact causes difficulty for both sides of the controversy, as the next section indicates.

The data on categorical perception of nonspeech, animal perception o phoneme contrasts, perceptual constancy for phonemes, and innateness o infant phoneme perception have thus failed to decide between the psychoacoustic and the phonetic explanations of infant speech perception. When an impasse as extensive as this is reached, it is important to consider whethe the difficulty is not with the research but rather with the logic of the tw theoretical views themselves.

3. QUESTIONING THE AUDITORY-PHONETIC QUESTION

As stated earlier, the tacit assumptions shared by the two sides of th dichotomy have generally been (a) that the units of speech perception ar phonemes (cf., Bertoncini & Mehler, 1981; Jusczyk, 1981a); (b) that som intraperceiver interpretive process must transform acoustic properties t phonemic percepts; and (c) that the process is mediated by some specialize neural mechanism(s). All three assumptions can be questioned, particular in relation to the infant's discovery of words. The first assumption require that infants segment phonemes from connected speech. Phonemic segmen tation has not been assessed in infants, and is not straightforward even whe the perceiver *is* a language-user, since young children seem unable to explic itly segment phonemes (E. J. Gibson & Levin, 1975), as do illiterate adul (Morais, Cary, Alegria, & Bertelson, 1979). The notion that infants perceiv phonemes can also be questioned on a linguistic level (see discussion in th next paragraph). The second assumption entails that the listener shed meaning on the presumed meaninglessness of the superficial acoustics of speech, that is, the meaning of the stimulus resides solely in the listener and not directly in the signal. According to the third assumption, the transformation is accomplished by specialized nervous system structures or information-processing stages. Thus, the psychoacoustic and the phonetic views regard speech perception as a mechanistic intraperceiver process.

We turn now to a more detailed examination of each position in the psychoacoustic-phonetic controversy. The main problem with the phonetic view of infant speech perception is that infants presumably do not have a language system, having not yet discovered words. Phoneme contrasts cannot be perceptually available to infants, since they are defined by a language system, being dependent on word meanings in that system. They represent abstract relations among speech sounds that are used by the language to convey semantic differences, as in "pat" versus "bat." Although infants, like adults, categorically distinguish between /p/ and /b/ (e.g., Eimas *et al.*, 1971), they do not necessarily perceive such differences as *phonemic* contrasts (recall that discrimination is an equivocal measure of *phoneme* perception).

The phonetic view also has difficulty accounting for how and why infants would adjust their perceptual categories to suit the language of their environment. Presumably, the evolutionary advantage of innate phoneme categories is that they would filter out irrelevant within-category acoustic variations and thereby relieve the perceiver of having to deal with those unnecessary details. Yet young infants fail to discriminate some phoneme contrasts existing in certain languages according to the adult categories. How and why would infants learning those languages later become able to focus on those innately filtered-out details, in order to adjust their category boundaries or develop new categories? According to researchers on genetic evolution (Jacob, 1977), on nervous system function (Rose, 1976), on perceptual development (Spelke, 1979; Trevarthen, 1979), and on speech perception (Studdert-Kennedy, 1981b), the most efficient evolutionary solution for developmental adaptation to stimulus environments, such as that provided by the native language, is *not* an array of innate mechanisms that are tightly tuned to specific stimulus values, but instead a more flexible attunement to detect the range of stimulus values that could occur. Thus, any specialization we have for perceiving speech would have to be sufficiently flexible to adapt to the specific phoneme contrasts of one's own particular language.

A major drawback of the psychoacoustic view is the argument that the object of speech perception is intrinsically meaningless speech-neutral acoustic data. In particular, this claim has difficulty accounting for the perceived constancy of phonemes spoken by different vocal tracts (vocal-tract normalization) and spoken in different contexts of surrounding phonemes (acoustic variability). The ability to recognize the invariance of a phoneme or word spoken in different contexts and by different people is crucial for language-learning infants. The phonemes they hear occur in a variety of phonemic contexts; and the vocal tracts of the older speakers they hear (and must eventually base their own vocalizations on) differ proportionally from each other, as well as from the infant's own vocal tract. Vocal-tract normalization and acoustic variability do not appear to cause perceptual difficulties for infants. Both infants and other animals apparently solve the normalization and acoustic variability problems in their discriminations among phonemes (e.g., Baru, 1975; Kuhl, 1979, 1980, 1981b; Lieberman, 1980). The psychoacoustic view does not adequately explain the infant's perceptual solution to those problems. The nontrivial acoustic variations involved have thus far defied speaker-independent and speech-neutral acoustic definitions for either phonemes or words. Thus, it cannot be assumed that the infant's solution focuses on speech-neutral acoustic information.

A problematic implication of the psychoacoustic view is that the infant must at some time move from perceiving speech in purely auditory terms to perceiving linguistic structures, such as phonemes (Jusczyk, 1981a, 1981b). What would lead the infant to take the cognitive step from meaningless acoustics to meaning at the level of either phonemes or words? One proposal would be the empiricist philosophical perspective that infants learn meanings by contextual association. However, the associationist solution is unsatisfactory on logical grounds (e.g., E. J. Gibson, 1977; J. J. Gibson, 1966; J. J Gibson & Gibson, 1955; Jenkins, 1974). Meaning cannot emerge from meaninglessness, so some meaningful element would have to predate the infant's first association.[1] The traditional empiricist perspective is that the elements of meaning are extrinsic to the individual, introduced by sensory stimulation However, the psychoacoustic view assumes that meaning is *not* intrinsic to the speech stimulus, and it has provided no argument that other stimulation is intrinsically meaningful. In other words, the psychoacoustic view gives us no reason to believe that sensory stimulation in any modality provides extrinsic elements of meaning.

If meaning cannot be assumed to derive from extrinsic stimulation, the traditional nativist perspective offers the alternative proposal that element of meaning are innate or intrinsic to the infant. The prime candidates fo innate elements of meaning in infant speech perception, of course, would be phonemes or phoneme contrasts. This is exactly the position of the phoneti

---

[1] The general underlying issue of meaning is complex and certainly cannot be settled here. Th source of meaning for our knowledge of the world is at the heart of a centuries-long debate i epistemology between phenomenologists and realists. Psychology has taken up this debate. N satisfactory solution, acceptable to all, has been reached in either field. In the context of thi chapter, the discussion reflects the author's view on the relation of the topic to how infant perceive speech.

viewpoint questioned previously, and is antithetical to the psychoacoustic viewpoint because it cannot be invoked for animal speech perception. A third possibility is the constructivist perspective that the infant cognitively constructs meaning for sensory stimulation that does not iself provide intrinsic meaning. But this would still depend on some mechanism that determined the nature of the meaning to be constructed, and again the most likely mechanism for speech would be phoneme-based.

A fourth, nontraditional perspective on the source of meaning in speech can be offered, however, which is not consistent with either the psychoacoustic or the phonetic viewpoints. This is the ecological perspective (e.g., J. J. Gibson, 1966) that meaning is directly available to perception in the active, adaptive relation between the perceiver and the objects/events being perceived. It will be discussed at greater length in the subsequent sections of the chapter.

The psychoacoustic position may also be troubled by its evolutionary proposal. It assumes that specialized perceptual "notches" in the sensitivity of the mammalian auditory system along certain phonemically relevant acoustic dimensions have imposed selective pressures on the phonemes that can be uttered by the human vocal tract (e.g., Aslin & Pisoni, 1980; Kuhl, 1981b; J. D. Miller, 1977; Stevens, 1972; Walley *et al.*, 1981). An alternative proposal is that although the anatomy of the human vocal tract places quantal limits on the sounds it can make (see Stevens, 1972); it is this fact that has placed selective pressures on the evolution of specialized notches in the auditory system.

Specializations of neural tissue, like any structural specializations, are naturally selected (evolve) because they have suited some purpose. Yet the psychoacoustic model fails to specify a purpose that could have selected for the speech-related perceptual notches or discontinuities in the mammalian auditory system. Certain basic properties of the auditory system probably are shared by all mammals, reflecting selective adaptation to the commonalities in their auditory environments such as the sounds of weather, vegetation, predators, and prey. However, individual mammalian species do develop more highly specialized sensitivities for certain acoustic properties that are uniquely suited to the particulars of their own ecological niche (e.g., the bat; Neuweiler, Bruns, & Schuller, 1980). For some mammals, notably humans, species-specific vocalizations are particularly important to the species' survival, and have probably placed selective pressures on the development of specialized responsivities of the auditory system. In these species, we would expect to find that specializations in auditory sensitivity have evolved to be uniquely responsive to the vocal characteristics of the species (see also Petersen, 1981; Zoloth, Petersen, Beecher, Green, Marler, Moody, & Stebbins, 1979), which is the converse of the psychoacoustic model of evolution. In the case of speech perception by humans and animals, recall that the specialized notches or discontinuities do indeed appear to be more sensitive and finely tuned in human adults than in the animals studied (see Section 2.4).

The general principle in evolution and in ontogeny of the nervous system has been that motor functions (e.g., vocalization) precede and often motivate the development of the correlated sensory functions (Bekoff, 1981; Horridge, 1968). Consistent with that principle, motor areas develop in advance of sensory areas at each level of the neuraxis (Jacobson, 1978), including the human neocortex (Marshall, 1968; Tuchmann-Duplessis, Aroux, & Haegel, 1975). Even more important, the motor-sensory precedence applies to the neural supports for human speech perception and production: the development and maturation of Broca's area (the motor speech cortex) precedes that for Wernicke's area (the receptive speech cortex) (Rabinowicz, 1979). The more likely evolutionary scenario, then, may be that the human auditory system's properties were selected for best responsiveness to the sound-producing abilities of the uniquely human vocal tract (e.g. Studdert-Kennedy, 1981c), rather than vice versa as the psychoacoustic view suggests.

In summary, both contemporary views of infant speech perception are flawed. The chapter's introduction pinpointed four knotty perceptual problems as requisites to discovering words in the speech stream: (a) disembedding words from connected speech; (b) recognizing word pattern invariances across different utterances and vocal tracts; (c) recognizing the variations that do specify different word patterns; and (d) hearing how to imitate a pattern made by another person. Neither the psychoacoustic nor the phonetic viewpoint offers the infant adequate means for discovering words or phonemes in speech.

But what is the alternative? The remaining sections on *speech source perception* focus on the organization of the speed medium for an answer. The vocal tract offers a structural and dynamic meaning that is intrinsic to speech itself, since as the source of speech it determines the shape of its acoustic product (e.g., Fant, 1973). It would be more parsimonious for perceivers to directly and actively attend to the available vocal-tract information that is intrinsic to speech, than to have the perceptual process mediated with a step involving the meaningful interpretation of meaningless superficial acoustics. According to the *speech source* view, meaning exists in perceiver's relation to speech at its source. This alternative approach derives from the general ecological approach to perception taken by James Gibson and his followers (e.g., J. J. Gibson, 1966; Fowler & Turvey, 1978; Studdert-Kennedy, 1981a, 1981d; Summerfield, 1978; Verbrugge, Rakerd, Fitch, Tuller, & Fowler, in press).

4. AN ECOLOGICAL PERSPECTIVE ON INFANT SPEECH PERCEPTION

Perception of the speech source implies that the infant attends to intrinsically specified information about the vocal tract and the articulatory events that shaped the speech medium. This premise is consistent with the ecological argument that perceiving organisms actively seek information about distal events, which is lawfully specified in the stimulus array (J. J. Gibson, 1966; E. J. Gibson, 1977). It stands in contrast to depictions of perception as the cognitive or neural *transformation* of proximal sensory data, which are intrinsically meaningless and informationally impoverished with respect to the distal event. The speech medium carries many parallel messages, some of which are defined within a particular language, and thus presumably not detected by infants, who do not yet recognize that phonemes can function to distinguish word meanings. Others are human universals, such as the paralinguistic messages of emotional affect, of regional accent, and of age or gender effects on vocal-tract size and configuration.

The speech characteristics of interest for the current discussion are the structural organization and articultory gestures of the vocal tract. According to speech source perception, to perceive these characteristics is to simultaneously apprehend the constant anatomical structure and the transforming positions of articulators in a speaking vocal tract (see also Schubert, 1974). This proposition is supported by the speed and accuracy with which adults and even young children can imitate speech sounds, in spite of the "normalization problem" (e.g., Alekin, Klass, & Chistovich, 1962; Ferguson & Farwell, 1975; Galunov & Chistovich, 1966; Kent & Forner, 1979; see Studdert-Kennedy, 1981a). However, the proposition may seem counterintuitive in two manners. First, common sense suggests that we can *see* the structure and movements of objects, but that we *hear* only "sounds." Thus, the claim that we hear structure and movements, especially the small hidden ones of vocal tracts, may seem unlikely. Second, it may seem implausible that structure and motion are captured at once in the same information, because the qualities of form and movement seem dissociable in our experience. However, an ecological appreciation of sound and hearing shows these "problems" to be false.

4.1. Ecological Acoustics and Speech

Acoustic energy is the radiation over time and space of a wave of rapid alternations in air pressure. It originates from the oscillatory motion of some object or surface that compresses and rarefies the distribution of air molecules around it. Both an object/surface and some oscillatory motion are necessary to produce acoustic energy. In fact, their contributions to sound cannot be dissociated. Structural properties constrain the sound-producing motions an object/surface can undertake; in turn those motions temporally deform the structure in a characteristic manner. It follows that the specific properties of an acoustic flow (e.g., time-varying frequencies and amplitude of the pressure wave) necessarily reflect those structural properties of the object and its vibratory movements that shaped the sound-production.

By the ecological view, the interdependence between structure and transformation is at the core of real events and therefore of their perception (Shaw & Pittenger, 1977). The nature of an object is revealed to the perceiver through event-determined, codefined information about *structural invariants* and *transformational invariants* in objects/events. These terms refer, respectively, to the structural identity of an object undergoing some transformation or change (e.g., by its movement or structural deformation, or by changes in the observer's orientation to it), and the transformations or changes it partakes in. Consider, for example, one person saying three different words as opposed to three people saying a single word. In the first instance, information common to the three words reflects the structural invariance of that speaker's vocal tract. In the second case, there is a transformational invariant in the articulation of the single word by three structurally different vocal tracts. Structural and transformational invariants lawfully shape the energy medium that carries their message (i.e., acoustical or optical energy). The ecological premise is that, through their modulation of the energy medium, transformations can perceptually specify an object's structure. In support of this, infants' visual recognition of objects and their structure is enhanced by watching the objects undergo various spatial-temporal transformations (E. J. Gibson, 1980; Ruff, 1980, 1982).

From the ecological perspective, auditory perception is the codetection of the transformations (motions) and structure of the sound source, which are veridically conveyed in the acoustic medium (see also Schubert, 1974; Warren & Verbrugge, in press). Thus, structure and motion are not only seen but heard. In the case of speech, the acoustic medium is better suited than optics for conveying the structure and transformations of the vocal tract, many of which are invisible in face-to-face communication as well as when the speaker is out of view. Articulatory gestures may be beyond the capabilities of vision in another way, since their speed and precision exceed the temporal and spatial resolution of the visually perceived manual American Sign Language (Studdert-Kennedy & Lane, 1980).

The unseen messages carried by natural speech reflect not only the origin of its acoustic energy (respiratory and laryngeal), but also the structural identity, biokinematic coupling, and specific movements of the speaker's supralaryngeal vocal tract.[2] In vocalizations and musical sounds (among others), the acoustic wave does not radiate freely from its oscillatory origin to the perceiver: the medium is also molded by the structure and transformations of an intervening resonating tube. The size and shape of a resonant cavity determine its natural resonating frequency (or frequencies) at which the air contained within its walls will oscillate when excited by a flow of air introduced from outside. If the extrinsic air flow is already oscillating (e.g., when acoustic energy from the vibrating larynx is introduced to the supralaryngeal vocal tract), then those oscillatory frequencies in the flow that match the resonant properties of the tube will be amplified in intensity; other frequencies that mismatch the resonant properties will be attenuated (filtered out). The larger a resonating cavity is, the lower its primary resonant frequency, which is also affected by the size and number and positions of openings in the cavity. As its shape deviates from perfectly spherical or cylindrical, particularly if there are corners or "side pockets," higher order resonant frequencies may be added. Surface properties, such as the smoothness and elasticity of the resonant tube's walls, largely determine the time course of intensity changes in the resonated frequencies.

In the case of speech, the critical sound-shaping properties of the resonant tube include the shape and elasticity of the cheeks, throat, lips, and tongue, and the position and rigidity of the teeth. The properties that allow the vocal tube to transform in shape are especially important for its articulatory gestures: the hinged movements of the jaw, and the moving and deforming obstructions of the tongue, lips, and velum (which opens the nasal passage to resonate for sounds like /m/ and /n/). These structural and transformational properties all shape the acoustic speech wave. The time-varying formants and other acoustic discontinuities (see Section 1.3) reflect, in particular, rapid transformations of the vocal-tract resonant configuration that are produced by movements of the tongue, lips, jaw, and velum, within the constraints posed by the enduring anatomical relations within the tract. (For more detailed discussions of the physical acoustics of speech, see Fant, 1973; Flanagan, 1973.)

Since these distal source properties determine the acoustic shape of speech, they should be available to perceivers (see also Studdert-Kennedy, 1981a, 1981d; Summerfield, 1978; Verbrugge *et al.*, in press). Infants and even animals should be able to detect at least some of the structural and

---

[2] Natural speech also identifies the speaker as a member of the human species. Synthetic speech, insofar as it "works" perceptually, must capture necessary information about human vocal-tract dynamics, and usually also about the structure of a generic vocal tract, often appropriate for an adult male of indeterminate age. Rarely, however, does synthetic speech capture sufficient "textural" detail about a natural vocal tract to sound like a live human speaker, even an unknown one.

transformational invariants in speech, although evolutionary and ontogenetic history will affect how well different perceivers are attuned to pick up the various messages. As discussed earlier, although animals and human infants do show phoneme boundary effects, their performance deviates significantly from human adults' categorical perception.

The speech source perception view may be further clarified by comparison and contrast with the psychoacoustic and the phonetic views. If auditory perception is the detection of information about source structure and transformations in the acoustic medium, then the perception of speech should be abstractly similar to the perception of other sounds, in agreement with the psychoacoustic view. However, the speech source perspective disagrees with the psychoacoustic notions that the perceiver's focus is on event- or speech-neutral acoustic parameters, and that these parameters need to be transformed by the system into percepts. The speech source proposal is also consistent in an important respect with the phonetic view. That is, speech perception even by the infant is considered "special" and uniquely human; however, that special perceptual quality is not agreed to be based on phonemes for infants. Moreover, the relative emphasis on "speech" and "perception" is different. Because the speech medium conveys its source to the perceiver, acoustic cues need not be transformed into percepts, whether by codes for phoneme categories or by neuromotor codes for phoneme production.

According to the speech source view, the specialness of speech perception derives from the unique structural and transformational properties of its sound source, the human vocal tract. It is unique in its complex anatomy, its biokinematic organization, and its particular dynamic gestures (e.g., Lieberman, 1967; Lieberman *et al.*, 1971), all of which are reflected in its acoustic productions. Moreover, humans have a privileged relation to speech as the tool of human-specific language communication.

Some of the advantages of this view for several important aspects of speech perception in adults and infants will be considered next.

4.2. Speech Source Perception in Adults

The major acoustic puzzles in speech perception research have been the problems of acoustic variability and normalization (see Section 1.3.). As a reminder, the acoustic variability problem refers to the sometimes quite striking variability in the acoustic properties of an invariantly perceived phoneme, which occurs primarily when it is produced in different contexts of surrounding phonemes. The variability is caused by coarticulation among phonemes as well as by the articulatory trajectories that interconnect adjacent phonemes. The acoustic properties of the phoneme are thus assimilated to the acoustic properties of its neighbors (e.g., in Figure 1, the differences among the /l/'s in "little" and "girl"). Another source of acoustic variability is the wide variety of acoustic features that can identify a given phoneme (e.g., Lisker, 1978). These sorts of acoustic variations pose a greater empirical puzzle for psychoacoustic accounts of consonant perception rather than vowel perception. They have much stronger effects on the formant trajectories and other acoustic features associated with consonants than on the formant frequencies in the nuclear portions of vowels (although the latter are also affected to considerable degree in conversational speech).

The normalization problem refers to another acoustic context effect, caused by variations in the dimensions of different vocal tracts. It causes greater difficulties for psychoacoustic explanations of vowel perception than consonant perception. The formant frequencies in vowel nuclei are more obviously affected by vocal-tract proportions than are the formant trajectory patterns associated with consonants. The different effects of these two types of acoustic variability in vowels and consonants suggest a difference in the information those two phoneme classes convey, which has been supported by several other lines of research on speech production (e.g., Fowler, 1980; Fowler *et al.*, 1980) and perception (e.g., Ades, 1977; Crowder, 1973; Cutting, 1974; Darwin, 1971; Pisoni, 1973; Studdert-Kennedy & Shankweiler, 1970).

That adults perceive an invariant identity underlying the acoustical variations of a given consonant (e.g., acoustic differences for /d/ in "dee" versus "dah" versus "doo") is difficult for the psychoacoustic view to explain, because it requires finding a unitary acoustic principle for the invariant percept. Although several auditory solutions have been proposed for the acoustic invariance problem, for example, perceptual "templates" for consonant-specific spectral (frequency) acoustic properties (Blumstein, 1980; Searle, Jacobson, & Rayment, 1979; Stevens & Blumstein, 1978), these do not hold well up under empirical test (Blumstein, Isaacs, & Mertus, 1982; Walley *et al.*, 1981) or logical scrutiny (e.g., Liberman, 1982; Studdert-Kennedy, 1981a,d). However, the acoustic variability is actually an advantage from the speech source perception view that speech acoustics convey information about the structural and transformational invariants of their vocal-tract source. That is, for the variability that derives from the coarticulation of adjacent phonemes, the form of that vocal-tract transformation should clarify rather than confuse the source properties that identify both elements. As for the variety of acoustic features that can specify a given consonant, these also result from, and thus may offer equivalent information about, the vocal-tract invariants identifying that consonant.

4.2.1. Consonant Perception

Research on phoneme context effects has found shifts in category boundary positions that are predictable from the coarticulatory effects of different neighboring phonemes. For example, a continuum between "s" and "sh" can be generated by varying only the center frequency of the fricative noise. But the frequency of natural fricatives is lower if the following vowel is "oo" rather than "ah." That is because the lip-rounding for "oo," which lengthens the vocal tract and therefore lowers its resonant frequencies, is coarticulated with the fricative (e.g., Bell-Berti & Harris, 1979). In support of the notion that perceivers detect coarticulatory information, the "soo-shoo" boundary occurs at a lower frequency than the "sah-shah" boundary (Mann & Repp, 1980). Similar coarticulatory context effects have been found for stop consonant place of articulation differences (Mann, 1980) and for stop consonant voicing boundaries (e.g., Summerfield, 1982).

Research on the perceptual unity of multiple acoustic properties for a consonant distinction offers converging support for the speech source position. The various acoustic properties are not perceived according to their acoustic differences but are perceived instead as equivalent information about the articulation of the same consonant (Best, Morrongiello, & Robson, 1981; Fitch, Halwes, Erickson, & Liberman, 1980). These perceptual equivalences and context effects are difficult to explain by speech-neutral psychoacoustic mechanisms (see Studdert-Kennedy, 1981d), and control studies have in fact failed to find analogous effects in nonspeech perception (Best *et al.*, 1981; Mann, Madden, Russell, & Liberman, 1981; Summerfield, 1982).

4.2.2. Vowel Perception

If vowel perception is accomplished by detecting the underlying vocal-tract configuration and transformations that remain invariant across the structural variations of different vocal tracts, then vocal-tract normalization would not be a problem for speech source perception. In contrast, the psychoacoustic account posits that a singular underlying neutral acoustic description of the vowel must be derivable by some formula. No such description has yet been found because proportional differences among vocal tracts, especially male versus female versus child, prevent a uniform scaling of vowel formant frequencies among speakers (Broad, 1981); that is, formant frequency ratios are speaker-specific. Nor can the problem be solved through some formula that partials out sex and age differences based on the value of some independent acoustic feature such as fundamental frequency of the voice, which is the most obvious formant-independent feature that could differentiate those speaker characteristics. The sex and age groups show considerable overlap in fundamental frequency, a laryngeal property that is imperfectly correlated with the variation in supralaryngeal configurations that affect formant properties. Furthermore, as the latter observation suggests, the perceived sex and age of a speaker depend on supralaryngeal characteristics rather than on fundamental frequency (Lehiste & Meltzer, 1973). Thus, the psychoacoustic approach is left in an untenable position: the acoustic normalization solution would appear to depend on *a priori* knowledge of the supralaryngeal vocal-tract properties whose influence it is trying to circumvent.

Of further relevance to the speech source view, identification performance is better for vowels spoken in CVC context than for isolated vowels, even though formant frequencies are more clearly differentiated among the isolated vowels (Strange, Verbrugge, Shankweiler, & Edman, 1976). Likewise, similarity judgments among vowels in CVCs are clearly differentiated along three dimensions, which correspond closely to vowel articulatory factors, whereas most perceivers differentiate isolated vowels along only one or two dimensions (Rakerd & Verbrugge, 1982). These contextual effects suggest that coarticulatory information aids vowel as well as consonant perception, consistent with the ecological premise that structural invariants (vocal tract configurations) are clarified by transformational (dynamic articulation) properties (see Verbrugge, Shankweiler, & Fowler, 1980). Studies with CVC syllables whose vowel nucleus has been replaced with silence, leaving only the syllable-initial and syllable-final formant transitions (in correct temporal relation), further support the ecological interpretation. Vowel identification under these conditions is remarkably well-preserved (Strange, Jenkins, & Edman, 1977), even if each remaining piece of coarticulatory information is taken from a different-sexed speaker (Verbrugge & Rakerd, 1980). These coarticulatory influences on vowel perception are no problem for the view that we detect vocal-tract source information, but would be difficult to explain via psychoacoustic mechanisms (Fowler & Shankweiler, 1978; Shankweiler, Strange, & Verbrugge, 1977; but see Howell, 1981).

These consonant and vowel findings suit the speech source interpretation, but at the same time they also suit a phonetic interpretation of *adult* speech perception. Certainly, the experimental tasks often required subjects to "recover phonemes from the speech stream" (Liberman, 1982). Since speech conveys language, the detection of "pure" (nonlinguistic) structural and transformational invariants of the vocal tract may rarely if ever be ends in themselves for adults, and instead serve as means to the *linguistic* ends of recognizing, for example, phonemes. This would not be the case for infants, however, since phonemes can be "recovered" from the speech stream only by listeners who already know they are there (Studdert-Kennedy, 1981d). Phonemes may indeed be at the "surface of language" (Liberman, 1982) for adults, but vocal-tract source information must be at the "surface of speech" for prelinguistic infants. We now turn to some recent research with prelinguistic infants, which offers strong support for speech source perception.

4.3. Speech Source Perception by Infants

4.3.1. Infants' Perceptual Constancy for Speech

Perceptual constancy is at the base of the acoustic variability and vocal-tract normalization puzzles. The findings just discussed suggest that the adult's solution lies in a perceptual focus on the structural and transformational invariants of a speaker's vocal tract. Even without the adult's linguistic motivations, infants also show perceptual constancy for vowels and consonants spoken by different people or in different phoneme contexts, as described earlier in the chapter (Holmberg *et al.*, 1977; Kuhl, 1979, 1980, 1981b). Therefore, the invariant features they apprehend must exist at the surface of speech and not only at the surface of language. The speech source view suggests that the properties of relevance to the infant lie in the vocal tract and not in the speech-neutral superficial acoustics.

A psychoacoustic interpretation of infant perceptual constancy works no better than it did for adults. We cannot assume that infants are guided by knowledge about phonemes in their solution of the two acoustic puzzles, so some independent source of guidance to the invariant acoustic features of vowels and consonants would be needed. One psychoacoustic solution to the normalization problem might be that although adults do not solve it by partialing out speaker differences based on fundamental frequency, linguistically naive infants do. However, this approach does not work, because infants as well as adults appear to rely on supralaryngeal information in the formant structure of speech rather than on fundamental frequency when perceiving speaker gender (C. L. Miller, Younger, & Morse, 1982). Nor does an acoustic template model (e.g., Blumstein) appear to give an adequate psychoacoustic explanation to the acoustic invariance problem of infants' perceptual constancy for consonants across varying vowel contexts (Studdert-Kennedy, 1981d; Walley *et al.*, 1981).

What *does* remain constant in the utterances of a vowel or consonant by different speakers or across different phoneme contexts is the underlying similarity in vocal-tract structure and articulator positioning. Speech source information would thus seem to offer a more straightforward metric than speech-neutral acoustic invariance for infant perceptual constancy, as was argued in the case of adult speech perception. Moreover, perception of speech source information would certainly be a more direct guide than speech-neutral acoustic patterns for the infant's attempts at vocal imitation of older speakers and eventual production of words provided by her native language environment.

Thus far, the central argument that the infant perceives speech source information has been oriented around the acoustic medium. However, as indicated in the next section on infants' recognition of auditory and visual commonalities in speech, this information is provided by sight as well as by sound. The intermodal perception of speech by infants provides strong support for the speech source perception view, and is particularly difficult to reconcile with the psychoacoustic and phonetic perspectives.

4.3.2. Infants' Intermodal Perception of Speech

When adults listen and watch someone speak, the acoustic and optic information about speech is not perceptually independent. Rather, the two seemingly disparate types of information are perceptually unified, implying that a common metric underlies them. Prelinguistic infants likewise recognize underlying commonalities between audio and visual presentations of speech that cannot be described in linguistic terms for them. The speech source perception view suggests that infants perceive speech intermodally by attending to the underlying articulatory events that provide the auditory and visual information.

To appreciate the contribution of visual information to speech perception, recall listening to someone speaking at the front of a room. It probably seemed easier to understand what was being said if you could also keep the speaker's face in view. This intuition has recently been empirically validated with adult listeners. Under difficult listening conditions, adults perceive speech more correctly when they can watch the speaker than when they must rely on their ears alone (Binnie, Montgomery, & Jackson, 1974; Dodd, 1977; Summerfield, 1979). These findings suggest that listeners obtain information about speech not only from the acoustic signal, but also from the optical information that results from articulatory maneuvers.

Of course, the major responsibility for speech perception is carried by the auditory modality. That blind adults successfully perceive speech, whereas the deaf have serious difficulty with lipreading, would seem to imply that auditory information is both necessary and sufficient for speech perception, although visual information plays a negligible role unless listening is particularly difficult. Speech researchers accepted this logic until recently, when MacDonald and McGurk (1978) and Summerfield (1979) reported that listeners fail to recognize phonemic conflicts in concurrent auditory and visual presentations of speech, instead perceiving a unified speech event. The percepts did not veridically reflect either the acoustic or the optic signal considered in isolation. For example, when perceivers watched a face silently articulating "ga" while a voice said "ba," they heard *"da."* These results indicate that in face-to-face speech perception, listening is not simply supplemented by arbitrary, learned associations between vocal-tract configurations and speech sounds. Rather, at the level of the speech event itself, the information provided by the two modalities shows an intermodal articulatory equivalence.

Two opposing views have been offered for adults' perception of acoustic-optic equivalence in speech. MacDonald and McGurk (1978) have suggested that the equivalence be described linguistically, in terms of abstract features of phonemes. The other view, proposed by Summerfield (1979) and based on the Fowler *et al.* (1980) ecological interpretation of speech production findings, argues that the equivalence is nonlinguistic and modality-free, arising from the dynamics of articulation. The first view has been criticized, in part, because it accounts for only a limited number of the speech percepts that result from audiovisual conflict (see Summerfield, 1979, for detailed discussion of these and other criticisms).

In light of recent findings, the latter view appears to best account for how prelinguistic infants perceive speech intermodally. Infants' sensitivity to acoustic-optic equivalences in speech have been demonstrated under two conditions. Under the first condition, infants were presented with acoustic-optic speech displays in which the overall synchrony and the specific articulatory details presented in the two modalities of information were confounded. In one study, 2½- to 4-month-old infants saw the mirrored reflection of a woman's face repeating nursery rhymes; the auditory signal was either in synchrony or delayed relative to the optic presentation by 400 milliseconds. The infants watched the reflection significantly longer when the visual and the auditory presentations were in natural synchrony (Dodd, 1979). In a second study, 4-month-old infants viewed two women speaking, in two adjacent video films, while the concurrent speech of one woman was played over a central loudspeaker. Infants preferred to look at the face that talked in synchrony to the audio speech presentation (Spelke & Cortelyou, 1981).

It is unclear whether the infants were responding to the general synchrony and/or the specific articulatory details of the optic and acoustic displays, however, since these two aspects of audiovisual match were confounded in both experiments. They might have only recognized the overall synchrony, for example, for syllable onsets, between the speech seen and heard. However, they might also have preferred watching the natural acoustic-optic concurrence of specific articulatory gestures. For example, infants might prefer to look at a speaker's lips being rounded and protruded for the production of the vowel "oo" as they heard an audio "oo," as opposed to looking at the speaker's lips being opened wider to produce "ah."

This prediction was recently tested experimentally (Kuhl & Meltzoff, 1982; MacKain, Studdert-Kennedy, Spieker, & Stern, 1981). Kuhl & Meltzoff (1982) presented 4 to 5-month-olds with two adjacent films of a woman's face synchronously articulating the vowels "ah" and "ee," while one of those vowels was presented auditorily and in synchrony over a central loudspeaker.

The infants preferred to watch the film whose articulatory details specified the vowel presented auditorily. In the MacKain *et al.* study, disyllables (e.g., "mama," "lulu") were presented audiovisually, under similar experimental conditions, to 5- to 6-month-old infants. The infants looked significantly longer at the video display whose articulatory dynamics matched the acoustic presentations, for the disyllables "mama," "baby," and "zuzu." These findings indicate that young infants recognize at least some auditory-visual equivalences of articulatory gestures, and are not only sensitive to general synchrony. Moreover, this intermodal recognition was accomplished in the presumed absence of a language system, making linguistically based explanations (e.g., MacDonald & McGurk) untenable for this age group. The commonality that the infants recognized between the acoustic and the optic information may best be described in nonlinguistic terms (Summerfield, 1978).

Given that the infant seems attuned to detect vocal-tract source information in speech, what may be the organization of the supporting perceptual system? A consideration of the biological basis of this attunement should move us closer to understanding the ease with which humans recognize auditory-visual equivalences in articulatory details, and the infant's apparent ease in learning to speak a first language. Research with adults suggests that the answer lies in the functional asymmetries of the left- and right-cerebral hemispheres of the human brain.

4.4. Left-Hemisphere Attunement for Articulatory Information

For the adult, the left-cerebral hemisphere shows a specialized advantage for the perception of speech, in contrast to a right-hemisphere advantage for the perception of music and certain other nonspeech sounds (e.g., Kimura, 1973). Even in young infants, the hemispheres are differentially responsive to human speech versus other sounds. Auditory evoked response asymmetries in young infants favor the left hemisphere when words or syllables are presented auditorily, whereas the right-hemisphere response is stronger when musical or other nonspeech sounds are presented (Molfese, Freeman & Palermo, 1975). In dichotic listening tests, consistent with the adult findings previously cited, infants as young as 2½ to 3 months show a right-ear advantage (REA) in discriminating among consonants, indicating a left-hemisphere superiority. Conversely, they show a left-ear advantage (LEA) in discriminating notes played by different musical instruments, indicating a right-hemisphere superiority (Best, Hoffman, & Glanville, 1982; Entus, 1977; Glanville, Best, & Levenson, 1977).

These functional asymmetries in infants indicate an early left-hemisphere attunement to information in speech, which could be an important biological support for the infant's perceptual discovery of the articulatory patterns of spoken words. But the data do not indicate exactly the sort of information in speech to which the left hemisphere is attuned (see Molfese, Nuñez, Seibert, & Ramaniah, 1976). Two recent findings suggest that the infant's left hemisphere is apparently attuned to information about the articulatory gestures of the vocal tract.

As an additional result of their intermodal speech perception study, MacKain, Studdert-Kennedy, Spieker, & Stern (1983) found a rightward attentional bias (implying left-hemisphere activation: Kinsbourne, 1973, 1982), which facilitated the infants' recognition of acoustic-optic commonality in the articulatory details of speech. In that experiment, infants attended primarily to either the right or the left video monitor during the synchronous audio presentation. An analysis of visual preferences indicated that the infants recognized auditory-visual matches versus mismatches in articulatory properties only when they were attending to the *right* video monitor. Since intermodal perception of speech appears to entail the recognition of its vocal-tract source properties, these results indicate the infant's recognition of that information is facilitated by a left-hemisphere attentional bias toward those properties.

The results of another study (Best, 1978) may further clarify which aspects of human speech are the object of the infant's left-hemisphere attunement. Adults show a consistent left-hemisphere advantage for consonant perception, whereas isolated vowels yield a nonsignificant perceptual asymmetry (e.g., Studdert-Kennedy & Shankweiler, 1970). Even vowels in CVC syllables show a weak and equivocal left-hemisphere advantage (e.g., Studdert-Kennedy & Shankweiler, 1970; Weiss & House, 1973). This vowel-consonant difference in hemispheric perceptual asymmetry may depend on the earlier discussed differences in the acoustic and articulatory properties of vowels and consonants.

The aim of the infant study (Best, 1978) was to determine whether 3½-month-olds show a similar hemispheric difference in perception of consonants versus vowels. The infants showed a clear left-hemisphere advantage for discriminating a set of synthetic consonants, as adults did. However, the infants also showed a clear *right*-hemisphere advantage for discriminating steady-state synthetic vowels, which differs from adult reports.

The psychoacoustic interpretation of the adult hemisphere differences is that the hemispheres are differentially specialized for processing the acoustic features that differ between consonants and vowels (e.g., Cutting, 1974; Schwartz & Tallal, 1980). However, this interpretation confounds acoustic and articulatory differences between vowels and consonants, and must be rejected on methodological grounds (Studdert-Kennedy & Shankweiler, 1980), as well as for the general criticisms against the psychoacoustic view. A speech source interpretation would instead consider articulatory differences between consonants and vowels. Indeed, consonant and vowel productions engage different coordinations of the articulatory musculature (Fowler, 1980). Moreover, vowel-consonant differences in intermodal speech perception effects (Summerfield, 1979; Summerfield, McGrath, & Forster, 1982) suggest that such articulatory differences may be influential in perception. The pattern of the phoneme class differences in production and in intermodal perception suggest that consonant information is conveyed in rapid articulatory changes, whereas vowel information is conveyed in relatively more slowly changing configurations of the tongue, lips, and jaw.

The implication of these two recent findings on infant hemispheric asymmetries is that the left-hemisphere attunement takes the form of an attentional bias toward information about rapid articulatory transformations. In complement to that attentional bias, the infant's right hemisphere may be better attuned to information about relatively more enduring structural properties of sound-making objects, such as the structural properties of instruments that determine their musical timbre and the configurations of the articulators in the human vocal tract that determine steady-state vowel color.

In conclusion, these specializations of the cerebral hemispheres for responding to different aspects of articulatory information in speech may offer biological support for the infant's discovery and production of words. In the final section, the relation of speech source perception to the discovery of words and the broader motivation provided by the context of communicative development will be briefly discussed.

5. THE BROADER CONTEXT OF COMMUNICATIVE DEVELOPMENT

The speech medium carries a number of parallel messages (Pike, 1959), and not only the sort of vocal-tract source information we have been focusing on at the surface of speech. To learn language, the infant must discover, or learn to recognize, many or all of those other messages as they are specified by convergent information in speech and in the context of its occurence. Some of the messages that must be discovered are linguistic, beginning with words or phrases. Although the linguistic messages are more abstract than speech source messages, their expression in the speech medium depends directly on speech source information. Words and phrases are conveyed in the medium as patterns of articulator configurations and transformations, which have invariant properties across speakers, speaking rates, and surrounding speech context. Therefore, the infant's eventual discovery of them depends on attention to informational invariants in vocal-tract shapes and gestures as they are patterned over time, and conveyed in both the acoustic and optic media.

As for phonemes, their discovery as invariant vocal patterns that convey differences in meaning appears to be *served by* the child's developing use of words rather than vice versa (Menn, 1980; Menyuk & Menn, 1979). Given that the function of phonemes in a language system is determined by word meanings and contrast, it is consistent with this interpretation that maternal speech to toddlers who produce words *does* include hyperdifferentiation in the productions of some phoneme contrasts, whereas maternal speech to prelinguistic infants does *not* include such hyperdifferentiation (Malsheen, 1980).

But the recognition of invariant vocal patterns, of course, is still not support enough for the discovery of words and phonemes. In order to comprehend and use language normally, the infant must also come to recognize the speaker's *communicative* messages. Some of these appear at the surface of speech, where infants can detect them. Emotional affect may be conveyed to the infant directly in the mother's speech, for example, through the level and modulations of her voice pitch and intensity (see Stern, Spieker, & MacKain, 1982). These communicative messages often gain converging support from information in the visual and even in the haptic modalities; for example, emotional affect often receives contextual support in the speaker's changing facial expressions and the way she or he touches or holds the infant. The argument here is that the infant's discovery of more abstract communicative messages, notably the referential meaning of words, *requires* such nonspeech contextual support. Word meanings cannot be revealed for the first time through the speech medium alone.

This observation brings us to the end of our discussion, moving as it does beyond both the prelinguistic period and the infant's perception of the surface of speech. In closing, however, it is suggested that the prelinguistic infant's ability to perceive in speech the structure and transformations of its vocal-tract source provide her or him a crucial tool for discovering the more abstract messages of language.

Acknowledgements

I would like to thank Steven S. Braddon, Kristine S. MacKain, Roy Pea, Carol Fowler, Michael Studdert-Kennedy, and the editors of this book for their thoughtful comments on an earlier version of this chapter. Special thanks are due to Kristine S. MacKain for her helpful discussion about intermodal speech perception by infants, and to Steven Braddon for discussion and encouragement throughout the project.

6. REFERENCES

Ades, A. E. Vowels, consonants, speech and nonspeech. *Psychological Review*, 1977, *84*, 524–530.

Alekin, R. O., Klaas, Y. A., & Chistovich, L. A. Human reaction time in the copying of aurally perceived vowels. *Soviet Physics-Acoustics*, 1962, *8*, 17–22.

Aslin, R. N., & Pisoni, D. B. Some developmental processes in speech perception. In G. Yeni-Komshian, J. F. Kavanaugh, & C. A. Ferguson (Eds.), *Child phonology. Vol. 2: Perception.* New York: Academic Press, 1980.

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

Baru, A. V. Discrimination of synthesized vowels [a] and [i] with varying parameters in dog. In G. Fant & M. A. A. Tatham (Eds.), *Auditory analysis and the perception of speech.* London: Academic Press, 1975.

Bekoff, A. Embryonic development of the neural circuitry underlying motor coordination. In W. M. Cowan (Ed.), *Studies in developmental neurobiology.* New York: Oxford University Press, 1981.

Bell-Berti, F., & Harris, K. S. Anticipatory coarticulation: Some implications from a study of lip-rounding. *Journal of the Acoustical Society of America*, 1979, *65*, 1268–1270.

Bernstein, J. Formant-based representation of auditory similarity among vowel-like sounds. *Journal of the Acoustical Society of America*, 1981, *69*, 1132–1144.

Bernstein, L. E. Developmental differences in labeling VOT continua with varied fundamental frequency. *Journal of the Acoustical Society of America*, 1979, *65*, S112 (A). (Supplement)

Bertoncini, J., & Mehler, J. Syllables as units in infant speech perception. *Infant Behavior and Development*, 1981, *4*, 247–260.

Best, C. T. *The role of consonant and vowel acoustic features in infant cerebral asymmetries for speech perception.* Unpublished doctoral dissertation, Michigan State University, 1978.

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

Best, C. T., Hoffman, H., & Glanville, B. B. Development of infant ear asymmetries for speech and music. *Perception and Psychophysics*, 1982, *31*, 75–85.

Binnie, C. A., Montgomery, A. A., & Jackson, P. L. Auditory and visual contributions to the perception of consonants. *Journal of Speech and Hearing Research*, 1974, *17*, 619–630.

Blumstein, S. E. Speech perception: An overview. In G. H. Yeni-Komshian, J. F. Kavanaugh, & C. A. Ferguson (Eds.), *Child phonology. Vol. 2: Perception.* New York: Academic Press, 1980.

Blumstein, S. E., Cooper, W. E., Zurif, E., & Caramazza, A. The perception and production of voice-onset time in aphasia. *Neuropsychologia*, 1977, *15*, 371–383.

Blumstein, S. E., Isaacs, E., & Mertus, J. The role of the gross spectral shape as a perceptual cue to place of articulation in initial stop consonants. *Journal of the Acoustical Society of America*, 1982, *72*, 43–50.

Broad, D. J. Piecewise planar vowel formant distributions across speakers. *Journal of the Acoustical Society of America*, 1981, *69*, 1423–1429.

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

Clifton, R. K., Morrongiello, B. A., Kulig, J. W., & Dowd, J. M. Developmental changes in auditory localization in infancy. In R. N. Aslin, J. R. Alberts, & M. R. Petersen (Eds.), *Development of perception: Psychobiological perspectives.* (Vol. 1.) New York: Academic Press, 1981.

Cooper, F. S. Some reflections on speech research. *Haskins Laboratories: Status Report on Speech Research,* 1981, *SR-65,* 201–215.

Crowder, R. G. Representation of speech sounds in precategorical acoustic storage. *Journal of Experimental Psychology,* 1973, *1,* 14–24.

Cutting, J. E. Two left hemisphere mechanisms in speech perception. *Perception and Psychophysics,* 1974, *16,* 601–612.

Cutting, J. E., & Eimas, P. D. Phonetic feature analyzers and the processing of speech in infants. In J. F. Kavanaugh & J. E. Cutting (Eds.), *The role of speech in language.* Cambridge: M.I.T. Press, 1975.

Cutting, J. E., & Rosner, B. S. Categories and boundaries in speech and music. *Perception and Psychophysics,* 1974, *16,* 564–570.

Darwin, C. J. Ear differences in the recall of fricatives and vowels. *Quarterly Journal of Experimental Psychology,* 1971, *23,* 46–62.

Dodd, B. The role of vision in the perception of speech. *Perception,* 1977, *6,* 31–40.

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

Dudley, H. The carrier nature of speech. *Bell System Technical Journal,* 1940, *19,* 495–515. (Reprinted in J. L. Flanagan & L. R. Rabiner (Eds.), *Speech synthesis.* Stroudsberg, Penn.: Dowden, Hutchinson & Ross, 1973.)

Eilers, R. E. Infant speech perception: History and mystery. In G. Yeni-Komshian, J. F. Kavanaugh, & C. A. Ferguson (Eds.), *Child phonology. Vol. 2: Perception.* New York: Academic Press, 1980.

Eilers, R. E., & Gavin, W. J. The evaluation of infant speech perception skills: Statistical techniques and theory development. In R. E. Stark (Ed.), *Language behavior in infancy and early childhood.* New York: Elsevier-North Holland, 1981.

Eilers, R. E., & Minifie, F. D. Fricative discrimination in early infancy. *Journal of Speech and Hearing Research,* 1975, *18,* 158–167.

Eilers, R. E., Wilson, W. R., & Moore, J. M. Discrimination of synthetic prevoiced labial stops by infants and adults. *Journal of the Acoustical Society of America,* 1976, *60,* S91 (A). (Supplement)

Eilers, R. E., Wilson, W. R., & Moore, J. M. Developmental changes in speech discrimination in infants. *Journal of Speech and Hearing Research,* 1977, *20,* 766–780.

Eilers, R. E., Oller, D. K., & Gavin, W. J. *A cross-linguistic study of infant speech perception.* Paper presented at the meeting of the Southeastern Conference on Human Development. Atlanta, Georgia, 1978.

Eilers, R. E., Gavin, W., & Wilson, W. R. Linguistic experience and phonemic perception in infancy: A cross-linguistic study. *Child Development,* 1979, *50,* 14–18.

Eimas, P. D. Linguistic processing of speech by young infants. In R. L. Schiefelbusch & L. L. Lloyd (Eds.), *Language perspectives—Acquisition, retardation, and intervention.* Baltimore: University Park Press, 1974. (a)

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

Eimas, P. D. Speech perception in early infancy. In L. B. Cohen & P. Salapatek (Eds.), *Infant perception: From sensation to cognition. Vol. 2: Perception of space, speech, and sound.* New York: Academic Press, 1975. (a)

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

Eimas, P. D., & Miller, J. L. Contextual effects in infant speech perception. *Science*, 1980, *209*, 1140–1142. (a)

Eimas, P. D., & Miller, J. L. Discrimination of information for manner of articulation. *Infant Behavior and Development*, 1980, *3*, 367–375. (b)

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

Emde, R. N., & Robinson, J. The first two months: Recent research in developmental psychobiology and the changing view of the newborn. In J. Noshpitz & J. Call (Eds.), *Handbook of child psychiatry: Vol. 1: Development*. New York: Basic Books, 1979.

Entus, A. K. Hemispheric asymmetry in processing of dichotically-presented speech and nonspeech stimuli by infants. In S. J. Segalowitz & F. A. Gruber (Eds.), *Language development and neurological theory*. New York: Academic Press, 1977.

Fant, G. *Acoustical theory of speech production*. The Hague: Mouton, 1960.

Fant, G. *Speech sounds and features*, Cambridge: M.I.T. Press, 1973.

Ferguson, C. A., & Farwell, C. B. Words and sounds in early language acquisition. *Language*, 1975, *51*, 419–439.

Fitch, H. L., Halwes, T., Erickson, D., & Liberman, A. M. Perceptual equivalence of two acoustic cues for stop-consonant manner. *Perception and Psychophysics*, 1980, *27*, 343–350.

Flanagan, J. L. *Speech analysis, synthesis and perception*. (2nd ed.). New York: Academic Press, 1973.

Fodor, J. A., Garrett, M. F., & Shapero, D. B. Discriminations among phones by infants. *Quarterly Progress Report: Research Laboratory of Electronics (M.I.T.)*, 1970, *96*, 180.

Fodor, J. A., Garrett, F. M., & Brill, S. L. Pi ka pu: The perception of speech sounds by prelinguistic infants. *Perception and Psychophysics*, 1975, *18*, 74–78.

Fowler, C. A. Coarticulation and theories of extrinsic timing. *Journal of Phonetics*, 1980, *8*, 113–133.

Fowler, C. A., & Shankweiler, D. P. Identification of vowels in speech and nonspeech contexts. *Journal of the Acoustical Society of America*, 1978, *63*, S4. (Supplement)

Fowler, C. A., & Turvey, M. T. Skill acquisition: An event approach with special reference to searching for the optimum of a function of several variables. In G. E. Stelmach (Ed.), *Information processing in motor control and learning*. New York: Academic Press, 1978.

Fowler, C. A., Rubin, P., Remez, R. E., & Turvey, M. T. Implications for speech production of a general theory of action. In B. Butterworth (Ed.), *Speech production*. New York: Academic Press, 1980.

Galunov, V. I., & Chistovich, L. A. Relationship of motor theory to the general problem of speech recognition. *Soviet Physics-Acoustics*, 1966, *11*, 357–365. (Review)

Garnica, O. The development of phonemic speech perception. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language*. New York: Academic Press, 1973.

Gibson, E. J. How perception really develops: A view from outside the network. In D. LaBerge & S. J. Samuels (Eds.), *Basic processes in reading: Perception and comprehension*. Potomac, Md.: Erlbaum, 1977.

Gibson, E. J. *Infants' perception of structural invariants of objects*. Psychology Department, Columbia University, New York, Spring 1980.

Gibson, E. J., & Levin, H. *The psychology of reading*. Cambridge: M.I.T. Press, 1975.

Gibson, J. J. *The senses considered as perceptual systems*. Boston: Houghton Mifflin, 1966.

Gibson, J. J., & Gibson, E. J. Perceptual learning: Differentiation or enrichment? *Psychological Review*, 1955, *62*, 32–41.

Glanville, B. B., Best, C. T., & Levenson, R. A cardiac measure of cerebral asymmetries in infant auditory perception. *Developmental Psychology*, 1977, *13*, 54–59.

Goldstein, U. G. Modeling children's vocal tracts. *Journal of the Acoustical Society of America*, 1979, *65*(Suppl. 1), S25. (Abstract)

Haith, M. M. Visual cognition in early infancy. In R. B. Kearsley & I. E. Sigel (Eds.), *Infants at risk: Assessment of cognitive function.* Hillsdale, N.J.: Erlbaum, 1979.

Hillenbrand, J., Minifie, F. D., & Edwards, T. J. Tempo of spectrum change as a cue in speech-sound discrimination by infants. *Journal of Speech and Hearing Research,* 1979, *22,* 147–165.

Holmberg, T. L., Morgan, K. A., & Kuhl, P. K. *Speech perception in early infancy: Discrimination of fricative consonants.* Paper presented at the meeting of the Acoustical Society of America, December 1977.

Horridge, G. A. *Interneurons.* London: W. H. Freeman, 1968.

Howell, P. Identification of vowels in and out of context. *Journal of the Acoustical Society of America,* 1981, *70,* 1256–1260.

Jacob, F. Evolution and tinkering. *Science,* 1977, *196,* 1161–1166.

Jacobson, M. *Developmental neurobiology.* New York: Plenum Press, 1978.

Jenkins, J. J. Remember that old theory of memory? Well, forget it! *American Psychologist,* 1974, *29,* 785–795.

Jenkins, J. J. Research in child phonology: Comments, criticism, and advice. In G. H. Yeni-Komshian, J. F. Kavanaugh, & C. A. Ferguson (Eds.), *Child phonology: Vol. 2: Perception.* New York: Academic Press, 1980.

Jusczyk, P. W. Perception of syllable-final stop consonants by two-month-old infants. *Perception and Psychophysics,* 1977, *21,* 450–454.

Jusczyk, P. W., & Thompson, E. Perception of a phonemic contrast in multisyllabic utterances by 2-month-old infants. *Perception and Psychophysics,* 1978, *23,* 105–109.

Jusczyk, P. W. Infant speech perception: A critical appraisal. In P. D. Eimas & J. L. Miller (Eds.), *Perspectives on the study of speech.* Hillsdale, N.J.: Erlbaum, 1981. (a)

Jusczyk, P. W. The processing of speech and nonspeech sounds by infants: Some implications. In R. N. Aslin, J. R. Alberts, & M. R. Petersen (eds.), *Development of perception: Psychobiological perspectives.* (Vol. 1.) New York: Academic Press, 1981 (b).

Jusczyk, P. W., Rosner, B. S., Cutting, J. E., Foard, C. F., & Smith, L. B. Categorical perception of nonspeech sounds by 2-month-old infants. *Perception and Psychophysics,* 1977, *21,* 50–54.

Jusczyk, P. W. Murray, J., & Bayly, J. *The perception of place of articulation of fricatives and stops by infants.* Paper presented at the meeting of the Society for Research on Child Development, San Francisco, March 1979.

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

Kaye, K. Why we don't talk "baby talk" to babies. *Journal of Child Language,* 1980, *7,* 489–507.

Kent, R. D., & Forner, L. L. Developmental study of vowel formant frequencies in an imitation task. *Journal of the Acoustical Society of America,* 1979, *65,* 208–217.

Kimura, D. The asymmetry of the human brain. *Scientific American,* 1973, *228,* 70–78.

Kinsbourne, M. The control of attention by interaction between the hemispheres. In S. Kornblum (Ed.), *Attention and performance IV.* New York: Academic Press, 1973.

Kinsbourne, M. Hemispheric specialization and the growth of human understanding. *American Psychologist,* 1982, *37,* 411–420.

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

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

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

Kuhl, P. K. Auditory category formation and developmental speech perception. In R. E. Stark (Ed.), *Language behavior in infancy and early childhood.* New York: Elsevier-North Holland, 1981. (a)

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

Kuhl, P. K., & Meltzoff, A. N. Bimodal speech perception in early infancy. *Journal of the Acoustical Society of America,* 1982, *71,* (Suppl.) S77–78. (Abstract)

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

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

Lehiste, I., & Meltzer, D. Vowel and speaker identification in natural and synthetic speech. *Language and Speech,* 1973, *16,* 356–364.

Lenneberg, E. H. *Biological foundations of language.* New York: Wiley, 1967.

Liberman, A. M. On finding that speech is special. *American Psychologist,* 1982, *37,* 148–167.

Liberman, A. M., Cooper, F. S., Shankweiler, D., & Studdert-Kennedy, M. The speech code. *Psychological Review,* 1967, *74,* 431–461.

Lieberman, P. *Intonation, perception and language.* Cambridge: M.I.T. Press, 1967.

Lieberman, P. On the development of vowel production in young children. In G. H. Yeni-Komshian, J. F. Kavanaugh, & C. A. Ferguson (Eds.), *Child phonology. Vol. 1: Production.* New York: Academic Press, 1980.

Lieberman, P., Harris, K. S., Wolff, P., & Russell, L. H. Human infant cry and nonhuman primate vocalization. *Journal of Speech and Hearing Research,* 1971, *14,* 718–727.

Lisker, L. Rapid vs. rabid: A catalogue of acoustic features that may cue the distinction. *Haskins Laboratories: Status Report on Speech Research,* 1978, *SR-54,* 127–132.

MacDonald, J., & McGurk, H. Visual influences on speech perception processes. *Perception and Psychophysics,* 1978, *24,* 253–257.

MacKain, K. S., Studdert-Kennedy, M., Spieker, S., & Stern, D. *Cross-modal coordination in infants' perception of speech.* Paper presented at the meeting of the 2nd International Conference on Child Psychology, Vancouver, August 1981

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

Malsheen, B. J. Two hypotheses for phonetic clarification in the speech of mothers to children. In G. H. Yeni-Komshian, J. F. Kavanaugh, & C. A. Ferguson (Eds.), *Child phonology. Vol. 2: Perception.* New York: Academic Press, 1980.

Mann, V. A. Influence of preceding liquid on stop consonant perception. *Perception and Psychophysics,* 1980, *28,* 407–412.

Mann, V. A., & Repp, B. H. Influence of vocalic context on perception of the esh-/s/ distinction. *Perception and Psychophysics,* 1980, *28,* 213–228.

Mann, V. A., Madden, J., Russell, J. M., & Liberman, A. M. Further investigation into the influence of preceding liquids on stop consonant perception. *Journal of the Acoustical Society of America,* 1981, *69*(Suppl.), S91 (A).

Marshall, W. A. *The development of the brain.* London: Oliver & Boyd, 1968.

McCulloch, W. S. *The embodiments of mind.* Cambridge: M.I.T. Press, 1965.

Mehler, J., & Bertoncini, J. Infants' perception of speech and other sounds. In J. Morton & J. Marshall (Eds.), *Psycholinguistics. Series II.* London: Elek Science, 1978.

Menn, L. Phonological theory and child phonology. In G. H. Yeni-Komshian, J. F. Kavanaugh, & C. A. Ferguson (Eds.), *Child phonology. Vol. 1: Production.* New York: Academic Press, 1980.

Menyuk, P., & Menn, L. Early strategies for the perception and production of words and sounds. In P. Fletcher & M. Garman (Eds.), *Studies in language acquisition*. Cambridge, England: Cambridge University Press, 1979.

Miller, C. L., & Morse, P. A. The "heart" of categorical speech discrimination in young infants. *Journal of Speech and Hearing Research*, 1976, *19*, 578–589.

Miller, C. L., Morse, P. A., & Dorman, M. F. Cardiac indices of infant speech perception: Orienting and burst discrimination. *Quarterly Journal of Experimental Psychology*, 1977, *29*, 533–545.

Miller, C. L., Younger, B. A., & Morse, P. A. The categorization of male and female voices in infancy. *Infant Behavior and Development*, 1982, *5*, 143–159.

Miller, J. D. Perception of speech sounds in animals: Evidence for speech processing by mammalian auditory mechanisms. In T. H. Bullock (Ed.), *Recognition of complex acoustic signals*. Berlin: Abakon Verlagsgesellschaft, 1977.

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

Miller, J. D., Henderson, B. C., Sullivan, H. T., & Rigden, G. K. Speech perception by the chinchilla: Learning functions for pairs of synthetic stimuli from various points along a VOT continuum. *Journal of the Acoustical Society of America*, 1978, *64* (Suppl.), S18 (A).

Miller, J. L., & Eimas, P. D. *Contextual perception of voicing by infants*. Paper presented at the meeting of the Society for Research in Child Development, Boston, April 1981.

Moffitt, A. R. Consonant cue perception by twenty- to twenty-four week old infants. *Child Development*, 1971, *42*, 717–731.

Molfese, D. L., & Molfese, V. J. VOT distinctions in infants: Learned or innate? In H. A. Whitaker & H. Whitaker (Eds.), *Studies in neurolinguistics*. (Vol. 4.) New York: Academic Press, 1979.

Molfese, D. L., Freeman, R. B., & Palermo, D. S. The ontogeny of brain lateralization for speech and nonspeech stimuli. *Brain and Language*, 1975, *2*, 356–368.

Molfese, D. L., Nuñez, V., Seibert, S. M., & Ramaniah, N. V. Cerebral asymmetry: Changes in factors affecting its development. *Annals of the New York Academy of Science*, 1976, *280*, 821–833.

Morais, J., Cary, L., Alegria, J., & Bertelson, P. Does awareness of speech as a sequence of phones arise spontaneously? *Cognition*, 1979, *7*, 323–331.

Morse, P. A. The discrimination of speech and nonspeech stimuli in early infancy. *Journal of Experimental Child Psychology*, 1972, *14*, 477–492.

Morse, P. A. Infant speech perception: Origins, processes, and *Alpha Centauri*. In F. Minifie, & L. L. Lloyd (Eds.), *Communicative and cognitive abilities—Early behavioral assessment*. Baltimore, Md.: University Park Press, 1978.

Morse, P. A., Eilers, R. E., & Gavin, W. J. The perception of the sounds of silence in early infancy. *Child Development*, 1982, *53*, 189–195.

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

Neuweiler, G., Bruns, V., & Schuller, G. Ears adapted for the detection of motion, or how echolocating bats have exploited the capacities of the mammalian auditory system. *Journal of the Acoustical Society of America*, 1980, *68*, 741–753.

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

Paget, R. *Human speech*. London: Routledge & Keegan-Paul, 1930.

Petersen, M. R. Perception of acoustic communication signals by animals: Developmental perspectives and implications. In R. N. Aslin, J. R. Alberts, & M. R. Petersen (Eds.), *Development of perception: Psychobiological perspectives.* (Vol. 1.) New York: Academic Press, 1981.

Pike, K. L. Language as particle, wave, and field. *The Texas Quarterly*, 1959, *2*, 37–55.

Pisoni, D. B. Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception and Psychophysics*, 1973, *13*, 253–260.

Pisoni, D. B. Auditory short term memory and vowel perception. *Memory and Cognition*, 1975, *3*, 7–18.

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

Pisoni. D. B. Speech perception. In W. K. Estes (Ed.), *Handbook of cognitive processes. Vol. 6: Linguistic functions in cognitive theory.* Hillsdale, N.J.: Erlbaum, 1978.

Rabinowicz, T. The differentiate maturation of the human cerebral cortex. In F. Falkner & J. M. Tanner (Eds.), *Human growth: Vol. 3: Neurobiology and nutrition.* New York: Plenum Press, 1979.

Rakerd, B., & Verbrugge, R. R. Individual differences in the perception of isolated vowels and vowels in a consonantal context. *Journal of the Acoustical Society of America*, 1982, *71*, S76 (A).

Repp, B. H. Categorical perception. *Haskins Laboratories: Status Report on Speech Research*, 1982, *SR-70*, 99–183.

Riedel, K. *Durational factors in the speech perception of aphasics.* Unpublished doctoral dissertation, City University of New York Graduate Center, 1981.

Robson, R. C., Morrongiello, B. A., Best, C. T., & Clifton, R. K. Trading relations in the perception of speech by five-year-old children. *Haskins Laboratories: Status Report on Speech Research*, 1982, *SR-71/72*, in press.

Rose, S. R. *The conscious brain.* New York: Vintage Books, 1976.

Rosen, S. M., & Howell, P. Plucks and bows are not categorically perceived. *Perception and Psychophysics*, 1981, *30*, 156–168.

Ruff, H. A. The development of perception and recognition of objects. *Child Development*, 1980, *51* 981–992.

Ruff, H. A. Effect of object movement on infants' detection of object structure. *Developmental Psychology*, 1982, *18*, 462–472.

Schubert, E. D. The role of auditory perception in language processing. In D. D. Duane & M. B. Rawson (Eds.), *Reading, perception and language.* Baltimore: York Press, 1974.

Schvachkin, N. K. The development of phonemic speech perception in early childhood. In C. A. Ferguson & D. I. Slobin (Eds.), *Studies of child language development.* New York: Holt, Rhinehart & Winston, 1973.

Schwartz, J., & Tallal, P. Rate of acoustic change may underlie hemispheric specialization for speech perception. *Science*, 1980, *207*, 1380–1381.

Searle, C. L., Jacobson, J. Z., & Rayment, S. G. Phoneme recognition based on human audition. *Journal of the Acoustical Society of America*, 1979, *65*, 799–809.

Shankweiler, D. P., Strange, W., & Verbrugge, R. R. Speech and the problem of perceptual constancy. In R. E. Shaw & J. Bransford (Eds.), *Perceiving, acting, and knowing: Toward an ecological psychology.* Hillsdale, N.J.: Erlbaum, 1977.

Shaw, R. E., & Pittenger, J. Perceiving the face of change in changing faces: Implications for a theory of object perception. In R. E. Shaw & J. Bransford (Eds.), *Perceiving, acting, and knowing: Toward an ecological psychology.* Hillsdale, N.J.: Erlbaum, 1977.

Simon, C., & Fourcin, A. J. Cross-language study of speech-pattern learning. *Journal of the Acoustical Society of America*, 1978, *63*, 925–935.

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

Spelke, E. S. Exploring audible and visible events in infancy. In A. D. Pick (Ed.), *Perception and its development: A tribute to Eleanor J. Gibson*. Hillsdale, N.J.: Erlbaum, 1979.

Spelke, E. S., & Cortelyou, A. Perceptual aspects of social knowing. In M. Lamb & E. Sherrod (Eds.), *Infant social cognition*. Hillsdale, N.J.: Erlbaum, 1981.

Stark, R. E. Stages of speech development. In G. H. Yeni-Komshian, J. F. Kavanaugh, & C. A. Ferguson (Eds.), *Child phonology. Vol. 1: Production*. New York: Academic Press, 1980.

Stern, D. N., Spieker, S., & MacKain, K. S. Intonation contours as signals in maternal speech to infants. *Developmental Psychology*, 1982, *18*, 717–735.

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

Stevens, K. N., & Blumstein, S. E. Invariant cues for place of articulation in stop consonants. *Journal of the Acoustical Society of America*, 1978, *64*, 1358–1368.

Strange, W., Verbrugge, R. E., Shankweiler, D. P., & Edman, T. R. Consonant environment specifies vowel identity. *Journal of the Acoustical Society of America*, 1976, *60*, 213–224.

Strange, W., Jenkins, J. J., & Edman, T. R. *Journal of the Acoustical Society of America*, 1977, *61*, S39 (A).

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

Studdert-Kennedy, M. The emergence of phonetic structure. *Cognition*, 1981, *10*, 301–306. (a)

Studdert-Kennedy, M. A note on the biology of speech perception. *Haskins Laboratories: Status Report on Speech Research*, 1981, *SSR-65*, 223–232. (b)

Studdert-Kennedy, M. Cerebral hemispheres: Specialized for the analysis of what? *Behavioral and Brain Sciences*, 1981, *4*, 76–77. (c)

Studdert-Kennedy, M., Perceiving phonetic segments. In T. Myers, J. Laver, & J. Anderson (Eds.), *The cognitive representation of speech*. Amsterdam: North-Holland, 1981. (d)

Studdert-Kennedy, M., & Lane, H. The structuring of language: Clues from the differences between signed and spoken language. *Haskins Laboratories: Status Report on Speech Research*, 1980, *SR-63/64*, 1–8.

Studdert-Kennedy, M. & Shankweiler, D. P. Hemispheric specialization for speech perception. *Journal of the Acoustical Society of America*, 1970, *48*, 579–594.

Studdert-Kennedy, M., & Shankweiler, D. P. Hemispheric specialization for language processes. *Science*, 1980, *211*, 960–961.

Studdert-Kennedy, M., Liberman, A. M., Harris, K. S., & Cooper, F. S. Motor theory of speech perception: A reply to Lane's critical review. *Psychological Review*, 1970, *77*, 234–249.

Summerfield, Q. Perceptual learning and phonetic perception. In *Interrelations of the communicative senses: Proceedings of the National Science Foundation conference at Asilomar, California, September–October 1978*. Washington, D. C.: National Science Foundation Publications, 1978.

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

Summerfield, Q. Differences between spectral dependencies in auditory and phonetic temporal processing: Relevance to the perception of voicing in initial stops. *Journal of the Acoustical Society of America*, 1982, *72*, 51–61.

Summerfield, Q., McGrath, M., & Forster, J. *Detection and resolution of audio-visual conflict in the perception of vowels*. Paper presented at the meeting of the Acoustical Society of America, Chicago, April 1982.

Swoboda, P. J., Kaas, J., Morse, P. A., & Leavitt, L. A. Memory factors in infant vowel discrimination in normal and at-risk infants. *Child Development*, 1978, *49*, 332–339.

Swoboda, P. J., Morse, P. A., & Leavitt, L. A. Continuous vowel discrimination in normal and at-risk infants. *Child Development,* 1976, *47,* 459–465.

Trehub, S. E. Infants' sensitivity to vowel and tonal contrasts. *Developmental Psychology,* 1973, *9,* 91–96.

Trehub, S. E. & Rabinovitch, M. S. Auditory-linguistic sensitivity in early childhood. *Developmental Psychology,* 1972, *6,* 74–77.

Trehub, S. E., Bull, D., & Schneider, B. A. Infant speech and nonspeech perception: A review and reevaluation. In R. L. Schiefelbusch & D. D. Bricker (Eds.), *Early language acquisition and intervention: Language intervention series.* Baltimore, Md.: University Park Press, 1981.

Trevarthen, C. Neuroembryology and the development of perception. In F. Falkner & J. M. Tanner (Eds.), *Human growth: Neurobiology and nutrition.* New York: Plenum Press, 1979.

Tuchmann-Duplessis, H., Aroux, M., & Haegel, P. *Illustrated human embryology. Vol. 3: Nervous system and endocrine glands.* New York: Springer-Verlag, 1975.

Verbrugge, R. R., & Rakerd, B. Talker-independent information for vowel identity. *Haskins Laboratories: Status Report on Speech Research,* 1980, *SR-62,* 205–215.

Verbrugge, R. R., Shankweiler, D. P., & Fowler, C. A. Context-conditioned specification of vowel identity. *Haskins Laboratories: Status Report on Speech Research,* 1980, *SR-62,* 217–224.

Verbrugge, R. R., Rakerd, B., Fitch, H., Tuller, B., & Fowler, C. A. The perception of speech events: An ecological perspective. In R. E. Shaw & W. Mace (Eds.), *Event perception.* Hillsdale, N.J.: Erlbaum, in press.

Walley, A. C., Pisoni, D. B., & Aslin, R. N. The role of early experience in the development of speech perception. In R. N. Aslin, J. R. Alberts, & M. R. Petersen (Eds.), *Development of perception: Psychobiological perspectives. Vol. 1.* New York: Academic Press, 1981.

Warren, W. H., & Verbrugge, R. R. Toward an ecological acoustics. In R. E. Shaw & W. Mace (Eds.), *Event perception.* Hillsdale, N.J.: Erlbaum, in press.

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

Weiss, M. S., & House, A. S. Perception of dichotically presented vowels. *Journal of the Acoustical Society of America,* 1973, *53,* 51–53.

Werker, J. F. *Language-specific developmental changes in the speech perception of prelinguistic infants.* Paper presented at the meeting of the Society for Research in Child Development, Detroit, April 1983.

Werker, J. F., & Tees, R. C. *Developmental changes in the perception of non-native speech sounds.* Papers presented at the meeting of the Canadian Psychological Association, Toronto, June 1981.

Werker, J. F., & Tees, R. C. *Evidence for perceptual reorganization during the first year of life.* Paper presented at the meeting of the Canadian Psychological Association, Montreal, June 1982.

Williams, L., & Bush, M. Discrimination by young infants of voiced stop consonants with and without release bursts. *Journal of the Acoustical Society of America,* 1978, *63,* 1223–1226.

Wood, C. C. Discriminability, response bias, and phoneme categories in discrimination of voice onset time. *Journal of the Acoustical Society of America,* 1976, *60,* 1381–1389.

Zlatin, M. & Koenigsknecht, R. Development of the voicing contrast: A comparison of voice onset time in stop perception and production. *Journal of Speech and Hearing Research,* 1976, *19,* 78–92.

Zoloth, S. R., Petersen, M. R., Beecher, M. D., Green, S., Marler, P., Moody, D. B., & Stebbins, W. Species-specific perceptual processing of vocal sounds by monkeys. *Science,* 1979, *204,* 870–873.