Published in final edited form as:

Brain Behav Evol. 1979 ; 16(5-6): 330–350.

On the Perception of Speech Sounds as Biologically Significant Signals^{1,2}

Dr. David B. Pisoni

Speech Perception Laboratory, Department of Psychology Indiana University of Bloomington,
Bloomington, IN 47405 (USA)

Abstract

This paper reviews some of the major evidence and arguments currently available to support the view that human speech perception may require the use of specialized neural mechanisms for perceptual analysis. Experiments using synthetically produced speech signals with adults are briefly summarized and extensions of these results to infants and other organisms are reviewed with an emphasis towards detailing those aspects of speech perception that may require some need for specialized species-specific processors. Finally, some comments on the role of early experience in perceptual development are provided as an attempt to identify promising areas of new research in speech perception.

Keywords

Speech perception; Synthetic speech; Identification; Discrimination; Species-specific acoustic signals; Perceptual development; Feature detectors; Perceptual constancy; Speech mode

Introduction

The purpose of this paper is to briefly review some of the evidence currently available in support of the view that the perception of speech sounds may require specialized biological mechanisms. Evidence has accumulated rapidly over the last 30 years to suggest that speech is a special type of acoustic signal that has species-specific properties unique to humans. While the bulk of the evidence for this view has come from work on the acoustic analysis of speech and perceptual experiments with adults using synthetic speech sounds, more recent findings have been obtained with prelinguistic infants and animals. This work has expanded our knowledge of speech perception in a number of directions and has shown the value of comparative experimentation for the evaluation of species-specific behaviors, particularly those associated with acoustic signaling systems.

Before turning to the main arguments of this paper and the evidence to be discussed, it would be useful first to briefly review the role of speech in language. One of the essential 'design features' of all spoken languages is what *Hockett* [1958] refers to as 'duality of patterning'. That is, at the message level all languages have two levels of representation, one consisting of the arrangement of meaningless elements (phonemes) and the other consisting of the arrangement of meaningful elements (morphemes or words). Differences between morphemes and hence differences in meanings are realized by variations in the sequencing

¹This paper is a substantially revised and expanded version of a talk given as part of a symposium entitled 'Perception of Species-specific Vocalizations: From Primates to Human Infants' at the Midwest Regional Meeting of the Animal Behavior Society in Chicago on November 13, 1976.

²Preparation of this paper was supported in part by NIMH Grant MH-24027-03 and NIH Grant NS-12179-02.

and arrangement of the constituent phonemes and their features. The phonology of a language represents the sound patterning or structural arrangement of the phonemes whereas the syntax represents the rules for arranging and ordering morphemes and words.

Although the formal analysis of natural languages has been carried out for quite some time and a fairly good understanding of the message units exists, the same cannot be said for the analyses of animal communication systems. Indeed, knowledge of the message level is still a somewhat problematical question. The distinction between message units and sound units or segments should be emphasized here because much of the work to be discussed below tacitly assumes the existence of the phoneme as the smallest unit of linguistic analysis. And while the phoneme may not have physical reality as represented in current analyses of the acoustic waveform, it does have psychological reality in terms of the functional sound systems employed by natural languages. In studying animal communication systems that use vocal signals, an investigator is unfortunately faced with the problem of not having access to knowledge of the units and code appropriate to the message level and therefore must focus his attention entirely on the physically realized sound segments of the signal and their possible correlation to some observable response by the organism. This preoccupation with the physical realization of the message creates the obvious problem, in many cases, of failing to recognize the functional role that a signal plays in the communication system of an organism. With these preliminary remarks completed we can now turn our attention to relevant issues in speech perception.

Human speech production imposes certain well-defined constraints on the resulting acoustic waveform. These constraints are derived from a consideration of the anatomy and physiology of the production mechanism and the associated resonant properties of the vocal tract [Stevens and House, 1955; Fant, 1960; Flanagan, 1972]. Although a good deal of work still remains to be done in understanding speech perception, there are a number of empirical findings which define the basic issues in the field of speech perception and set it apart from other related areas such as auditory psychophysics and general auditory perception. The nature of these problems form the basis, in part, for suspecting that speech perception may require the use of specialized mechanisms for perceptual analysis.

The research in speech perception may be described in terms of two general lines of investigation: (1) studies aimed at establishing the acoustic cues to the perception of speech sound segments and, (2) studies aimed primarily at demonstrating the effects of manipulating syntactic and semantic variables on speech perception. This contribution will be concerned primarily with research that falls into the first category. Three sets of findings will be discussed: (1) early research on the acoustic cues for stop consonants, (2) experiments on the identification and discrimination of speech and speech-like sounds, and (3) research on developmental aspects of speech perception in young infants. The material to be reviewed in these sections is quite selective in order to highlight some of the major problems and theoretical issues that have been studied in speech perception over the last few years and to see how these have influenced the kinds of theoretical approaches that are currently prominent in the field.

The study of speech perception differs in several important ways from the study of auditory perception and psychophysics. First, the signals typically used to study the functioning of the auditory system are simple, discrete and usually well defined mathematically. Moreover, they typically vary along only a single dimension. In contrast, speech sounds involve complex spectral relations that vary as a function of time; changes that occur in a single parameter often affect the perception of other attributes of the stimulus. Secondly, most of the research in auditory psychophysics over the last two decades has been concerned with the discriminative capacities or resolving power of the transducer and the peripheral

auditory mechanism. In the perception of speech, the relevant mechanisms are, for the most part, centrally located. Moreover, experiments in auditory psychophysics have commonly focused on experimental tasks involving discrimination rather than absolute identification. This is rarely the situation of listeners when they perceive and understand speech. In fact, the listener must almost always attempt to identify, on an absolute basis, a particular stretch of speech. As a consequence, it is generally believed that a good deal of what we have learned from traditional auditory psychophysics is unfortunately only marginally relevant to the study of the complex cognitive processes that are involved in speech perception.

In addition to differences in the signal, there are also marked differences in the way speech sounds are processed by listeners. For the most part, when people are presented with speech signals they respond to them as linguistic entities rather than simply as auditory events in the environment. Speech signals are categorized and labeled almost immediately with reference to the listener's linguistic background. Moreover, as we shall see, a listener's ability to discriminate certain speech sounds is often a function of the extent to which the particular acoustic distinction under study plays a functional role in the listener's linguistic system. That is, speech sounds are interpreted by listeners as biologically significant acoustic signals which have an important functional role in the communicative process.

Perceptual Constancy and Acoustic Cues for Stop Consonants

One of the most firmly established findings in the speech perception literature is that the acoustic correlates of a number of consonantal features are highly dependent on context. This is especially true for the stop consonants (b, d, g; p, t, k) which show the greatest amount of contextual variability. As a result of co-articulation effects, one sound segment often carries information about two or more successive phonemes in an utterance. And, conversely, a single phoneme often exerts an influence on several successive sound segments in the acoustic signal. Although the physical signal often varies continuously, listeners perceive speech as consisting of a sequence of discrete segments arrayed in time.

The earliest experimental studies of speech perception were aimed at uncovering the relation or mapping between attributes of the acoustic signal and the linguistic units derived from perceptual analysis. The outcome of these initial studies indicated that there were very few discrete isolatable and invariant sound segments in the physical signal that correspond uniquely to perceived phonemes. This lack of correspondence between attributes of the acoustic signal and units of linguistic analysis has been, and still currently is, one of the most important and controversial issues in speech perception. Because of the prominence of this single issue, it is appropriate to briefly review some of the early findings.

The initial work on the acoustic cues for phonemes involved two related procedures. First, spectrographic analyses were carried out on minimal pairs to identify the potentially important acoustic attributes that distinguished these utterances. Then synthesis experiments were conducted to verify the significance of these acoustic cues in perception. In the first study to use these combined methods, *Lieberman et al.* [1952] examined the relation between the frequency of a noise burst and the perception of the voiceless consonants p, t and k. An examination of sound spectrograms of real speech showed that the voiceless stops could potentially be distinguished by the frequency of a brief burst of noise, the acoustic counterpart of the articulatory explosion at the release of stop closure. *Lieberman et al.* [1952] systematically varied the frequency of a synthetic noise burst before a number of different two formant vowels and observed its effect on perception. Their results showed that subjects' identification of a particular stop consonant varied not only according to the frequency of the noise burst but also in terms of the relation of the burst to the vowel with which it was paired.

The results of this initial study with synthetic speech were replicated with natural speech stimuli in a study by *Schatz* [1954]. She found that release bursts spliced from /ki/, /ka/ and /ku/ were perceived as different stops depending upon the following vowel context. For example, the release burst excised from /ki/ is perceived as /t/ when spliced before /a/ but is perceived as /p/ before /u/. The findings from both of these experiments indicate that identification of the consonant does not depend exclusively on the absolute frequency of the burst but rather depends on the attributes of the burst in relation to the vowel which follows.

In another study, *Cooper et al.* [1952] studied the role of formant transitions in the perception of stop consonants. Variations in the first formant transition were found to provide cues to voicing and manner whereas variations in the second formant transition were found to provide cues to place of articulation among the stops /b,d,g/ and /p,t,k/. *Cooper et al.* [1952] also studied a range of second formant transitions with the same vowels that were used in the previous burst experiment. Subjects were required to identify the stimuli as b, d or g. The results of this study indicated that while most subjects heard rising transitions as /b/ in almost all vowel contexts, perception of the falling transitions as either /d/ or /g/ varied as a function of the following vowel. Thus, as in the previous burst experiments, the perception of the formant transitions as a particular phoneme also depended on the following vowel.

Three important conclusions have been drawn from the results of these early perceptual experiments with synthetic speech. First, with regard to stop consonants, no invariant acoustic cues could be identified which corresponded uniquely to the same phoneme in all environments. Second, because burst and transition cues depend, to a large extent, on properties of the following vowel, the minimal *acoustic* unit seemed more likely to be about the size of a consonant-vowel syllable than an isolated phoneme. Indeed, as *Cooper et al.* [1952] remark, 'one may not always be able to find the phoneme in the speech wave, because it may not exist there in free form.' Finally, within the context of these experiments which have employed relatively simple synthetic stimuli, it was apparent that the acoustic information for a particular phoneme was encoded into the sound stream in a complex way; no one-to-one correspondence could be found between perceived phoneme and acoustic segment. These findings made it very unlikely at the time that the recognition of speech, particularly its phonological structure, could be carried out passively by simple template matching or filtering of the acoustic waveform. Indeed, these methods have yet to be successful as candidates for recognition schemes even some 25 years later. The earliest experiments on speech perception as well as those which followed made it very clear that the linguistic information in the message was restructured and encoded in the waveform in a complex and nontrivial way. Understanding the process by which the human listener carries out this task has been one of the fundamental problems in the field of speech perception. At the present time there is still no adequate account of the processes by which this is accomplished or the perceptual mechanisms involved.

Although these early findings failed to uncover the invariant acoustic attributes for phonemes (e.g., stop consonants) and emphasized a complex relation between acoustic attribute and phonetic unit, numerous investigators have continued, nevertheless, to search for a description of the acoustic signal which would reveal invariant attributes for phonemes [*Fant*, 1960, 1962, 1973; *Stevens*, 1967, 1973]. The experimental literature on this topic during the 1950s, 1960s and even the 1970s is extensive and cannot be reviewed here [see *Pisoni*, 1978].

It is worth reemphasizing here that the invariance problem in speech perception has, by no means, been resolved yet by anyone. The work reviewed in this section has been limited to only the stop consonants in initial position in stressed CV syllables. But some idea of the

magnitude of the invariance problem in speech perception can be obtained by considering the contextual effects that appear for stop consonants in other phonetic environments such as medial and final position of syllables as well as consonant clusters. Moreover, the problem becomes enormous when we add to it the contextual variability found for other classes of speech sounds such as fricatives, liquids, nasals and vowels as well as the inherent variability associated with phonetic context, speaking rate and individual talker differences.

The Speech Mode and Categorical Perception of Speech Signals

The earliest perceptual experiments with speech stimuli showed that listeners respond to these sounds quite differently from other auditory signals. *Liberman* and his colleagues at Haskins Laboratories found that listeners perceived synthetic speech stimuli varying between /b/, /d/ and /g/ as members of distinct categories [*Liberman et al.*, 1957]. When these same listeners were required to discriminate pairs of these sounds, they could discriminate stimuli drawn from different phonetic categories but could not discriminate stimuli drawn from the same phonetic category. The obtained discrimination functions showed marked discontinuities at places along the stimulus continuum that were correlated with changes in identification.

The ideal case of this form of perception, ‘categorical perception’, is illustrated in figure 1. In an experiment such as this, two or more phonetic segments are selected to represent end points and a continuum of synthetic stimuli is generated spanning this range. Subjects are required to carry out two tasks: identification and discrimination. In the identification task, stimuli are selected from the continuum and presented one-at-a-time in random order for labeling into categories defined by the experimenter. In the discrimination task, pairs of stimuli are selected from the continuum and presented to listeners for some discriminative response.

The basic finding of the categorical perception experiments is that listeners can discriminate between two speech sounds which have been identified as different phonemes much better than two stimuli which have been identified as the same phoneme even though the acoustic differences are comparable. At the time, the categorical perception results were considered by *Liberman* and others to be quite unusual when compared with the results typically obtained in most psychophysical experiments with nonspeech stimuli. In general, stimuli that lie along a single continuum are perceived continuously resulting in discrimination functions that are monotonic with the physical scale. As is well known, there are capacity limitations on information transmission in terms of absolute identification [*Miller*, 1956]. Listeners can discriminate many more acoustic stimuli than they can identify in absolute terms [*Pollack*, 1952, 1953]. However, in the case of categorical perception the situation is quite different. The listener’s differential discrimination appears to be no better than his absolute identification. In the extreme case of categorical perception, a listener’s discrimination performance can be predicted from his identification function under the strong assumption that the listener can discriminate between two stimuli only to the extent that these stimuli are identified as different on an absolute basis [*Liberman et al.*, 1957].

These initial findings with stop consonants led to a similar experiment with synthetic vowels which varied in acoustically equal steps through the range /I/, /e/ and /æ/. *Fry et al.* [1962], reported that these stimuli were perceived continuously, much like nonspeech stimuli. That is, the discrimination functions did not yield discontinuities along the stimulus continuum which were related to changes in identification but were relatively flat across the whole continuum. Moreover, it was observed that vowels were, in general, more discriminable than stop consonants indicating that listeners could perceive many more intraphonemic differences.

The differences in perception between stop consonants and steady-state vowels have been assumed to reflect two basic modes of perception, a categorical mode and a continuous mode. Categorical perception reflects a mode of perception in which each acoustic pattern is *always* and *only*, perceived as a token of a particular phonetic type [Studdert-Kennedy, 1974]. Listeners can discriminate between two different acoustic patterns if the stimuli have been categorized into different phonetic categories, but they cannot discriminate two different acoustic patterns that have been categorized into the same phonetic category. Information about the acoustic properties of these stimuli appears to be unavailable for purely auditory judgments as a consequence of phonetic classification. What remains available to the decision process is a more abstract and permanent code based on the listener's interpretation of the stimulus event [see Pisoni, 1911; Pisoni and Tash, 1974].

Although the stimulus generalization reflected in categorical perception might seem at first glance to be a more primitive form of stimulus control, it may provide, on the other hand, a more efficient mode of response for absolute and rapid decisions concerning the presence or absence of particular attributes such as those required in the processing of connected speech. Indeed, the great interest expressed in categorical perception of speech presumably derives from the assumption that listeners do indeed make categorical decisions when listening to continuous speech.

Continuous perception, on the other hand, may be thought of as reflecting an auditory mode of perception where discrimination is independent of category assignment. Although listeners can assign acoustically different stimuli to the same category, they may still discriminate between tokens selected from the same category. Thus, an auditory, nonphonetic basis for discrimination is available to the listener.

For a number of years the categorical perception results were assumed to be a unique aspect of speech perception and primarily a consequence of phonetic or linguistic categorization. Indeed, the differences in perception between consonants and vowels lead Liberman [1970] to argue strongly for a specialized mode of perception, a 'speech mode', to characterize the way these acoustic stimuli are perceived by humans. Other findings have suggested that a specialized perceptual mechanism – a 'special speech decoder' may exist for processing speech sounds [Studdert-Kennedy and Shankweiler, 1970].

The interpretation that categorical perception reflects a specialized mode of perception unique to speech has come under strong criticism from a number of directions in the last few years. Several investigators have argued that the differences in perception between consonants and vowels reflect differences in the psychophysical properties of the acoustic cues which distinguish these two classes of speech sounds [Lane, 1965; Pisoni, 1971; Studdert-Kennedy, 1974]. For the stop consonants there is a relatively complex relation between phoneme and its representation as sound; the essential acoustic cues are contained in the rapidly changing spectrum at onset (i.e., release burst and formant transitions) which is weak, relatively brief in duration (30–50 msec) and transient in nature. On the other hand, the cues to the vowels, at least the ones used in these experiments, involve changes in the steady-state frequencies of the first three formants which have a relatively long duration and more uniform spectral properties as well as greater intensity. In support for this, Fujisaki and Kawashima [1969, 1970] and Pisoni [1971, 1975] have shown that the differences in perception between consonants and vowels are due, in part, to the duration of their respective acoustic cues. Vowels of very short duration (i.e., 40–50 msec) are perceived more categorically than identical stimuli having longer durations.

Other findings have shown that categorical perception is also due, in part, to encoding processes in short-term memory that are a consequence of the particular type of

discrimination task used in these experiments [Pisoni, 1971, 1973, 1975]. The ABX procedure has been used in almost all of the speech perception experiments demonstrating categorical perception. In this task the subject is presented with three sounds successively, ABA or ABB. A and B are always acoustically different and the subject has to indicate whether the third sound is identical to the first or second sound. This is basically a recognition memory paradigm. In order to solve the discrimination task, the subject is forced to encode the individual stimuli in temporal succession and then base his decision on the encoded representations that have been maintained in short-term memory rather than to respond to the magnitudes of difference between stimuli within an ABX triad. In a number of experiments, Pisoni [1971, 1973] has shown that differences between categorical and continuous modes of perception are crucially dependent on the memory requirements of the particular discrimination procedure and the level of encoding required to solve the task.

Several recent experiments employing nonspeech stimuli have also suggested that categorical perception may not be peculiar to speech sounds or a specialized 'speech' mode of perception as once supposed, but may be a more general property relevant to processing complex signals that involves categorization and coding of the stimulus input and later storage in memory. For example, Cutting and Rosner [1974] recently demonstrated categorical perception for nonspeech musical sounds varying in rise-time that could be labeled easily by listeners as a 'pluck' or a 'bow'. Miller *et al.* [1976] have shown comparable categorical perception effects for nonspeech stimuli varying in the onset of a noise preceding a buzz. And in another study, Pisoni [1977] has reported similar results for stimuli differing in the relative onset-time of two component tones. In each case, these nonspeech experiments showed that discrimination was better for pairs of stimuli selected from different perceptual categories than pairs of stimuli selected from the same category. Moreover, discrimination of stimuli selected from within a category was very nearly close to chance performance as predicted by the categorical perception assumption.

The results obtained with nonspeech stimuli have provided some important insights into the underlying basis of categorical perception for speech stimuli. These nonspeech experiments have succeeded in demonstrating categorical perception when previous attempts have failed primarily for three reasons. First, the investigators employed relatively complex acoustic stimuli in which only a single component was varied relative to the remainder of the stimulus complex. In most of the previous nonspeech experiments only simple stimuli were used. Second, while these complex stimuli may be characterized as varying in linear steps along some nominally physical continuum, on both psychophysical and perceptual grounds, the stimulus continuum that is generated results in several distinctive perceptual attributes or qualities that are present for some stimuli but not others. These perceptual attributes, in turn, define quantal regions along the stimulus continuum that are separated by natural psychophysical boundaries: within these regions sensitivity is low whereas between these regions it is high. Finally, because the stimulus continuum can be partitioned into several perceptually distinctive classes, subjects can easily employ a set of labels or codes to represent these attributes in short-term memory. These codes can then be assigned to stimuli presented in the subsequent ABX discrimination task.

Categorical perception of both speech and complex nonspeech signals, therefore, can be explained, in part, by the presence of well-defined psychophysical boundaries which separate stimuli into distinctive perceptual categories and by the use of verbal labels which can be used to encode these attributes in short-term memory. Thus, this account of categorical perception involves two distinct components, a sensory component and a memory or labeling component involving some interpretative process.

But what is the basis for the distinctive perceptual attributes of speech sounds and what biological significance would they have? One approach to this problem can be found in the recent work of *Stevens* [1972] on the Quantal Theory of Speech. According to *Stevens*, phonetic features are grounded in a close ‘match’ between articulatory and auditory capacities of human speech perception and production. The acoustic attributes common to a phonetic category are determined, in part, by articulatory constraints on the speech production mechanism and, in part, by the distinctiveness of the resulting acoustic signals in perception. One piece of evidence for a perceptual match between speech perception and production according to *Stevens*, is the categorical perception results. The acoustic correlates of certain phonetic features that show quantal properties in production are also precisely those features that show categorical-like discrimination in perceptual experiments. Thus, there is some very close relation between attributes of production and distinctions made in perception.

It should be pointed out, however, that the category boundaries separating phonetic features cannot be inherently fixed since their precise location and alignment on a stimulus continuum shifts as a function of linguistic experience. Indeed, as *Popper* [1972] has suggested, ‘people who speak different languages may tune their auditory systems differently’ (p. 218). Cross-language research has shown, in fact, that the categorizations imposed on synthetic speech stimuli are based on both the particular acoustic attributes of the stimuli *and* the linguistic experience of the listener. To take one example, *Abramson and Lisker* [1965] generated a continuum of synthetic stimuli varying in voice onset time between /da/ and /ta/ and presented them for labeling to listeners of three different language backgrounds. The labeling functions for English, Spanish and Thai subjects are shown in figure 2. The results indicate that these listeners categorized the same stimuli in quite different ways depending on the phonological structure of their language. The phoneme boundaries are not only placed somewhat differently along the continuum in each case, but the Thai subjects show an additional category in their labeling behavior. This result was expected of course since in Thai a phonological distinction exists between the voiceless aspirated stop [t^h] and the voiceless unaspirated stop [t].³ This phonetic difference is not realized in either English or Spanish and consequently fails to play a role in the listener’s identification and discrimination. The importance of these findings is that the phonological systems of different languages make use of the acoustic and phonetic distinctions that exist between different speech sounds in different ways and these distinctions must be learned from the local environment.

In summary, several implications can be drawn from the categorical perception research reviewed in this section. First, the perception of speech sounds appears to have certain quantal properties because listeners treat acoustically different sounds as functionally the same. Second, the categorical perception results can be thought of as representing a phonetic mode of perception in which the listener responds to speech signals in terms of the auditory features deployed in his own linguistic system. Thus, categorical perception of speech by humans may represent a species-specific form of coding acoustic signals for subsequent linguistic processes. The extent to which the mechanisms underlying these perceptual processes are innately determined or modified by the environment has been a topic of recent interest in speech perception as shown below.

³Phonetic differences are represented by brackets whereas phonological or phonemic differences are represented by slashes. This notation is used throughout the remainder of the paper. Differences in the notation are used to call attention to the differences between phonetic and phonological levels of linguistic structure. The phonetic level corresponds to a level at which the articulatory gestures and their acoustic attributes can be represented as static or discrete states each corresponding to a set of distinctive features. In contrast, the phonological level corresponds to a somewhat more abstract level of perceptual analysis where only linguistically significant information would be represented. The phonetic level is assumed to represent universal properties of the vocal tract common to all speakers whereas the phonological level is necessarily by definition confined to language-specific distinctions.

Speech Perception in Infants

Much of what we currently know about the development of language and the acquisition of phonology is based on data obtained in studies of speech production [see *McNeill*, 1970; *Menyuk*, 1971; *Jakobson*, 1968]. One conclusion that has been drawn by a number of investigators is that the developmental process proceeds from the general to the specific and gradually involves a greater and greater differentiation of language skills. Within the last 5 years a number of pioneering studies by *Eimas* and others have demonstrated that infants as young as 1 month of age are capable of making fine discriminations among a number of the distinctive attributes of speech sounds [see *Eimas et al.*, 1971; *Eimas*, 1974, 1975]. These results obviously call into question the validity of the differentiation assumption. While increasing differentiation may very well be true of the development of productive language skills, it may not be true of the perception of speech. The recent evidence from studies of infant speech perception points to a loss of discriminative abilities over time for certain speech sounds if specific experience with these distinctions fails to take place in the local environment.

The procedure used in these speech perception experiments involved a discrimination paradigm in which the infant was first familiarized with a particular stimulus and then shifted to another stimulus. If the infant showed an increased response rate after the shift, it was assumed that the infant could discriminate the difference between the two stimuli [see *Morse*, this volume]. In the first experiment using this procedure, *Eimas et al.* [1971] studied the voicing feature which distinguishes /b/ from /p/. The stimuli were synthetically produced CV syllables that varied in acoustically equal steps of voice-onset time (VOT) between [ba] and [p^ha]. The results revealed two important findings. First, infants could discriminate between two speech sounds selected from different phonetic categories. Second, infants could *not* reliably discriminate between two acoustically different stimuli selected from within the same phonetic category. The latter finding is particularly important since it permitted *Eimas* and his collaborators to argue that their infants perceived the voicing distinction in a more nearly categorical manner and, therefore, in a linguistic mode comparable to that found with adults.

Two cross-language studies of infant speech perception have also been carried out recently on the voicing distinction using similar synthetic stimuli and methodology. In one study, *Lasky et al.* [1975] found that infants from Spanish-speaking environments could discriminate three categories along the voicing continuum. One boundary occurred between +20 and +60 msec and another between -20 and -60 msec. The first boundary is consistent with the discrimination findings of *Lisker and Abramson* [1970] for English-speaking adults and the previous results of *Eimas et al.* [1971] with infants and suggests a possible innate or sensory basis to the voicing distinction. However, the presence of a second boundary in the discrimination data was of particular interest. Spanish-speaking adults distinguish between only two categories of voicing and, based on the adult discrimination data of *Lisker and Abramson* [1970], their phoneme boundary does not correspond to either of the two VOT boundaries found with these infants. The implication of these findings is that infants are capable of perceiving three major voicing distinctions in the absence of any specific experience with these particular voicing contrasts in the environment.

In another related study, *Streeter* [1976] found that Kikuyu infants are capable of discriminating voicing differences between labial stops which are not used phonologically by the adults in their language environment. In Kikuyu, a Bantu language spoken in Kenya, there is only one labial stop with a VOT value in the range of -60 msec (i.e., a prevoiced stop). However, the Kikuyu infants could discriminate differences between three voicing categories corresponding roughly to the same ones found in the *Lasky et al.* [1975] study.

Thus, the discrimination of these voicing contrasts can also be made in the absence of relevant linguistic experience with the specific sound contrasts. The results of the cross-language experiments suggest that infants may be predisposed, in some sense, to deal with these acoustic attributes with only a very limited exposure to the specific sounds and obviously well before any experience in producing these distinctions.

The developmental course of speech perception seems to be somewhat different from other forms of perceptual development in which it is assumed that environmental experience serves primarily to sharpen the discriminative capacities of an organism [Gibson, 1969]. Since the child is capable of making relevant discriminations between the important distinctive acoustic attributes of speech sounds at a very early age, the effects of linguistic experience may be restricted primarily to learning that particular distinctions are *not* functional within a child's language environment. Thus, the course of development may not involve learning to make finer and finer discriminations among stimulus attributes but may be more analogous to the effects of acquired similarity or equivalence where there is a loss of discriminative abilities [Gibson and Gibson, 1955; Liberman et al., 1961]. As Eimas [1978] has recently suggested 'the course of development of phonetic competence is one characterized by a loss of abilities over time if specific experience is not forthcoming.' Like the adult, if the phonetic distinctions are not used functionally in the language, sensitivity to the relevant acoustic attributes is lowered or even lost and the child will fail to process or interpret them. One proposal to account for the infant results is considered below in terms of a detector system with specialized feature detectors each sensitive to a restricted range of acoustic information in the speech signal. These detectors are assumed to be available innately to the infant for processing the relevant acoustic attributes of speech although they can be modified and tuned substantially by specific linguistic experience in the environment.

To explain the results of the categorical-like discrimination found in infants, Eimas [1974] proposed an approach to speech perception based on the idea of specialized feature detectors which are finely tuned to restricted ranges of acoustic information in the speech signal. This particular idea did not originate with Eimas as a number of other investigators have remarked on the possibility of some sort of feature detecting mechanism in speech perception [see, for example, Whitfield, 1965; Liberman et al., 1967; Abbs and Sussman, 1971; Lieberman, 1970; Stevens, 1972]. However, it was left to Eimas and Corbit [1973] to introduce an experimental paradigm – selective adaptation, to speech perception that could reveal the workings of these hypothesized detectors in some detail [see Cooper, 1975, for an extensive review]. In selective adaptation, repetitive presentation of a stimulus alters the perception of a set of test stimuli. For example, in the initial study, Eimas and Corbit [1973] investigated the voicing feature and showed that adaptation with the syllable [ba] caused the locus of the phonetic category boundary between [ba] and [p^ha] to shift towards the [ba] end of the continuum. Stimuli near the boundary which were identified as [ba] when the listener was in an unadapted state were subsequently labeled as [p^ha] after adaptation with [ba]. Similar findings were obtained when [p^ha] was used as an adaptor; the locus of the phonetic boundary shifted toward the [p^ha] end of the stimulus continuum.

Eimas and Corbit [1973] interpreted the selective adaptation findings as support for the hypothesis that the perception of voicing in stop consonants involves two distinct types of feature detectors organized as opponent pairs, a voiced detector (+V) and a voiceless detector (–V). Each detector is assumed to be selectively tuned to a range of partially overlapping voice onset-time values. When a stimulus containing a particular VOT value is presented repetitively, it fatigues the detector most sensitive to that range of the feature and, accordingly, its sensitivity is reduced. After adaptation, the opponent or unadapted detector provides a greater output to the decision process in identification than the adapted detector and, accordingly, produces a shift in the locus of the phonetic category boundary.

One of the intriguing questions that the infant speech perception work has raised is the extent to which environmental input determines the development and sensitivity of these hypothesized feature detectors. There is now an extensive literature on the role of early experience in the development of the visual system which indicates that early environmental experience can modify the selectivity of cortical cells in kittens [e.g., *Hirsch and Spinelli*, 1970; *Blakemore and Cooper*, 1970]. The analogy to this developmental work has already been drawn by *Eimas* [1976] who argues that the lack of experience with specific phonetic distinctions in the local environment during language acquisition has the effect of modifying the appropriate detectors by reducing their sensitivity. Some detectors originally designed to process certain phonetic features may be captured or subsumed by other detectors after exposure to specific acoustic stimuli from the linguistic environment. The nonspecific detectors might, therefore, assume the specificity for only those attributes present in the stimuli to which they are exposed. The poor within category discrimination of speech sounds found with adults and infants in the categorical perception experiments discussed earlier may not only be due to phonetic coding of these signals but may also be a consequence of the modification of the needed discrimination mechanism. In considering the recent infant findings it may well be the case that the general program of development of speech perception is genetically determined with experience in the environment playing only a role in the tuning and alignment of the system. Moreover, the presence of linguistic universals, particularly in terms of the relatively small number of phonetic features found across many different languages, lends some support to this contention and implies that at least some of the structural mechanisms underlying speech perception are part of the biological endowment of the organism. This, of course, would not be a surprising conclusion for a biologist or ethologist but it is one that psychologists are only recently coming to appreciate and take seriously.

References

- Abbs JH, Sussman HM. Neurophysiological feature detectors and speech perception: A discussion of theoretical implications. *J Speech Hear Res.* 1971; 14:23–36. [PubMed: 4994487]
- Abramsori, AS.; Lisker, L. Voice onset time in stop consonants: acoustic analysis and synthesis. *Proc. 5th Int. Congr. of Acoustics*; Liege. 1965.
- Blakemore C, Cooper GF. Development of the brain depends on the visual environment. *Nature.* 1970; 228:477–478. [PubMed: 5482506]
- Cooper FS, Delattre PC, Liberman AM, Borst JM, Gerstman LJ. Some experiments on the perception of synthetic speech sounds. *J acoust Soc Am.* 1952; 24:597–606.
- Cooper, WE. Selective adaptation to speech. In: Restle, Shiffrin; Castellan, Lindman; Pisoni, editors. *Cognitive theory.* Vol. 1. Erlbaum Ass; Potomac: 1975.
- Cutting JE, Rosner BS. Categories and boundaries in speech and music. *Percept Psychophys.* 1974; 16:564–570.
- Eimas PD. Auditory and linguistic processing of cues for place of articulation by infants. *Percept Psychophys.* 1974; 16:513–521.
- Eimas, PD. Cohen and Salapatek, *Infant perception.* Academic Press; New York: 1975. Speech perception in early infancy.
- Eimas, PD. Held, Leibowitz and Teuber, *Handbook of sensory physiology: perception.* Springer; New York: 1978. Developmental aspects of speech perception.
- Eimas PD, Corbit JD. Selective adaptation of linguistic feature detectors. *Cognitive Psychol.* 1973; 4:99–109.
- Eimas PD, Siqueland ER, Jusczyk P, Vigorito J. Speech perception in infants. *Science.* 1971; 171:303–306. [PubMed: 5538846]
- Fant, G. *Acoustic theory of speech production.* Mouton; The Hague: 1960.
- Fant CGM. Descriptive analysis of the acoustic aspects of speech. *Logos.* 1962; 5:3–17. [PubMed: 13891546]

- Fant, CGM. Speech sounds and features. MIT Press; Cambridge: 1973.
- Flanagan, JL. Speech analysis, synthesis and perception. 2. Academic Press; New York: 1972.
- Fry DB, Abramson AS, Eimas PD, Liberman AM. The identification and discrimination of synthetic vowels. *Lang Speech*. 1962; 5:171–189.
- Fujisaki, H.; Kawashima, T. Annual Report of the Engineering Research Institute. Vol. 28. Faculty of Engineering, University of Tokyo; Tokyo: 1969. On the modes and mechanisms of speech perception; p. 67-73.
- Fujisaki, H.; Kawashima, T. Annual Report of the Engineering Research Institute. Vol. 29. Faculty of Engineering, University of Tokyo; Tokyo: 1970. Some experiments on speech perception and a model for the perceptual mechanism; p. 207-214.
- Gibson, EJ. Principles of perceptual learning and development. Appleton Century Crofts; New York: 1969.
- Gibson JJ, Gibson EJ. Perceptual learning: differentiation or enrichment? *Psychol Rev*. 1955; 62:32–1. [PubMed: 14357525]
- Hirsch HVB, Spinelli DN. Visual experience modifies distribution of horizontally and vertically oriented receptive fields in cats. *Science*. 1970; 168:869–871. [PubMed: 5444065]
- Hockett, CF. A course in modern linguistics. MacMillan; New York: 1958.
- Jakobson, R. Child language, aphasia and phonological universals. Mouton; The Hague: 1968.
- Lane HL. The motor theory of speech perception: a critical review. *Psychol Rev*. 1965; 72:275–309. [PubMed: 14348425]
- Lasky RE, Syrdal-Lasky A, Klein RE. VOT discrimination by four to six and a half month old infants from Spanish environments. *J exp Child Psychol*. 1975; 20:215–225. [PubMed: 1194843]
- Liberman, AM. Some characteristics of perception in the speech mode. Hamburg, Perception and its disorders; Proc. of ARNMD; Baltimore: Williams & Wilkins; 1970. p. 238-254.
- Liberman AM, Cooper FS, Shankweiler DP, Studdert-Kennedy M. Perception of the speech code. *Psychol Rev*. 1967; 74:431–461. [PubMed: 4170865]
- Liberman AM, Delattre PC, Cooper FS. The role of selected stimulus variables in the perception of the unvoiced stop consonants. *Am J Psychol*. 1952; 65:497–516. [PubMed: 12996688]
- Liberman AM, Harris KS, Hoffman HS, Griffith BC. The discrimination of speech sounds within and across phoneme boundaries. *J exp Psychol*. 1957; 54:358–368. [PubMed: 13481283]
- Liberman AM, Harris KS, Kinney JA, Lane HL. The discrimination of relative onset time of the components of certain speech and non-speech patterns. *J exp Psychol*. 1961; 61:379–388. [PubMed: 13761868]
- Lieberman P. Towards a unified phonetic theory. *Linguistic Inquiry*. 1970; 1:307–322.
- Lisker, L.; Abramson, AS. The voicing dimension: some experiments in comparative phonetics. Proc. 6th Int. Congr. Phonetic Sciences; Prague. 1967; Prague: Academia; 1970. p. 563-567.
- McNeill, D. The acquisition of language. Harper & Row; New York: 1970.
- Menyuk, P. The acquisition and development of language. Prentice-Hall; Englewood Cliffs: 1971.
- Miller GA. The magic number seven plus or minus, two: Some limits on our capacity for processing information. *Psychol Rev*. 1956; 63:81–97. [PubMed: 13310704]
- Miller JD, Wier CC, Pastore R, Kelly WJ, Dooling RJ. Discrimination and labeling of noise-buzz sequences with varying noise-lead times: an example of categorical perception. *J acoust Soc Am*. 1976; 60:410–417. [PubMed: 993463]
- Pisoni, DB. Supplement to Status Report on Speech Research, SR-27. Haskins Laboratories; New Haven: Nov. 1971 On the nature of categorical perception of speech sounds.
- Pisoni DB. Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Percept; Psychophys*. 1973; 13:253–260.
- Pisoni DB. Auditory short-term memory and vowel perception. *Memory Cognit*. 1975; 3:7–18.
- Pisoni DB. Identification and discrimination of the relative onset of two component tones: implications for the perception of voicing in stops. *J acoust Soc Am*. 1977; 61:1352–1361. [PubMed: 881488]
- Pisoni, DB. Estes, Handbook of learning and cognitive processes. Vol. 6. Erlbaum Ass; Hillsdale: 1978. Speech perception.

- Pisoni DB, Tash JB. Reaction times to comparisons within and across phonetic categories. *Percept Psychophys.* 1974; 15:285–290.
- Pollack I. The information in elementary auditory displays. *J acoust Soc Am.* 1952; 24:745–749.
- Pollack I. The information in elementary auditory displays II. *J acoust Soc Am.* 1953; 25:765–769.
- Popper RD. Pair discrimination for a continuum of synthetic voiced stops with and without first and third formants. *J Psycholing Res.* 1972; 1:205–219.
- Schatz C. The role of context in the perception of stops. *Language.* 1954; 30:47–56.
- Stevens, KN. Acoustic correlates of certain consonantal features. Paper presented at Conference on Speech Communication and Processing; Cambridge: MIT; 1967.
- Stevens, KN. The quantal nature of speech: evidence from articulatory-acoustic data. In: David; Denes, editors. *Human communication: a unified view.* McGraw-Hill; New York: 1972.
- Stevens, KN. Q Prog Rep No 108. Research Laboratory of Electronics, MIT; Cambridge: 1973. Further theoretical and experimental bases for quantal places of articulation for consonants; p. 247-252.
- Stevens KN, House AS. Development of a quantitative description of vowel articulation. *J acoust Soc Am.* 1955; 27:484–493.
- Streeter LA. Language perception of 2-month-old infants shows effects of both innate mechanisms and experience. *Nature.* 1976; 259:39–41. [PubMed: 1256541]
- Studdert-Kennedy, M. Sebeok, Current trends in linguistics. Vol. XII. Mouton; The Hague: 1974. The perception of speech.
- Studdert-Kennedy M, Liberman AM, Harris KS, Cooper FS. Motor theory of speech perception: a reply to Lane's critical review. *Psychol Rev.* 1970; 77:234–249. [PubMed: 5454133]
- Studdert-Kennedy M, Shankweiler D. Hemispheric specialization for speech perception. *J acoust, Soc Am.* 1970; 48:579–594. [PubMed: 5470503]
- Whitfield, IC. 'Edges' in auditory information processing. *Proc. XXIIIrd Int. Congr. of Physiological Sciences; Tokyo.* 1965. p. 245-247.

Fig. 1. Idealized form of categorical perception showing the identification function (left ordinate) and the discrimination function (right ordinate). From *Studdert-Kennedy et al.*, 1970, with permission.

Fig. 2. Labeling functions for a set of synthetic speech stimuli varying in voice onset time (VOT) that were presented to speakers of English, Spanish and Thai. From *Abramson and Lisker, 1965*, with permission.