Categorical Perception: Issues, Methods, Findings

BRUNO H. REPP

Haskins Laboratories
New Haven, Connecticut

(N. J. Lass, Editor)

I. INTRODUCTION

Ever since the beginning of language—and perhaps even earlier—human beings have classified things and events into categories. Categorization occurs when we focus on important properties that are common to different objects and ignore irrelevant detail. Although such an act of attention is commonly accompanied by verbal statements, categorization may also occur covertly. However, the fact that most categories do have names is definitely advantageous in communication. For example, the name of an object or event may still be recalled when memories of physical details have long faded. It is not surprising, therefore, that category names form the core of our vocabulary.

Many of the categories we have are natural—they reflect obvious physical partitions among things in the world, and there is little question or choice as to what is included in a particular category and what is not. Other categories, however, are less transparent, and may reflect special knowledge or conventions. Some scientific categories fall in this class; for example, the zoologist's category of "fish" excludes dolphins and whales but includes eels and sea horses, whereas a prescientific, shape-oriented category of "fish" might include the former but exclude the latter. In addition, there are cases, such as those involving aesthetic judgment or preference, where individuals are free to draw the boundaries between categories. Categories based on relative judgment (size, weight, speed, etc.) are totally situation-specific and essentially arbitrary.

The categories of speech—which include the phonetic segments, or phonemes—play an important part in linguistic theory and are implicated in the development and continued use of alphabetic writing. However, illiterates have little awareness of them (Morais, Cary, Alegria, & Bertelson, 1979); nonlinguists know them only in a vague fashion, commonly mistaking letters for phonemes; and even among specialists, there are disputes about their precise nature and description. Did linguists merely invent these categories for the purpose of abstract description, or did they discover an important, though not very transparent, principle of discrete organization that underlies human speech production and perception? And if the latter, do the proposed descriptive categories map directly onto the functional categories of active speech communication? These questions are aspects of the more general question about the psychological reality of the products of linguistic analysis—an issue that lies at the heart of modern psycholinguistics.

Categorical perception research in the speech domain is concerned with the perceptual reality of phonetic segments—that is, with the role of phonetic categories in perceptual processing regardless of whether the perceiver has any awareness of them. Although categorical perception research is, in principle, a rather broad area of inquiry permitting a variety of methods, over the years it has become identified with a particular laboratory paradigm. That paradigm has

generated a large amount of useful research that presents a challenge to theories of speech perception. However, in recent years, there have been some signs of exhaustion. This seems a good time to review some of the history, methods, and problems of categorical perception research and to try to see where we stand. We will begin with a historical overview. The studies mentioned will be discussed in greater detail in later sections.

II. HISTORICAL OVERVIEW

A. The Early Haskins Research

Categorical perception research began at Haskins Laboratories not long after the construction of the first research-oriented speech synthesizer, the Pattern Playback. Liberman, Harris, Hoffman, and Griffith (1957) used this new tool to construct a series of syllables spanning the three categories /b/, /d/, and /g/ preceding a vowel approximating /e/. Although these stimuli formed a physical continuum (obtained by increasing the onset frequency of the second formant in equal steps), listeners classified them into three rather sharply divided categories. To test whether the physical differences among the stimuli within a category could be detected by listeners, Liberman *et al.* employed an ABX discrimination task. (This task requires subjects to indicate whether the last of three successive stimuli matches the first or the second, which are always different from each other.) The results showed that stimuli classified as belonging to different categories were easily discriminated, whereas stimuli perceived as belonging to the same category were very difficult to tell apart, even though the physical differences seemed comparable. This characteristic pattern of results came to be called "categorical perception" (see Section III,A). By assuming that listeners have no information beyond the phonetic category labels (an assumption later often referred to as the "Haskins model"), Liberman *et al.* (1957) were able to generate a fair prediction of discrimination performance from known labeling probabilities; however, performance was somewhat better than predicted, suggesting that the subjects did have some additional stimulus information available.

The pioneering experiment of Liberman *et al.* (1957) set the pattern for a number of similar studies exploring different kinds of phonetic contrasts. Thus, Liberman, Harris, Kinney, and Lane (1961) reported categorical perception of the /d/ versus /t/ contrast cued by "first-formant cutback"; Liberman, Harris, Eimas, Lisker, and Bastian (1961) found similar results for the intervocalic /b/ versus /p/ distinction cued by closure duration; and Bastian, Eimas, and Liberman (1961) demonstrated that stop manner cued by closure duration (/slɪt/–/split/) was likewise categorically perceived. These findings contrasted with those of Fry, Abramson, Eimas, and Liberman (1962) and Eimas (1963), who showed

that synthetic vowels forming an /ɪ/–/ɛ/–/æ/ continuum were discriminated equally well within and between phonetic categories—a result referred to as "continuous perception." Continuous perception was obtained also with other properties of vowels such as duration (Bastian & Abramson, 1964) and intonation contour (Abramson, 1961), as well as with nonspeech stimuli that had certain critical features in common with categorically perceived speech stimuli (e.g., Liberman, Harris, Eimas, Lisker, & Bastian, 1961; Liberman, Harris, Kinney, & Lane, 1961). Thus, categorical perception seemed to be specific to speech (excluding isolated vowels), and to stop consonants in particular.

These early findings provided one of the pillars for the motor theory of speech perception set forth by the Haskins group (Liberman, 1957; Liberman, Cooper, Harris, MacNeilage, & Studdert-Kennedy, 1967; Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). The basic tenet of the motor theory is that speech perception and articulatory control involve the same (or closely linked) neurological processes. When different phonetic categories are distinguished by essentially discrete articulatory gestures (as with stop consonants differing in voicing or place of articulation), perception of stimuli from a physical continuum spanning these categories will be categorical; on the other hand, when continuous articulatory variations between phonetic categories are possible (as with the vowels), perception will be continuous (cf. Liberman, Harris, Eimas, Lisker, & Bastian, 1961). In other words, the motor theory takes categorical perception to be a direct reflection of articulatory organization.

For a number of years, categorical perception research stayed at Haskins Laboratories—a situation that changed only in the 1970s, when appropriate speech synthesizers became available in other laboratories. The only pertinent research outside Haskins in the early years was conducted by Harlan Lane and his collaborators at the University of Michigan, who examined categorical perception from a psychophysical viewpoint, focusing on the question of whether a similar phenomenon could be produced with nonspeech stimuli under comparable experimental conditions. The results of that not very successful effort were summarized in Lane's (1965) critical review of the early Haskins research. Lane's criticisms anticipated some of the concerns of later researchers, but they had little impact at the time because they were backed up by rather weak data. However, they provoked a forceful, if somewhat belated, reply by Studdert-Kennedy, Liberman, Harris, and Cooper (1970) that remains the classic statement of the Haskins view of categorical perception (see Section III,A).

Categorical perception research continued at Haskins during the 1960s. Abramson and Lisker (1970) showed that the voiced–voiceless distinction for utterance–initial stop consonants, as cued by voice onset time (VOT), was categorically perceived by speakers of two languages with different voicing boundaries, Thai and English. Another early cross-language study was conducted by Stevens, Liberman, Öhman, and Studdert-Kennedy (1969) with Swed-

ish and English vowels. Although perception of these vowels was not quite as continuous as in the earlier study by Fry *et al.* (1962), there seemed to be no connection between identification and discrimination, suggesting noncategorical perception. The categorical perception of the place of articulation distinction for voiced stop consonants (Liberman *et al.*, 1957) was replicated by several studies, including one by Mattingly, Liberman, Syrdal, and Halwes (1971), who for the first time included stop consonants in utterance-final position as well as several nonspeech controls that were not categorically perceived.

B. The Information-Processing Approach

In the meantime, two Japanese scientists became interested in the Haskins findings and began to experiment along similar lines. The work of Fujisaki and Kawashima (1968, 1969, 1970, 1971), presented in a series of limited-circulation progress reports, remained virtually unknown in the West until Pisoni (1971, 1973, 1975) discussed and extended it. The work of these authors, Pisoni in particular, brought categorical perception into the mainstream of contemporary psychology. Whereas up to this time the focus had been on categorical perception as a pure **phenomenon,** on its relation to articulatory behavior, and on the effects of learning on auditory sensitivity, attention now turned to perceptual **processes** and to stimulus and task variables involved in categorical perception experiments.

Fujisaki and Kawashima (1969, 1970, 1971) formulated a dual-process model for the discrimination of speech stimuli that explicitly distinguished between categorical phonemic judgments and judgments based on auditory memory for acoustic stimulus attributes (see Section III,B). Thus, the model attempted to account for the commonly observed difference between the categorical predictions of the Haskins model and actual discrimination performance—a difference that was treated as an uninteresting nuisance in the early Haskins research (unless it was sufficiently large to be interpreted as "continuous" perception). Fujisaki and Kawashima also explored new classes of speech stimuli (synthetic fricatives, semivowels, and liquids) and showed that their perception was somewhat less categorical than that of stop consonants, although not as continuous as that of isolated vowels. They further experimented with vowels of varying duration, with or without added context, and showed that even vowels may be perceived quite categorically when conditions are unfavorable for auditory memory. The imaginative (though somewhat fragmentary) work of Fujisaki and Kawashima has served as a stimulus for further research to the present day (see Sections IV,A and V,A).

Several ideas of the Japanese researchers were elaborated and tested by Pisoni (1971, 1973, 1975; Pisoni & Lazarus, 1974), who applied the dual-process model to a variety of discrimination paradigms, showing that the categoricalness

of perception depends, to some extent, on how much use can be made of auditory memory in a task. He further confirmed this point by varying stimulus duration, the duration of interstimulus intervals, and by introducing interfering sounds between the stimuli to be discriminated. Pisoni and Tash (1974) were the first to use same–different reaction times as an indicator of subjects' sensitivity to acoustic stimulus differences within phonetic categories. This analytic research began a trend of increasing interest in subjects' ability to discriminate subphonemic (within-category) acoustic differences between speech stimuli—a trend that shifted the emphasis from categorical perception as a mere phenomenon to the psychoacoustics and psychophysical methodology of speech discrimination.

C. Offsprings of Categorical Perception Research

The early 1970s spawned several significant research developments that grew out of categorical perception research and have since become highly active areas semiindependent from (but, of course, intimately related to) the traditional approach to categorical perception, with which they share the use of the classic experimental paradigm requiring identification and discrimination of synthetic speech sounds from a physical continuum. The diversification proceeded on three fronts: new subjects, new tasks, and new stimuli.

One of the new enterprises was research on **infant speech perception.** In a now classic paper, Eimas, Siqueland, Jusczyk, and Vigorito (1971) reported that 1- and 4-month-old human infants responded to stimuli from a VOT (/ba/–/pa/) continuum in a way similar to adults: The infants discriminated stimuli from opposite sides of the adult category boundary (as indicated by an increase in the rate of nonnutritive sucking in response to a stimulus change), but not physically different stimuli from the same category. This exciting finding has since been replicated several times and has been extended to a variety of different stimuli. Infant speech perception research has been following closely on the heels of the research on adult speech perception and, in general, it has revealed that infants' perceptual capabilities are remarkably similar to those of adults, though without the influence of specific linguistic experience. Important research is now underway to determine the role played by exposure to a specific language in the course of perceptual development (see Section VI,C).

A second development concerns studies of **animal speech perception.** Although few in number, they have attracted much attention through Kuhl and Miller's (1975, 1978) finding that chinchillas divide a VOT continuum into the same categories as adult humans. There is increasing activity today in this methodologically difficult but fascinating area (see Section VI,D).

On the methodological side, researchers began to experiment with a variety of discrimination paradigms and different response measures, including rating

scales, reaction time, and even evoked potentials (see Section IV,B). The phenomenon of categorical perception held up remarkably well under this onslaught. A vigorous strand of research was started by Eimas and Corbit (1973), who applied the technique of **selective adaptation** to continua of synthetic speech stimuli. By presenting one or the other endpoint stimulus over and over, it was possible to shift the location of the phonetic category boundary and even to shift the associated discrimination peak with it. Numerous studies, including some of the most elegant work in speech perception, have tried to unravel the sources and mechanisms of the adaptive shifts. Unfortunately, the returns have been somewhat disappointing, for it is now quite clear that the adaptation effect does not take place at the level of "phonetic feature detectors," as originally believed, but is a purely auditory phenomenon (Roberts & Summerfield, 1981; Sawusch & Jusczyk, 1981). Although the selective adaptation technique continues to be useful for probing into the auditory processes of speech perception, this research is tangential to the concerns of this review and will not be discussed in detail (for reviews, see Ades, 1976; Cooper, 1975; Diehl, 1981; Eimas & Miller, 1978).

Categorical perception research also continued along more traditional lines with adult human subjects. Encouraged by the increasing sophistication of speech synthesis, however, researchers explored phonetic categories other than those of stop consonants and vowels. More or less categorical perception was demonstrated for the affricate–fricative distinction (Cutting & Rosner, 1974) and for continua of liquid consonants (McGovern & Strange, 1977; Miyawaki, Strange, Verbrugge, Liberman, Jenkins, & Fujimura, 1975), of nasal consonants (Larkey, Wald, & Strange, 1978; J. L. Miller & Eimas, 1977), and of the oral–nasal distinction (J. L. Miller & Eimas, 1977), among others. With certain qualifications, this research showed that virtually all consonantal distinctions are categorically perceived (see Section V,B).

D. The Psychophysical Approach

In the early Haskins research and in Lane's (1965) critical review of it, a good deal of attention was paid to the possibility that categorical perception was caused by general auditory processes. The conclusion from the early Haskins studies (notwithstanding Lane's objections, which had only weak empirical support) had been that categorical perception was specific to speech and to (stop) consonants in particular. Interest in the psychoacoustics of categorical perception reawakened in the mid-1970s, when the earlier conclusion was shattered by several demonstrations of apparently categorical perception of nonspeech sounds. Thus, Cutting and Rosner (1974) claimed to have found categorical perception of complex tones varying in rise time (the *pluck–bow* distinction); J. D. Miller, Wier, Pastore, Kelly, and Dooling (1976) reported categorical perception of noise–buzz sequences intended to be analogous to a VOT continuum; and

Pisoni (1977) found similar results for two tones varying in relative onset time. In Section V,C, we will examine these and other studies in considerable detail.

The demonstrations of categorical perception of nonspeech sounds stimulated some psychophysicists to take a closer look at categorical perception, and some speech researchers to take a closer look at psychophysics. Thus, Macmillan, Kaplan, and Creelman (1977) attempted to fit categorical perception into the framework of signal detection theory; Ades (1977) made a cautious (and still largely unexplored) connection with the related psychophysical work of Durlach and Braida (1969; Braida & Durlach, 1972); Pastore (1981) reviewed psychoacoustic factors that may be relevant to categorical perception; and Schouten (1980) went so far as to propose that all of speech perception could be explained by psychoacoustic principles.

Psychophysical theories were further encouraged by several reports of successful speech discrimination training. Whereas earlier studies had focused on the role of learning in categorical perception and had attempted (with limited success) to produce the phenomenon by training subjects in the use of category labels for nonspeech stimuli (e.g., Cross, Lane, & Sheppard, 1965; Parks, Wall, & Bastian, 1969), Carney, Widin, and Viemeister (1977), for example, took the converse approach: They showed that categorical perception of speech may be attenuated by training listeners to pay attention to acoustic stimulus properties. These findings suggested that categorical perception is essentially a function of experience and attentional strategies (see Section VI,A).

Underlying these psychophysical approaches is a single-process (or "common-factor") view of categorical perception that assumes that linguistic categories are essentially psychoacoustic in nature (J. D. Miller et al., 1976; Pastore, Ahroon, Baffuto, Friedman, Puleo, & Fink, 1977). This view has emerged in recent years as a serious competitor for the dual-process model proposed by Fujisaki and Kawashima (see Section III,D). The antagonism between these two models has become tied up with the more general controversy about whether it is necessary to postulate a special phonetic mode of perception at all (cf. Liberman, 1982; Repp, 1982; Schouten, 1980).

The psychophysical trend stimulated researchers at Haskins Laboratories and elsewhere to illustrate the complexity of phonetic perception in new experiments. The emphasis of much of this new research is on the complex many-to-one relationship between acoustic stimulus properties and phonetic percept, demonstrated experimentally as phonetic "trading relations" or other contextual interactions among several different acoustic cues. Since many of these studies use the methodology of categorical perception research (i.e., identification and discrimination of stimuli from synthetic speech continua), they may be viewed as dealing with the categorical perception of stimuli varying along two or more dimensions (e.g., Best, Morrongiello, & Robson, 1981; Fitch, Halwes, Erickson, & Liberman, 1980), with particular attention to the distinction between

auditory and phonetic modes of perception. This research has led to various contemporary versions of the motor theory (e.g., Bailey & Summerfield, 1980; Repp, Liberman, Eccardt, & Pesetsky, 1978). Several recent studies have been particularly successful in constructing appropriate nonspeech analogs to examine the presumed speech-specificity of the demonstrated cue trading relations (Best *et al.*, 1981; Summerfield, 1982). We will discuss some of these studies later; for detailed reviews, however, see Liberman (1982) and Repp (1982).

Investigators have also shown an increased interest in one aspect of the methodology of categorical perception: contextual dependencies among successive stimuli in a labeling or discrimination task (Crowder, 1982a; Healy & Repp, 1982; Repp, Healy, & Crowder, 1979; see Section III,C). Related work has grown out of the research on selective adaptation (Diehl, Elman, & McCusker, 1978; Sawusch & Nusbaum, 1979). This is likely to be an area of considerable activity in the near future.

We have come to the end of this brief historical review, in the course of which I hope to have mentioned all major trends and landmarks. In the following, more detailed review, I focus in sequence on the several different factors that contribute to the phenomenon called "categorical perception." Discussions of theoretical and methodological issues (Sections III and VII) precede and follow the core sections (IV, V, and VI) that are dedicated to the review of data.

III. EMPIRICAL ASSESSMENT OF CATEGORICAL PERCEPTION: MODELS AND METHODS

A. Defining Categorical Perception: The Classical Haskins View

The preceding section has provided a broad answer to the question, What constitutes categorical perception? Now we shall examine this issue in somewhat more detail. First, it is useful to point out that the term "categorical" may be understood in at least three different ways, which may be called "literal," "phenomenal," and "empirical."

Literally speaking, categorical perception refers to the use of categories by individuals in responding to their environment. In this sense, it is a ubiquitous phenomenon not restricted to speech and, in particular, there is no implication that the perceiver is unaware of stimulus variations within a category. This is not the way in which the term has been used by speech researchers, but others have occasionally interpreted and used it that way.

Phenomenally speaking, categorical perception refers to the experience of discontinuity as a continuously changing series of stimuli crosses a category

boundary, together with the absence of clearly perceived changes within a category. It must be emphasized here that categorical perception is a very striking and readily demonstrated phenomenon. All persons who sit down and listen to one of the standard series of stop consonants varying in VOT or formant transitions, provided they are able to hear the synthetic sounds as speech, will experience abrupt perceptual changes at certain places on the continuum. The continuing attraction of categorical perception to both the novice and the seasoned investigator lies in its permanent and replicable vividness in the listener's experience.

However, subjective experience alone is not enough to satisfy the rigors of scientific investigation, and we must therefore turn to categorical perception as an empirical concept, describing a particular pattern of data in an experiment. It is here that the situation becomes more complex, because ideal categorical perception (where category labels are the sole determinant of performance) is rarely, if ever, encountered in the laboratory. Empirical data typically deviate more or less from this ideal, and some criterion must be applied for deciding whether they do or do not provide evidence for categorical perception. In fact, to capture different amounts of deviation, it may be necessary to speak of degrees of categorical perception (cf. Studdert-Kennedy et al., 1970), although this violates the strict definition of categorical perception proposed by the Haskins group:

> Categorical perception refers to a mode by which stimuli are responded to, and *can only be responded to, in absolute terms*. Successive stimuli drawn from a physical continuum are not perceived as forming a continuum, but as members of discrete categories. They are identified absolutely, that is, independently of the context in which they occur. Subjects asked to discriminate between pairs of such "categorical" stimuli are able to discriminate between stimuli drawn from different categories, but not between stimuli drawn from the same category. In other words, discrimination is limited by identification: subjects can only discriminate between stimuli that they identify differently. (Studdert-Kennedy et al., 1970, p. 234; emphasis theirs)

A typical experiment might proceed as follows: In an identification (labeling) test, stimuli from a physical continuum spanning two categories unambiguously represented by the endpoint stimuli are presented repeatedly in randomized order to subjects for classification into one or the other category. In a subsequent (sometimes preceding) discrimination test, typically using the ABX paradigm, adjacent or more widely separated stimuli from the continuum are presented for discrimination. The identification data are summarized in the form of labeling functions that relate response percentages to stimulus location on the continuum. The discrimination data yield one or more discrimination functions, which relate a measure of discrimination accuracy (usually percent correct) for stimulus pairs of equivalent physical separation to stimulus location. Ideal categorical perception in this standard design exhibits four semiindependent characteristics:

1. Labeling probabilities change abruptly somewhere along the continuum; in other words, the identification functions have a rather steep slope. The point of maximum slope is the **category boundary** (equivalently defined as the point at which responses in two adjacent categories are equiprobable).

2. Discrimination functions show a peak at the category boundary; that is, stimuli are more easily discriminated when they fall on opposite sides of the boundary than when they fall on the same side.

3. Discrimination performance within each category is at or near chance level.

4. Discrimination functions are perfectly predictable from the labeling probabilities (using one of the simple formulae provided by the Haskins model—see Pollack & Pisoni, 1971). This implies that (a) the discrimination peak is in exactly the right place and of the right height, and (b) the labeling probabilities are appropriate; that is, they apply independently of the context in which they were observed. [These two corollaries show that criterion (4) is not directly implied by criteria (1), (2), and (3).]

As we have already observed, the actual data are rarely perfect. They may fit the ideal description more or less well. In evaluating the data, more importance is attached to some criteria than to others. For example, the criterion of steepness of labeling functions is a very weak one. Given that stimulus continua do contain ambiguous stimuli in the category boundary region, the steepness of labeling functions depends in part on how closely the stimuli are spaced along the continuum (see the discussion of this issue by Lane, 1965, and Studdert-Kennedy *et al.*, 1970). A much more important criterion is the presence of a peak in the discrimination function that coincides with the location of the phoneme boundary—a feature of the data later christened the **phoneme boundary effect** (Wood, 1976a). It is the essential defining characteristic of categorical perception, although it may not be sufficient if the other criteria are grossly violated. A certain amount of deviation is usually tolerated for both of the remaining criteria (near-chance performance within categories and match of predicted and obtained discrimination functions).

A statistical criterion of whether some data do or do not represent categorical perception is provided by the goodness of fit of the predictions (Healy & Repp, 1982; cf. Pisoni, 1971). In practical usage, however, the striking contrast between the results for stop consonants and isolated vowels (or nonspeech stimuli) has often supported the categorical–continuous dichotomy irrespective of any deviations from the ideal patterns of categorical or continuous perception. Later research, however, has yielded a number of intermediate cases that can no longer be accurately characterized by this simple dichotomy.

The question of what constitutes admissible evidence for categorical perception was discussed in detail by Studdert-Kennedy *et al.* (1970) in their reply to

Lane's (1965) critical review. Lane had focused on criterion (1) (described earlier) and had revealed its weakness, and he had criticized criterion (4) on the basis that corollary (4b) may not be satisfied (see Section III,C for further discussion of his arguments). Although the Haskins authors were remarkably effective in rebutting Lane's methodological objections, there remained one prime weakness in their presentation. It stemmed, in large measure, from viewing categorical perception as a monolithic phenomenon, and from a resulting unwillingness to consider in detail the different factors that enter the experimental situation defining categorical perception. In a perceptive commentary, Haggard (1970) noted that ''the controversy between Lane and the Haskins group stems from a failure to enumerate levels or aspects of the perceptual process and make separate statements about them'' (p. 6).

B. Speech Perception as a Two-Component Process: The Dual-Process Model

Speech perception was conceived by the Haskins group of the 1950s and 1960s as a modular process that, for a given phonetic distinction, is either categorical or continuous. The origin of the two types of phonetic perception was hypothesized to lie in the articulatory continuity or discontinuity of the segmental distinctions perceived; that is, in whether articulations intermediate between those typical of two segments occur in natural speech (or are anatomically possible at all). Both types of phonetic perception were thought to be mediated by an articulatory representation of the input, in accord with the motor theory, although the similarity of continuous speech perception and nonspeech perception was evident.

This essentially unidimensional view of speech perception contrasts with the dual-process model introduced by Fujisaki and Kawashima (1969, 1970) and elaborated by Pisoni (1971, 1973, 1975). Rather than assuming that only a single perceptual mode is active at any given time, they proposed that two modes are active simultaneously (or in rapid sequence). One of them is strictly categorical and represents phonetic classification and the associated verbal short-term memory. The other mode is completely continuous and represents processes common to all auditory perception, including auditory short-term memory. The results of any particular speech discrimination experiment are assumed to reflect a mixture of both component processes: The part of performance that can be predicted from labeling probabilities (using the Haskins model) is attributed to categorical judgments, whereas the remainder (the deviation from ideal categorical perception) is assigned to memory for acoustic stimulus properties.

The dual-process model partially abandons the articulatory rationale for categorical perception by explicitly equating continuous with auditory (i.e., nonspeech) perception. Accordingly, the difference in categoricalness between, for

example, stop consonants and vowels is hypothesized to derive not from the different articulatory properties of these segments, but from the different strengths of their representations in auditory memory. By augmenting the Haskins prediction model with a free parameter representing the contribution of auditory memory, Fujisaki and Kawashima also introduced a way of quantifying different degrees of categorical perception that, unfortunately, has not been adopted by other researchers.

It is obvious that the dual-process model opened up new avenues for research. It now became possible to ask how subjects in an experiment utilize the two sources of information (categorical and continuous, or phonetic and auditory) and what factors might lead them to rely more on one than on the other. Since the continuous component was identified with general auditory memory, several standard experimental techniques became available to weaken or strengthen that memory and to observe the subsequent changes in speech discrimination performance. Attention turned from categorical perception as a somewhat mysterious, "special" speech phenomenon to an analysis of the experimental situation—of the task factors, stimulus factors, and subject factors that conspire to generate a particular pattern of results.

C. Problems of Prediction: Context Sensitivity versus Phonetic Mediation

At this point, a brief digression into the methodology of predicting discrimination performance is in order since the prediction test is the most widely used formal criterion of categorical perception. The Haskins model derives its predictions of perfectly categorical discrimination from labeling probabilities obtained in an independent identification task in which the individual stimuli are presented in random order (see Pollack & Pisoni, 1971, for computational techniques). This procedure was criticized by Lane (1965) on two grounds. First, he argued, the phonetic categories assumed to be employed covertly in the discrimination task may not be identical with the ones employed overtly in the labeling task. Second, even if the same categories were used, the probabilities of classifying the stimuli into the different categories may not be the same in the two tasks because the labeling probabilities may be sensitive to context (i.e., they may be influenced by immediately preceding or following stimuli), and the context of individual stimuli is different in the two tasks. Of course, these arguments applied only to cases of apparently noncategorical perception; they reflected Lane's contention that categorical perception was not specific to speech and could be acquired in the laboratory (see Section V,C).

The first objection is the less serious of the two. For many continua of speech sounds, there are no plausible alternative phonetic categories to the ones intended and suggested to the subjects by the experimenter. In other cases, the objection

may be valid but could be met by not restricting the subjects' response set in the labeling task. However, although individual differences in the number and kind of categories used may come to the fore in a free-response situation, subjects are also rather willing to adopt categories suggested by the experimenter, even if they are not the standard ones (see Carden, Levitt, Jusczyk, & Walley, 1981, for a recent striking example). Therefore, it seems that a mismatch of phonetic categories in identification and discrimination tasks has not been a serious problem in categorical perception research. (A related, but more subtle, problem that cannot be so easily dismissed is that subjects may devise phonetic **subcategories** in a discrimination task, based on different degrees of confidence in their phonetic judgments—e.g., "good /b/" versus "poor /b/"; see Liberman, Harris, Eimas, Lisker, & Bastian, 1961, for an early documented example. We will encounter this issue again later in this review.)

The second objection, that of context effects in labeling, deserves closer attention. Studdert-Kennedy *et al.* (1970) responded to it by insisting that "categorical perception entails context-free perception" (p. 246). In other words, if context effects are present and lead to a mismatch of predicted and obtained discrimination performance, that is simply evidence that perception is not categorical. Lane (1965) suggested that the predictions be derived by having subjects label the stimuli in the same context in which they are presented for discrimination. (For early applications of this method, see Cross and Lane, 1964, cited in Lane, 1965, and also Fujisaki & Kawashima, 1969.) However, Studdert-Kennedy *et al.* (1970) dismissed this procedure on the grounds that "by 'acknowledging context,' we predict discrimination from discrimination" (p. 247).

This response is characteristic of the unidimensional view of categorical perception espoused by the Haskins group at that time. Their sole concern was to determine whether or not perception of a given set of stimuli was categorical. Although they acknowledged that ideal categorical perception is rarely encountered, they were not particularly interested in the causes of the deviations from the ideal. However, an explanation of these deviations is likely to increase our understanding of categorical perception, particularly since there are many instances of noncategorical perception that are far from continuous. It is possible to distinguish three such situations (Healy & Repp, 1982): (1) There may be context effects in (covert) phonetic labeling but the subjects may nevertheless rely exclusively on category labels in discriminating different stimuli. (This is certainly a form of categorical perception, though not the absolute one of the Haskins definition.) (2) Labeling may be independent of context but subjects may utilize auditory stimulus information in discrimination and thereby exceed the predictions of the Haskins model. (In this case, perception is absolute without being categorical.) (3) The deviations from the categorical ideal may be due to both contextual effects in labeling and auditory memory in discrimination.

These considerations suggest that **phonetic mediation** (reliance on category

labels) in discrimination and **context sensitivity** in labeling are two logically distinct aspects of the experimental situation that can (and should) be assessed separately. To assess phonetic mediation, the predictions of discrimination performance are derived from "in-context" labeling probabilities, that is, from subjects' labeling responses to stimuli presented in the exact sequence used also in the discrimination task; any remaining discrepancies between predicted and obtained performance may then be unambiguously attributed to auditory memory. The magnitude of context effects in labeling, on the other hand, may be inferred directly from the in-context labeling responses by examining contextual contingencies (Fujisaki & Kawashima, 1969; Healy & Repp, 1982; Repp et al., 1979).

The separation of context sensitivity and phonetic mediation is essentially an elaboration of the dual-processing hypothesis. It provides more realistic estimates of labeling probabilities and, thereby, a more accurate assessment of the relative contributions of (covert) categorical judgments and auditory memory to discrimination. Indeed, it appears that the small advantage of obtained over predicted discrimination scores, which are customarily obtained with stop consonants, may be entirely due to contrast effects in (covert) labeling and not to any direct access to auditory memory (Healy & Repp, 1982). Context effects may themselves have a dual-process explanation: They may either represent a form of response bias at the level of phonetic categorization (see, e.g., Diehl et al., 1978; Shigeno & Fujisaki, 1980) or they may derive from an interaction of auditory memory traces akin to lateral inhibition (Crowder, 1978, 1981) or both factors may be at work simultaneously.

D. Psychoacoustics and Categorical Perception: The Common-Factor Model

The dual-process hypothesis of Fujisaki and Kawashima contains the assumption that categorical perception derives entirely from the phonetic component in the model, that is, from the application of linguistic categories. The auditory component is assumed to be essentially continuous. There is an alternative possibility, however: It could be that some auditory dimensions of speech are **not** continuous and that there are psychoacoustic thresholds that may coincide with the phonetic category boundaries on a speech continuum. In other words, categorical perception may be a phenomenon of auditory perception, in part or in toto. Pastore et al. (1977) introduced the term "common-factor model" for the hypothesis that "a single (common) factor [other than phonetic categorization] causes both a peak in the discrimination function and a categorical dichotomy and thus the correlation between the two" (p. 686). This proposal was encouraged by the early findings of seemingly categorical speech discrimination in human infants (Eimas et al., 1971) and in nonhuman animals (Kuhl & Miller,

1975), and of certain nonspeech stimuli by human adults (Cutting & Rosner, 1974; J. D. Miller *et al.*, 1976); it has come to play a central role in contemporary speech perception research. It is so important because it promises not only to explain the speech perception capabilities of infants and animals, but also to provide a principled account of the demarcation and evolution of linguistic categories.

According to the common-factor model, the discrimination peak that characterizes categorical perception (the "phoneme boundary effect") comes about because, given a psychoacoustic threshold on a continuum, different subthreshold stimuli are mutually indiscriminable, sub- and suprathreshold stimuli are easy to tell apart, and different suprathreshold stimuli are discriminated according to Weber's law, which predicts increasingly poorer performance as stimulus differences of constant absolute size move away from the threshold (cf. J. D. Miller *et al.*, 1976). The difficulty with the common-factor model does not lie in its proposal that discrimination peaks **can** come about in this way (for they obviously can, as several studies of nonspeech continua have shown—see Section V,C), but in the difficulty of showing that they **do** have a strictly psychoacoustic basis in the case of speech continua that are categorically perceived.

To obtain support for this hypothesis, some authors have employed signal detection theory or related methods to derive the "perceptual spacing" of stimuli on a speech continuum, characteristically finding that stimuli are spaced further apart in the boundary region than within categories (Elman, 1979; Macmillan *et al.*, 1977; Oden & Massaro, 1978; Perey & Pisoni, 1978). However, this result merely amounts to a redescription of the data; it does not answer the question of why stimuli are spaced in this way in perception. As we will see in later sections, the various attempts at proving that specific auditory thresholds underlie particular phonetic boundaries have not been uniformly successful, although some have produced encouraging results.

Another problem for the common-factor model is that there are cases of boundary effects on continua that quite clearly do not straddle any psychoacoustic thresholds. These include continua of isolated vowels (e.g., Pisoni, 1971), isolated fricative noises (Fujisaki & Kawashima, 1970), or musical intervals (e.g., Burns & Ward, 1978). The results of these studies suggest (as does some of the research reviewed in Section VI) that a discrimination peak may be caused simply by the existence of appropriate categories. On the other hand, we do have some rather strong evidence for psychoacoustic discontinuities on certain speech continua (see Pastore, 1981). Perhaps what is needed is a **modified dual-process model**—one that admits the possibility of significant nonlinearities in auditory perception while, at the same time, assuming a separate contribution of phonetic category labels in the process of discrimination.

This modified dual-process model might be considered unparsimonious by

some, but it does appear to accommodate the existing evidence, as the following review will attempt to show. The model also bears a certain resemblance to the two-factor model of Durlach and Braida (1969; Braida & Durlach, 1972), although their model was developed to account for discrimination of sound intensity (a true psychoacoustic continuum over most of its range). The Durlach–Braida model assumes two components, a "sensory-trace mode" and a "context-coding mode," that jointly contribute to discrimination accuracy and differ in their relative permanence. The relevance of this model to categorical perception was pointed out by Ades (1977). If two processes are necessary to account for simple intensity resolution, it can hardly be unparsimonious to postulate two separate processes in speech perception.

It can be seen from the foregoing discussion that theoretical reasoning in categorical perception research has not progressed very far. The models proposed so far are simple and few in number. They contrast with the richness and occasional complexity of the data, to which we now turn. The following three sections are dedicated to a review of research on categorical perception within the confines of the standard identification–discrimination paradigm. Some relevant research using unconventional methods will be mentioned in the concluding section. The organization of the three sections is based on the view that categorical perception, as a pattern of experimental results, is a joint function of three major factors: task variables, stimulus variables, and subject variables. Categorical perception is **not** a property attached to a particular stimulus set. Rather, it is a way in which a particular individual responds to particular stimuli in a particular experimental situation. Accordingly, Sections IV–VI divide the evidence into pieces relating to task, stimulus, and subject factors. Although it would be logical to begin with the most important section (that on stimulus factors), it seemed more convenient to treat task factors first in order to avoid prolonged discussions of methodology in the following sections.

IV. TASK FACTORS IN CATEGORICAL PERCEPTION

In this section, we will examine to what extent categorical perception is a function of the task used to assess discrimination. There are two ways of pursuing that question: Either one starts with stimuli that are not very categorically perceived (e.g., isolated vowels) and tries to make their perception more categorical by modifying the task or, conversely, one starts with stimuli whose perception is highly categorical and attempts to make their perception less categorical. Both approaches have been used in the past. Within the framework of the dual-process model, they amount to either decreasing or increasing the audi-

tory memory component in subjects' performance. The contribution of the categorical component is assumed to be either constant or inversely proportional to that of auditory memory.

A. Procedures for Increasing Categorical Perception

There are two ways of reducing auditory memory without changing the stimuli themselves or their relationship (see Section V,A for effects of stimulus manipulations). One is to introduce interference in the form of noise or by interpolating irrelevant sounds between the stimuli to be discriminated. The other way is to increase the temporal separation of the stimuli so that auditory memory for the first stimulus has decayed by the time the second stimulus arrives.

1. Interference with Auditory Memory

In the earliest vowel discrimination study, Fry *et al.* (1962) found no discrimination peaks at category boundaries, but this was probably due to a ceiling effect, coupled with the use of imperfectly controlled stimuli. Most later studies (e.g., Fujisaki & Kawashima, 1969, 1970; Pisoni, 1971; Stevens *et al.*, 1969) have found fairly clear peaks on vowel continua, so there is good reason to believe that there is a phonetic component in vowel discrimination. Cross and Lane (1964, cited in Lane, 1965) actually used the original tapes of Fry *et al.* and added noise in the form of an additional, irrelevant resonance. Although it seems that phonetic identification should have suffered considerably, Lane (1965) nevertheless reports that marked discrimination peaks were observed at the category boundaries.

Fujisaki and Kawashima (1969, 1970) included a condition in which a constant /a/ vowel immediately followed each of the test stimuli (vowels from an /i/−/e/ continuum presented in ABX triads for identification and discrimination). They claimed to have found more nearly categorical perception in that condition than when the fixed context was omitted, and they attributed that difference to the context serving as a "perceptual reference." By this they presumably meant that it facilitated categorization and also, perhaps, that it interfered with auditory memory. Their data are less than clear, however, and this is compounded by the fact that different data are reported in their 1969 and 1970 papers for ostensibly the same experiment. The 1970 data, in particular, show a narrowing of the discrimination peak coupled with an **increase** in within-category discrimination performance. Thus, the context did not seem to interfere with auditory memory, although it may have aided categorization.

Fujisaki and Kawashima also reported that adding a constant vocalic context to fricative noise stimuli from a /ʃ/−/s/ continuum had little effect on discrimination performance (which, curiously, was highly categorical even for isolated

fricative noises), although closer inspection of their results again reveals that within-category discrimination was improved by the presence of context. These results contrast with recent data that suggest that a following vowel reduces the discriminability of fricative noises, even in subjects who are able to perceptually segregate the noise from the vowel (Repp, 1981c), and that isolated noises are not categorically perceived (Healy & Repp, 1982; Repp, 1981c).

Pisoni (1975, Exp. III) examined the role of a fixed context in more detail. He argued that if the context stimuli serve as a perceptual anchor, as hypothesized by Fujisaki and Kawashima, then it should not matter whether the context precedes or follows the test stimuli. If, on the other hand, the context interferes with auditory memory, one might expect that a following context will produce more interference than a preceding one. In addition, Pisoni hypothesized that the similarity of context and test stimuli would determine the amount of interference. To test this last hypothesis, Pisoni used four different sounds (a 1000-Hz pure tone, a burst of white noise, and the vowels /a/ or /ɛ/) as contexts for stimuli from an /i/–/ɪ/ continuum. The context immediately preceded or followed each test stimulus in labeling and ABX discrimination tests, with a no-context control condition included. The results supported the similarity hypothesis: Discrimination scores were lowest in the /ɛ/-vowel context, although all contexts lowered performance somewhat. There was also more of a decrement when the context followed, rather than preceded, the test stimuli, although the difference was small.

Pisoni made no attempt to assess the degree of categorical perception in the various context conditions, nor did he report whether labeling probabilities were influenced by the various contexts. To examine these issues, Repp *et al.* (1979) presented pairs of vowels from an /i/–/ɪ/–/ɛ/ continuum in a same–different discrimination task. The interval between the two stimuli on a trial was either silent or partially filled by an irrelevant vowel sound (/y/). The intervening stimulus produced a clear decrement in discrimination performance, and a comparison with predictions from standard identification data led to the conclusion that perception had become more categorical. However, Repp *et al.* also had their subjects label the stimuli in pairs and computed in-context predictions of discrimination performance (see Section III,C). These predictions matched the obtained scores much better than did the standard predictions and, significantly, the match was equally good whether or not an interfering sound was present, even though discrimination scores (as well as the predictions) were much lower in the presence of interference. Evidently, the interpolated sound affected both in-context labeling and discrimination. The effect on labeling was evident in a drastic reduction of contrast effects between the members of a stimulus pair (i.e., of the tendency to assign them different labels).

These results permit two interpretations. The one preferred by Repp *et al.* (1979; see also Crowder, 1981) was that auditory memory had its effect before

phonetic categorization, in the form of contrastive interactions between auditory stimulus traces, and that discrimination was subsequently based in large part on phonetic labels, even though the stimuli were isolated vowels. To account for the remaining difference between predicted and obtained discrimination performance (which was considered negligible by Repp *et al.* but turned out to be rather large in a later, similar study by Healy & Repp, 1982), it seems necessary to appeal either to the covert use of additional phonetic categories in discrimination or to some more permanent form of auditory memory that is immune to interference [such as Massaro's (1975), "synthesized auditory memory"]. The other interpretation is that labeling and discrimination were both based directly on auditory stimulus representations, so that interference with auditory memory affected both equally. In this view, which is congenial to psychophysical theories and seems more parsimonious, labeling is viewed simply as a form of coarse-grained discrimination, and contrast effects in labeling are the consequence, not the cause, of accurate discrimination. However, the presence of peaks in the discrimination function indicated that phonetic categories did influence the subjects' "same–different" decisions at some stage.

Whichever interpretation is preferred, the Repp *et al.* data clearly demonstrated that interference with auditory memory has a large effect in a categorical perception task. They are also consistent with the research on the so-called "suffix effect"—the increase in recall errors for the last item in a word list when that list is followed by another, irrelevant item (Crowder, 1971, 1973a, 1973b; Crowder & Morton, 1969). The traditional interpretation of this effect has been that the suffix disrupts a precategorical auditory trace lasting a few seconds—a trace that retains primarily vocalic information because of its higher distinctiveness (Crowder, 1971; Darwin & Baddeley, 1974). Vowel discrimination tasks probably tap the same kind of memory.

2. Decay of Auditory Memory

Let us now turn to studies that attempted to manipulate auditory memory by changing the temporal interval (interstimulus interval = ISI) between stimuli to be discriminated. In the context of categorical perception research, this method was first applied by Pisoni (1971, 1973), who introduced variable ISIs (0–2 sec) in a same–different discrimination task using both vowels (/i/–/ɪ/) and stop consonants (/bæ/–/dæ/, /ba/–/pa/). There was a clear decrement in vowel discrimination performance as the interval increased (except for reduced scores at the zero interval), whereas there was little effect on stop-consonant discrimination performance. A breakdown of the data into within-category and between-category discrimination scores revealed that both scores decreased for vowels, whereas only a slight decrease in between-category performance could be seen for stop consonants. (Within-category discrimination of stop consonants was close to chance.) Very similar results were obtained in a replication by Cutting,

Rosner, and Foard (1976) and, in related studies, by Cowan and Morse (1979) and Repp *et al.* (1979) for vowels, and by Frazier (1976) for consonants.

Since between-category discrimination of vowels was thought to be based on category labels, Pisoni concluded from the uniform decline in performance that an increase in temporal delay resulted in a decay not only of auditory memory (of which there was very little for stop consonants), but also of phonetic memory. However, it seems unlikely that phonetic short-term memory for a single label would decay at all over 2 sec (cf. Fujisaki & Kawashima, 1971). Therefore, all decrements observed were probably due to auditory memory decay.

One question not answered by these studies is whether the memory decay has any asymptote (performance continued to decline up to 2 sec). The question of the time course of memory decay for vowel stimuli was investigated by Crowder (1982a), who varied the ISI in pairs of vowels in a same–different discrimination task, covering the range from 0 to 5 sec. He found that performance declined up to about 3 sec and then remained stable. In a second experiment of his, the subjects' task was not to respond "same" or "different," but instead to identify the second vowel in each pair. The result was similar: The contextual (contrastive) influence of the first vowel on the second, assumed to be mediated by auditory memory, went away at about 3 sec of separation. (However, see Fujisaki and Shigeno, 1979, for a contradictory finding.) Crowder's results converge with those from suffix effect experiments, where a similar decay rate of auditory memory has been found (Crowder, 1969; however, see Watkins & Todres, 1980). The hypothesis that suffix effects and vowel discrimination are mediated by the same memory store was further supported in a recent study by Crowder (1982b), in which he showed that individual differences in the magnitude of the suffix effect correlated reliably with the same subjects' vowel discrimination performance when the interstimulus intervals were short (500 msec) but not when they were long (3 sec).

In summary, these studies leave little doubt that auditory memory plays a role in vowel discrimination tasks, and the parallelism with the suffix effect results suggests that the auditory memory store employed for isolated vowels may also be functional in other tasks involving more complex speech stimuli. The same auditory memory also appears to be responsible for contrastive influences of one stimulus on identification of a following stimulus. (Note, however, that there is also retroactive contrast.) One question that is still not resolved is whether vowel discrimination at delays beyond 3 sec is based entirely on phonetic labels, or whether there is another, more permanent form of auditory memory that aids discrimination at longer delays. Crowder's (1982a) data indicated that the decline in vowel discrimination performance as a function of temporal delay was relatively small while, at the same time, contrast effects in vowel labeling disappeared completely. This suggests that, even at the longest intervals, obtained discrimination performance probably exceeded the in-context predictions (which

Crowder did not calculate). Crowder's results appear consistent with the just-mentioned data of Repp *et al.* (1979), which showed that contrast effects nearly disappeared at a long (filled) interval, whereas obtained discrimination scores were still higher than predicted.

Thus, an explanation of vowel discrimination may ultimately require a three-process model, including two kinds of auditory memory: a fast-decaying one of the kind discussed by Crowder, which mediates contrast effects, and a slower-decaying one that may be utilized in discrimination. The latter corresponds to the "context-coding mode" of Durlach and Braida (1969) and to the "synthesized auditory memory" of Massaro (1975).

The third process, of course, is phonetic categorization. This process is needed in the model to account for the phoneme boundary effects in vowel discrimination, for they could hardly be caused by psychoacoustic thresholds. However, it is possible that these effects, like those on true nonspeech continua (Kopp & Livermore, 1973), and unlike those on stop-consonant continua (Elman, 1979; Popper, 1972; Wood, 1976a, 1976b), are entirely due to response bias and not to increased perceptual sensitivity at category boundaries. In other words, there may be no direct phonetic mediation in vowel discrimination; rather, the phonetic labels may merely bias **auditory** judgments. In view of the relative auditory salience of vowel differences, this would not be surprising. One might think of auditory and phonetic decisions being engaged in a race, with auditory decisions winning when the stimuli are isolated vowels but losing when the stimuli are stop consonants. Thus, the influence of phonetic categorization on vowel discrimination may occur by hindsight, as it were, whereas it may be truly mediational in consonant discrimination.

B. Procedures for Reducing Categorical Perception

We turn now to a review of studies that approached the problem of auditory memory from the other side: Instead of reducing discrimination performance (and increasing categorical perception) by decreasing auditory memory, these studies attempted to **increase** performance (and thereby decrease categorical perception), either by enhancing the auditory memory component or by providing the subjects with finer-grained scales on which to respond. These efforts concentrated on a class of speech sounds that, in the standard experimental setting, were highly categorically perceived and showed little evidence of auditory memory: stop consonants differing in voicing (VOT) or place of articulation (formant transitions).

1. More Sensitive Discrimination Paradigms

Early studies of categorical perception had suggested that stop consonants might not have any representation in auditory memory at all. Although discrimi-

nation performance was usually somewhat higher than predicted by the Haskins model, the difference was relatively small and tended to be ignored. Stop consonants were regarded by the Haskins group as abstract perceptual categories stripped of all auditory information, and as the prime example of "encoded" speech sounds whose perception requires the operation of a special speech processor (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Liberman, Mattingly, & Turvey, 1972). Therefore, a demonstration of the existence of some memory for acoustic properties of stop consonants would have been an important contribution.

The ABX discrimination paradigm was used in all early categorical perception studies and remains popular to this day. This paradigm was preferred because it requires a forced choice and, at the same time, absolves the experimenter from specifying the dimension on which the stimuli differ (which, in the case of speech, may be difficult to convey to naive subjects). However, it has often been suggested that ABX is not the most sensitive paradigm, the reason cited being the presumed necessity to compare A and X, with the resulting demand on memory (e.g., J. D. Harris, 1952; Pisoni, 1971). Pisoni (1971) tried out a different procedure, the 4IAX paradigm, which shares with the simpler AX (same–different) task the advantage of using pairs rather than triads of stimuli, and with the ABX task the advantage of requiring a forced choice. (In the 4IAX task, the subject must decide which of two stimulus pairs contains a difference.) In Experiment IV of his dissertation, Pisoni found that discrimination of steady-state vowels was improved considerably in the 4IAX paradigm, as compared to the ABX paradigm. In his Experiment V, he compared stop consonants from a place of articulation (/bæ/–/dæ/–/gæ/) continuum in the same two tasks. Performance in the 4IAX paradigm was only slightly better than in the ABX paradigm, and then only for two-step comparisons but not for one-step comparisons. These data did not offer very striking support for an auditory memory component in stop-consonant discrimination, although both ABX and 4IAX scores differed reliably from the Haskins model predictions.

In another study using the same two paradigms, Pisoni and Lazarus (1974) examined stop consonants from a VOT (/ba/–/pa/) continuum. This study also included a condition in which the subjects were not given the standard labeling test, but received instead the /ba/–/pa/ continuum repeatedly in fixed order before performing the discrimination test. This procedure was expected to sensitize the listeners to acoustic stimulus differences. Indeed, there was some increase in performance due to both the 4IAX procedure and the prior experience with the stimulus continuum. However, prior experience appears to have been the critical factor, for Pisoni and Glanzman (1974) failed to find any difference between the ABX and 4IAX paradigms when no pretraining was provided. It should also be noted that in these experiments the difference between the two paradigms was confounded with differences in interstimulus intervals: In the ABX paradigm, there was a 1-sec interval between stimuli in a triad, whereas in

the 4IAX paradigm, the stimuli within a pair were separated by only 150 or 250 msec, with a 1-sec interval between the two stimulus pairs that constituted one trial. The small size of the difference between the two paradigms is consistent with the finding (Pisoni, 1971, 1973) that temporal separation has little effect on stop-consonant discrimination.

A direct comparison of the ABX and AX paradigms with speech stimuli was performed recently by Crowder (1982b), who used vowels from an /i/–/ɪ/ continuum and computed d' indices according to the tables published by Kaplan, Macmillan, and Creelman (1978), which make a fair comparison between the two tasks possible. Crowder also made the interstimulus intervals in the two tasks comparable by having the same short (500 msec) or long (3 sec) delays between the B and X items of the ABX triads and between the A and X items of the AX pairs. (The A–B interval in ABX triads was fixed at 250 msec.) The results showed not only that the AX paradigm was more sensitive than the ABX paradigm, but also that it yielded much more stable results, as measured by split-half reliability indices. In Crowder's (1982b) words, "this result does suggest some caution for investigators choosing the ABX task lest they be making it hard for themselves to demonstrate experimental effects in a sensitive way" (p. 481).

Suspicions that the ABX paradigm encourages categorical perception had existed for some time, and researchers increasingly used alternative paradigms, including oddity (which probably shares all the disadvantages of ABX), AXB (essentially an economical version of 4IAX), 4IAX, and AX. MacKain, Best, and Strange (1981) compared the AXB and oddity paradigms using an /r/–/l/ continuum and found AXB to be superior. A comparison of more than two paradigms for speech discrimination in a single study still remains to be done. However, an extensive comparison of different paradigms for **nonspeech** discrimination (pure tone frequency or phase relationships) was conducted by Creelman and Macmillan (1979). In contrast to the results with speech, they found greater sensitivity to frequency differences in the (variable standard) ABX task than in the AX task, with 4IAX performance in between. (However, no differences at all were found between the three paradigms when the task was phase discrimination, suggesting that stimulus factors may interact with task factors in determining discrimination performance.) Another result of the Creelman and Macmillan study was that fixed-standard paradigms (in which only the X stimulus varies from trial to trial) were found to be superior to variable-standard paradigms. Fixed-standard tasks have not been used in speech perception research until fairly recently; since they are usually employed in conjunction with discrimination training, we will review these studies in a later section (VI,A).

We should note that it is not quite clear why certain discrimination paradigms are superior to others. Psychophysical theory predicts certain differences for ideal observers (Creelman & Macmillan, 1979), but real subjects are typically far from this ideal. To give a psychological explanation of performance differences,

we need a model of the perceptual strategies employed in different tasks, especially in the more complex ones. An unpublished study by Pastore, Friedman, and Baffuto (1976) was directly concerned with that issue. Pastore *et al.* found for intensity discrimination, as did Creelman and Macmillan for frequency discrimination, that ABX was superior to AX, and that fixed-standard tasks were superior to variable-standard tasks. What is of interest here is that Pastore *et al.* examined different models of subject strategies in the ABX task and found that the results were best explained by the assumption that only B and X were compared, with A merely serving to "reduce uncertainty." Thus, the data of Pastore *et al.* do not support the assumption commonly made by speech researchers that listeners compare A and X as well as B and X. However, both sides may be right. The subjects in speech experiments are typically inexperienced, whereas those in psychophysical experiments are highly practiced. Therefore, it should not be surprising that the latter subjects adopt a more effective strategy. Unless subject strategies also depend on whether the stimuli are speech or nonspeech (as indeed they may), the results available suggest that the ABX paradigm is inferior to the AX paradigm with naive subjects but not with experienced subjects. In Section VI,A, we will discuss the effects of discrimination training on categorical perception. Without such training, it appears that the perception of stop consonants remains fairly categorical even when more sensitive discrimination paradigms are used.

2. Rating Scales and Reaction Times

Several researchers have attempted to obtain evidence for subjects' sensitivity to subphonemic detail by modifying the single-item identification task so as to permit the subjects to transmit more information about perceived stimulus differences. One of the earliest studies in that vein was published by Barclay (1972). He presented listeners with a /bæ/–/dæ/–/gæ/ continuum but permitted only two labels, "b" and "g." If subjects' perception had been truly categorical, all stimuli perceived as "d" (as determined in a separate test) should have been assigned to the "b" or "g" categories on a random basis. However, listeners were found to be more likely to apply the label "b" to the more "b"-like instances of /dæ/, and the label "g" to the more "g"-like instances. Thus, listeners showed some sensitivity to acoustic stimulus properties in the center of the continuum. Barclay proposed that categorical perception is primarily a memory phenomenon, observed only when successive stimuli are to be compared. However, Haggard (1970) has pointed out that Barclay's stimuli lacked a third formant, which may have created considerable ambiguity in the /dæ/ region. If the intended /dæ/ tokens could indeed be heard as either /bæ/ or /gæ/, Barclay's results would seem trivial.

An alternative approach is to provide subjects with a numerical scale on which to rate the individual stimuli. The possibility that categorical perception is merely

a consequence of the limited number of phonetic categories available to the perceiver was first investigated by Conway and Haggard (1971; see also Haggard, Summerfield, & Roberts. 1981), who gave their subjects a 9-point rating scale to judge stimuli from five-member /bɪl/–/pɪl/ and /gɪl/–/kɪl/ (VOT) continua. The functions relating average stimulus ratings to position on the continuum were distinctly sigmoid in shape, with the largest change in ratings occurring across the phoneme boundary and virtually no change within categories. If perception had been continuous, the functions should have been linearly increasing. Thus, these results not only provided strong evidence for categorical perception, but also offered no indication that a more fine-grained response scale enabled listeners to make distinctions within phonemic categories. In a second, similar study, Conway and Haggard (1971) obtained more continuous-looking functions, but the stimuli spanned only a small range in the vicinity of the boundary, where even the two-category labeling function is nearly linear. Therefore, these data were consistent with categorical perception.

The rating scale of Conway and Haggard had no special relation to the stimuli on the continuum and may have been used by the subjects merely to indicate their degree of confidence in their categorical judgments (as noted by Haggard *et al.*, 1981). Since the endpoints of the scale were explicity identified with phonetic categories, it is perhaps not surprising that categorical perception was obtained. An alternative method is to establish a one-to-one correspondence between stimuli and responses—the task called **absolute identification.** This task was employed by Sachs (1969), whose subjects used the numbers 1–8 to identify eight stimuli from a /bɑdəl/–/bædəl/ continuum, as well as eight stimuli from two /ɑ/–/æ/ continua with different stimulus durations. Despite the procedure used, and despite the fact that the distinction was located in the vowel, perception of the word continuum was quite categorical, and so, to some extent, was the perception of the short-duration vowels (see Section V,A for a discussion of effects of phonetic context and duration on vowel discrimination). These results provided strong evidence that absolute identification does not prevent or even attenuate categorical perception. Later, Cooper, Ebert, and Cole (1976) had their subjects use a 7-point scale to identify stimuli from seven-member /ba/–/wa/ and /ga/–/ja/ (formant transition duration) continua. Once again, the average numerical responses changed most rapidly across the phoneme boundary, and there was no indication that stimuli strictly within a category (which really applied only to the /ba/ end of the /ba/–/wa/ continuum) were distinguished by the subjects.

Using the same procedure, Perey and Pisoni (1978) compared absolute identification of stimuli from /ba/–/pa/ and /i/–/ɪ/ continua. Once again, the stop-consonant data showed categorical perception, whereas the vowel ratings were more nearly continuous, though not a strictly linear function of stimulus number. Perey and Pisoni showed, however, that stop-consonant (and vowel) discrimina-

tion in a subsequent ABX test could be predicted more accurately from the rating data than from simple binary labeling probabilities, suggesting that some sub-phonemic differences were picked up by subjects in the rating task. Still, perception of stop consonants was far from continuous.

Rating scales or absolute identification have been used in many other studies, all of which obtained the basic phenomenon of categorical perception of stop consonants (e.g., Elman, 1979; McNabb, 1976a; Rosen, 1979; Sawusch, 1976). Another variant, the method of direct magnitude scaling, was employed by Port and Yeni-Komshian (1971, cited in Strange, 1972) and Strange (1972). Strange's subjects responded to individual stimuli (stop consonants from a VOT continuum) by positioning a pointer within a bounded interval. Still, perception remained categorical unless a fair amount of training was provided, in which case some subjects responded more nearly continuously (see Section VI,A).

Yet another approach was recently taken by Samuel (1982). His intention was to locate, for each listener, the "best /ga/" on a narrowly spaced /ga/–/ka/ (VOT) continuum, presupposing that subjects would be able to distinguish between different stimuli within the /ga/ category. The subjects in this study could control stimulus presentation, step repeatedly through the continuum, and zero in on the preferred stimulus. Although Samuel did not determine the reliability of his subjects' estimates of the prototypical /ga/, he did find individual differences that correlated with the magnitude of boundary shifts obtained in a subsequent selective-adaptation experiment. However, since prototype location correlated neither with the location of the phoneme boundary nor with prototype estimates derived by several other procedures (Samuel, 1979), the results must be viewed with some caution.

Studdert-Kennedy, Liberman, and Stevens (1963) found that labeling reaction times for stimuli from stop-consonant and vowel continua exhibited a peak at the category boundary—a finding that has often been replicated (e.g., Pisoni & Tash, 1974; Repp, 1975, 1981a; however, see Hanson, 1977) and is also obtained with nonspeech continua (Cross et al., 1965). Since reaction times indicate the subjects' uncertainty in making phonetic decisions, they are long for ambiguous stimuli and short for unambiguous ones. However, the prototype concept, introduced to speech perception by Oden and Massaro (1978) and Repp (1976a), suggests that, even for stimuli that are consistently placed in the same category, there might be a gradient of reaction times reflecting their perceptual distance from the category prototype. The only attempt so far to test this hypothesis for stop consonants (Samuel, 1979) appears to have been unsuccessful. In other studies, too, labeling reaction times to different stop-consonant stimuli strictly within the same category (if several such stimuli existed on a continuum) have tended to be equivalent (e.g., Pisoni & Tash, 1974).

Numerical ratings and reaction times have also been collected in discrimination tasks. Vinegrad (1972) conducted a direct magnitude scaling study with stop

consonants (/bɛ/–/dɛ/–/gɛ/), vowels (/i/–/ɪ/–/ɛ/), and pure tones varying in frequency. The stimuli were presented in AXB triads, and the subjects' task was to locate X in relation to A and B by marking a point on a line. Stimuli A and B were always the endpoints of the continuum, which made the procedure highly similar to that of Strange (1972), who presented only the middle stimuli. The results were very clear-cut: The stop consonants exhibited strongly categorical perception; different stimuli from within the same category were located in the same place. Vowels, on the other hand, gave more continuous results, as expected. The results for the tones were similar to those for the vowels; however, neither were perfectly continuous (see Section V,C).

Category boundary effects for isolated vowels have also been obtained in studies where the subjects' task was to rate the perceived similarity of stimuli drawn from a continuum (e.g., Golusina, cited in Chistovich, 1971; Van Valin, 1976). Unless subjects are very carefully instructed to base their judgments on auditory stimulus properties alone, this task is likely to elicit a phonetic strategy.

Following an earlier study by Strange and Halwes (1971), Pisoni and Glanzman (1974) obtained confidence ratings for discrimination judgments of stop consonants (/ba/–/pa/) presented in AXB and 4IAX formats. There was a very straightforward monotonic relation between discrimination accuracy and confidence; in other words, subjects accurately postdicted their own success on each trial. Although performance was not any better with confidence ratings than without, the correlation obtained does suggest, as Conway and Haggard (1971) had observed earlier, that subjects have at least statistical information about acoustic stimulus differences in the form of subjective uncertainty. Seen in this way, the Pisoni and Glanzman results are equivalent to a previous demonstration by Studdert-Kennedy, Liberman, and Stevens (1964) that reaction times in a stop-consonant ABX task were shortest for between-category comparisons, where discrimination was easiest, and longest for within-category comparisons. These observations also raise the possibility that, rather than directly accessing some auditory memory representations, subjects might base decisions about stimulus differences on estimates of their subjective uncertainty in phonetic categorization.

Most of the studies discussed in this section demanded an **overt** indication of subjects' awareness of intraphonemic stimulus differences. The results provided relatively little evidence of such awareness as far as stop consonants are concerned. On the other hand, there is overwhelming evidence that acoustic stimulus properties do have perceptual effects of which listeners are not directly conscious. Some of this evidence comes from same–different reaction time studies (which will be reviewed in Section V,A) together with the role played by the perhaps most obvious factor influencing the detectability of acoustic differences—the physical size of the difference itself (i.e., the "step size" on a continuum). Other studies have shown that the magnitude of the selective adapta-

tion effect depends on the precise acoustic properties of the adapting stimulus (e.g., McNabb, 1976a; J. L. Miller, 1977, 1981; J. L. Miller & Connine, 1980; Samuel, 1979) and that the perception of fused dichotic stimuli is sensitive to similar acoustic variables (e.g., J. L. Miller, 1977; Repp, 1976a, 1977). These and other studies show that the auditory properties of stop-consonant stimuli play a significant role at early, precategorical stages of processing (as they must).

It remains for us to mention several studies that assessed listeners' sensitivity to within-category differences by monitoring some more immediate response of the organism than overt labeling. Studies of vocal imitation fall in this category because immediate repetition does not require categorization of a stimulus. K. S. Harris, Bastian, and Liberman (1961) showed long ago that imitation of stimuli from a /slɪt/–/splɪt/ continuum was strongly categorical; that is, subjects were unable to reproduce the precise closure durations of the stimuli and instead produced only two types of utterances. Of course, this result may reflect articulatory limitations or habits rather than (or as well as) an influence of categorical perception on the articulatory response. (The motor theory does not even distinguish these two possibilities because categorical perception is hypothesized to derive from articulation.) For this reason, perhaps, imitiation has rarely been used in later studies of categorical perception. A phoneme boundary effect in the imitation of isolated vowels was reported by Chistovich, Fant, de Serpa-Leitão, and Tjernlund (1966), whereas imitations of vowel durations by American listeners (Bastian & Abramson, 1964) showed no effect of phonetic categorization (see also Section V,B,5).

A more covert physiologic response to auditory stimuli may be obtained from the surface of the skull in the form of evoked potentials. Dorman (1974) presented listeners with stop-consonant–vowel stimuli differing in VOT. At varying times during a train of stimuli, the standard stimulus (/ba/) changed to a different stimulus either within the same category or in a different category (/pa/). The N1–P2 component of the evoked potential (100–200 msec after stimulus onset) was significantly larger for between-category shifts than for within-category shifts, and the response to the latter did not differ from that to a no-change control. Dorman interpreted his results as reflecting immediate phonetic recoding.

Curiously, Dorman's results were not mentioned by Molfese (1978), who reinvestigated the problem using principal components analysis of evoked-potential waveforms. His subjects listened to stimuli from a /ba/–/pa/ continuum and identified each stimulus by pressing one of two keys. The results were complex but suggested that within- as well as between-category differences affected the electric brain response. This basic finding was replicated with /ga/–/ka/ stimuli in 4-year-old children (Molfese & Hess, 1978) and 2- to 5-month-old infants (Molfese & Molfese, 1979). The evoked potentials of these young subjects also exhibited a component that responded only to between-category differences,

whereas those of newborn infants did not (Molfese & Molfese, 1979), and those of adults (Molfese, 1978) followed a somewhat more complex pattern. These findings are intriguing, although they are not without methodological problems; at the simplest level of interpretation, they suggest that neuroelectric correlates of both auditory and phonetic processing may be found.

Changes in evoked potentials for within-category differences occur without the subject's awareness. However, some striking evidence that listeners **can** gain conscious access to subphonemic acoustic stimulus differences comes from several studies that provided extensive training for the listeners. Although these results would fit in the present section on paradigms, we prefer to discuss them in Section VI, which deals with subject factors in categorical perception, one of which is experience.

V. STIMULUS FACTORS IN CATEGORICAL PERCEPTION

In this section, we will review various relevant factors residing in the stimuli themselves (rather than in their arrangement or in the kinds of responses given by subjects). In Section V,A, we will examine the effects of variables operating within a given set of stimuli, the most important ones being physical separation (step size) and duration. In Section V,B, we will review differences in the degree of categorical perception among different stimulus sets, focusing on stimuli other than the ubiquitous stop consonants and vowels. This will lead us to a detailed consideration of the perception of "nonspeech analogs" of speech stimuli, together with findings of categorical perception of other kinds of nonspeech stimuli (Section V,C).

A. Stimulus Factors and Auditory Memory

1. Step-Size Effects

The variable most obviously related to the ease of discriminating two stimuli is the magnitude of the physical difference. Several levels of this variable, in the form of different step sizes in comparisons drawn from a continuum, have been included in most studies of categorical perception, including the earliest ones. It is a commonplace finding that two-step discrimination performance is higher than one-step discrimination performance, three-step is higher than two-step, and so on. One might think that here is prima facie evidence that listeners are sensitive to subphonemic physical differences between the stimuli. However, the issue is not that simple: Stimuli that are more widely separated on the physical continuum generally are more likely to be classified into different categories, and

under the assumption that discrimination is mediated by category labels, discrimination accuracy is **predicted** to increase with step size. Therefore, an effect of step size cannot be taken to reflect auditory (rather than phonetic) discrimination unless it is significantly larger than predicted from (in-context) labeling probabilities.

This point was given systematic attention by Healy and Repp (1982), who computed the differences between predicted (in-context) and obtained same–different discrimination performance at three different step sizes for four different stimulus continua (stop-consonant–vowel syllables, isolated vowels, isolated fricative noises, and complex tones varying in timbre). The idea was that, given a linear measure of performance (d' in their case; percentages are not suitable because of their inherent nonlinearity), the predicted–obtained differences should increase with step size if listeners are indeed sensitive to acoustic differences; otherwise, the step-size effect should be fully accounted for by the in-context predictions from labeling performance. Healy and Repp found that a residual step-size effect was present for vowels and tones, and probably for fricative noises as well (a ceiling effect prevented statistical significance), but not for stop consonants. Since stop-consonant discrimination was generally slightly **worse** than predicted (a seemingly unusual result that, however, reflected the effective partialling-out of contrast effects in labeling), the results provided strong support for the hypothesis that stop-consonant discrimination was based exclusively on phonetic labels. Apparently, the subjects in the Healy–Repp experiment retained no distinctive acoustic details of stop-consonant stimuli but did make use of auditory information with the other stimulus classes.

However, these results do not warrant the conclusion that acoustic properties of stop consonants do not enter auditory memory at all. Rather, their auditory traces may be so weak as to influence performance only under very special conditions. One sufficiently sensitive measure of performance appears to be reaction time in a same–different task. Pisoni and Tash (1974) adapted to speech perception a procedure used by Posner (e.g., Posner & Mitchell, 1967) in his well-known letter-matching studies: A "same" judgment for two physically identical stimuli ("physical match") might be faster than a "same" judgment for two physically different stimuli from the same category ("name match") if any auditory information is retained from the first stimulus in the pair. Similarly, "different" reaction times to two stimuli from opposite sides of a category boundary might be faster when the physical separation between the two stimuli is large than when it is small. Both results were reported by Pisoni and Tash (1974) for syllables from a /ba/–/pa/ continuum presented in pairs with 250-msec ISIs: When two stimuli from the same category were separated by two steps on the continuum, "same" responses were significantly slower than for pairs of identical stimuli; at the same time, subjects were not any more likely to say "different" to two-step pairs than to identical pairs, so that, overtly, perception was

highly categorical. "Different" response latencies to stimuli crossing the boundary and separated by two steps were longer than for stimuli separated by four or six steps. However, there was no significant difference between four- and six-step "different" pairs and, moreover, the likelihood of incorrect "same" responses was highest for two-step pairs, so that the "different" reaction times may have reflected uncertainty in phonetic, rather than auditory, judgments.

On the basis of their results, Pisoni and Tash (1974) proposed a two-stage model for same–different comparisons, according to which a comparison of auditory stimulus properties precedes the comparison of phonetic labels, the second stage being used only if the auditory difference falls neither below the "same" nor above the "different" criterion adopted by the listener. This ordering of stages is reversed with respect to the Fujisaki–Kawashima dual-process model for ABX discrimination, which puts the phonetic comparison first. However, unlike the Pisoni–Tash model, the Fujisaki–Kawashima model was not intended to describe real-time information processing; rather, it merely captured the fact that phonetic categories loom large in the listener's awareness, and actually permits either order of deployment of the two component processes.

The demonstration by Pisoni and Tash that some acoustic properties of stop consonants are retained in memory inspired other researchers to ask whether these memory traces, like those of isolated vowels, decay over time. Several studies addressing this question have yielded mixed results. Eimas and Miller (1975) presented pairs of stimuli from a /ba/–/da/ (formant transition) continuum at three ISIs (50, 200, and 800 msec). Since the distinctive information was located at stimulus onset, stimulus onset asynchrony (SOA) is a more appropriate measure of temporal separation; the SOAs were 310, 460, and 1060 msec. "Same" latencies were significantly faster for physically identical stimulus pairs than for physically different pairs, but only at the two shorter SOAs. At the shortest SOA (310 msec), subjects actually detected the physical within-category difference on 22.8% of the trials, as compared to 2.8% at the 460-msec SOA. A partial replication of these results was obtained in a second study by Eimas and Miller (1975) with a /ra/–/la/ continuum. These findings provided rather striking support for a rapidly decaying auditory memory that after 460 msec no longer afforded conscious detection of within-category differences but still generated a reaction time difference that disappeared after 1060 msec. The fast decay of the memory relative to the 3-sec asymptote found in studies with vowels (see Section IV,A,2) may reflect the initial weakness of the auditory trace (i.e., the general auditory similarity of the stimuli in the set—cf. Darwin & Baddeley, 1974). It should be added that the data of Eimas and Miller, like those of Pisoni and Tash, did not yield any unambiguous evidence for any involvement of auditory memory in "different" judgments.

Negative results were obtained in two unpublished studies by Repp (1975, 1976b). Repp (1975) used /ba/–/pa/ stimuli similar to those of Pisoni and Tash

(1974) and presented them to different ears at a number of SOAs ranging from 0 to 3.3 sec. The listeners were given two types of instruction: Either they were told to make their same–different judgments on the basis of stimulus categories only (phonetic matching condition) or they were given some experience with the stimulus continuum (following the example of Pisoni & Lazarus, 1974) and then tried to make auditory same–different judgments (physical matching condition). The expected effect of physical mismatch on "same" latencies was only weakly present in the phonetic matching condition and did **not** systematically decline with SOA; it was totally absent in the auditory matching condition, where subjects, surprisingly, proved less sensitive to physical differences than in the phonetic matching condition. Thus, this study provided no evidence whatsoever for auditory memory. Perhaps presentation of the stimuli to different ears prevented the efficient use of auditory memory. In an attempt to examine this possibility, Repp (1976b) presented stimuli either binaurally or to different ears at one of two SOAs, 500 or 2000 msec. By using only four different stimuli (/bæ/, two versions of /dæ/, and /gæ/), Repp controlled for the effect of labeling uncertainty on reaction times, thereby making "different" latencies a potentially unconfounded indicator of auditory memory. However, the results of this study were entirely negative: There were no significant step-size effects in either "same" or "different" latencies.

Another study in the same vein, and the only one to be published, was conducted by Hanson (1977). Like Repp (1975), she used a /ba/–/pa/ continuum and two different sets of instructions (phonetic matching and physical matching). Unlike Repp, she presented her stimuli binaurally and had only two SOAs, 550 and 870 msec, which were varied between subjects. Although Hanson was successful in eliciting better discrimination performance through physical matching instructions (see Section VI,A,2), step-size effects were absent in the physical matching task and only weakly present in the phonetic matching task. Hanson's study must be viewed with caution because of high error rates and because it is the only study in the literature that failed to find a reaction time peak at the category boundary in a simple labeling task.

In summary, same–different reaction time studies have yielded some rather clear instances of listener sensitivity to within-category differences among stop-consonant stimuli, but there are also failures to obtain such effects. Although the causes of the negative findings remain obscure, the positive results do strengthen the hypothesis that all aspects of speech signals are represented in auditory memory.

2. Stimulus Duration

We turn now to a group of studies that attempted to either increase or decrease categorical perception by directly manipulating the stimuli, with the purpose of thereby modifying the strength of their auditory memory representations. One

manipulation that promised to have some effect was to vary stimulus duration. In the case of homogeneous stimuli, such as the steady-state vowels used in a number of experiments, a reduction in stimulus duration might weaken the auditory trace and thereby lead to more nearly categorical perception.

The first study to test this hypothesis was conducted by Fujisaki and Kawashima (1968). They presented vowels from an /i/–/e/ continuum (there is no /ɪ/ category in Japanese) in identification and ABX discrimination tasks, with stimulus duration set at either 25, 50, or 100 msec. A subsequent paper (Fujisaki & Kawashima, 1969) reports data from a similar experiment with shorter vowel durations—1, 3, or 6 pitch pulses, corresponding to durations of 8, 23, and 46 msec. Finally, Fujisaki and Kawashima (1970) presented what seem to be new data for single-pulse (8 msec) and 100-msec vowels. In all three reports, the figures show that discrimination performance was (paradoxically) higher for the short vowels, whereas the accompanying text consistently states the opposite. These inconsistencies in the Fujisaki–Kawashima papers were apparently not noticed by other authors concerned with the same issue: Pisoni (1971, 1973, 1975) paid attention only to the text, whereas Tartter (1982) paid attention only to the figures. In light of Pisoni's later findings, the only plausible explanation is that Fujisaki and Kawashima kept using incorrect figure legends and that their data really showed what they claimed to have found—namely, poorer discrimination and more nearly categorical perception of short vowel stimuli.

Pisoni (1971) investigated the matter more systematically. In his Experiment III, he presented short (50 msec) and long (300 msec) vowels from an /i/–/ɪ/ continuum in identification and ABX discrimination tasks. Although this preliminary study involved only five subjects, it did yield significantly (but not dramatically) higher discrimination scores for the long vowels. A replication with a larger number of subjects was reported by Pisoni (1975, Exp. I). Again, performance was slightly higher for the long vowels, but the difference reached significance only for one-step, and not for two-step, comparisons.

In another experiment, Pisoni (1971, Exp. IV) presented short (50 msec) and long (300 msec) vowels from an /i/–/ɪ/–/ɛ/ continuum in identification, ABX, and 4IAX tasks. Besides getting substantially higher and virtually continuous discrimination performance in the 4IAX paradigm, he also obtained consistent differences in favor of the long vowels, which were especially clear in the 4IAX test. A replication using an /i/–/ɪ/ continuum was conducted by Pisoni (1975, Exp. II), which again yielded sizeable effects of vowel duration (although they were, surprisingly, reported to be statistically nonsignificant).

Vowels of different duration were also used in Pisoni's (1971, Exp. VI; 1973) study of same–different discrimination at different temporal delays, and although there was little difference on between-category trials, performance for long vowels was clearly higher on within-category trials, where auditory memory was presumed to be the prime source of distinctive information. Similar

results were obtained by Sachs (1969), who used 150-msec and 250-msec /a/–/æ/ vowels in an absolute identification task. Tartter (1982), in a recent critical review, overlooked these data when she concluded that changes in vowel duration have equal effects across a vowel continuum and that, therefore, the dual-process model should be rejected. Although the data reviewed in the preceding two paragraphs indeed showed fairly uniform effects of vowel duration across a continuum, those just cited do support the dual-process model by showing that perception of short vowels is more nearly categorical (especially at long interstimulus intervals) than perception of long vowels. Because the gradual transitions between categories make it difficult to achieve a clear separation of between- and within-category pairs on a vowel continuum, the inconsistencies in the literature with regard to the uniformity or nonuniformity of performance decrements across a continuum can hardly justify the rejection of a model as conceptually sound as the dual-process model. It is possible, however, that the influence of phonetic categorization on vowel discrimination is more indirect than is generally assumed (see Section IV,A,2).

Vowel duration effects have also been obtained in verbal memory research: Crowder (1973a) found that the suffix effect was smaller for lists of short vowels than for lists of long vowels. It has also been reported that shortened vowels exhibit a right-ear advantage in dichotic presentation, whereas long vowels do not (Godfrey, 1974). All these results strongly suggest that auditory memory strength depends on the duration of a (homogeneous) stimulus.

A more radical modification of vowel duration was recently performed by Tartter (1981). She started with stimuli from an /ɪ/–/ɛ/ continuum 260 msec in duration, and obtained typical identification and oddity discrimination functions. Then she preceded the stimuli with 40-msec formant transitions appropriate for /b/. In one condition, the transitions for each vowel started at the same frequencies; in a second condition, they started at different frequencies that covaried with the vowel steady-state frequencies so that transition slopes remained constant. Neither manipulation had any effect on vowel discrimination—not an unexpected finding in view of the poor auditory memory for transitional cues on stop-consonant continua (e.g., Pisoni, 1971). In a subsequent condition, however, Tartter removed the vocalic steady states, leaving only the 40-msec transitional portions. The vowels were still identified quite accurately from these truncated /b/-vowel syllables, but discrimination performance suffered considerably. For both sets of transitions, perception was virtually categorical and the results exhibited the pattern typical for stop-consonant continua. This finding strongly suggests that rapidly changing acoustic information is poorly retained in auditory memory, regardless of whether it conveys consonantal or vocalic distinctions, and that the noncategorical perception of isolated vowels is due to their steady-state characteristics and their resulting salience in auditory memory, not to any special perceptual status of vowels as phonological segments.

This conclusion is further supported by the results of studies on the perception of vowels in context (Sachs, 1969; Stevens, 1968). The stimuli in these studies were not simply steady-state vowels embedded in some acoustic context (as they are sometimes described in the literature), but synthetic words with little (Sachs) or no (Stevens) steady-state vocalic portion. In Stevens's (1968) study, the continuum ranged from /bil/ (a nonsense word) to /bɪl/ and was obtained by interpolating between formant patterns obtained from natural utterances. Listeners actually perceived three categories (*beel, bill,* and *bell*) but, in an ABX test, showed sharp discrimination peaks at both category boundaries, indicating strongly categorical perception. A matched continuum of isolated steady-state vowels was included as control and yielded results typical of noncategorial perception.

Sachs (1969) employed a /bɑdəl/–/bædəl/ (or *bottle–battle*) continuum together with two matched steady-state /ɑ/–/æ/ continua of different durations. Measuring discrimination by computing d′ indices for pairs of adjacent stimuli from the results of an absolute identification task, he found a pronounced peak at the category boundary for the word continuum, a somewhat less pronounced peak for the short vowels, and even less of a peak for the long vowels. Although neither Stevens nor Sachs compared their discrimination data to predictions generated by the Haskins model, the pattern of their results suggests fairly categorical perception of vowels in word context. A recent study by Sawusch, Nusbaum, and Schwab (1980) yielded similar results. They used /i/–/ɪ/, /sis/–/sɪs/, and /bit/–/bɪt/ continua and obtained more nearly (though not completely) categorical results for the latter two. The fact that they observed no difference between the two context conditions, one of which merely put steady-state vowels in a fixed fricative-noise context whereas the other contained time-varying vocalic portions, suggests that auditory memory may be weakened by either dynamic change or by the presence of irrelevant context.

The finding of increased categorical perception for shortened or dynamically varying vowels suggests that the short duration and rapidly changing nature of the critical cues for initial stop consonants may be at least partially responsible for their categorical perception. One way to investigate this hypothesis with stop-consonant stimuli is to lengthen (and, thereby, also to slow down) the formant transitions that distinguish different places of articulation. This was done in two nearly simultaneous but independent studies by Dechovitz and Mandler (1977) and by Keating and Blumstein (1978). Dechovitz and Mandler extended the F2 and F3 transitions of a /ba/–/da/–/ga/ continuum from 30 to 135 msec. It was known from informal observations that a syllable with such extended transitions sounds rather similar to the original as long as the F1 transition remains constant. This impression was confirmed by the results of identification and same–different discrimination tests that showed no difference between the original and extended-transition stimuli: Perception of both sets of stimuli was strikingly categorical.

Keating and Blumstein (1978) used a /da/–/ga/ continuum with three lengths of F2 and F3 transitions (45, 95, and 145 msec). The three sets of stimuli yielded similar results in identification and 4IAX discrimination tests although there were some significant differences, primarily due to the stimuli with intermediate transition length, which were discriminated best. Within-category discrimination in this study was significantly better than predicted (perhaps due to the sensitive 4IAX paradigm), particularly with the longer transitions. Therefore, the Keating and Blumstein results are not entirely negative, but they do suggest that the short duration of F2 or F3 transitions is not a major determinant of categorical perception.

A very interesting result was recently reported by Tartter (1981). She removed the steady-state vocalic portions of /ba/–/da/ stimuli, leaving only the initial 40 msec that contained the formant transitions. Compared to the full syllables, this resulted in a distinct improvement in within-category discrimination (an oddity task was used), whereas stop-consonant identification was just as accurate as when the steady states were present. This finding strongly suggests that the formant transitions have a representation in auditory memory that can be accessed when the redundant steady state is eliminated. Thus, the vocalic portion of a stop-consonant–vowel syllable, although it aids phonetic perception, appears to interfere with the preservation of consonantal cues at a precategorical level. The overriding auditory salience of an irrelevant stimulus portion may be a major factor causing categorical perception.

3. Other Stimulus Parameters That May Affect Categorical Perception

One parameter that generally has received little attention in speech-perception research is amplitude. However, recent studies by Syrdal-Lasky (1978), Dorman and Dougherty (1981), and Van Tasell and Crump (1981) have shown that the identification of synthetic stop consonants varying along a place of articulation continuum may exhibit large shifts with changes in playback level. Syrdal-Lasky also presented her stimuli in an oddity discrimination task and found different discrimination functions at different signal levels. However, it seems from an inspection of her figures that, if the changes in labeling probabilities are taken into account, perception was about equally categorical in all conditions. It is tempting to speculate that auditory discrimination along some physical dimension might be improved when that dimension is highlighted by increasing its amplitude relative to nondistinctive signal components. However, so far there are no data pertaining to this hypothesis.

Another parameter that does not seem to have much effect on categorical perception is whether a stimulus is periodic or aperiodic, other things being equal. Fujisaki and Kawashima (1968) synthesized an /i/–/e/ continuum with either periodic or aperiodic excitation. There was a shift in the category boundary (more /i/ responses were given to the aperiodic vowels), and ABX discrimina-

tion functions showed a corresponding peak shift but did not differ in overall level. Highly similar (though not completely identical) data were reported by Fujisaki and Kawashima (1969). Thus, periodicity, like overall amplitude, seems to affect categorical perception only to the extent that labeling probabilities are affected; these variables do not seem to have any direct influence on the strength of the auditory trace. This conclusion was further supported by a recent study by May and Repp (1982), who failed to find any difference in auditory memory for periodic and aperiodic nonspeech stimuli (single-formant resonances).

One stimulus factor that has not been systematically investigated but may well play a role in categorical perception is naturalness. Poorly synthesized stimuli may be expected to be less categorically perceived (given that they are sufficiently distinct acoustically) than good synthetic stimuli or natural speech. The reason for this is that poor stimuli may make it easier for listeners to adopt auditory strategies in discrimination, whereas highly realistic stimuli may elicit a phonetic strategy (more about strategies in Section VI,A).

B. Different Classes of Speech Sounds

The large majority of studies concerned with categorical perception and related topics have used as materials either the two standard sets of prevocalic stop consonants (VOT or place of articulation continua) or isolated steady-state vowels. In this subsection, we will review studies that examined other types of speech contrasts or used less common varieties of stop-consonant or vowel continua. We will pay some attention to the specific stimulus parameters that were varied to obtain a continuum, as these may have a bearing on the strength of the auditory memory trace.

1. Stop Consonants

a. Voicing Continua. The earliest voicing continua were generated on the Haskins Laboratories Pattern Playback by the procedure called "F1 cutback"— increasing delays in the onset of F1 relative to the onsets of the higher formants. Perception of these stimuli was highly categorical (Liberman, Harris, Kinney, & Lane, 1961). During the following years, Abramson and Lisker developed the now commonly used procedure for varying VOT, which combines a delay in the onset of F1 with the substitution of aperiodic for periodic energy in the higher formants during the period of the delay. These stimuli, too, show highly categorical perception in the standard experimental setup (Abramson & Lisker, 1970; Lisker & Abramson, 1970). The original Abramson–Lisker stimuli, which have been used in many different studies, included variations in VOT on the "negative" side: Different degrees of **prevoicing** were simulated by preceding the stop release with varying amounts of low-energy buzz from the periodic source of the

synthesizer. This region of the continuum is of interest because prevoicing is not distinctive in English (and native speakers of English are very poor in discriminating differences in prevoicing—cf. Abramson & Lisker, 1970) although it is distinctive in some other languages (see Section VI,B).

In acoustic terms, the Abramson–Lisker VOT continuum is really not one continuum, but two: The acoustic variations used to achieve different degrees of prevoicing (voicing lead) are quite different from those used to generate different degrees of aspiration (voicing lag). On the "positive" side, as increasing amounts of aspiration are substituted for voicing, there is at first a correlated spectral change as the F1 transition (always rising) is cut back more and more, so that the onset of F1 occurs at increasingly higher frequencies and amplitudes. Spectral cues, particularly from the F1 region, are relevant to the perception of voicing, as several studies have shown (Lisker, Liberman, Erickson, Dechovitz, & Mandler, 1977; Stevens & Klatt, 1974; Summerfield & Haggard, 1977). As voicing onset is delayed beyond the region of the formant transitions (the first 30–70 msec), the spectral covariation ceases, but the duration of the periodic portion decreases as the aspirated position increases. This negative covariation has been given little attention in the past, although it may play a role when VOTs get rather long and the periodic portions get short enough for the temporal variations to exceed the detection threshold (cf. Wood, 1976a). An alternative, and perhaps preferable, way of synthesizing VOT continua in the long positive range would be to hold the duration of the periodic portion constant (cf. Repp, 1981b).

A procedure for generating VOT continua (in the positive VOT range) by cross-splicing pitch periods and aspiration from natural-speech tokens was devised by Lisker (1976) and described in detail by Ganong (1980). There is little doubt that such stimuli are perceived categorically: Repp (1981b, Exp. 3) presented stimuli from a natural-speech VOT continuum in a fixed-standard AX task and obtained extremely poor within-category discrimination performance.

The highly categorical perception of stop-consonant voicing in initial position may be contrasted with the less categorical perception of the same phonetic distinction in final position. This comparison is important because it shows that categorical perception is not only a function of phonological status, but also of the acoustic stimulus dimensions varied. One important cue for consonant voicing in postvocalic position (in English) is the duration of the vocalic portion. Using variations in vowel duration to generate a variety of voiceless–voiced continua (including final fricatives and stop-fricative clusters as well as final stops), Raphael (1972) found that oddity discrimination was much better than predicted, given a sufficiently large physical difference. There also appeared to be a discrimination peak at the category boundary, making the data similar to those typically obtained with isolated vowels. Although there have been numerous studies of the various cues to the voicing distinction in postvocalic position, Raphael's remains the only study to date that includes discrimination tests.

The voicing contrast for stops in intervocalic position may be cued by variations in the duration of the (silent) closure interval. Liberman, Harris, Eimas, Lisker, and Bastian (1961) synthesized a /ræbɪd/–/ræpɪd/ continuum in this way and presented it in identification and ABX discrimination tasks. The results provided an interesting instance of perception that was neither very categorical nor very continuous: Discrimination performance was considerably better than predicted, but showed a peak at the boundary. A second peak was noted within the "p" category and was attributed to subjects' use of a covert third category, "unnatural 'p.'" However, even revised predictions based on three categories did not reach the level of the obtained discrimination performance. Here is a case, it seems, where the contribution of phonetic and auditory processes to discrimination was in approximate balance.

b. Place of Articulation Continua. Early studies used two-formant stimuli in which the F2 transition was the sole cue to place of articulation (Liberman *et al.,* 1957; Mattingly *et al.,* 1971). Despite the relative crudeness of the stimuli, the perception of these syllable-initial stops was invariably quite categorical. Later experiments in which stimuli also had a varying F3 transition yielded similar results (e.g., Pisoni, 1971). Numerous studies have employed variants of /b/–/d/–/g/ continua, and the categorical discrimination of these stimuli is one of the most consistently replicated results in speech-perception research, notwithstanding Barclay's (1972) findings (see Section IV,B). All of these studies used formant transitions as the sole cue to place of articulation; so far, the discriminability of variations in release burst spectrum (another important cue for stop-consonant place of articulation) has not been tested. In addition, there are very few studies that have employed continua of voiceless stops (/p/–/t/–/k/). What data there are (Syrdal-Lasky, 1978, used F1 cutback without aspiration) suggest categorical perception.

Syllable-final stops varying in place of articulation were synthesized by Mattingly *et al.* (1971) by varying the final F2 transition in two-formant stimuli (/ab/–/ad/–/ag/). The oddity discrimination function for these sounds showed no clear peaks at phonetic boundaries, which the authors attributed to the poor quality of the stimuli. Subsequently, Popper (1972) found a well-defined peak on an /ab/–/ad/ continuum, but within-category same–different discrimination was better than predicted by the Haskins model. Recently, J. L. Miller, Eimas, and Zatorre (1979) obtained similar results with /ab/–/ad/ stimuli in an oddity discrimination task: There was a discrimination peak at the category boundary, but also unexpectedly high performance within the /ad/ category, which the authors were unable to explain. Taken together, these results suggest that syllable-final stops are not perceived as categorically as syllable-initial stops. One likely reason is that the distinctive information, being in final position, is better retained in auditory memory (cf. the importance of offset frequency in determining the pitch of nonspeech frequency glides—e.g., P. T. Brady, House, & Stevens, 1961;

Schwab, 1981). However, one study that directly compared initial and final stops (Larkey *et al.*, 1978) using stimuli that were acoustic mirror images, found equally categorical perception for both.

c. Manner Continua. One primary cue for the perceived presence or absence of a stop consonant in medial position is the presence or absence of an appropriate closure interval. Bastian *et al.* (1961) constructed a continuum from /slɪt/ to /splɪt/ by inserting increasing amounts of silence after the /s/ noise of a natural-speech token of /slɪt/. The stimuli were presented in identification and oddity discrimination tasks, and the listeners' responses proved to be highly categorical, with obtained discrimination scores only slightly exceeding the predictions of the Haskins model. These results were essentially replicated in a recent study by Fitch *et al.* (1980) with a synthetic /slɪt/–/splɪt/ continuum, although these authors did not conduct a direct comparison of predicted and obtained discrimination scores. Even more recently, Best *et al.* (1981) presented a synthetic /sei/–/stei/ continuum, generated similarly by varying silent closure duration, in oddity and same–different tasks; Best *et al.* also computed the Haskins model predictions. The discrimination functions showed pronounced peaks at the category boundary, but performance in both tasks was a good deal better than predicted, particularly within categories. Thus, in this study, the listeners did seem to pick up some auditory differences. Moreover, Repp (1981b) recently obtained rather good within-category discrimination of closure duration differences in /splɪt/ and /stei/ stimuli in a fixed-standard AX task.

A related stop manner contrast is that between a fricative and an affricate (effectively, stop + fricative). In intervocalic position, this difference may be cued by silence preceding the fricative noise (e.g., Gerstman, 1957). By employing stimuli from a *say shop–say chop* continuum in a fixed-standard AX discrimination task, Repp (1981b) obtained fairly high within-category discrimination, which adds to the mounting evidence that within-category differences in temporal stimulus structure are detected more readily than differences in spectral structure. Another way of cueing the fricative–affricate distinction is by means of fricative noise duration (Gerstman, 1957), but no discrimination data for this cue are in the literature. A third important cue is the amplitude rise time of the noise; this cue has been investigated in initial position by Cutting and Rosner (1974, 1976). They generated synthetic /tʃa/–/ʃa/ and /tʃæ/–/ʃæ/ continua by varying the rise time of the fricative noise, and presented the stimuli in identification and ABX discrimination tasks. The results showed fairly categorical perception, even though fricative noise duration apparently covaried with rise time.

2. Nasal Consonants

Nasal consonants are relative latecomers in categorical perception experiments because it took some time before convincing nasals could be produced synthetically. Initial studies by Garcia (1966, 1967a, 1967b) still suffered from stimulus

problems. Garcia (1966) converted a two-formant /bɛ/–/dɛ/–/gɛ/ continuum into a /mɛ/–/nɛ/–/ŋɛ/ continuum by simply preceding the stimuli by a constant synthetic nasal murmur. An /ɛm/–/ɛn/–/ɛŋ/ continuum was obtained by playing the stimuli backwards. It turned out that the nasals were labeled rather poorly, especially in initial position. Discrimination performance was also rather poor, but it did show some evidence of peaks at category boundaries for subjects who labeled the final nasals consistently. Somewhat more consistent data were obtained in a replication with three-formant stimuli (Garcia, 1967a, 1967b); they suggested fairly categorical perception.

Much cleaner results were obtained by J. L. Miller and Eimas (1977), who compared a /ba/–/da/ with a /ma/–/na/ continuum obtained by adding initial nasal resonances and by flattening the F1 transition. Although the nasal categories were not quite as sharply separated as the stop categories, discrimination of both stimulus sets was equally categorical in an oddity task, with obtained scores only slightly better than predicted. A careful replication of Garcia's work was undertaken by Larkey et al. (1978), who not only used all three nasal categories in initial and final position (with the vowel /æ/), but also compared their perception with that of matched stop-consonant continua. The results showed highly categorical perception of all stimulus sets, with somewhat better within-category discrimination for final than for initial nasals. In the meantime, Miller and Eimas also extended their study to syllable-final nasals (J. L. Miller et al., 1979) and obtained categorical perception, except for high levels of discrimination within the /n/ category. In view of the Larkey et al. data, this is likely to have been a stimulus artifact of some sort.

Given the consistently categorical results for both stop consonants and nasals, the results of experiments using stop–nasal (oral–nasal) continua would seem highly predictable. Yet these studies are not trivial, for the acoustic dimension cueing the oral–nasal distinction (amplitude or duration of nasal resonance) is considerably less complex and, therefore, perhaps more readily discriminable than the spectral changes cueing place of articulation distinctions. Thus, oral–nasal continua offer an opportunity for noncategorical perception, even though the phonetic boundary may coincide with the auditory detection threshold for the presence of nasal murmur. The first study was conducted by Mandler (1976), who synthesized /ba/–/ma/ and /da/–/na/ continua by two different methods, using either the oral branch or the nasal branch of a serial resonance synthesizer. In each case, the amplitude of the simulated nasal resonance was varied in a number of steps. The labeling functions for these continua were not very steep, but same–different discrimination scores showed a peak in the boundary region, suggesting categorical perception.

Rather similar results were obtained by J. L. Miller and Eimas (1977) for synthetic /ba/–/ma/ and /da/–/na/ continua obtained by simultaneously varying the duration of nasal murmur and F1 onset frequency (which is higher for nasal

than for oral stops). Again, labeling functions were rather gradual, but oddity discrimination functions exhibited peaks. Discrimination was somewhat better than predicted. (An unusually high level of discrimination performance in comparisons involving the most stoplike stimulus was traced to a stimulus artifact and eliminated in a supplementary experiment, described in the same paper.) Equally categorical perception was found for syllable-final /ab/−/am/ and /an/−/ad/ continua (acoustic mirror-images of the original stimuli) by J. L. Miller *et al.* (1979).

A possibility suggested by the motor theory of speech perception is that categorical-like perception might be caused by a nonlinear relation of an acoustic continuum to changes along the corresponding articulatory dimension. In the case of the oral−nasal distinction, this problem was addressed by Abramson, Nye, Henderson, and Marshall (1981), who created a /da/−/na/ continuum on an articulatory synthesizer by directly controlling the degree of velar opening. The amplitude of nasal murmur was determined to be a negatively accelerated function of the velopharyngeal port area, which was varied in equal steps. Although the category boundary was once again not very sharp, AXB discrimination functions showed clear peaks that unmistakably pointed toward categorical perception, even though no predictions were calculated (however, see Beddor & Strange, 1982). Thus, the observed nonlinear relation between articulation and acoustic output was not responsible for categorical perception in this instance.

3. Liquids and Semivowels

In a study primarily intended to demonstrate effects of linguistic experience (see Section VI,B), Miyawaki *et al.* (1975) synthesized a /ra/−/la/ continuum by varying the onset frequency of F3, which, in this instance, had an initial 50-msec steady state followed by a 75-msec transition. American listeners perceived the stimuli fairly categorically: Oddity discrimination scores showed a clear peak at the boundary, but within-category discrimination was significantly better than predicted, particularly within the /la/ category. Clearly, perception was less categorical than that of stop consonants. McGovern and Strange (1977) subsequently conducted experiments with synthetic, mirror-image /ri/−/li/ and /ir/−/il/ continua and obtained results very similar to those of Miyawaki *et al.* So did MacKain *et al.* (1981) with a /rak/−/lak/ continuum in AXB and oddity discrimination tests.

Fujisaki and Kawashima (1970) obtained a (Japanese) /wa/−/ra/ continuum by varying the frequency of the (rather slow) F2 transition. The ABX discrimination functions showed a broad peak at the category boundary—considerably broader than predicted. Thus, perception of this continuum was not highly categorical. More nearly categorical results were obtained by Frazier (1976), who synthesized an acoustic continuum from /wɛ/ to /lɛ/ to /yɛ/ by varying the initial steady state (90 msec) and transition (60 msec) of F2. A mirror-image

/ɛw/–/ɛl/–/ɛy/ continuum was also used. The stimuli were presented in identification and same–different discrimination tests at two different ISIs (57 msec and 1 sec). The results revealed highly categorical perception in all conditions. The ISI seemed to have no effect on performance.

J. L. Miller (1980) has reported essentially categorical perception of stimuli from a stop–semivowel continuum (/ba/–/wa/) obtained by varying the duration of the initial formant transitions (J. L. Miller & Liberman, 1979). This study also demonstrated a shift in the discrimination peak along with a shift in the category boundary when the duration of the steady-state vocalic portion was extended. (However, this shift may have a purely psychoacoustic reason—see Carrell, Pisoni, & Gans, 1980.) More recently, Godfrey and Millay (1981) found somewhat less categorical perception of a /bɛ/–/wɛ/ continuum due to rather high discrimination scores within the /b/ category.

4. Fricatives

Fricative consonants offer a better opportunity for noncategorical perception than any speech sounds discussed so far in this section. Fricative vowel stimuli contain a noise portion that is nearly homogeneous, lasts for 100 msec or more, and has a characteristic pitch. Moreover, stimuli along a synthetic fricative continuum tend to be rather widely spaced, so that even one-step differences should exceed the auditory detection threshold.

The first categorical perception study with fricatives was conducted by Fujisaki and Kawashima (1968). They synthesized a /ʃ/–/s/ continuum by varying the frequencies of two fricative poles (formants) and presented these noises either in isolation or followed by a vowel (probably /e/—cf. Fujisaki & Kawashima, 1970). The ABX discrimination results were rather variable and showed fairly good within-category discrimination, especially at the /ʃ/ end, but there was also a peak at the category boundary. The vocalic context depressed discrimination scores somewhat without changing the shape of the discrimination function. Fujisaki and Kawashima (1969) report slightly different data from the same experiment. (Perhaps subjects had been added.) However, there was no consistent effect of vowel context. Finally, Fujisaki and Kawashima (1970) display yet another set of data, again showing peaks at the boundary, but with **better** within-category discrimination in vocalic context. Thus, although the effect of context is not clear at all, the data consistently show moderately categorical perception of fricative noises in context and in isolation. The finding for isolated noises contrasts starkly with results obtained by Healy and Repp (1982), who found discrimination in a same–different task to be essentially continuous. However, Healy and Repp used larger step sizes than Fujisaki and Kawashima, and a ceiling effect may have obscured a possible discrimination peak at the boundary. The high scores achieved by subjects at larger step sizes show quite clearly, however, that acoustic differences between isolated fricative noises are not diffi-

cult to detect (cf. also Repp, 1981c). The perception of these stimuli appears to be at least as noncategorical as that of isolated vowels.

Fricatives in vocalic context also have yielded conflicting results. A dissertation by Hasegawa (1976) examined noises from a /ʃ/–/s/ continuum in postvocalic position following either /i/ or /u/. The subjects were first given considerable training in ABX discrimination of vowels. Their fricative discrimination was essentially continuous; there was not even a hint of a peak at the category boundary. May (1981), on the other hand, obtained fairly categorical perception for three fricative continua presented to Egyptian listeners in a 4IAX paradigm. The continua ranged from /ʃ/ to /s/, from /x/ to /ħ/, and from /ɣ/ to /ʕ/, always in intervocalic context (/ə-ə/). Although discrimination performance was better than predicted, all three continua showed a discrimination peak at the boundary. Repp (1981c) recently synthesized /ʃa/–/sa/ and /ʃu/–/su/ continua and presented them in AXB and fixed-standard AX tasks. In both tasks, the majority of subjects perceived the stimuli quite categorically: Although within-category discrimination was better than predicted, the peaks at the category boundary were extremely pronounced. However, there were some subjects whose discrimination scores were far superior and probably continuous. (A ceiling effect prevented any peaks from appearing.) These subjects apparently followed a radically different perceptual strategy (see Section VI,A for further discussion). Fricative stimuli seem to be especially suited for the application of different strategies so that they may be perceived fairly categorically in one situation but continuously in another. This may explain the conflicting results in the literature.

5. Vowels

Most of the vowel studies in the literature have already been reviewed in Section IV or will be reviewed in Section VI. We note here that the finding of a discrimination peak at the category boundary is the rule rather than the exception; the earliest study by Fry *et al.* (1962) is one of the few that did not find a peak. We also note that most studies used continua of high front vowels (the /i/–/e/ range). The instability of vowel category boundaries and the magnitude of context effects in labeling may be due, in part, to the inclusion of categories, such as /ɪ/, that do not normally apply to isolated vowels (cf. Strange, Edman, & Jenkins, 1979). Although the primary reason for the noncategorical perception of isolated vowels is undoubtedly their inherent high discriminability and good auditory retention, it is also true that the acoustic homogeneity that confers these perceptual advantages is not very typical of vowels in natural speech. Thus, in addition to favoring an auditory mode of processing, isolated vowels, by their very unnaturalness, may discourage phonetic processing and, in extreme cases, lose their speechlike quality altogether.

It remains for us to mention some categorical perception studies that varied

properties of vowels other than their phonetic quality. One such property is duration, which carries some distinctive phonetic information in English, but much more in certain other languages, such as Thai. Bastian and Abramson (1964) created a continuum from /baat/ to /bat/ (meaningful words in Thai) by removing pitch pulses from the center of a natural token of /baat/. Oddity discrimination scores were quite continuous for both Thai and American listeners, showing no evidence of a phoneme boundary effect. These results were further confirmed in a vocal imitation task where the duration of the responses was found to be a nearly linear function of the duration of the stimuli. (Thai subjects did show a slight effect of categorization here, but since Bastian and Abramson did not dwell on it, it was probably nonsignificant.) We have already mentioned (Section V,B,1) the study by Raphael (1972), who showed that variations in vowel duration are not categorically perceived even when they cue a consonantal distinction (final consonant voicing).

Another property of vowels that carries phonemic significance in many languages, but not in English, is their pitch contour. Thai, for example, has five distinctive tones. Abramson (1961) generated a synthetic continuum between two of these on the fixed carrier /naa/. The ABX discrimination results provided some evidence for a phoneme boundary effect in Thai listeners, but the results rested on a comparison of Thai and American listeners since stimulus problems prevented a direct interpretation of discrimination functions. A subsequent study by Chan, Chuang, and Wang (see Wang, 1976) found evidence of a category boundary effect for Chinese subjects listening to a continuum of Mandarin tones. The effect disappeared, however, after practice in ABX discrimination. Abramson (1979) reinvestigated the issue using a new continuum of Thai tones that consisted simply of flat frequency contours varying in level. The 4IAX discrimination of these stimuli by Thai listeners was entirely continuous. Taken together, these three studies suggest that moving pitch contours, but not static frequency levels, may elicit a tendency toward categorical perception.

6. Summary

A brief summary is in order after reviewing so many different studies. It is evident that the large majority of experiments obtained results consistent with categorical perception. Thus, categorical perception is not only characteristic of stop consonants, but also of nasals and, to some lesser degree, of liquids, semivowels, and fricatives. The perception of liquids, semivowels, and fricatives is clearly less categorical than that of stops, and that of fricatives, at least, may become entirely continuous under certain conditions. Vowels, too, show a phoneme boundary effect in most conditions, and may even be perceived fairly categorically when embedded in context. Indeed, there are few experiments in the literature that present conclusive evidence for perfectly continuous discrimination of a speech continuum.

C. Perception of Nonspeech Stimuli

From the very beginnings of categorical perception research, the comparison of speech and nonspeech stimuli has been of central interest. Initially, the purpose of these comparisons was to determine whether categorical perception was due to "acquired similarity" of different sounds from the same category (in which case nonspeech discrimination should be easier than within-category speech discrimination), "acquired distinctiveness" of sounds from different categories (in which case between-category speech contrasts should be easier to discriminate than nonspeech), or both (e.g., Liberman, Harris, Eimas, Lisker, & Bastian, 1961). As interest in this issue faded (Mattingly et al., 1971), it was replaced by a search for possible psychoacoustic bases of linguistic category boundaries and discrimination peaks. This required nonspeech stimuli as similar as possible to the speech stimuli with which they were to be compared, but sufficiently dissimilar so as not to elicit speechlike percepts. Finding the right balance between these two requirements has been a major (and, perhaps, insurmountable) methodological obstacle.

1. Perception of Continua Unrelated to Speech

In the early stages of categorical perception research, it was important to make sure that perception of simple nonspeech continua was really continuous in the standard categorical perception paradigm. It seemed possible, after all, that categorical perception was an artifact of the procedures used, which differed in certain respects from those of psychophysical research.

An appropriate comparison was undertaken by Eimas (1963). He included, along with vowel and stop-consonant continua, a continuum of noise bursts varying in duration and a visual continuum of different levels of reflectance (Munsell grey scale). Both nonspeech continua were presented in labeling and ABX tests. The labels were "long" or "short" for the noises, and "light," "medium," or "dark" for the visual stimuli. Although both nonspeech continua were consistently labeled by the subjects, discrimination was far better than predicted and quite continuous. Thus, discrimination of the nonspeech stimuli was clearly not limited by categorization but, since discrimination scores were at or near the ceiling, Eimas did not provide a strong test of whether labels can have any influence on nonspeech discrimination.

Indeed, Cross et al. (1965), by employing a visual continuum of sectored circles, found results not unlike categorical perception. Their subjects were first trained to give verbal labels to the stimuli. A subsequent ABX discrimination test revealed a clear peak at the category boundary. However, discrimination of within-category contrasts was considerably better than predicted on the basis of labeling performance, so that the data showed only "a degree of categorical perception typical of vowels" (Studdert-Kennedy et al., 1970, p. 242), not of

stop consonants. Unfortunately, two independent replications of the Cross *et al.* study failed to find similar effects. Liberman, Studdert-Kennedy, Harris, and Cooper (1965), in a detailed critique of Cross *et al.*, reported that they could not find any discrimination peaks before or after categorization training. It may be countered that they provided less formal training and that discrimination performance was too high to reveal any peaks. However, a second, almost exact replication of Cross *et al.* by Parks, Wall, and Bastian (1969) revealed no consistent category boundary effects and no influence of categorization training.

More recently, Pastore (1976) also reported a failure to obtain a discrimination peak at the "alternation" versus "movement" boundary for the visual Phi phenomenon (two lights alternating at varying rates). However, Kopp and Udin (1969) and Kopp and Livermore (1973) found a clear discrimination peak (in ABX and same–different tasks, respectively) on a continuum of pure tones varying in frequency, following classification training (see Vinegrad, 1972, for corresponding results in a magnitude scaling study). Kopp and Livermore performed a signal detection analysis of their data and found that the discrimination peak was entirely due to response bias, so that an unbiased measure of sensitivity was constant across the whole continuum. This finding contrasts with Wood's (1976a, 1976b) similar analyses of stop-consonant discrimination, which showed both bias and sensitivity changes to contribute to the phoneme boundary effect (cf. also Elman, 1979; Popper, 1972).

Healy and Repp (1982) recently constructed a nonspeech continuum consisting of brief, steady-state, single-formant resonances varying in frequency (timbre). The stimuli were presented in same–different and labeling tasks whose order was counterbalanced. Prior labeling experience did not seem to have any effect on discrimination performance, which exhibited a peak at the category boundary.

The data just reviewed suggest that category labels may influence nonspeech discrimination under certain circumstances. We might expect these circumstances to be those that make it difficult to rely on auditory memory—that is, when the differences to be detected are small to begin with. A role for some form of categorical encoding in discrimination is also predicated by the psychophysical dual-coding theory of Durlach and Braida (1969). In all nonspeech studies mentioned, however, within-category discrimination was substantially better than predicted by the Haskins model; perception was never truly categorical.

The studies discussed so far looked for category boundary effects on obviously continuous physical dimensions; therefore, if such effects were found, they must have been due either to response bias introduced by the subjects' category labels or to procedural artifacts. On the other hand, some recent studies have demonstrated category boundary effects on continua that straddle a psychophysical threshold. These findings are hardly surprising; the point of these studies was, however, to lend plausibility to the hypothesis that boundary effects on speech

continua might likewise be caused by psychophysical discontinuities, not by categorization per se.

Some pertinent data were reported by Pastore *et al.* (1977). In one experiment, they flashed a light at various rates centered around the flicker fusion threshold. The subjects were able to label the stimuli consistently as "flicker" or "fusion," and ABX discrimination results showed a peak at the boundary and poor discriminability within categories. In a second experiment intended to have some relevance to speech perception, Pastore *et al.* varied the intensity of a pure tone that alternated with a constant reference tone of the same frequency. The ABX discrimination scores showed a peak at the boundary between the two (arbitrary) categories used by subjects in the labeling task. In a control condition, the reference tone was omitted and the discrimination peak disappeared. Pastore *et al.* mention, however, that they failed to replicate these results using noise stimuli, and their data for tones seem fairly variable. For these reasons, the claim of Pastore *et al.* that a fixed reference stimulus generates a sharp boundary and a corresponding discrimination peak must be accepted with caution. It is also clear from their discussion that good within-category discrimination would have been possible at larger step sizes, so that perception was not truly categorical.

In all the cases discussed in this subsection, the categories were not particularly familiar and were sometimes even arbitrary. This is also true for the majority of the various nonspeech analogs of speech, to be discussed next. However, there are also nonspeech domains associated with highly overlearned categories; two of them (color and music) will be considered in the final subsection (V,C,5).

2. Nonspeech Analogs of Voice Onset Time (VOT)

The primary cue for the voicing distinction in initial stop consonants is temporal—the delay of the onset of voicing relative to the stop release. On the positive (voicing lag) side, this temporal delay results in correlated spectral changes: The interval prior to voicing onset is filled with aperiodic noise (except in the earliest studies, where only F1 cutback was manipulated), there is no energy in the region of the first formant before the onset of voicing, and at voicing onset, the formants (F1 in particular) start at frequencies close to those of the following vocalic portion. These spectral correlates of VOT all are relevant to the perception of the voicing distinction, but most studies have focused on the temporal aspect of VOT only.

The first attempt to devise nonspeech analogs of VOT was undertaken by Liberman, Harris, Kinney, and Lane (1961). They synthesized a /do/–/to/ continuum by delaying the onset of F1 in varying amounts. A matched nonspeech continuum was obtained by playing the stimuli with the frequency scale inverted so that F1 was in the region previously occupied by F3 and vice versa. (This was literally possible on the Haskins Laboratories Pattern Playback.) In addition, the initial transition of the new F1 (previously F3) was modified to assure that the

stimuli would not sound speechlike. Although ABX discrimination of the speech stimuli was highly categorical, that of the nonspeech stimuli was extremely poor and barely exceeded chance even at the largest step size used. In other words, speech discrimination was vastly superior to nonspeech discrimination. Liberman *et al.* interpreted this finding as evidence for the acquired distinctiveness (rather than acquired similarity) of speech sounds. They did acknowledge, however, that there were a number of differences between speech and nonspeech stimuli that may have been responsible for the poor performance with the latter.

Liberman *et al.* did not ask their subjects to label the nonspeech stimuli. Lane and Schneider (1963, cited in Lane, 1965) found that some subjects could be trained to label them as accurately as the speech stimuli. In a subsequent ABX test, these subjects produced above-chance discrimination scores with a peak at the boundary. This report was questioned, however, by Studdert-Kennedy *et al.* (1970), whose detailed examination of the Lane and Schneider data revealed that they were extremely variable and hardly conclusive. Studdert-Kennedy *et al.* also reported a failure to replicate the results with five subjects, none of whom could be trained to label the nonspeech stimuli in a consistent way.

The /do/–/to/ control stimuli may have been too complex for listeners to detect the relevant differences without extensive training. Later studies used stimuli of a simpler acoustic structure. Hirsh's (1959) finding of a threshold in the vicinity of 20 msec for determining the temporal order of two auditory events stimulated the thought (Liberman, Harris, Kinney, & Lane, 1961) that this threshold might be related to the category boundary on a VOT continuum. This suggestion makes good sense when applied to speech stimuli generated by the method of F1 cutback, where the onset of low-frequency energy may indeed either precede or follow the onset of high-frequency energy. However, it loses some of its appeal when aspiration enters the scene (as it does in more sophisticated—and more appropriate—VOT synthesis), for aspiration always precedes the onset of voicing and provides a powerful cue to the voicing distinction. It has also been long known that VOT boundaries tend to be at rather longer onset asynchronies (especially for alveolar and velar stops) than the temporal order threshold (Lisker & Abramson, 1970). Nonetheless, a good deal of research has been generated by this presumed analogy.

Stevens and Klatt (1974) synthesized stimuli consisting of a 5-msec broadband noise burst followed by a variable silent interval and steady-state formants roughly appropriate for the vowel /ɛ/. According to these authors, "none of the stimuli could be readily interpreted as speech events" (Stevens & Klatt, 1974, p. 654). Listeners were asked to label the stimuli according to whether or not they heard a silent interval between the noise and the vowel. The category boundary fell at about 20 msec of "VOT" (measured from the onset of the burst), which matched the time obtained by Hirsh (1959) with tones. However, no discrimina-

tion data were obtained for these stimuli, and their analogy to VOT in speech may be questioned because of the absence of aspiration noise. Their relation to Hirsh's findings is equally doubtful, for the task did not require temporal order judgments but detection of a gap.

These objections do not apply equally to a subsequent study by J. D. Miller *et al.* (1976), who presented white noise and a square-wave buzz at varying noise–buzz lead times in labeling ("no-noise" versus "noise") and oddity discrimination tasks. The listeners were experienced in psychoacoustic experiments. Their category boundaries varied widely (from 4 to 31 msec of noise lead time), but they showed clear discrimination peaks that, in all cases but one, coincided with the boundary. Control results obtained with isolated noises did not reveal any discrimination peaks. Miller *et al.* compared their results with those of Abramson and Lisker (1970) for VOT and found a striking similarity of the average discrimination functions. However, they neglected to point out that at least three of their eight listeners had category boundaries at substantially shorter values of noise lead time (4–8 msec) than are ever obtained with speech stimuli varying in VOT. Such a wide range of individual differences in boundary locations is quite atypical of speech and presumably reflects variations in auditory acuity or response criteria since all listeners were quite experienced. Therefore, although Miller *et al.* have shown (as have Pastore *et al.*, 1977) that results resembling categorical perception can be obtained with nonspeech stimuli straddling a psychophysical threshold, they have not presented a convincing case for any direct correspondence of the category boundaries in speech and nonspeech.

Of course, it could always be argued that the supposed nonspeech analogs of VOT simply fell short of the mark. As we have pointed out, if the analogs are made too speechlike, there is the danger that they are perceived as speech. Wood (1976a) accepted this risk when he decided simply to excise most of the steady-state vowels of stimuli from a /ba/–/pa/ continuum (ranging from −50 to +70 msec of VOT) and to use the initial 120 msec as nonspeech analogs. According to Wood, who interviewed his subjects carefully, these truncated stimuli were not spontaneously categorized as (or even recognized as being related to) speech. (They were not presented for identification at all.) Same–different discrimination results for full and truncated syllables were similar at short VOTs, but at long VOTs, the scores for the truncated stimuli were rather high, which obscured the discrimination peak that may otherwise have been obtained. Most likely, the reduction in the duration of the periodic portion with increasing VOT became detectable at long VOTs in the truncated stimuli. Wood also mentions that identical results were obtained in a subsequent, unpublished experiment, in which subjects were instructed to hear the short syllables either as speech or as nonspeech. He concluded that "the phoneme boundary effect for VOT does not depend exclusively upon phonetic categorization but may reflect acoustic and

auditory properties which are independent of phonetic processing'' (Wood, 1976a, p. 1388). Unfortunately, Wood's results cannot be considered conclusive because of the confounding of VOT with vowel duration in the truncated stimuli.

Following a previous, unpublished attempt by Ades (1973), Pisoni (1977) employed a temporal order judgment task to examine how much it might have in common with VOT perception (cf. also Pastore, Harris, & Kaplan, 1982). He varied the relative onset times of two pure tones similar in frequency to F1 and F2 of a neutral vowel and trained subjects to classify these stimuli into two categories exemplified by the extreme (50 msec) low-tone lead and lag stimuli. As it happened, the category boundary of most subjects fell not at the point of simultaneous onset, but at short, low-tone lags (where, accepting the analogy with F1 cutback, the VOT boundary is located). Discrimination peaks at the subjects' boundaries were obtained in a subsequent ABX task with feedback. In a second experiment, the ABX test was presented without prior training in labeling. Some subjects showed results similar to the first experiment, whereas others showed two discrimination peaks, at approximately 20-msec lead and lag times of the lower tone. The double peaks suggested that there were two natural boundaries on the continuum, one corresponding to the detection threshold for low-tone leads and the other to that for low-tone lags. This hypothesis was strengthened by an additional experiment in which subjects were successfully taught to classify the stimuli into three categories.

Pisoni (1977) concluded, on the basis of these data, that a ''basic limitation on the ability to process temporal-order information'' (p. 1360) underlies the perception of VOT, acknowledging at the same time that the location of the voicing boundary is influenced by a variety of other factors, ranging from spectral signal properties to the subjects' linguistic background (cf. Section VI,B). However, Pisoni's conclusion provides, at best, an incomplete account of VOT perception, for the voiced–voiceless distinction for syllable-initial stops in English rests as much on the perceived presence of aspiration or of a high F1 onset as on the temporal cue of delay of voicing onset. In addition, it is not clear how factors such as linguistic experience might modify the location of a strictly psychoacoustic boundary. It seems more likely that psychoacoustic and linguistic boundaries coexist.

That the tone onset time (TOT) continuum used by Pisoni is not a very close analog of VOT is suggested by several recent findings. Pisoni (1980a) himself failed to find a selective adaptation effect of TOT stimuli on syllables from a VOT continuum or vice versa, which suggests that the two types of stimuli do not engage the same auditory mechanisms. Rather convincing evidence for a fundamental difference between VOT and TOT was obtained by Summerfield (1982), who used, in addition, noise–buzz stimuli similar to those of J. D. Miller *et al.* (1976). All three sets of stimuli were composed of two steady-state components analogous to F1 and F2 and closely matched in frequency and amplitude

across the three sets. Summerfield investigated the influence of the frequency of the lower-frequency component (F1 or its analog) on the location of the boundary. On the VOT continuum (labeled "g" or "k"), he found, in accordance with previous results (Summerfield & Haggard, 1977), a shift of the boundary toward longer values as F1 frequency was raised. However, there were no comparable effects on the two nonspeech continua (labeled "simultaneous onset" or "successive onset"). Even granting that the use of phonetic labels only for the speech stimuli may have contributed to the difference, these results seriously weaken the proposal that the VOT boundary is merely a temporal order threshold (or even, for that matter, a noise-detection threshold).

It appears, however, that the last word on this issue has not yet been spoken. Hillenbrand (1982) recently reported an effect of the duration of a simulated F1 transition on the TOT boundary. Although the details of this study are not available at this time, it seems possible that Hillenbrand's stimuli, which contained frequency transitions in both tones, were sufficiently speechlike to elicit a phonetic mode of processing (cf. Grunke & Pisoni, 1982; Schwab, 1981). We might also take note of Molfese's (1978, 1980) analysis of evoked potentials to VOT and TOT stimuli. For both kinds of stimuli, a right-hemisphere component was found that distinguished between short-lag and long-lag stimuli and also between different extents of long lags, but not of short lags. This component seems consistent with a temporal order threshold. It is evident that the question about the psychoacoustic bases of VOT perception is far from resolved.

3. Nonspeech Analogs of Formant Transition Cues

The critical cues for distinguishing different places of articulation in synthetic stop-consonant continua are the transitions of F2 and F3. In the earliest continua, only two formants (F1 and F2) were used. This suggested an obvious nonspeech control: to omit the constant signal portions (F1 and perhaps also the steady state of F2) and to present F2 (or only the F2 transition) by itself. Several studies have investigated the perception of these isolated transitions ("chirps") or transitions plus steady state ("bleats"). It should be noted that although chirps sound rather nonspeech-like, they may be associated with speech sounds when subjects are provided with appropriate labels (Nusbaum, Schwab, & Sawusch, 1981). Bleats have some resemblance to strongly nasalized stop–vowel syllables and therefore are problematic as a nonspeech control. Studies employing these stimuli, however, invariably report that naive listeners do not perceive them as speech.

Kirstein (1966) was the first to present bleats in an ABX discrimination task. These isolated second formants were derived from the two-formant /be/– /de/–/ge/ continuum of Liberman et al. (1957) by omitting the constant F1. Although the speech stimuli had been discriminated fairly well (at the level predicted by the Haskins model or better), discrimination of the bleats was at chance at all step sizes used. However, when the bleats were played backward,

so that the transition was at the end, discrimination was better than chance and improved as step size increased.

A more comprehensive study along the same lines was conducted by Mattingly *et al.* (1971). They used both bleats and chirps that were derived from continua of initial and final stops. Oddity discrimination scores for chirps and bleats were rather similar and noncategorical, and discrimination was easier when the transitions were at the end (more precisely, when offset frequencies varied, rather than onset frequencies), which confirmed Kirstein's results and was in agreement with existing psychophysical data (P. T. Brady *et al.*, 1961). Due to peaks in the boundary regions, discrimination of syllable-initial stops was superior to discrimination of the corresponding nonspeech stimuli. The relationship was reversed for syllable-final stops whose discrimination function was also more similar to those for the corresponding nonspeech stimuli. However, Popper (1972) employed F2 bleats with final transitions and three-formant vowel–consonant syllables and found that, although the overall discriminability of speech and nonspeech was similar, the speech discrimination function showed a broad peak at the boundary, whereas the nonspeech function did not.

In another, related study, Syrdal-Lasky (1978) presented F2 chirps in an oddity discrimination task at three different intensities. Whereas at the two higher intensities the discrimination functions were nearly flat, at the lowest intensity there were two discrimination peaks. The peaks resembled those obtained with a simple /pæ/–/tæ/–/kæ/ continuum consisting of the chirps followed by a steady-state F1–F2 pattern. These data deserve to be replicated for they are the only instance so far of boundary effects on a chirp continuum.

Pisoni (1971, Exp. II) used bleats with initial transitions as stimuli in a training experiment intended to test Lane's (1965) proposition that categorical perception of nonspeech stimuli could be acquired in the laboratory. The stimuli were derived from a /bæ/–/dæ/ continuum, and listeners were given these labels to use. Although training did improve both labeling consistency and discrimination accuracy, there was no evidence that it introduced any consistent phoneme boundary effects. Moreover, discrimination following training was generally much better than predicted by the Haskins model, suggesting noncategorical perception. In a later replication, however, Pisoni (1976a) obtained not only very steep labeling functions, but also discrimination peaks at the category boundary for most listeners. It is not clear what caused this difference in results. Pisoni (1976a) states only that his earlier study was "not entirely satisfactory for a number of reasons" (p. 125), and he does not discuss the possibility that the bleats were heard as speech (/mæ/–/næ/) by the subjects. However, that possibility seems very real, and one is led to wonder whether the same results would have been obtained had arbitrary labels, or the same labels in reverse assignment, been used.

Isolated F3 resonances were presented in two studies of the /r/–/l/ contrast

(McGovern & Strange, 1977; Miyawaki *et al.*, 1975). Although located at higher frequencies than F2 bleats derived from stop-consonant continua, they are easier to discriminate because they have a distinctive steady state and slower transitions. As with bleats, however, discrimination is easier when the distinctive information is located at the end (as it is in vowel-liquid stimuli) than when it occurs at the beginning (McGovern & Strange, 1977). In both studies cited, F3 discrimination results showed no resemblance to /r/–/l/ discrimination.

So far, there is no convincing evidence that chirps or bleats yield a boundary effect when they are perceived as nonspeech. To avoid the objection that chirps and bleats are poor analogs of speech because so much of the original acoustic context (F1, F3) has been removed, Bailey, Summerfield, and Dorman (1977) constructed "sine-wave analogs" of speech stimuli: The first three formants of /bo/–/do/ and /be/–/de/ continua were mimicked by three pure tones (cf. Cutting, 1974). The interesting fact about sine-wave analogs is that they may be heard as speech with experience or with appropriate instructions but sound like nonspeech whistles to naive subjects. (Although this is also true, to some extent, for chirps and bleats, the phonetic and nonphonetic interpretations of sine-wave analogs appear to be more disparate in the listener's experience, which makes introspections a reliable source of information about perceptual modes.) Bailey *et al.* presented their speech and nonspeech stimuli in AXB identification (i.e., classification without labels) and discrimination tasks. The sine-wave stimuli were presented twice, first without and then with instructions to hear them as speech. The speech continua had been chosen to yield boundaries in different locations, one to the left and one to the right of the center of the stimulus range. Although classification accuracy was not very high, the expected difference in boundaries was obtained for the speech stimuli as well as for the sine-wave stimuli under speech instructions. However, under nonspeech instructions, the boundaries on the two continua coincided in the center of the stimulus range. The discrimination functions for the two sine-wave continua showed corresponding differences in the speech condition but no difference in the nonspeech condition. Unfortunately, the discrimination scores were rather low and did not show pronounced peaks, probably due to the poor labeling performance. In a second experiment, Bailey *et al.* used a /ba/–/da/ continuum and its sine-wave analog and divided subjects into speech and nonspeech groups on the basis of postexperimental interviews. Again, the category boundary on the sine-wave continuum resembled that on the speech continuum when the sine-wave stimuli were heard as speech but not when they were heard as nonspeech.

The significant work of Bailey *et al.* has remained unpublished and still awaits replication, particularly as far as the discrimination results are concerned. Together with the earlier chirp and bleat data, however, it strongly suggests that the location of the category boundary as well as the shape of the discrimination function are not determined by acoustic stimulus properties alone. The contribu-

tion of Bailey *et al.* lies, in part, in their attention to listeners' introspections as an indicator of perceptual modes. Pisoni (1976b), in an interesting pilot study, may have failed to take this aspect into consideration. He synthesized sine-wave analogs of a /ba/–/da/–/ga/ continuum omitting the steady-state portion, so that only the initial 50-msec transitions remained. Three experienced listeners generated ABX discrimination functions that exhibited two peaks approximately where the phoneme boundaries would lie on the corresponding speech continuum. Pisoni took this as support for the hypothesis that psychoacoustic discontinuities related to phonetic boundaries existed on the sine-wave transition continuum. However, in view of recent demonstrations that initial formant transitions without a following steady-state vowel can be quite accurately labeled as stop consonants (Blumstein & Stevens, 1980; Jusczyk, Smith, & Murphy, 1981; Tartter, 1981), it seems not impossible that Pisoni's experienced listeners were able to achieve this also with the sine-wave analogs.

However, Pisoni's (1976b) results receive support from another unpublished study (Wood, 1976b). Wood presented the initial 40 msec of synthetic stimuli from a /bæ/–/dæ/–/gæ/ continuum in a same–different task and obtained clear indications of increased perceptual sensitivity (in terms of a bias-free measure) at the points where the category boundaries for the full syllables were located. Significantly, Wood interviewed his subjects very carefully and determined that they did not relate the truncated stimuli in any way to the full syllables. The plausibility of this finding is increased by a comparison of Wood's results with Tartter's (1981): Using similar stimuli under speech instructions, Tartter obtained better discrimination performance for truncated than for full syllables, whereas Wood obtained the opposite, suggesting that Wood's subjects indeed did not hear the stimuli as speech. (However, Wood goes on to mention that in a subsequent study he did not find any effect of instructions, which is puzzling.)

Given the excellent reputation of both Pisoni and Wood as careful researchers, their findings may be taken as highly suggestive of psychoacoustic boundaries on a place of articulation continuum. However, it is difficult to reach a firm conclusion on the basis of unpublished and partially conflicting (Bailey *et al.*, 1977) evidence.

4. Nonspeech Analogs of Closure Cues

Nonspeech analogs of the closure duration cue for intervocalic stop voicing were constructed by Liberman, Harris, Eimas, Lisker, and Bastian (1961). The stimuli consisted of two noise bursts whose durations (about 200 and 80 msec) and amplitude envelopes matched those of the pre- and postclosure portions of speech stimuli (/ræbɪd/–/ræpɪd/); they were separated by varying intervals of silence (30–120 msec). The ABX discrimination of silence in this nonspeech context was consistently inferior to its discrimination in speech context, and there were no pronounced peaks in performance. At the time, these results were

welcomed as support for the "acquired distinctiveness" hypothesis. Further support came from a study by Baumrin (1974), who found, in an information-theoretic analysis, that less information was transmitted on a nonspeech continuum of silence durations than on a corresponding speech continuum.

Perey and Pisoni (1980) examined the discrimination of silence embedded between two 250-msec, three-tone complexes (imitating the first three formants of /ə/-like vowels) with or without simulated formant transitions into and out of the closure. Although the subjects were first taught to classify the stimuli into two categories, subsequent ABX discrimination was extremely poor and entirely continuous. Although both this study and that of Liberman et al. suffered from a (somewhat unnecessary) floor effect, they certainly demonstrated striking differences in listeners' sensitivity to silence duration in and out of speech context.

Silence is also an important cue for stop manner. A second cue in prevocalic position is a rapidly rising F1 transition. These two cues can be traded off against each other, within limits: For example, less silence is needed to hear *stay* rather than *say* when the onset of F1 in the vocalic portion is low than when it is high. Best et al. (1981) examined whether this trading relation is found in sine-wave analogs of *say–stay* stimuli consisting of an initial noise burst followed by a variable silent interval and a three-tone complex with variable onset frequency of the lowest (F1 analog) tone. The results of labeling and oddity discrimination tasks provided a positive answer, but only for those subjects who reported that they perceived the sine-wave stimuli as speech. The remaining subjects, who reported various nonspeech impressions, fell into two groups: those that appeared to pay attention to the temporal cue (gap duration) and those that paid attention to the spectral cue (onset quality of the simulated vocalic portion). The discrimination results for these two groups differed radically: The scores of the temporal listeners were somewhat lower than those of the speech listeners and exhibited two unpredicted peaks (at about 20 and 65 msec of silence, respectively) that warrant further investigation. The scores of the spectral listeners, on the other hand, were extremely high and much superior to those of the speech listeners. Those listeners who interpreted the stimuli as speech adopted neither of these selective-attention strategies but instead seemed to integrate the two cues into a single (phonetic) percept that, as the comparison with the nonspeech listeners shows, at the same time aided and hindered discrimination. These findings of Best et al. provide some of the most convincing evidence for the existence of separate modes of perception for speech and nonspeech.

To provide a potential nonspeech analog for the fricative–affricate contrast, one important cue for which is amplitude rise time, Cutting and Rosner (1974, 1976) varied the rise times of tonal stimuli (sawtooth or sine waves). These stimuli had the special distinction of conveying a manner contrast important in music, "pluck" versus "bow." Thus, unlike any of the other nonspeech controls discussed so far, these stimuli spanned two natural musical categories.

Comparing affricate–fricative (/tʃa/–/ʃa/, /tʃæ/–/ʃæ/) and pluck–bow continua in standard identification and discrimination tasks, Cutting and Rosner found categorical perception for both. This result suggested, more than any other, that a speech contrast had been built on a preexisting auditory threshold, and it became one of the most widely cited and replicated findings of recent years (e.g., Cutting, 1978; Cutting *et al.*, 1976; Jusczyk, Rosner, Cutting, Foard, & Smith, 1977; Remez, Cutting, & Studdert-Kennedy, 1980). All replications, however, used the original pluck–bow stimuli provided by Cutting and Rosner. It was embarrassing, therefore, when Rosen and Howell (1981) analyzed these stimuli and found them to be not equally spaced along the rise-time continuum. They conducted a series of very careful experiments and failed to find categorical perception with equally spaced stimuli; on the whole, rise-time discrimination followed Weber's law, and there was no effect of prior labeling experience. These results were replicated by Kewley-Port and Pisoni (1982). Thus, it appears that the findings of Cutting and his colleagues must be dismissed as artifactual.

In summary, despite a few suggestive results, there is no conclusive evidence so far for any significant parallelism in the perception of speech and nonspeech. What seems to matter is not whether the stimuli **are** speech or nonspeech but how listeners **interpret** (i.e., "hear") them (see also Section VI,A). Categorical perception appears to be a function not so much of the physical properties of the stimuli as of the frame of reference adopted by a listener.

5. Categorical Perception of Color and Music

A brief excursion is in order into domains that, like speech, employ highly overlearned categories. Here the question arises, as it does for speech, whether the category distinctions have a psychophysical basis or whether they are essentially arbitrary and determined by cultural convention. Although the role of cultural factors and experience in speech perception will be discussed in Section VI,B, we will touch on these topics as we briefly discuss some relevant findings on color and music perception.

To determine whether color discrimination performance covaries with color categorization, Lane (1967) compared data from earlier color labeling and discrimination studies and discovered that discrimination performance indeed showed peaks at the boundaries between the major categories (violet, blue, green, yellow, red). This finding was replicated by Kopp and Lane (1968) with two American subjects and compared to data obtained from two speakers of a Mexican Indian language (Tzotzil) whose color categories divide the wavelength continuum in a different fashion. Kopp and Lane interpreted their data as showing an influence of linguistic habits on discrimination, but a review of their figures makes their conclusion seem unwarranted. To the extent that one can conclude anything from comparing groups of two subjects each, the discrimination functions of American and Tzotzil subjects seemed not fundamentally differ-

ent. There appears to be little other evidence in favor of Kopp and Lane's thesis in the literature; on the contrary, there are studies showing that linguistic habits have no influence on the accuracy of color discrimination (Heider & Olivier, 1972). This suggests that the peaks in the color discrimination function have a psychophysical, rather than a cultural, basis.

Further support for this hypothesis comes from studies of color discrimination in infants. Using a habituation procedure, Bornstein, Kessen, and Weiskopf (1976) found that 4-month-old infants were more sensitive to hue differences across (adult) category boundaries than within categories. There is also anthropological evidence that the basic color categories are similar throughout the world, although some cultures use a greater array of categories than others (Berlin & Kay, 1969). All this ties in with extensive physiological evidence for two opponent-process mechanisms in the neural coding of color, so that the peaks in color discrimination are likely to have a direct physiological explanation. Bornstein (1973) has even proposed that certain cross-cultural differences in color naming can be explained by known racial variations in visual anatomy. We should mention that color perception was never a serious candidate for true categorical perception, for although it shows discontinuities in discrimination, many different hues can be distinguished within color categories. Color perception exhibits a category boundary effect, but it is far from categorical.

Results closer to true categorical perception have been obtained with musical stimuli. Musicians encounter a variety of explicit or implicit categories relating to intervals, chords, scales, timbres, attacks, etc. The ill-fated research on the pluck–bow distinction (Cutting & Rosner, 1974) has already been mentioned; this contrast, at least, does not seem to be categorically perceived. Most other research has been concerned with musical intervals (i.e., successive tones) or chords (i.e., simultaneous tones). One interesting aspect of music perception research is that familiarity with the distinctions involved varies enormously in the general population. Unlike speech, musical stimuli do not "name themselves." Comparisons of practicing musicians with nonmusicians provide information similar to that gained from comparing speech with nonspeech controls. (This author knows of no experiments conducted outside the reaches of traditional Western music.)

Siegel and Siegel (1977a) showed that musicians can accurately label intervals drawn from a continuum ranging from unison to a major triad, whereas nonmusicians exhibit very inconsistent labeling performance. In a subsequent study, Siegel and Siegel (1977b) obtained musicians' magnitude estimates for intervals ranging from a fourth to a fifth. They obtained plateaus and reduced variability within the three interval categories (fourth, tritone, fifth) and rapid changes with high variability at the boundaries. This finding suggested categorical perception, although no standard discrimination test was administered.

The classical methods of assessing categorical perception were applied to

musical intervals by Burns and Ward (1978). They presented intervals ranging from a major second to a tritone in labeling and in two-interval, forced-choice (2IFC) tasks (the pitch of the first note of each interval varied randomly). The discrimination functions were strongly categorical and closely matched the predictions generated by the Haskins model, although within-category discrimination was somewhat better than predicted. Although they varied the interstimulus interval between two successive intervals from 300 msec to 3 sec, they did not find any change in performance, which is reminiscent of the similar (near-)absence of an effect of temporal delay with stop consonants (Pisoni, 1973). Subsequently, Burns and Ward determined 2IFC difference limens by using a staircase method and testing their subjects until they reached asymptote. The results showed improved and more nearly continuous discrimination. The discrimination performance of a group of musically untrained subjects was much poorer but essentially continuous, which led Burns and Ward to conclude that musical intervals are learned, not natural, categories.

The categorical perception of simultaneous intervals or chords was first investigated by Locke and Kellar (1973). They presented chords consisting of three tones, with the frequency of the middle tone varying. The chords spanned the range from a minor triad to a major triad, but the subjects were not provided with these labels and instead classified the stimuli by matching them to a standard (one of the two endpoint stimuli). There was considerable individual variability, and the nonmusicians' performance was very poor. Musicians, on the other hand, showed a clear category boundary together with pronounced peaks in same–different discrimination scores; within-category discrimination, however, was much higher than predicted. A closer fit between predicted and obtained scores was obtained by Blechner (1977), who presented chords from a minor–major continuum in standard labeling and oddity discrimination tasks. Those subjects who were able to label the stimuli consistently as "minor" or "major" also showed fairly categorical discrimination, although scores were somewhat higher than predicted. A number of subjects were unable to label the chords consistently; their discrimination scores were low and showed no peak. Blechner also included a control consisting of only the middle tones of the chords. These stimuli were identified without difficulty as "low" or "high" by all subjects, and discrimination performance was noncategorical, although higher for trained musicians. Zatorre and Halpern (1979) essentially replicated Blechner's results for chords, using two-tone simultaneous intervals (from minor third to major third).

Categorical perception of stimuli varying in rhythm was reported by Raz and Brandt (1977). The stimuli consisted of three consecutive tones, with the temporal position of the second tone varying. However, since only an abstract of their study is available, it is not clear how categorical the results really were.

In summary, the musical results contrast with the color results—apart from the

difference in modality—in that the former seem to reflect learned categories whereas the latter reflect natural, physiologically based categories. Although category boundary effects are obtained in either case, perception is, interestingly, more nearly categorical in the case of the learned categories. Of course, their acquiredness does not necessarily mean that they do not have a physical basis: Musicians may learn to **discover** acoustic categories (e.g., simple frequency ratios) that simply are not registered by nonmusicians. Still, the fact that these categories must be established through experience and that they have an effect on perception once they have been learned is highly relevant to our understanding of speech perception. Specifically, it supports the hypothesis that categorical perception of speech is a product of categories acquired in the context of a particular language, and not of prewired psychoacoustic sensitivities (see Section VI,B).

VI. SUBJECT FACTORS IN CATEGORICAL PERCEPTION

In this section we will consider the contribution that the listener makes to categorical perception. Here we will encounter evidence that is of vital importance to understanding the phenomenon. In Section VI,A, we will first review the effects of experience and extensive practice on speech discrimination, as well as the roles played by expectations and strategies. Section VI,B discusses the important and rapidly expanding research comparing listeners of different language backgrounds or attempting to teach unfamiliar phonetic distinctions to subjects. Section VI,C briefly comments on infant speech perception. In the final section, VI,D, the topic will be the small and somewhat controversial literature on categorical perception in nonhuman animals.

A. Practice and Strategies

1. Effects of Discrimination Training

In Sections IV,B,1 and V,A,1, we have reviewed several studies showing that within-category discrimination on a stop-consonant continuum can be improved somewhat by using more-sensitive discrimination paradigms such as 4IAX (e.g., Pisoni & Lazarus, 1974). One of the largest increases in discrimination performance was obtained by Hanson (1977), who provided **feedback** throughout a same–different reaction time task, together with careful instructions to detect physical differences between stimuli (which contrasted with phonetic matching instructions in a second condition). The effectiveness of feedback is illustrated by a comparison of Hanson's results with those of Repp (1975), who used essentially the same task and instructions but did not provide any feedback: His subjects failed to show any improvement.

The exact effect of instructions on the degree of categorical perception is not quite clear. It is possible that inexperienced subjects do not always understand the meaning of "physical differences" among speech sounds, and some excessively categorical results in the literature may reflect that fact. What is more likely is that naive subjects do not know what sort of physical difference to listen for (see Pastore, 1981; Pisoni, 1980b). Some training with feedback may be necessary to direct their attention to the relevant auditory qualities, which are often difficult to convey by instructions alone.

Another procedural change that seems to improve performance is to restrict the discrimination task (or part of it) to within-category comparisons, only. The mixing of between- and within-category contrasts in the same block of trials, which has been the standard procedure in nearly all the studies reviewed so far, may place an attentional burden on the subjects that prevents them from focusing effectively on nonphonetic stimulus attributes. In addition to biasing subjects toward using a phonetic criterion, this mixing of different stimulus comparisons increases subject uncertainty, which, in turn, is known to increase psychophysical discrimination thresholds (Pastore, 1981).

A first attempt to improve VOT discrimination through extensive training was undertaken by Strange (1972). However, although she provided feedback, Strange used the standard oddity paradigm and a wide range of stimuli, which may have hindered her purpose. After a number of training sessions, discrimination performance had improved only slightly, primarily in the region of short voicing lags. A shift of labeling boundaries to shorter VOTs was also noted, which may account for the changes in discrimination performance. Although this shift may itself be taken to indicate an increased sensitivity to voicing lags, Strange's training study was considered unsuccessful both by herself and by later authors (Pisoni, Aslin, Perey, & Hennessy, 1982). It seems likely that the high-uncertainty discrimination paradigm prevented the accurate detection of acoustic differences (see also Section VI,B,2).

A fixed-standard AX task without feedback or extensive training was recently used by Repp (1981b) to assess the discriminability of within-category differences on several different speech continua. He found rather good performance on continua that varied silence duration (*say–stay, say shop–say chop*) but poor discrimination of VOT within the voiceless stop category. Repp (1981c), using the same paradigm, also found poor and seemingly categorical discrimination of fricative-vowel syllables by naive subjects. Thus, without training and/or feedback, low-uncertainty tasks do not lead to a dramatic improvement in discrimination performance. The secret lies in combining these procedures.

A fixed-standard AX task with feedback, using only two different stimuli in a whole block of trials, was employed first by Sachs and Grant (1976), who determined difference limens ($d' = 1$) on a /ga/–/ka/ VOT continuum. They reported threshold values of less than 2 msec with a 10-msec VOT standard and

of 10 msec with a 60-msec standard, which clearly is far superior to any within-category performance obtained in previous studies. In addition, the magnitude of the threshold increased monotonically with the VOT of the standard; that is, there was no phoneme boundary effect—a somewhat atypical result that was perhaps due to the use of subjects that were highly experienced in psychoacoustic tasks.

Ganong (1977) used a similar procedure to determine the discriminability of 15-msec VOT differences within the /pa/ category of a /ba/–/pa/ continuum. He found d′ scores close to 1.0, which is obviously better than chance, although not quite as good as the Sachs and Grant difference limens for experienced subjects. Interestingly, Ganong's subjects were equally accurate (following AX discrimination training) in an absolute identification task in which the standard and comparison stimuli were presented singly and randomly, separated by several seconds. Thus, it appears that the subjects eventually achieved discrimination not by physically comparing the stimuli but by referring to some long-term internal representations.

A third study using the fixed-standard AX procedure (and the first to be published) was conducted by Carney et al. (1977). These authors paired all stimuli from a /ba/–/pa/ continuum (including negative as well as positive VOTs) with selected standards and obtained discrimination functions before and after extensive training with feedback. A conventional oddity discrimination task was also administered. In both discrimination tasks, performance was fairly categorical before training but vastly improved after training. Discrimination was still best in the category boundary region, but secondary peaks emerged within categories, particularly around 20 msec of prevoicing—a little-noted finding that is in accord with Pisoni's (1977) results for TOTs. Phonetic labeling remained unaffected by training, and discrimination accuracy was equally high when subjects were required to provide labels following each "same–different" response. Finally, the trained subjects were even able to establish a new, arbitrary category boundary (at −50 msec of VOT) through identification training with feedback.

In a continuation of the research of Carney et al., Edman, Soli, and Widin (1978) observed that subjects trained on a labial VOT continuum could transfer their discrimination skills without any loss to a velar VOT continuum, and vice versa (see also Edman, 1979). However, discrimination remained most accurate in the boundary regions of both continua. In an application of the same techniques to place of articulation continua, Edman (1979) trained subjects on either a /bæ/–/dæ/–/gæ/ or a /pæ/–/tæ/–/kæ/ continuum and obtained excellent within-category discrimination and almost complete transfer to the other stimulus series.

Samuel (1977) demonstrated that a substantial improvement in within-category discrimination on a VOT continuum (/da/–/ta/, positive VOTs only) may also be obtained by training subjects in the ABX format, given that a fixed standard

and feedback are provided. The performance increase occurred primarily in the /da/ category, suggesting that discrimination of very short voicing lags was not limited by a simultaneity–successiveness threshold. A discrimination peak at the category boundary remained, which Samuel ascribed to phonetic categorization. By espousing a two-factor model, Samuel contrasts with Carney *et al.*, who favor a single-factor view, ascribing the boundary effect to psychoacoustic factors.

Several other training studies will be discussed in Section VI,B since they were concerned more with establishing a new phonetic contrast than with improving within-category discrimination. We have also omitted from discussion several studies that tested adults in low-uncertainty paradigms to provide comparison data for infants or animals run under the same conditions; some of these studies obtained rather good within-category discrimination (e.g., Aslin, Pisoni, Hennessy, & Perey, 1981; Sinnott, Beecher, Moody, & Stebbins, 1976). The spectacular success of the training studies reviewed in this subsection constitutes conclusive evidence that "specific feedback and fixed standards in a same–different task constitute an effective procedure for the learning of acoustic cues [and that] the utilization of acoustic differences between speech stimuli may be determined primarily by attentional factors" (Carney *et al.*, 1977, pp. 968–969).

2. Strategies and Expectations

a. Switching Modes. We have seen that feedback and/or many hours of training are necessary to achieve a high level of within-category discrimination on a stop-consonant continuum. Obviously, the acoustic differences on these continua are subtle and unfamiliar. Not only is it necessary to direct the subjects' attention to them but also subjects' discrimination accuracy needs to be sharpened by practice. There are other continua of speech sounds, however, where the acoustic differences are (or can be made) larger and more easily accessible. One might expect that little training would be necessary for acoustic discrimination of these differences and that it would be sufficient to direct the subjects' attention to the relevant auditory dimension.

Such a case was recently investigated by Repp (1981c), who employed an /ʃ/–/s/ fricative noise continuum followed by a vocalic context. When these stimuli were presented in AXB and fixed-standard AX tasks, most subjects perceived them fairly categorically, although within-category performance was better than expected. However, five subjects (two inexperienced and three experienced listeners) were extremely accurate in making within-category discriminations without any specific training. Two attempts were made to teach this skill to other subjects. In one condition, the subjects were given isolated fricative noises to discriminate before listening to the fricative-vowel syllables. Although all subjects were quite accurate in detecting spectral differences in the isolated

noises, their performance level dropped back to categorical levels when the noises occurred in vocalic context. In a second condition, the subjects heard a pair of noises immediately followed by exactly the same two noises in a constant vocalic context. The subjects were told to judge the isolated noises and then to verify the difference heard (if any) in the fricative-vowel syllables. Following this 25-min training period, the subjects listened to pairs of fricative-vowel syllables only, and most subjects performed noncategorically and with high accuracy.

The success of this last procedure, together with introspections of the experienced listeners, suggested that the skill involved lay in perceptually segregating the noise from its vocalic context, which then made it possible to attend to its pitch. Without this segregation, the phonetic percept was dominant. Once the auditory strategy has been acquired, it is possible to switch back and forth between auditory and phonetic modes of listening, and it seems likely (as Carney *et al.*, 1977, have shown) that both strategies could be pursued simultaneously (or in very rapid succession) without any loss in accuracy. These results provide good evidence for the existence of two alternative modes of perception, phonetic and auditory—a distinction supported by much additional evidence (see Sections V,C,3 and V,C,4; Bailey *et al.*, 1977; Best *et al.*, 1981; Liberman, 1982; Repp, 1982; Schwab, 1981). We may presume that the perception of other speech continua with relatively large auditory differences will likewise be susceptible to different strategies without much training.

b. Auditory Strategies. Several studies have indicated that subjects listening to speechlike stimuli may apply different auditory strategies, given that they are operating in the auditory mode. In the phonetic mode, listeners have no choice but to integrate all the relevant acoustic information into a phonetic percept. (However, there are often individual differences in the weights given to individual cues—see, e.g., Raphael, 1981.) Once in the auditory mode, however, it is possible either to selectively attend to individual auditory dimensions or to divide attention among several of them. Thus, Best *et al.* (1981) found two kinds of subjects among the listeners who heard sine-wave stimuli as nonspeech: "temporal listeners" and "spectral listeners" (see Section V,C,4). However, in a recent study using speech stimuli varying along similar dimensions, Repp (1981b) found that subjects took both temporal and spectral cues into account. This divided-attention strategy was encouraged by the task, which required auditory within-category discrimination (rather than auditory classification, as in Best *et al.*, 1981).

To mention another recent example, Rosen and Howell (1981) commented on individual differences in subjects' attention to spectral and temporal cues in the discrimination of amplitude rise time. It is not known whether there is any correlation between attentional preferences for certain cues in the auditory mode

and the weights given to the same cues in phonetic perception; this seems an interesting question for future research. The availability of a variety of auditory strategies is one of the reasons why training with feedback may be required to focus subjects' attention on particular cues. However, one strategy that subjects do **not** have available in the auditory mode is that of integrating the various cues into a single coherent percept; given that it is possible to divide attention among several cues, they remain separately perceived dimensions. Integration of psychoacoustically separable cues into a unitary percept is what characterizes the phonetic mode (Repp, 1981a, 1981b, 1982). However, there are also acoustic properties that are automatically integrated in auditory perception, such as the different formants of the spectrum (Stevens & Blumstein, 1978), and that do not normally permit selective attention strategies.

c. Phonetic Strategies. It is also possible to adopt different strategies while operating in the phonetic mode. Such strategies take the form of shifts in the phonetic frame of reference, achieved by adding or dropping categories or even by switching to a different set altogether. Staying within the confines of a single language (see Section VI,B for cross-linguistic research), the phonetic frame of reference for a given set of stimuli may differ from listener to listener or it may vary within a single listener, either spontaneously or as a consequence of instructions. Of course, such variations are facilitated if the stimuli are somewhat ambiguous. There is a lot of circumstantial evidence supporting these statements, but relatively little data. However, what data there are deserve close attention because they are relevant to the question of whether or not perceptual sensitivity in a discrimination task is determined by phonetic categorization. If it is possible to shift, create, or eliminate a discrimination peak merely by applying different phonetic categories, then that peak surely cannot have a solid psychoacoustic basis.

One instructive demonstration was conducted informally by investigators at Haskins Laboratories some years ago, and although it has not found its way into the literature, it has become part of the lore. A /ba/–/da/ continuum was presented in standard identification and discrimination tasks, and the usual pronounced peak at the category boundary was obtained. Then the tests were repeated, with one minor change. That change consisted of giving the subjects the additional response category /ða/, based on the observation that synthetic syllables ambiguous between /ba/ and /da/ often sound like /ða/. (The voiced fricative /ð/ has a place of articulation intermediate between /b/ and /d/ and, in natural speech, a very weak aperiodic component that is of little perceptual significance—cf. K. S. Harris, 1958.) With the additional category (which listeners almost never use spontaneously), listeners had two category boundaries and two associated discrimination peaks, neither of which coincided with the original peak. These results provided (admittedly anecdotal) evidence for an influence of phonetic

categorization per se on discrimination performance. And although it is possible to induce a similar change in categorization on a nonspeech continuum by permitting an "ambiguous" category, it is unlikely that discrimination performance will be much affected by this change (cf. Pisoni, 1977).

A recent study by Carden et al. (1981) was based on the acoustic affinity of /ba/, /da/ and /fa/, /θa/. The distinction between the two fricative categories is cued almost entirely by the vocalic formant transitions; the frication in natural productions is weak and nondistinctive (cf. K. S. Harris, 1958). Carden et al. preceded stimuli from a synthetic /ba/–/da/ continuum with a neutral noise, thus converting it into a /fa/–/θa/ continuum. The category boundaries on the two continua were significantly different. To counter the possible (though rather far-ry fetched) objection that the neutral noise may somehow have modified the auditory perception of the formant transitions, Carden et al. decided to hold the stimuli constant and to vary only the instructions. They first presented both continua in identification and oddity discrimination tasks and then repeated these procedures, requiring the listeners to apply the stop categories to the fricative stimuli and vice versa. The subjects were not only able to follow these instructions but also shifted their category boundaries in accordance with the categories used and exhibited a corresponding shift in the discrimination peak.

The results of Carden et al. provided strong evidence that the locations of the boundary and of the associated discrimination peak were not determined by psychoacoustic factors but mainly (if not exclusively) by the phonetic criteria adopted by the listeners. If there were any psychoacoustic boundaries at all on the continuum used, they seemed to be irrelevant to performance as long as the subjects operated in the phonetic mode. What seemed to matter, instead, was the relation of the stimuli to the listeners' internal prototypes of the relevant phonetic categories (however difficult it may be to conceptualize the mental representation of these prototypes). The difference between the /ba/–/da/ and /fa/–/θa/ boundaries is explained by the nonidentical places of articulation of these stops and fricatives, which result in characteristic differences in formant transitions. Most interestingly, it has been reported that even human infants show this boundary difference (Jusczyk, Murray, & Bayly, 1979, cited in Jusczyk, 1981). Thus, even at an early age, speech perception may not be governed solely by physical variables but may reflect an emerging (perhaps partially innate) referential system within the individual (see Section VI,C).

B. The Role of Linguistic Experience

Given that the degree of categorical perception in a particular experiment is largely a matter of stimulus, task, and subject factors, the central phenomenon to be explained is the phoneme boundary effect (cf. Carney et al., 1977). Cross-language research provides further valuable information on whether this effect is

auditory or phonetic in origin—a question that may have no general answer and therefore must be posed separately for each particular phonetic distinction. If the effect were due to a psychoacoustic threshold, then it should not only constrain (or even pin down) the phonetic boundary locations in different languages, but it should also be associated with a discrimination peak regardless of whether or not the threshold coincides with a linguistic boundary. If the two do not coincide and perception is strongly categorical, such a peak may not be immediately evident, but it should be possible to reveal it through discrimination or classification training. On the other hand, if the phoneme boundary effect is due to phonetic categorization only, then it should occur wherever a linguistic boundary happens to be, and efforts to reveal a peak at some other fixed location should fail. It is entirely possible that phoneme boundary effects on different speech continua require different types of explanation (cf. Ades, 1977).

One obvious question one might ask is: Where are the phoneme boundaries located when subjects with different language backgrounds listen to the same continuum of synthetic stimuli? There is ample evidence from comparative phonology that category distinctions present in one language may be absent in another. Some well-known examples that will concern us next are the absence of the [ba]–[pa] (prevoiced versus devoiced, or voiceless unaspirated) distinction in English, which is present in Thai, and the absence of the /r/–/l/ distinction in Japanese, which is present in English. However, there is less systematic information on the locations of boundaries between phonologically equivalent contrasts in different languages (which often differ in phonetic detail) and even less on discrimination functions corresponding to such boundaries. Since a number of relevant studies have been reviewed by Strange and Jenkins (1978), the present discussion will be brief and focus on work conducted since their article was written.

1. Cross-Linguistic Differences

By far the largest amount of cross-language work has been done on the voicing contrast for initial stop consonants, as cued by VOT. For example, Abramson and Lisker (1970; Lisker & Abramson, 1970) presented full VOT continua (containing voicing lead as well as voicing lag times) for all three places of articulation to speakers of English and Thai. The Thai subjects showed two category boundaries (prevoiced–devoiced–aspirated) and two corresponding discrimination peaks, whereas American listeners had only one (unaspirated– aspirated). The American and Thai results were similar on the voicing lag side (i.e., for the unaspirated–aspirated distinction common to both languages), but American listeners showed no indication of a discrimination peak on the voicing lead side, unlike Thai subjects. Similar results were obtained in a replication by Strange (1972).

Abramson and Lisker (1973) presented the same continua to speakers of

Spanish, a language that distinguishes only between prevoiced and devoiced stops. The Spanish category boundaries were surprisingly close to the English ones, though at somewhat shorter voicing lag times. A major discrimination peak was obtained in the same region, together with several secondary peaks. These data contrast with a replication by Williams (1977), who found the Spanish category boundary and the associated discrimination peak for labial stops to be in the vicinity of 0 msec VOT, with a secondary peak at about +25 msec of VOT, where the English /ba/–/pa/ boundary is located. Although the discrepancy between these two studies remains unexplained, Williams' results—which appear more reliable—are interesting for two reasons: First, they show that Spanish listeners can accurately discriminate among VOT values in the very short lead–lag range where, according to psychophysical arguments (Pisoni, 1977), they should be limited to near-chance performance by the simultaneity– successiveness threshold. Second, the secondary peak at short lag times suggests that these listeners were able to discriminate unaspirated from aspirated stops, presumably on an auditory basis. If so, then discrimination at very short VOTs was either entirely phonetic in nature (i.e., based on subjective uncertainty of phonetic judgments) or based on spectral signal properties (cf. Samuel, 1977), whereas the secondary peak at short lag times may have represented the temporal order threshold postulated by Pisoni (1977). The ability of Spanish listeners to discriminate unaspirated from aspirated stops contrasts with English-speaking listeners' inability to spontaneously discriminate prevoiced from devoiced stops. Presumably, the presence of prevoicing is less salient at the psychoacoustic level than the presence of aspiration (with its higher amplitude and concomitant spectral changes in the signal).

In a recent study of Polish, whose stop categories resemble those of Spanish, Keating, Mikoś, and Ganong (1981) found a VOT boundary in the short lag range (close to 0 msec VOT), together with a very broad discrimination peak that was skewed toward longer lag times. They also found that the boundary could be shifted toward longer voicing lags by adjusting the stimulus range so it included more aspirated tokens. These results suggest, in accord with the Spanish findings, that the presence of aspiration is a rather salient auditory event. Williams (1977) also found a broad discrimination peak similar to the Polish one for several Spanish–English bilinguals.

One phenomenon that has attracted the attention of researchers for some time is the inability of Japanese subjects to distinguish (and to correctly produce) American English /r/ and /l/, neither of which occurs in Japanese. (The Japanese /r/ is a dental flap—see Price, 1981.) These difficulties often persist for individuals that are quite fluent in English (Goto, 1971). An experimental demonstration was provided by Miyawaki et al. (1975), who showed that Japanese subjects performed very poorly when labeling or discriminating stimuli from a synthetic /ra/–/la/ continuum that were perceived fairly categorically by Ameri-

can listeners. However, when the distinctive third formants of these stimuli were presented in isolation as a nonspeech control, Japanese and American listeners gave almost identical results, with discrimination performance clearly above chance. This result suggested that the effect of linguistic experience was restricted to perception in the speech mode.

Little direct cross-language research has been done on other phonetic contrasts. For example, virtually nothing is known about the effect of linguistic background on the perception of stop-consonant place of articulation. Stevens et al. (1969) compared American and Swedish listeners' perception of steady-state vowels. Although there were differences in the locations of category boundaries, they were not reflected in the discrimination functions, which were very similar for the two groups of listeners. This study is well worth repeating in view of consistent findings of discrimination peaks at vowel category boundaries. Thus, for example, the Japanese subjects of Fujisaki and Kawashima (1969, 1970) show a single discrimination peak on an /i/–/e/ continuum, whereas American listeners show two peaks on a very similar continuum (Pisoni, 1971, Exp. 1), on which they distinguish three categories (/i/, /ɪ/, /ɛ/). Beddor and Strange (1982) have recently reported cross-linguistic differences in the perception of the oral–nasal distinction in vowels.

A cross-language difference in fricative perception may be gleaned from a comparison of data by Kunisaki and Fujisaki (1977) for Japanese listeners, and by Repp (1981c) for American listeners. Both studies used rather similar /ʃ/–/s/ continua, but the locations of the Japanese and American boundaries are different, and both are associated with marked discrimination peaks (cf. Fujisaki & Kawashima, 1969). Other comparisons of this sort between separate studies conducted in different countries could probably be found.

2. Acquisition of a New Phonetic Contrast

Students of a foreign language encounter the problem of learning to perceive and produce unfamiliar phonetic contrasts. Considering the importance of this problem, it is surprising how little laboratory research it has generated. The few studies in the literature were again concerned with either VOT or the /r/–/l/ contrast.

Given listeners' apparent sensitivity to the presence of aspiration in syllable-initial stops, it should be easy to teach Spanish or Polish listeners to discover the unaspirated–aspirated distinction. Lisker (1970) trained Russian listeners to discriminate labial stops ranging in VOT from +10 to +60 msec, all of which they normally label "p." The subjects learned to attach different labels to the endpoints of this range, but when labeling the stimuli in between, they showed a rather gradual change, with a midrange boundary that did not correspond to the American boundary (which is at about 25 msec). No discrimination tests were administered. Lisker concluded that Russian and American listeners used differ-

ent criteria for judging the same stimuli, with the Russians exhibiting either continuous perception or a different natural boundary in the voicing lag region. Pisoni *et al.* (1982) later criticized Lisker's study for not having employed feedback, thereby perhaps not directing the subjects' attention to the "correct" acoustic cues. They cite a study by Lane and Moore (1962), who successfully employed training with feedback to teach an aphasic patient the reacquisition of the English voicing contrast, using the /do/–/to/ (F1 cutback) continuum of Liberman, Harris, Kinney, and Lane (1961). Unfortunately, there have been no further studies with Russian subjects.

Several studies have attempted to teach American listeners the prevoiced–devoiced distinction for which they show little spontaneous sensitivity. After having relatively little success with extensive training in oddity discrimination, Strange (1972) first taught listeners to associate arbitrary labels with a clearly prevoiced (-100 msec VOT) and a clearly devoiced ($+10$ msec VOT) stop before administering standard identification and oddity discrimination tests using the negative VOT range only. The subjects showed fairly orderly labeling functions and improved discrimination scores following training, but the location of the category boundary was variable, and so were the shapes of the discrimination functions. Moreover, there was no transfer of training from an alveolar to a labial VOT continuum. Comparably variable results were obtained in a second study that provided training in judging VOT stimuli on a continuous scale.

Pisoni *et al.* (1982) resumed the task abandoned by Strange, with quite different results. They quite simply asked naive subjects to use "three response categories corresponding to [b], [p] and [pʰ]" (Pisoni *et al.*, 1982, p. 301) and obtained surprisingly consistent labeling in the prevoicing region, even without any special training (although training improved labeling consistency). What may have been responsible for their success but, curiously, was not mentioned by Pisoni *et al.* (but see McClasky, Pisoni, & Carrell, 1980), was that the categories used by the subjects were in fact "mba," "ba," and "pa." Apparently, it helped a great deal to associate the unfamiliar prevoicing distinction with a familiar phonemic contrast (even though initial nasal-stop clusters do not occur in English). In ABX discrimination tests, two peaks were found: a major one at the regular category boundary at short voicing lags ($+20$ msec of VOT) and a minor one in the short voicing lead region (-20 msec of VOT). Interestingly, both peaks were obtained regardless of whether or not the subjects had any prior labeling experience, either with two or with three categories. This finding contrasts with previous data that had found no discrimination peak in the voicing lead region. One factor that may have played a role here is the amplitude of the prevoicing, which may have been higher in the Pisoni *et al.* stimuli (no amplitudes are mentioned in any of the studies). There is no doubt that the detectability and discriminability of prevoicing will increase with its amplitude.

It is by no means clear that the new category distinction acquired by the

subjects of Pisoni *et al.*, even though it was apparently precipitated by the use of phonetic labels, was indeed a phonetic one (or, if it was, that it was the pre-voiced–devoiced rather than the nasal+stop–stop distinction). The "mba" label may simply have served to direct the subjects' attention to the relevant auditory dimension. A subsequent demonstration by McClasky *et al.* (1980) of virtually perfect transfer of the acquired distinction to an alveolar stop ("nda"– "da") continuum proves little, for the prevoiced portion is acoustically independent of the place of articulation of the stop consonant. The critical question is whether subjects who are able to perceive the prevoicing distinction in the laboratory will subsequently be able to use this skill in a natural-language context, for example, in learning a foreign language such as Thai. Until such transfer has been demon-strated, it is prudent to assume that the subjects of Pisoni *et al.*, rather than acquiring a new phonetic contrast, merely learned to make certain auditory discriminations.

The importance of conducting discrimination training in a way that facilitates transfer to a more naturalistic situation was stressed by MacKain *et al.* (1981), who reexamined Japanese listeners' perception of the English /r/–/l/ distinction. They found several individuals who were able to identify and discriminate stim-uli from a /rɑk/–/lɑk/ (*rock–lock*) continuum almost as well (i.e., as cate-gorically) as American subjects. It turned out that these subjects had not only had extensive experience with English, but with English conversation in particular, suggesting that transfer from the real world to the laboratory may be easier than the other way around. The continuing research in this area promises to yield useful insights into the process of second-language acquisition.

C. Categorical Perception in Human Infants

Since the rather extensive literature on infant speech perception has been reviewed repeatedly in recent years (Eilers, 1980; Juscyzk, 1981, in press; Kuhl, 1979b; Mehler & Bertoncini, 1979; Morse, 1979; Walley, Pisoni, & Aslin, 1981), only a very brief summary is needed here. It is now well known that infants as young as a few weeks do exhibit categorical discrimination. Although, for obvious methodological reasons, this result is usually established with a much smaller number of different stimuli than are used in corresponding studies with adult subjects, the pattern is generally clear: Pairs of stimuli crossing the adult (American English) boundary are discriminated more readily than pairs of stimuli from within an adult category. This has been shown for the voicing lag (unaspirated–aspirated) contrast in initial stop consonants (Eimas *et al.*, 1971; however, see Molfese & Molfese, 1979), for the place of articulation contrast in voiced initial stop consonants (Eimas, 1974), for the /ra/–/la/ distinction (Eimas, 1975), and for the /ba/–/wa/ distinction (Eimas & Miller, 1980). Iso-lated vowels, on the other hand, appear to be continuously discriminated by infants (Swoboda, Kass, Morse, & Leavitt, 1978).

In addition, there are a number of studies that, although not testing for within-category discrimination, have demonstrated the infant's ability to discriminate a variety of phonetic contrasts in natural or synthetic speech (e.g., Jusczyk, 1977; Jusczyk, Copan, & Thompson, 1978; Jusczyk & Thompson, 1978). Categorical-like discrimination has also been found for Pisoni's (1977) TOT continuum (Jusczyk, Pisoni, Walley, & Murray, 1980), whereas isolated third formants from a /ra/–/la/ continuum (Miyawaki et al., 1975) were perceived continuously by infants (Eimas, 1975). With the exception of occasional negative findings due to procedural factors (see Morse, 1979) or to the difficulty of certain phonetic contrasts (e.g., /f/–/θ/, Eilers, Wilson, & Moore, 1977), these results show the infant's perceptual capabilities to be remarkably developed and broadly similar to those of adults.

One important difference, however, is that infants have only minimal linguistic experience. It is generally considered unlikely that a few weeks or months of passive exposure to a particular language could have any significant effect on the infant's perceptual response to speech stimuli. Thus, infants reared in different language environments are expected to behave similarly, and this expectation has been confirmed in several cross-linguistic studies. What makes these studies especially interesting is that they show infants to be sensitive to certain distinctions that are not phonemic in their future language. Thus, American infants apparently can discriminate the prevoiced–devoiced contrast (Aslin et al., 1981; Eimas, 1975), whereas Kikuyu (Streeter, 1976) and Spanish infants (Eilers, Gavin, & Wilson, 1979; Lasky, Syrdal-Lasky, & Klein, 1975) can discriminate the unaspirated–aspirated contrast, which does not figure in their respective languages. Although it has not been established that infants perceive these unfamiliar distinctions in a truly categorical fashion (cf. Aslin et al., 1981; Morse, 1979), these results, at the very least, demonstrate high sensitivity to certain auditory stimulus properties—a sensitivity that adults seem to suppress unless these properties become associated with a phonetic distinction.

Additional evidence for American infants' superiority over adults in discriminating foreign-language contrasts has been obtained by Trehub (1976) for vowel nasalization and fricative palatalization, by Werker, Gilbert, Humphrey, and Tees (1981) for the dental–retroflex and aspirated voiced–voiceless contrasts, and by Werker (1982) for the dental–retroflex and velar–uvular contrasts. The work of Werker (1982) is especially intriguing in that it has provided longitudinal evidence that the ability to discriminate these contrasts disappears as early as 8–10 months of age, a time at which recognizable phonetic segments emerge in babbling. This startling finding was confirmed in a longitudinal study of individual infants (Werker, 1982). Unfortunately, no direct correlation with the onset of babbling could be established.

Of course, these findings should not be interpreted as showing that infants' auditory sensitivity is superior to that of adults. In fact, the opposite is likely to be the case; for example, higher TOT thresholds have been obtained with infants

than with adults (Jusczyk *et al.*, 1980) and, in a recent comparison of VOT discrimination thresholds obtained with identical procedures (Aslin *et al.*, 1981), adults proved to be far superior to infants. However, infants are free to attend to auditory properties of speech, whereas adults, being constrained by linguistic experience, are not. Once adults' attention is properly directed to auditory stimulus attributes (see Section VI,A,2), their discrimination performance is likely to be superior to that of infants.

The infant research has also revealed instances of phonetic distinctions that are **not** discriminated at an early age but are contrastive in the language. One such distinction is that between short negative and short positive VOTs, which crosses a phoneme boundary in Spanish but not in English (Lasky *et al.*, 1975). Presumably, infants in a Spanish-speaking environment must learn this distinction as they grow older, while learning to disregard other distinctions that are not phonemic in their language. Thus, this research again attests to the profound influence that linguistic experience exerts on speech perception. What is not yet clear is whether the infant's perceptual predispositions are purely auditory in nature or whether they already reflect specifically linguistic propensities. Recent research on trading relations between different acoustic speech cues in infants suggests the possibility of some innate linguistic mechanisms (J. L. Miller & Eimas, 1983), as does the finding of different boundaries on /ba/–/da/ and /fa/–/θa/ continua (Jusczyk *et al.*, 1979, cited in Jusczyk, 1981). Just how specific these mechanisms are and how they interact with later experience remains to be investigated in more detail. For excellent discussions of issues in the development of speech perception, see Aslin and Pisoni (1980) and Jusczyk (1983).

D. Categorical Perception in Nonhuman Animals

The question of whether human infants are endowed with any specific genetic predispositions for phonetic perception is usefully addressed by comparing their speech perception with that of nonhuman animals. Unless an animal has had extensive experience with human speech (and probably even then), its ability to discriminate speech sounds should reflect solely psychoacoustic factors. Provided that its auditory system is similar to the human one (which is true for the two species studied most closely, macaques and chinchillas), the results from the animal laboratory should reveal how much of the human infant's performance can be attributed to pure psychoacoustics.

Because of obvious methodological difficulties, animal research on speech perception has made only slow progress. A recent article (Kuhl, 1981) cites only four earlier studies concerned with categorical perception. Morse and Snowdon (1975) measured changes in macaques' heart rate in response to changes in speech stimuli drawn from Pisoni's (1971, Exp. I) /bæ/–/dæ/–/gæ/ continuum.

The monkeys exhibited good discrimination between categories and some sensitivity to within-category differences, although the latter finding rested primarily on an unexplained heart-rate acceleration in the no-change control condition. Sinnott *et al.* (1976) tested macaques and humans on a /ba/–/da/ continuum by using a key-press response and a fixed-standard paradigm. Although the results for humans were not very categorical (humans were actually better than monkeys in detecting within-category differences), those for the monkeys did not suggest categorical perception either. Because of differences in procedure, these results are not easily compared with those of Morse and Snowdon. Waters and Wilson (1976) used avoidance training to test macaques' discrimination of stimuli from a VOT continuum. Their data, like those of Sinnott *et al.*, yielded only the equivalent of labeling functions obtained with several different ranges of VOT. The monkeys' category boundary was found to be highly range-dependent, which suggests continuous perception. Since the boundary was consistently located in the voicing lag region, it seems likely that the animals paid attention to the presence of aspiration noise or to spectral differences in the F1 region. Of these three studies, only that by Morse and Snowdon (1975) provides some indication of a category boundary effect in monkeys. Clearly, those data need to be replicated if they are to stand on solid ground. However, a highly successful demonstration of category boundary effects in monkeys has recently been reported (Kuhl & Padden, 1983).

Animals would be expected to show categorical perception of speech only when a speech continuum straddles a psychoacoustic threshold. This may be true for the VOT continuum. In a widely cited study, Kuhl and Miller (1978) reported almost identical "labeling functions" (i.e., generalization gradients) for chinchillas and for humans on three VOT continua /ba/–/pa/, /da/–/ta/, and /ga/–/ka/. For both groups of subjects, the boundaries shifted toward longer values of VOT as place of articulation changed from labial to alveolar to velar, even though the range of VOTs remained constant. These results strongly suggested a psychoacoustic reason for the boundary shift, probably due to the spectral concomitants of VOT. No attempt was made to test whether the chinchilla boundary is as stable with changes in stimulus range as the human boundary (cf. S. Brady & Darwin, 1978; Keating *et al.*, 1981) or as unstable as the monkey boundary (Waters & Wilson, 1976).

Discrimination data for chinchillas were recently reported by Kuhl (1981). After training the animals to avoid shock by responding to differences between successive stimuli, she used a staircase procedure to determine VOT difference limens at various points along a /da/–/ta/ continuum. She found the highest accuracy in the region between 30 and 40 msec of VOT, where both the human and the chinchilla boundaries are also located. A previous unpublished study by J. D. Miller, Henderson, Sullivan, and Rigden (1978) had shown superior discrimination of stimuli crossing the boundary on a /ga/–/ka/ continuum. These

results provide rather strong evidence of a psychoacoustic boundary in the voicing lag region for chinchillas (and, presumably, for humans as well). Similar results have recently been obtained with monkeys (Kuhl & Padden, 1982). What remains uncertain is the role of these psychoacoustic factors in human speech perception. We agree with Pisoni's (1980b) reservation that findings on animal speech perception "are incapable, in principle, of providing any further information about how these signals might be 'interpreted' or coded within the context of the experience and history of the organism" (p. 304).

VII. CONCLUDING COMMENTS: BEYOND THE CATEGORICAL PERCEPTION PARADIGM

The research reviewed in the preceding sections has operated almost exclusively within a single experimental paradigm. Although there have been a great many variations in procedural detail, the essential common factor has been the use of (typically synthetic) continua of speech sounds. This concluding section offers some comments on the limitations of this approach and on its relation to categorical perception in the real world.

A. On Articulatory Realism

The possibility of constructing a continuum from one phonetic category to another is intriguing. However, the stimuli on such a continuum are not all equally realistic. Although the endpoint stimuli of a synthetic continuum are already removed from real speech by virtue of their stylized acoustic properties, this is even more true for stimuli from the middle of the continuum, which were never intended to model real speech but were obtained by mere parameter interpolation. In some cases, utterances resembling these stimuli may actually be impossible to produce by a human vocal tract.

Although this argument may be used to downgrade categorical perception research for its lack of ecological realism, it has not been traditionally considered a disadvantage. Indeed, it is part and parcel of the "motor-theoretic" view of categorical perception: Perception is categorical where the articulatory space (in a given language) is relatively discontinuous—in other words, when the stimuli from the middle of a continuum are less realistic than those from the ends. Seen in this way, the motor theory is not so much a theory as a statement of (though often poorly documented) fact. The mechanisms by which perceptual processes might "refer to" articulation have always remained obscure, which has led many researchers to dismiss the motor theory altogether. Nobody would deny, however, that perception is shaped by experience and that this shaping is due to events that occur frequently. Therefore, the phonetic categories that constitute

the frame of reference for speech perception must directly reflect the structure of speech—a structure that is imposed by the articulatory system within the conventions specific to a given language. Consequently, it is a truism that speech perception is intimately related to speech production. How this relationship is instantiated and solidified in the brain is a question for the philosopher and the neurophysiologist to answer (for some interesting developments in the latter direction, see Anderson, Silverstein, Ritz, & Jones, 1977). The difficulty of finding an answer should not prevent us, however, from recognizing that the specific systemic properties of speech are equally reflected in production and perception.

Several theorists have argued that, when listening to speech, we directly perceive what the articulators are doing (e.g., Gibson, 1966; Neisser, 1976; Summerfield, 1979). Essentially, this hypothesis is a contemporary version of the motor theory, though it denies any role of "mediation" or "reference" in perception. As far as natural speech is concerned, the hypothesis must be true, for speech is what the articulators are doing, as conveyed by sound. However, this cannot be said of the stimuli from synthetic continua. To the extent that they are unlikely products of articulation, they should be perceived either as non-speech or be perceptually assimilated to existing schemata of articulatory action, which are instantiated by the phonetic categories of a language. The phenomenon of categorical perception suggests that as long as the stimuli capture some salient properties of speech, they are perceived as the articulatory event most compatible with their structure, and this seems consistent with theories of direct perception, particularly with Neisser's (1976) formulation.

B. On Category Boundaries

The view of categorical perception as an acquired, language-specific, attentional phenomenon seems to contradict the hypothesis that categorical perception is caused by psychophysical boundaries on a stimulus continuum. However, the contradiction is more apparent than real. There is extensive evidence, reviewed earlier, that categorical perception may be caused either by categorization alone or by a psychophysical discontinuity and that both factors may be operating simultaneously for a single set of stimuli (although the former seems much more important in speech perception than the latter). Problems arise only when an attempt is made to reduce these two causes to a single one by assuming that auditory thresholds are plastic and shift with language experience (see, e.g., Aslin & Pisoni, 1980). This hypothesis (which is forced by the common-factor theory of categorical perception) is empty if the auditory thresholds in question are assumed to be entirely specific to speech, that is, if they are essentially equated with phonetic boundaries; and it is most likely wrong if auditory thresholds are understood in a more general sense. In the second case, for example, the

thresholds for certain nonspeech distinctions should show language-specific variations along with the phonetic boundaries that they are presumed to underlie—a prediction for which there is currently no positive evidence whatsoever. It seems much more likely that auditory thresholds and phonetic boundaries coexist, with the former limiting the possible locations of the latter only in the sense that what sounds the same cannot be phonetically distinctive.

One true shortcoming of the categorical perception paradigm is that it has overemphasized the importance of the boundaries between phonetic categories. After all, the categories, and not the boundaries between them, are the important functional elements of speech and language. The boundaries themselves are a mere epiphenomenon, apparent only in a particular experimental situation. Within the limits of the categorical perception paradigm, it may often not be clear whether the boundary is there because of the categories or whether the categories are there because of the boundary (although it should be possible, at least in principle, to decide this issue empirically in each case). However, beyond the realm of artificial speech continua, the boundary concept has little to offer.

It is appropriate to mention at this point some interesting research concerned with the basis of linguistic categories per se, disregarding the question of boundaries. For example, Fodor, Garrett, and Brill (1975) reinforced infants to respond with head turns to two (out of three) consonant–vowel syllables that either did or did not share the initial consonant, the vowels always being different. The infants showed more evidence of learning when the consonants were shared, indicating some ability to detect invariant acoustic properties (cf. Stevens & Blumstein, 1978) or, perhaps, even to conduct some sort of segmental analysis (Fodor *et al.*, 1975). Kuhl (1979a) demonstrated that infants are able to respond differentially to two vowel categories (/a/ and /i/) in the presence of a wide variety of distracting variability (different talkers). Similar perceptual constancy for vowels, at least, has been demonstrated in dogs (Baru, 1975) and chinchillas (Burdick & Miller, 1975). Perceptual classification techniques of this kind have also been used with adults to examine the possible psychoacoustic basis for the perceived similarity of stop consonants in initial and final position (Grunke & Pisoni, 1982; Schwab, 1981) or across different vocalic contexts (Jusczyk *et al.*, 1981), as well as listeners' awareness of phonological features (Healy & Levitt, 1980). These and related methods promise to provide useful information, particularly about the emergence of phonetic categories in human infants, without undue emphasis on the boundaries between categories.

C. On Dual Processing

Several recent reviews have argued that the dual-process hypothesis of categorical perception should be abandoned in favor of single-process models (e.g., Crowder, 1981; Macmillan *et al.*, 1977; Tartter, 1982). Although it is true that

the results of particular experiments are sometimes difficult to decompose into separate contributions of phonetic and auditory judgments, the basic distinction between the two modes of processing is logically unassailable (Pisoni, 1980b; Repp, 1982). To classify stimuli into the categories characteristic of the language is simply different from judging stimuli as long or short, constant or changing, continuous or interrupted, etc. We have reviewed several experiments showing that listeners can switch between phonetic and auditory modes, often with strikingly different results. There is no reason to doubt the original suggestion of Fujisaki and Kawashima (1969, 1970) that both modes may be employed simultaneously in a discrimination task; whether they are depends on the specific situation.

Categorical perception of speech is, first and foremost, an experimental demonstration that listeners persist in their normal perceptual habits in the laboratory even when given the opportunity to relinquish those habits. There is nothing surprising about the categorical nature of speech perception, which was known long before the discovery of the laboratory phenomenon of categorical perception. The interest of the phenomenon lies largely in subjects' strong resistance to adopt a mode of listening that enables them to detect subphonemic detail. That this resistance can be overcome by appropriate methods and training is one of the most significant findings reviewed here. An important question for future research will be whether analytic perceptual skills acquired in the laboratory can be transferred to real-life situations. However, the question immediately comes to mind, Having trained subjects to overcome their language habits and to pay some attention to the sound of speech, of what use could that esoteric skill be to them in the real world?

There are two (related) real-life endeavors that require the (more or less conscious) apprehension of subphonemic distinctions. One is phonetic transcription, the other is acquisition of a foreign language. Phonetic transcription is a skill that phoneticians acquire through training. However, even in its more narrow varieties, it is essentially categorization according to a fine-grained scheme, instantiated by the International Phonetic Alphabet. Thus, rather than paying attention to auditory properties of speech, phoneticians simply use a larger number of internalized phonetic categories than the ordinary individual. However, phoneticians are usually also able to make some fairly accurate judgments about the auditory quality of speech sounds. That such an ability could be cultivated to a high degree is presupposed in Pilch's (1979) proposal of a science of "auditory phonetics," which involves the systematic description, using a purely auditory vocabulary, of "the partitions of auditory space imposed by different phonemic systems" (Pilch, 1979, p. 157). Although, for purposes of communication, the auditory description once again makes use of categories, these categories are intended to be decidedly nonphonetic. How successful this approach will be, given the twin difficulties of attending to auditory properties of

speech in a natural setting and of finding the proper terms for their description, remains to be seen. It is possible, however, that laboratory training of the sort employed in several recent categorical-perception studies (e.g., Carney *et al.*, 1977) will be helpful in developing the auditory phonetician's skills. Such skills may also be useful to speech pathologists.

A similar (and more commonly encountered) problem faces the individual learning a foreign language. In order to detect certain novel phonetic distinctions and to realize them in production, some sensitivity to subphonemic detail is required (cf. Flege, 1981; Flege & Hammond, 1982). Note, however, that at no time does the language learner need to describe this detail in auditory terms or to detect differences that are subphonemic in both the new and old languages. The task is restricted to the acquisition of new phonetic categories—a process that may not involve the auditory mode of perception at all, at least not at the level of consciousness. The possibility that an increased awareness of the auditory properties of speech might facilitate the acquisition of new phonetic contrasts outside the laboratory certainly deserves continued attention, but we should perhaps not be overly optimistic. So far, there is no convincing evidence that new phonetic contrasts can be taught directly in the laboratory by the simple techniques discussed here. A fruitful connection between categorical perception research and foreign language instruction still needs to be made.

The prospect of gaining some insight into the processes of both first- and second-language acquisition will keep interest in the phenomenon of categorical perception alive. It is to be expected, however, that the traditional methodology will eventually give way to new approaches that more directly address the important theoretical and practical problems raised by communication in the real world. Indeed, it seems that this process is now well under way.

Acknowledgments

References

Abramson, A. S. Identification and discrimination of phonemic tones. *Journal of the Acoustical Society of America*, 1961, 33, 842. (Abstract)

Abramson, A. S. The noncategorical perception of tone categories in Thai. In B. Lindblom & S. Öhman (Eds.), *Frontiers of speech communication research*. New York: Academic Press, 1979. Pp. 127–134.

Abramson, A. S., & Lisker, L. Discriminability along the voicing continuum: Cross-language tests. *Proceedings of the sixth international congress of phonetic sciences*. Prague: Academia, 1970. Pp. 569–573.

Abramson, A. S., & Lisker, L. Voice-timing perception in Spanish word-initial stops. *Journal of Phonetics*, 1973, 1, 1–8.

Abramson, A. S., Nye, P. W., Henderson, J. B., & Marshall, C. W. Vowel height and the perception of consonantal nasality. *Journal of the Acoustical Society of America*, 1981, 70, 329–339.

Ades, A. E. *Categorical perception and the speech mode*. Unpublished manuscript, Massachusetts Institute of Technology, Cambridge, 1973.

Ades, A. E. Adapting the property detectors for speech perception. In R. J. Wales & E. Walker (Eds.), *New approaches to language mechanisms*. Amsterdam: North-Holland, 1976.

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

Anderson, J. A., Silverstein, J. W., Ritz, S. A., & Jones, R. S. Distinctive features, categorical perception, and probability learning: Some applications of a neural model. *Psychological Review*, 1977, 84, 413–451.

Aslin, R. N., & Pisoni, D. B. Some developmental processes in speech perception. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child phonology* (Vol. 2). New York: Academic Press, 1980. Pp. 67–96.

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

Bailey, P. J., & Summerfield, Q. Information in speech: Observations on the perception of [s]-stop clusters. *Journal of Experimental Psychology: Human Perception and Performance*, 1980, 6, 536–563.

Bailey, P. J., Summerfield, Q., & Dorman, M. On the identification of sine-wave analogues of certain speech sounds. *Haskins Laboratories Status Report on Speech Research*, 1977, SR-51/52, 1–25.

Barclay, J. R. Noncategorical perception of a voiced stop: A replication. *Perception & Psychophysics*, 1972, 11, 269–273.

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

Bastian, J., & Abramson, A. S. Identification and discrimination of phonemic vowel duration. *Speech research and instrumentation* (Tenth final report). New York: Haskins Laboratories, 1964. (Appendix L)

Bastian, J., Eimas, P., & Liberman, A. M. Identification and discrimination of a phonemic contrast induced by silent interval. *Journal of the Acoustical Society of America*, 1961, 33, 842. (Abstract)

Baumrin, J. M. Perception of the duration of a silent interval in nonspeech stimuli: A test of the Motor Theory of speech perception. *Journal of Speech and Hearing Research*, 1974, 17, 294–309.

Beddor, P. S., & Strange, W. Cross-language study of perception of the oral-nasal distinction. *Journal of the Acoustical Society of America*, 1982, 71, 1551–1561.

Berlin, B., & Kay, P. *Basic color terms: Their universality and evolution*. Berkeley: University of California Press, 1969.

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

Blechner, M. J. *Musical skill and the categorical perception of harmonic mode*. Unpublished doctoral dissertation, Yale University, 1977.

Blumstein, S. E., & Stevens, K. N. Perceptual invariance and onset spectra for stop consonants in

different vowel environments. *Journal of the Acoustical Society of America*, 1980, 67, 648–662.

Bornstein, M. H. Color vision and color naming: A psychophysiological hypothesis of cultural difference. *Psychological Bulletin*, 1973, 80, 257–285.

Bornstein, M. H., Kessen, W., & Weiskopf, S. Color vision and hue categorization in young human infants. *Journal of Experimental Psychology: Human Perception and Performance*, 1976, 2, 115–129.

Brady, P. T., House, A. S., & Stevens, K. N. Perception of sounds characterized by a rapidly changing resonant frequency. *Journal of the Acoustical Society of America*, 1961, 33, 1357–1362.

Brady, S., & Darwin, C. J. Range effect in the perception of voicing. *Journal of the Acoustical Society of America*, 1978, 63, 1556–1558.

Braida, L. D., & Durlach, N. I. Intensity perception. II. Resolution in one-interval paradigms. *Journal of the Acoustical Society of America*, 1972, 51, 483–502.

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.

Burns, E. M., & Ward, W. D. Categorical perception—phenomenon or epiphenomenon: Evidence from experiments in the perception of melodic musical intervals. *Journal of the Acoustical Society of America*, 1978, 68, 456–468.

Carden, G., Levitt, A., Jusczyk, P. W., & Walley, A. Evidence for phonetic processing of cues to place of articulation: Perceived manner affects perceived place. *Perception & Psychophysics*, 1981, 29, 26–36.

Carney, A. E., Widin, G. P., & Viemeister, N. F. Noncategorical perception of stop consonants differing in VOT. *Journal of the Acoustical Society of America*, 1977, 62, 961–970.

Carrell, T. D., Pisoni, D. B., & Gans, S. J. Perception of the duration of rapid spectrum changes: Evidence for context effects with speech and nonspeech. *Research on speech perception* (Prog. Rep. No. 6). Bloomington: Indiana University, Department of Psychology, 1980. Pp. 421–436.

Chistovich, L. A. Problems of speech perception. In L. L. Hammerich, R. Jakobson, & E. Zwirner (Eds.), *Form and substance*. Copenhagen: Akademisk Forlag, 1971. Pp. 83–93.

Chistovich, L. A., Fant, G., de Serpa-Leitão, A., & Tjernlund, P. Mimicking and perception of synthetic vowels. *Quarterly progress and status report* (Vol. 2). Stockholm: Royal Technical University, Speech Transmission Laboratory, 1966. Pp. 1–18.

Conway, D. A., & Haggard, M. P. New demonstrations of categorical perception. *Speech synthesis and perception* (Rep. No. 5). Cambridge: University of Cambridge, Psychology Laboratory, 1971. Pp. 51–73.

Cooper, W. E. Selective adaptation of speech. In F. Restle, R. M. Shiffrin, N. J. Castellan, H. Lindman, & D. B. Pisoni (Eds.), *Cognitive theory*. Hillsdale, N.J.: Erlbaum, 1975. Pp. 23–54.

Cooper, W. E., Ebert, R. E., & Cole, R. A. Perceptual analysis of stop consonants and glides. *Journal of Experimental Psychology: Human Perception and Performance*, 1976, 2, 92–104.

Cowan, N., & Morse, P. A. Influence of task demands on the categorical versus continuous perception of vowels. In J. J. Wolf & D. H. Klatt (Eds.), *Speech communication papers*. New York: Acoustical Society of America, 1979. Pp. 443–446.

Creelman, C. D., & Macmillan, N. A. Auditory phase and frequency discrimination: A comparison of nine procedures. *Journal of Experimental Psychology: Human Perception and Performance*, 1979, 5, 146–156.

Cross, D. V., Lane, H. L., & Sheppard, W. C. Identification and discrimination functions for a visual continuum and their relation to the motor theory of speech perception. *Journal of Experimental Psychology*, 1965, 70, 63–74.

Crowder, R. G. Improved recall for digits with delayed recall cues. *Journal of Experimental Psychology*, 1969, 82, 258–262.

Crowder, R. G. The sound of vowels and consonants in immediate memory. *Journal of Verbal Learning and Verbal Behavior*, 1971, 10, 587–596.

Crowder, R. G. Precategorical acoustic storage for vowels of short and long duration. *Perception & Psychophysics*, 1973, 13, 502–506. (a)

Crowder, R. G. Representation of speech sounds in precategorical acoustic storage. *Journal of Experimental Psychology*, 1973, 98, 14–24. (b)

Crowder, R. G. Mechanisms of auditory backward masking in the stimulus suffix effect. *Psychological Review*, 1978, 85, 502–524.

Crowder, R. G. The role of auditory memory in speech perception and discrimination. In T. Myers, J. Laver, & J. Anderson (Eds.), *The cognitive representation of speech*. Amsterdam: North-Holland, 1981. Pp. 167–180.

Crowder, R. G. Decay of auditory memory in vowel discrimination. *Journal of Experimental Psychology: Human Learning and Memory*, 1982, 8, 153–162. (a)

Crowder, R. G. The communality of auditory sensory storage in perception and immediate memory. *Perception & Psychophysics*, 1982, 31, 477–483. (b)

Crowder, R. G., & Morton, J. Pre-categorical acoustic storage (PAS). *Perception & Psychophysics*, 1969, 5, 365–373.

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

Cutting, J. E. There may be nothing peculiar to perceiving in a speech mode. In J. Requin (Ed.), *Attention and performance VII*. Hillsdale, N.J.: Erlbaum, 1978. Pp. 229–244.

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

Cutting, J. E., & Rosner, B. S. Discrimination functions predicted from categories in speech and music. *Perception & Psychophysics*, 1976, 20, 87–88.

Cutting, J. E., Rosner, B. S., & Foard, C. F. Perceptual categories for musiclike sounds: Implications for theories of speech perception. *Quarterly Journal of Experimental Psychology*, 1976, 28, 361–378.

Darwin, C. J., & Baddeley, A. D. Acoustic memory and the perception of speech. *Cognitive Psychology*, 1974, 6, 41–60.

Dechovitz, D., & Mandler, R. Effects of transition length on identification and discrimination along a place continuum. *Haskins Laboratories Status Report on Speech Research*, 1977, SR-51/52, 119–130.

Diehl, R. L. Feature detectors for speech: A critical reappraisal. *Psychological Bulletin*, 1981, 89, 1–18.

Diehl, R. L., Elman, J. L., & McCusker, S. B. Contrast effects on stop consonant identification. *Journal of Experimental Psychology: Human Perception and Performance*, 1978, 4, 599–609.

Dorman, M. F. Auditory evoked potential correlates of speech sound discrimination. *Perception & Psychophysics*, 1974, 15, 215–220.

Dorman, M. F., & Dougherty, K. Shifts in phonetic identification with changes in signal presentation level. *Journal of the Acoustical Society of America*, 1981, 69, 1439–1440.

Durlach, N. I., & Braida, L. D. Intensity perception. I. Preliminary theory of intensity resolution. *Journal of the Acoustical Society of America*, 1969, 46, 372–383.

Edman, T. R. Discrimination of intraphonemic differences along two place of articulation continua. In J. J. Wolf & D. H. Klatt (Eds.), *Speech communication papers*. New York: Acoustical Society of America, 1979. Pp. 455–458.

Edman, T. R., Soli, S. D., & Widin, G. P. Learning and generalization of intraphonemic VOT discrimination. *Journal of the Acoustical Society of America*, 1978, 63 (Supplement No. 1), S19. (Abstract)

Eilers, R. E. Infant speech perception: History and mystery. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child phonology* (Vol. 2). New York: Academic Press, 1980. Pp. 23–39.

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.

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.

Eimas, P. D. The relation between identification and discrimination along speech and non-speech continua. *Language and Speech,* 1963, 6, 206–217.

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

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

Eimas, P. D., & Corbit, J. D. Selective adaptation of linguistic feature detectors. *Cognitive Psychology,* 1973, 4, 99–109.

Eimas, P. D., & Miller, J. L. Auditory memory and the processing of speech. *Developmental studies of speech perception* (Prog. Rep. No. 3). Providence, R.I.: Brown University, 1975. Pp. 117–135.

Eimas, P. D., & Miller, J. L. Effects of selective adaptation on the perception of speech and visual patterns: Evidence for feature detectors. In H. L. Pick & R. D. Walk (Eds.), *Perception and experience.* New York: Plenum, 1978.

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

Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. Speech perception in early infancy. *Science.* 1971, 171, 304–306.

Elman, J. L. Perceptual origins of the phoneme boundary effect and selective adaptation to speech: A signal detection theory analysis. *Journal of the Acoustical Society of America,* 1979, 65, 190–207.

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

Flege, J. E. The phonological basis of foreign accent: A hypothesis. *TESOL Quarterly,* 1981, 15, 443–455.

Flege, J. E., & Hammond, R. M. Mimicry of non-distinctive phonetic differences between language varieties. *Studies in Second Language Acquisition.* 1982, 5, 1–18.

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

Frazier, L. What can /w/, /l/, /y/ tell us about categorical perception? *Haskins Laboratories Status Report on Speech Research,* 1976, SR-48, 235–256.

Fry, D. B., Abramson, A. S., Eimas, P. D., & Liberman, A. M. The identification and discrimination of synthetic vowels. *Language and Speech,* 1962, 5, 171–189.

Fujisaki, H., & Kawashima, T. The influence of various factors on the identification and discrimination of synthetic speech sounds. *Reports of the 6th international congress on acoustics* (No. 2). Tokyo: 1968. Pp. 95–98.

Fujisaki, H., & Kawashima, T. On the modes and mechanisms of speech perception. *Annual Report of the Engineering Research Institute, Faculty of Engineering, University of Tokyo,* 1969, 28, 67–73.

Fujisaki, H., & Kawashima, T. Some experiments on speech perception and a model for the perceptual mechanism. *Annual Report of the Engineering Research Institute, Faculty of Engineering, University of Tokyo,* 1970, 29, 207–214.

Fujisaki, H., & Kawashima, T. A model of the mechanisms for speech perception: Quantitative

analysis of categorical effects in discrimination. *Annual Report of the Engineering Research Institute, Faculty of Engineering, University of Tokyo,* 1971, 30, 59–68.

Fujisaki, H., & Shigeno, S. Context effects in the categorization of speech and non-speech stimuli. In J. J. Wolf & D. H. Klatt (Eds.), *Speech communication papers.* New York: Acoustical Society of America, 1979. Pp. 5–8.

Ganong, W. F. III *Selective adaptation and speech perception.* Unpublished doctoral dissertation, Massachusetts Institute of Technology, 1977.

Ganong, W. F. III Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance,* 1980, 6, 110–125.

Garcia, E. The identification and discrimination of synthetic nasals. *Haskins Laboratories Status Report on Speech Research,* 1966, SR-7/8, 3.1–3.16.

Garcia, E. Labelling of synthetic nasals (II). *Haskins Laboratories Status Report on Speech Research,* 1967, SR-9, 4.1–4.17. (a)

Garcia, E. Discrimination of three-formant nasal-vowel syllables (III). *Haskins Laboratories Status Report on Speech Research,* 1967, SR-12, 143–153. (b)

Gerstman, L. *Cues for distinguishing among fricatives, affricates, and stop consonants.* Unpublished doctoral dissertation, New York University, 1957.

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

Godfrey, J. J. Perceptual difficulty and the right ear advantage for vowels. *Brain and Language,* 1974, 1, 323–336.

Godfrey, J. J., & Millay, K. K. Discrimination of the "tempo of frequency change" cue. *Journal of the Acoustical Society of America,* 1981, 69, 1446–1448.

Goto, H. Auditory perception by normal Japanese adults of the sounds "L" and "R." *Neuropsychologia,* 1971, 9, 317–323.

Grunke, M. E., & Pisoni, D. B. Some experiments on perceptual learning of mirror-image acoustic patterns. *Perception & Psychophysics,* 1982, 31, 210–218.

Haggard, M. P. Theoretical issues in speech perception. *Speech synthesis and perception* (Rep. No. 4). Cambridge: University of Cambridge, Psychology Laboratory, 1970. Pp. 1–16.

Haggard, M. P., Summerfield, Q., & Roberts, M. Psychoacoustical and cultural determinants of phoneme boundaries: Evidence from trading Fo cues in the voiced-voiceless distinction. *Journal of Phonetics,* 1981, 9, 49–62.

Hanson, V. L. Within-category discrimination in speech perception. *Perception & Psychophysics,* 1977, 21, 423–430.

Harris, J. D. Remarks on the determination of a differential threshold by the so-called ABX technique. *Journal of the Acoustical Society of America,* 1952, 24, 417.

Harris, K. S. Cues for the discrimination of American English fricatives in spoken syllables. *Language and Speech,* 1958, 1, 1–7.

Harris, K. S., Bastian, J., & Liberman, A. M. Mimicry and the perception of a phonemic contrast induced by silent interval: Electromyographic and acoustic measures. *Journal of the Acoustical Society of America,* 1961, 33, 842. (Abstract)

Hasegawa, A. *Some perceptual consequences of fricative coarticulation.* Unpublished doctoral dissertation, Purdue University, 1976.

Healy, A. F., & Levitt, A. G. Accessibility of the voicing distinction for learning phonological rules. *Memory & Cognition,* 1980, 8, 107–114.

Healy, A. F., & Repp, B. H. Context independence and phonetic mediation in categorical perception. *Journal of Experimental Psychology: Human Perception and Performance,* 1982, 8, 68–80.

Heider, E. R., & Olivier, D. C. The structure of the color space in naming and memory for two languages. *Cognitive Psychology,* 1972, 3, 337–354.

Hillenbrand, J. Auditory perception experiments with sine-wave analogs to the voice-onset time

dimension. *Journal of the Acoustical Society of America*, 1982, 71 (Supplement No. 1), S74–75.

Hirsh, I. J. Auditory perception of temporal order. *Journal of the Acoustical Society of America*, 1959, 31, 759–767.

Jusczyk, P. W. Perception of syllable-final stop consonants by 2-month-old infants. *Perception & Psychophysics*, 1977, 21, 450–454.

Jusczyk, P. W. Infant speech perception: A critical appraisal. In P. D. Eimas & J. L. Miller (Eds.), *Perspectives in the study of speech*. Hillsdale, N.J.: Erlbaum, 1981. Pp. 113–164.

Jusczyk, P. W. On characterizing the development of speech perception. In J. Mehler & R. Fox (Eds.), *Neonate cognition: Beyond the blooming, buzzing confusion*. Hillsdale, N.J.: Erlbaum, 1983.

Jusczyk, P. W., Copan, H., & Thompson, E. Perception by 2-month-old infants of glide contrasts in multisyllabic utterances. *Perception & Psychophysics*, 1978, 24, 515–520.

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.

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

Jusczyk, P. W., Smith, L. B., & Murphy, C. The perceptual classification of speech. *Perception & Psychophysics*, 1981, 30, 10–23.

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

Kaplan, H. L., Macmillan, N. W., & Creelman, C. D. Tables for d' from matching and oddity experiments. *Behavior Research Methods & Instrumentation*, 1978, 10, 796–813.

Keating, P. A., & Blumstein, S. E. Effects of transition length on the perception of stop consonants. *Journal of the Acoustical Society of America*, 1978, 64, 57–64.

Keating, P. A., Mikoś, M. J., & Ganong, W. F., III A cross-language study of range of voice onset time in the perception of initial stop voicing. *Journal of the Acoustical Society of America*, 1981, 70, 1261–1271.

Kewley-Port, D., & Pisoni, D. B. Discrimination of rise time in nonspeech signals: Is it categorical or noncategorical? *Journal of the Acoustical Society of America*, 1982, 71 (Supplement No. 1), S36.

Kirstein, E. Perception of second-formant transitions in non-speech patterns. *Haskins Laboratories Status Report on Speech Research*. 1966. SR-7/8, 9.1–9.3.

Kopp, J., & Lane, H. Hue discrimination related to linguistic habits. *Studies in language and language behavior* (Prog. Rep. No. VI). Ann Arbor: University of Michigan, 1968. Pp. 11–21.

Kopp, J., & Livermore, J. Differential discrimination or response bias? A signal detection analysis of categorical perception. *Journal of Experimental Psychology*, 1973, 101, 179–182.

Kopp, J., & Udin, H. Identification and discrimination functions for pure tone frequencies. *Psychonomic Science*, 1969, 16, 95–96.

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. (a)

Kuhl, P. K. The perception of speech in early infancy. In N. J. Lass (Ed.), *Speech and language: Advances in basic research and practice* (Vol. 1). New York: Academic Press, 1979. Pp. 1–47. (b)

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, 70, 340–349.

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.

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

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

Kunisaki, O., & Fujisaki, H. On the influence of context upon perception of voiceless fricative consonants. *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics, University of Tokyo*, 1977, 11, 85–91.

Lane, H. L. Motor theory of speech perception: A critical review. *Psychological Review*, 1965, 72, 275–309.

Lane, H. L. A behavioral basis for the polarity principle in linguistics, *Language*, 1967, 43, 494–511.

Lane, H. L., & Moore, D. J. Reconditioning a consonant discrimination in an aphasic: An experimental case history. *Journal of Speech and Hearing Disorders*, 1962, 27, 232–243.

Larkey, L. S., Wald, J., & Strange, W. Perception of synthetic nasal consonants in initial and final syllable position. *Perception & Psychophysics*, 1978, 23, 299–312.

Lasky, R. E., Syrdal-Lasky, A., & Klein, R. E. VOT discrimination by four to six and a half month old infants from Spanish environments. *Journal of Experimental Child Psychology*, 1975, 20, 215–225.

Liberman, A. M. Some results of research on speech perception. *Journal of the Acoustical Society of America*, 1957, 29, 117–123.

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

Liberman, A. M., Cooper, F. S., Harris, K. S., MacNeilage, P. F., & Studdert-Kennedy, M. Some observations on a model for speech perception. In W. Wathen-Dunn (Ed.), *Models for the perception of speech and visual form*. Cambridge, Mass.: MIT Press, 1967. Pp. 68–87.

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

Liberman, A. M., Harris, K. S., Eimas, P. D., Lisker, L., & Bastian, J. An effect of learning on speech perception: The discrimination of durations of silence with and without phonemic significance. *Language and Speech*, 1961, 54, 175–195.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, 1957, 53, 358–368.

Liberman, A. M., Harris, K. S., Kinney, J. A., & Lane, H. The discrimination of relative onset time of the components of certain speech and nonspeech patterns. *Journal of Experimental Psychology*, 1961, 61, 379–388.

Liberman, A. M., Mattingly, I. G., & Turvey, M. Language codes and memory codes. In A. W. Melton & E. Martin (Eds.), *Coding processes and human memory*. Washington, D.C.: Winston, 1972. Pp. 307–334.

Liberman, A. M., Studdert-Kennedy, M., Harris, K. S., & Cooper, F. S. A reply to "Identification and Discrimination Functions for a Visual Continuum and their Relation to the Motor Theory of Speech Perception" by Cross, Lane, and Sheppard. *Haskins Laboratories Status Report on Speech Research*, 1965, SR-3, 3.1–3.14.

Lisker, L. On learning a new contrast. *Haskins Laboratories Status Report on Speech Research*, 1970, SR-24, 1–15.

Lisker, L. Stop voicing production: Natural outputs and synthesized inputs. *Haskins Laboratories Status Report on Speech Research*, 1976, SR-47, 155–164.

Lisker, L., & Abramson, A. A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, 1964, 20, 384–422.

Lisker, L., & Abramson, A. S. The voicing dimension: Some experiments in comparative phonetics.

Proceedings of the 6th international congress of phonetic sciences. Prague: Academia, 1970. Pp. 563–567.

Lisker, L., Liberman, A. M., Erickson, D. M., Dechovitz, D., & Mandler, R. On pushing the voice-onset-time (VOT) boundary about. *Language and Speech*, 1977, 20, 209–216.

Locke, S., & Kellar, L. Categorical perception in a non-linguistic mode. *Cortex*, 1973, 9, 353–367.

MacKain, K. S., Best, C. T., & Strange, W. Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics*, 1981, 2, 369–390.

Macmillan, N. A., Kaplan, H. L., & Creelman, C. D. The psychophysics of categorical perception. *Psychological Review*, 1977, 84, 452–471.

Mandler, R. Categorical perception along an oral-nasal continuum. *Haskins Laboratories Status Report on Speech Research*, 1976, SR-47, 147–154.

Massaro, D. W. Language and information processing. In D. W. Massaro (Ed.), *Understanding language*. New York: Academic Press, 1975. Pp. 3–28.

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

May, J. G. Acoustic factors that may contribute to categorical perception. *Language and Speech*, 1981, 24, 273–284.

May, J. G., & Repp, B. H. Periodicity and auditory memory: A pilot study. *Haskins Laboratories Status Report on Speech Research*, 1982, SR-69, 145–149.

McClasky, C. L., Pisoni, D. B., & Carrell, T. D. Effects of transfer of training on identification of a new linguistic contrast in voicing. *Research on speech perception* (Prog. Rep. No. 6). Bloomington: Indiana University, Department of Psychology, 1980. Pp. 205–234.

McGovern, K., & Strange, W. The perception of /r/ and /l/ in syllable-initial and syllable-final position. *Perception & Psychophysics*, 1977, 21, 162–170.

McNabb, S. D. Must the output of phonetic feature detectors be binary? *Research on speech perception* (Prog. Rep. No. 2). Bloomington: Indiana University, Department of Psychology, 1976. Pp. 166–179. (a)

McNabb, S. D. *Using confidence ratings to determine the sensitivity of phonetic feature detectors.* Paper presented at the meeting of the Midwestern Psychological Association, Chicago, May 1976. (b)

Mehler, J., & Bertoncini, J. Infants' perception of speech and other acoustic stimuli. In J. Morton & J. C. Marshall (Eds.), *Psycholinguistics 2: Structures and processes*. Cambridge, Mass.: MIT Press, 1979. Pp. 67–105.

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. No. 1), S18.

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. L. Properties of feature detectors for VOT: The voiceless channel of analysis. *Journal of the Acoustical Society of America*, 1977, 62, 641–648.

Miller, J. L. Contextual effects in the discrimination of stop consonant and semivowel. *Perception & Psychophysics*, 1980, 28, 93–95.

Miller, J. L. Phonetic perception: Evidence for context-dependent and context-independent processing. *Journal of the Acoustical Society of America*, 1981, 69, 822–831.

Miller, J. L., & Connine, C. M. Psychophysical tuning curves for phonetically relevant acoustic information. *Journal of the Acoustical Society of America*, 1980, 67 (Supplement No. 1), S52. (Abstract)

Miller, J. L., & Eimas, P. D. Studies on the perception of place and manner of articulation: A comparison of the labial-alveolar and nasal-stop distinctions. *Journal of the Acoustical Society of America*, 1977, 61, 835–845.

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

Miller, J. L., Eimas, P. D., & Zatorre, R. J. Studies of place and manner of articulation in syllable-final position. *Journal of the Acoustical Society of America*, 1979, 66, 1207–1210.

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

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

Molfese, D. L. Neuroelectrical correlates of categorical speech perception in adults. *Brain and Language*, 1978, 5, 25–35.

Molfese, D. L. Hemispheric specialization for temporal information: Implications for the perception of voicing cues during speech perception. *Brain and Language*, 1980, 11, 285–299.

Molfese, D. L., & Hess, T. M. Hemispheric specialization for VOT perception in the preschool child. *Journal of Experimental Child Psychology*, 1978, 26, 71–84.

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.

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 infancy of infant speech perception: The first decade of research. *Brain, Behavior, and Evolution*, 1979, 16, 351–373.

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

Neisser, U. *Cognition and reality*. San Francisco: Freeman, 1976.

Nusbaum, H. C., Schwab, E. C., & Sawusch, J. R. *Duplex perception: Dichotic integration or chirp identification?* Paper presented at the 52nd meeting of the Eastern Psychological Association, New York, April 1981.

Oden, G. C., & Massaro, D. W. Integration of featural information in speech perception. *Psychological Review*, 1978, 85, 172–191.

Parks, T., Wall, C., & Bastian J. Intercategory and intracategory discrimination for one visual continuum. *Journal of Experimental Psychology*, 1969, 81, 241–245.

Pastore, R. E. Categorical perception: A critical re-evaluation. In S. K. Hirsh, D. H. Eldredge, I. J. Hirsh, & S. R. Silverman (Eds.), *Hearing and Davis: Essays honoring Hallowell Davis*. St. Louis, Mo.: Washington University Press, 1976. Pp. 253–264.

Pastore, R. E. Possible psychoacoustic factors in speech perception. In P. D. Eimas & J. L. Miller (Eds.), *Perspectives in the study of speech*. Hillsdale, N.J.: Erlbaum, 1981. Pp. 165–205.

Pastore, R. E., Ahroon, W. A., Baffuto, K. J., Friedman, C., Puleo, J. S., & Fink, E. A. Common-factor model of categorical perception. *Journal of Experimental Psychology: Human Perception and Performance*, 1977, 3, 686–696.

Pastore, R. E., Friedman, C. J., & Baffuto, K. J. *A comparative evaluation of the AX and two ABX procedures*. Paper presented at the 92nd meeting of the Acoustical Society of America, San Diego, November 1976.

Pastore, R. E., Harris, L. B., & Kaplan, J. K. Temporal order identification: Some parameter dependencies. *Journal of the Acoustical Society of America*, 1982, 71, 430–436.

Perey, A. J., & Pisoni, D. B. Dual processing vs. response-limitation accounts of categorical perception: A reply to Macmillan, Kaplan, and Creelman. *Research on speech perception* (Prog. Rep. No. 4). Bloomington: Indiana University, Department of Psychology, 1978. Pp. 197–216.

Perey, A. J., & Pisoni, D. B. Identification and discrimination of durations of silence in nonspeech signals. *Research on speech perception* (Prog. Rep. No. 6). Bloomington: Indiana University, Department of Psychology, 1980. Pp. 235–269.

Pilch, H. Auditory phonetics. *Word*, 1979, 29, 148–160.

Pisoni, D. B. *On the nature of categorical perception of speech sounds*. Unpublished doctoral dissertation, University of Michigan, 1971.

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

Pisoni, D. B. The role of auditory short-term memory in vowel perception. *Memory & Cognition*, 1975, 3, 7–18.

Pisoni, D. B. Some effects of discrimination training on the identification and discrimination of rapid spectral changes. *Research on speech perception* (Prog. Rep. No. 3). Bloomington: Indiana University, Department of Psychology, 1976. Pp. 122–141. (a)

Pisoni, D. B. Discrimination of brief frequency glissandos. *Research on speech perception* (Prog. Rep. No. 3). Bloomington: Indiana University, Department of Psychology, 1976. Pp. 155–169. (b)

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. Adaptation of the relative onset time of two-component tones. *Perception & Psychophysics*, 1980, 28, 337–346. (a)

Pisoni, D. B. The perception of speech versus nonspeech. In E. Fischer-Jørgensen and N. Thorsen (Eds.), *Proceedings of the ninth international congress of phonetic sciences* (Vol. 3). Copenhagen: University of Copenhagen, Institute of Phonetics, 1980. Pp. 301–312. (b)

Pisoni, D. B., Aslin, R. N., Percy, A. J., & Hennessy, B. L. Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. *Journal of Experimental Psychology: Human Perception and Performance*, 1982, 8, 297–314.

Pisoni, D. B., & Glanzman, D. L. Decision processes in speech discrimination as revealed by confidence ratings. *Research on speech perception* (Prog. Rep. No. 1). Bloomington: Indiana University, 1974. Pp. 114–128.

Pisoni, D. B., & Lazarus, J. H. Categorical and non-categorical modes of speech perception along the voicing continuum. *Journal of the Acoustical Society of America*, 1974, 55, 328–333.

Pisoni, D. B., & Tash, J. Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics*, 1974, 15, 285–290.

Pollack, I., & Pisoni, D. B. On the comparison between identification and discrimination tests in speech perception. *Psychonomic Science*, 1971, 24, 299–300.

Popper, R. D. Pair discrimination for a continuum of synthetic voiced stops with and without first and third formants. *Journal of Psycholinguistic Research*, 1972, 1, 205–219.

Posner, M. I., & Mitchell, R. F. Chronometric analysis of classification. *Psychological Review*, 1967, 74, 392–409.

Price, P. J. *A cross-linguistic study of flaps in Japanese and in American English*. Unpublished doctoral dissertation, University of Pennsylvania, 1981.

Raphael, L. J. Preceding vowel duration as a cue to the perception of the voicing characteristics of word-final consonants in American English. *Journal of the Acoustical Society of America*, 1972, 51, 1296–1303.

Raphael, L. J. Durations and contexts as cues to word-final cognate opposition in English. *Phonetica*, 1981, 38, 126–147.

Raz, I., & Brandt, J. F. Categorical perception of nonspeech stimuli by musicians and nonmusicians. *Journal of the Acoustical Society of America*, 1977, 62 (Supplement No. 1), S60. (Abstract)

Remez, R. E., Cutting, J. E., & Studdert-Kennedy, M. Cross-series adaptation using song and string. *Perception & Psychophysics*, 1980, 27, 524–530.

Repp, B. H. *Categorical perception, dichotic interference, and auditory memory: A "same-different" reaction time study*. Unpublished manuscript, 1975. (Available from the author)

Repp, B. H. Identification of dichotic fusions. *Journal of the Acoustical Society of America*, 1976, 60, 456–469. (a)

Repp, B. H. "Posner's paradigm" and categorical perception: A negative study. *Haskins Laboratories Status Report on Speech Research*, 1976, SR-45/46, 153–161. (b)

Repp, B. H. Dichotic competition of speech sounds: The role of acoustic stimulus structure. *Journal of Experimental Psychology: Human Perception and Performance*, 1977, 3, 37–50.

Repp, B. H. Perceptual equivalence of two kinds of ambiguous speech stimuli. *Bulletin of the Psychonomic Society*, 1981, 18, 12–14. (a)

Repp, B. H. Phonetic and auditory trading relations between acoustic cues in speech perception: Preliminary results. *Haskins Laboratories Status Report on Speech Research*, 1981, SR-67/68, 165–190. (b)

Repp, B. H. Two strategies in fricative discrimination. *Perception & Psychophysics*, 1981, 30, 217–227. (c)

Repp, B. H. Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. *Psychological Bulletin*, 1982, 92, 81–110.

Repp, B. H., Healy, A. F., & Crowder, R. G. Categories and context in the perception of isolated steady-state vowels. *Journal of Experimental Psychology: Human Perception and Performance*, 1979, 5, 129–145.

Repp, B. H., Liberman, A. M., Eccardt, T., & Pesetsky, D. Perceptual integration of temporal cues for stop, fricative, and affricate manner. *Journal of Experimental Psychology: Human Perception and Performance*, 1978, 4, 621–637.

Roberts, M., & Summerfield, Q. Audiovisual presentation demonstrates that selective adaptation in speech perception is purely auditory. *Perception & Psychophysics*, 1981, 30, 309–314.

Rosen, S. M. Range and frequency effects in consonant categorization. *Journal of Phonetics*, 1979, 7, 393–402.

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

Sachs, R. M. Vowel identification and discrimination in isolation vs. word context. *Quarterly Progress Report* (No. 93). Cambridge, Mass.: MIT, Research Laboratory of Electronics, 169. Pp. 220–229.

Sachs, R. M., & Grant, K. W. Stimulus correlates in the perception of voice onset time (VOT): II. Discrimination of speech with high and low stimulus uncertainty. *Journal of the Acoustical Society of America*, 1976, 60 (Supplement No. 1), S91. (Abstract)

Samuel, A. G. The effect of discrimination training on speech perception: Noncategorical perception. *Perception & Psychophysics*, 1977, 22, 321–330.

Samuel, A. G. *Speech is specialized, not special*. Unpublished doctoral dissertation, University of California at San Diego, 1979.

Samuel, A. G. Phonetic prototypes. *Perception & Psychophysics*, 1982, 31, 307–314.

Sawusch, J. R. Selective adaptation effects on end-point stimuli in a speech series. *Perception & Psychophysics*, 1976, 20, 61–65.

Sawusch, J. R., & Jusczyk, P. Adaptation and contrast in the perception of voicing. *Journal of Experimental Psychology: Human Perception and Performance*, 1981, 7, 408–421.

Sawusch, J. R., & Nusbaum, H. C. Contextual effects in vowel perception. I. Anchor-induced contrast effects. *Perception & Psychophysics*, 1979, 25, 292–302.

Sawusch, J. R., Nusbaum, H. C., & Schwab, E. C. Contextual effects in vowel perception II. Evidence for two processing mechanisms. *Perception & Psychophysics*, 1980, 27, 421–434.

Schouten, M. E. H. The case against a speech mode of perception. *Acta Psychologica*, 1980, 44, 71–98.

Schwab, E. C. *Auditory and phonetic processing for tone analogs of speech*. Unpublished doctoral dissertation, State University of New York at Buffalo, 1981.

Shigeno, S., & Fujisaki, H. Context effects in phonetic and non-phonetic vowel judgments. *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics, University of Tokyo*, 1980, 14, 217–224.

Siegel, J. A., & Siegel, W. Absolute identification of notes and intervals by musicians. *Perception & Psychophysics*, 1977, 21, 143–152. (a)

Siegel, J. A., & Siegel, W. Categorical perception of tonal intervals: Musicians can't tell *sharp* from *flat. Perception & Psychophysics*, 1977, 21, 399–407. (b)

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.

Stevens, K. N. On the relations between speech movements and speech perception. *Zeitschrift für Phonetik, Sprachwissenschaft, und Kommunikationsforschung*, 1968, 21, 102–106.

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.

Stevens, K. N., & Klatt, D. H. Role of formant transitions in the voiced-voiceless distinction for stops. *Journal of the Acoustical Society of America*, 1974, 55, 653–659.

Stevens, K. N., Liberman, A. M., Öhman, S. E. G., & Studdert-Kennedy, M. Crosslanguage study of vowel perception. *Language and Speech*, 1969, 12, 1–23.

Strange, W. *The effects of training on the perception of synthetic speech sounds: Voice onset time.* Unpublished doctoral dissertation, University of Minnesota, 1972.

Strange, W., Edman, T. R., & Jenkins, J. J. Acoustic and phonological factors in vowel identification. *Journal of Experimental Psychology: Human Perception and Performance*, 1979, 5, 643–656.

Strange, W., & Halwes, T. Confidence ratings in speech perception research: Evaluation of an efficient technique for discrimination testing. *Perception & Psychophysics*, 1971, 9, 182–186.

Strange, W., & Jenkins, J. J. Role of linguistic experience in the perception of speech. In R. D. Walk & H. L. Pick, Jr. (Eds.), *Perception and Experience*. New York: Plenum, 1978. Pp. 125–169.

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

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.

Studdert-Kennedy, M., Liberman, A. M., & Stevens, K. N. Reaction time to synthetic stop consonants and vowels at phoneme centers and at phoneme boundaries. *Journal of the Acoustical Society of America*, 1963, 35, 1900. (Abstract)

Studdert-Kennedy, M., Liberman, A. M., & Stevens, K. N. Reaction time during the discrimination of synthetic stop consonants. *Journal of the Acoustical Society of America*, 1964, 36, 1989. (Abstract)

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., & Haggard, M. On the dissociation of spectral and temporal cues to the voicing distinction in initial stop consonants. *Journal of the Acoustical Society of America*, 1977, 62, 436–448.

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

Syrdal-Lasky, A. Effects of intensity on the categorical perception of stop consonants and isolated second formant transitions. *Perception & Psychophysics*, 1978, 23, 420–432.

Tartter, V. C. A comparison of the identification and discrimination of synthetic vowel and stop consonant stimuli with various acoustic properties. *Journal of Phonetics*, 1981, 9, 477–486.

Tartter, V. C. Vowel and consonant manipulations and the dual-coding model of auditory storage: A re-evaluation. *Journal of Phonetics*, 1982, 10, 217–224.

Trehub, S. E. The discrimination of foreign speech contrasts by adults and infants. *Child Development*, 1976, 47, 466–472.

Van Tasell, D. J., & Crump. E. S. A. Effects of stimulus level on perception of two acoustic cues in speech. *Journal of the Acoustical Society of America*, 1981, 70, 1527–1529.

Van Valin, R. D., Jr. Perceived distance between vowel stimuli. *Journal of Phonetics*, 1976, 4, 51–58.

Vinegrad, M. D. A direct magnitude scaling method to investigate categorical versus continuous modes of speech perception. *Language and Speech*, 1972, 15, 114–121.

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. Alberts, & M. R. Peterson (Eds.), *Sensory and perceptual development*. New York: Academic Press, 1981.

Wang, W. S.-Y. Language change. *Annals of the New York Academy of Science*, 1976, 280, 61–72.

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

Watkins, M. J., & Todres, A. K. Suffix effects manifest and concealed: Further evidence for a 20-second echo. *Journal of Verbal Learning and Verbal Behavior*, 1980, 19, 46–53.

Werker, J. F. *The development of cross-language speech perception: The effect of age, experience, and context on perceptual organization*. Unpublished doctoral dissertation, University of British Columbia, 1982.

Werker, J. F., Gilbert, J. H. V., Humphrey, K., & Tees, R. C. Developmental aspects of cross-language speech perception. *Child Development*, 1981, 52, 349–355.

Williams, L. The perception of stop consonant voicing by Spanish-English bilinguals. *Perception & Psychophysics*, 1977, 21, 289–297.

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. (a)

Wood, C. C. *Discriminability, response bias, and phoneme categories in speech discrimination. II. Place of articulation*. Paper presented at the 91st meeting of the Acoustical Society of America, Washington, D.C., April 1976. (b)

Zatorre, R. J., & Halpern, A. R. Identification, discrimination, and selective adaptation of simultaneous musical intervals. *Perception & Psychophysics*, 1979, 26, 384–395.