Categorical perception
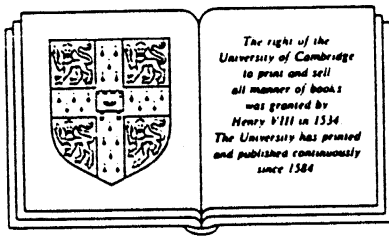
The groundwork of cognition

Edited by

STEVAN HARNAD

Behavioral and Brain Sciences
Princeton, New Jersey

The right of the
University of Cambridge
to print and sell
all manner of books
was granted by
Henry VIII in 1534
The University has printed
and published continuously
since 1584

CAMBRIDGE UNIVERSITY PRESS

Cambridge
New York   New Rochelle   Melbourne   Sydney

3   Phonetic category boundaries are flexible

Bruno H. Repp and Alvin M. Liberman

In this chapter we review the various factors that may influence the location of phonetic category boundaries on physical-stimulus continua of the kind widely used in speech-perception research. These factors range from the context provided by other stimuli in a test (which gives rise to effects such as sequential contrast, range-frequency shifts, and selective adaptation) to the internal structure of a single speech stimulus (effects of other cues or features present, of adjacent phonetic segments, of speaking rate and speaker characteristics) to the listener's linguistic experience and expectations (effects of semantic and syntactic structure, and of cross-language phonetic differences). We conclude that phonetic category boundaries are flexible in a way that suggests that speech perception is constrained by tacit knowledge of what a vocal tract does when it makes linguistically significant gestures.

In the grammatical domains of language we find no gradients, only categories. Thus, gradations of, for example, tense (present–past), form class (noun–verb), or even word (night–day) are everywhere absent. Indeed, they are impossible, for syntactic, morphologic, and phonologic devices do not permit continuous variation. At the surface of language, however, the situation is different. There, in the relation between phonetic structure and sound, the role of the segments is categorical – a segment is, for example, [d] or [g], not something in between – but the sound can vary continuously. That being so, at least in synthetic speech, we can ask whether the phonetic segments are categorical, not only in their linguistic function, but also in the way they are perceived. The answer is a qualified "yes." Other things being equal, stimuli belonging to the same phonetic category are more difficult to discriminate than stimuli on opposite sides of a phonetic boundary. This phenomenon has long been known as "categorical perception" (CP) (Studdert-Kennedy, Liberman, Harris, & Cooper, 1970). The research it has generated, which was recently reviewed by one of us (Repp, 1984), is largely concerned with the ability of listeners to detect stimulus differences within the categories – that is, with the degree to which perception is perfectly categorical – and with the conditions under which that ability can be made to vary. Our concern in this chapter is rather with the conditions under which the *locations* of the categories on a continuum can be shown to vary, and with the implications of that variation for a theory about the nature of the categories. More particularly, we will be concerned with the boundaries between

the categories (and with their movement), so before considering the relevance to theory, we should justify our concern with the boundaries.

We take the boundary to be the point along the appropriate (acoustic) stimulus continuum at which subjects classify stimuli into alternative categories with equal probability. In the typical case of two (adjacent) categories, this is simply the point corresponding to the 50% crossover of the response function. If more than one stimulus dimension is varied, category boundaries may be represented by contours in a multidimensional space (see, e.g., Oden & Massaro, 1978). The standard method of obtaining category boundaries is to present a set of stimuli repeatedly (and in random order) for identification as members of one class or another. Several alternative methods – for example, a method of adjustment – have been used, but all yield similar boundaries (Ganong & Zatorre, 1980).

Why do we take account only of the boundaries? After all, it is the categories themselves, rather than the boundaries between them, that play the important role in speech communication. Why not, then, deal with some appropriate exemplar – the prototype, as it were – of the category? A sufficient reason is that, until recently, no one had used methods designed to identify the prototypes. Worse yet, the application of such methods has so far failed to yield entirely satisfactory results (Samuel, 1979, 1982). The measurement of boundaries, on the other hand, has long been common in research on speech, so the data are plentiful. Moreover, the boundaries do inform us about the categories and, under some specifiable conditions, about their positions on the appropriate acoustic continua. And, finally, as we will argue in this chapter, it is the boundaries, not the prototypes, that are central to the assumptions underlying at least one of the important theories about the categories.

Still, it is important to keep in mind that the location of a category boundary is determined not only by the listeners' internal representations (the prototypes) of the categories, but also by the criterion they adopt for deciding between two competing categories, which makes the boundary vulnerable to biasing influences of various kinds. In principle, at least, a change in the location of a boundary may result either from a change in one or the other (or both) of the category prototypes, or from a criterion shift.

It is important to know whether, and under what conditions, the boundaries between phonetic categories are flexible, because the question bears on two very different hypotheses about the processes that underlie the categorization. According to one hypothesis, the perceived categories result from psychophysical discontinuities that directly reflect the characteristics of the auditory system. Thus, given an acoustic-stimulus continuum appropriate for some phonetic distinction, a category boundary is assumed to fall naturally at a point on the continuum where, owing to the way the ear works, differential sensitivity undergoes a sudden change. Perhaps the most general implication of this hypothesis is that auditory categories are the stuff of which phonetic categories are made. Put another way, the implication is that articulatory gestures are so governed as to produce sounds that fit within the categories that the auditory system happens to provide. Accordingly, we will refer to this

as the "auditory" hypothesis. By any name, it is the hypothesis that deals directly with the boundaries of the categories rather than their ideal exemplars or prototypes. As to movement of category boundaries – this is allowed under this hypothesis, but only as a result of psychoacoustic factors that apply to auditory perception in general, and only to the extent that such factors can actually modify the patterns of differential sensitivity on which the auditory boundaries rest.

The other hypothesis is that the boundaries are determined by category prototypes that reflect typical *productions* of the relevant speech segments. Accordingly, the prototypes and the boundaries between them need not conform to boundaries set by discontinuities in the auditory system, but are instead free to be precisely as flexible as the acoustic consequences of the articulatory gestures require. Considerable flexibility may in fact be demanded. The efficiency of phonetic communication depends crucially on the ability of the several articulators to produce successive phonetic segments at the same time (or with considerable overlap), and also to accommodate in other ways to changes in phonetic context and rate. These maneuvers can produce systematic changes in the way a particular phonetic segment is represented in the sound. If the perceiving apparatus were not flexibly responsive to those changes, communication would break down, or so it seems. Moreover, the inventory of phones will itself change as language changes, and this, too, requires flexibility in the prototypes. Our hypothesis is that a link between perception and production (in most general terms) enables the category prototypes to respond appropriately to articulatory or coarticulatory adjustments, and so to mirror the speaker's phonetic intent. Needing a convenient name to refer to this hypothesis, and wishing to distinguish it from the "auditory" hypothesis we described first, we will call it "phonetic."[1]

Our aim in this chapter is to bring together the many data that demonstrate flexibility of a kind the phonetic hypothesis leads us to expect. These pertain to the influences on perceived phonetic boundaries of such factors as phonetic context, speaking rate, the mix of acoustic cues, and linguistic experience. There are other effects on the perceived boundaries, however, about which the auditory and phonetic theories are neutral. These include the consequences of varying the range, frequency, and order of the stimuli, as well as such phenomena as contrast and adaptation. Because effects of that kind need to be distinguished from those that are more directly relevant to the auditory and phonetic theories, we will consider them first. We will note, however, that even these "simple" effects sometimes follow patterns that seem difficult to reconcile with a purely auditory theory, and that suggest that speech-specific perceptual criteria may play a role in certain situations. Our review will be selective and focus especially on these instances.

Stimulus-sequence effects

Under the heading of stimulus-sequence effects we consider influences on the perception of speech stimuli exerted by other, similar stimuli preceding or following

them in a sequence. These effects need to be distinguished from the "stimulus-structure effects" discussed in another section, which concern perceptual dependencies within a single coherent speech stimulus or influences entirely due to factors within the listener.[2]

It is generally agreed that vowel identification – of isolated steady-state vowels, at least – is highly susceptible to all sorts of stimulus-sequence effects. On the other hand, the identification of consonants, and of stop consonants in particular, is more stable and less sensitive to stimulus context. This difference parallels the well-known difference between these two stimulus classes in the extent of CP; indeed, the criterion of "absoluteness" (i.e., independence of surrounding stimuli) constituted part of the classical definition of CP (Studdert-Kennedy et al., 1970). "Context sensitivity" in a sequence may be distinguished on logical grounds, however, from the extent of the subject's reliance on category labels in discriminating between stimuli (Lane, 1965; Repp, Healy, & Crowder, 1979), and these two aspects of CP can, to some extent, be dissociated experimentally (Healy & Repp, 1982).

Local sequential effects

Local sequential effects – typically, influences of a preceding stimulus on the identification of a following stimulus – may occur in any random test sequence. These effects are pervasive in absolute identification, magnitude estimation, and other psychophysical tasks involving nonspeech stimuli. Surprisingly, there have been very few attempts to determine the extent of sequential effects in standard speech-identification tests, where stimuli are presented in random order. Of course, there is an indirect test in the shape of the labeling function, as it can be steep only if sequential effects are relatively small.

In several studies of speech-sound identification, however, the stimuli have been presented in balanced arrangements specifically designed for the assessment of sequential-context effects. In one of the earliest of these studies, Eimas (1963) called for identification of stimuli presented in ABX triads of the sort often used in discrimination tasks and found large context effects for isolated vowels (see also Fry, Abramson, Eimas, & Liberman, 1962) and smaller, but by no means negligible, effects for both the voicing and place dimensions of stop consonants. All effects were contrastive – that is, a stimulus tended to be classified into a category different from that of the stimulus it was paired with – and the magnitude of the effect increased with the acoustic distance between adjacent stimuli. Comparable results have been obtained more recently by, among others, Healy and Repp (1982).

Although sequential effects are generally considered to be common to speech and nonspeech stimuli, there are some intriguing differences. For example, it has been found in several studies that the magnitude of the contrast effect is greater for continua of isolated vowels than for nonspeech continua such as pitch or duration (Eimas, 1963; Healy & Repp, 1982; Fujisaki & Shigeno, 1979; Shigeno & Fujisaki, 1980). Although it is possible that the difference is to be accounted for by the more

complex acoustic (and auditory) nature of the vowels (and there are also problems with comparing the magnitudes of contrast effects across different stimulus continua), it may, with equal plausibility, be taken to reflect a flexibility of categorization peculiar to the class of vowel sounds, a class that happens to carry the major burden of dialectal variation and language change.

If two or more stimuli in a sequence must be held in memory before a response is permitted, as in the procedure of Eimas (1963) described at the beginning of this section, the effects of the stimuli on each other are retroactive as well as proactive. Interestingly, retroactive effects tend to be larger than proactive effects for isolated vowels, while the opposite tends to be the case for all other types of stimuli examined, whether speech or nonspeech (Diehl, Elman, & McCusker, 1978; Healy & Repp, 1982; Shigeno & Fujisaki, 1980). This finding, like the one having to do with the magnitude of contrast, may be explicable by acoustic-stimulus properties alone, or it may reflect a specific tendency, derived perhaps from experience with fluent speech, to revise tentative decisions about vowel categories in the light of later information.

One reason we consider that even simple sequential effects may exhibit speech-specific patterns is that these effects almost certainly take place in two quite distinct ways, one reflecting a sensory effect, the other a judgmental effect (see Simon & Studdert-Kennedy, 1978). That is, there may be an effect of a preceding stimulus on the sensory representation of a following stimulus (as well as the reverse, if both are held in a precategorical memory store), but the judgment of a stimulus may also be affected by the response that was assigned to the preceding or following stimulus, usually in a contrastive fashion. Whereas the purely sensory effects are presumably shared by speech and nonspeech stimuli and are sensitive to factors like spectral similarity and temporal proximity (Crowder, 1981, 1982), the special structure and function of phonetic categories may produce criterion shifts in the response domain that are specific to speech. Although a clear separation of stimulus and response effects has rarely been achieved in speech experiments, separate studies provide evidence for each type. Thus, Crowder (1982) has shown that proactive-contrast effects for isolated vowels decrease with temporal separation over about 3 seconds in a manner that parallels the decay of auditory sensory storage in other paradigms. On the other hand, Sawusch and Jusczyk (1981) found that sequential contrast depended more on the perceived category of the preceding stimulus than on its acoustic structure. Judgmental effects may depend in part on whether or not a response to the contextual stimulus is required: A comparison of Crowder's (1982) data with those of Repp et al. (1979) for isolated vowels suggests that proactive-contrast effects are reduced when only the second stimulus in a pair requires a response. (It goes almost without saying that retroactive-contrast effects would be reduced or eliminated if only the first stimulus in a pair were responded to.)

The distinction between sensory and judgmental components of sequential effects is also familiar in nonspeech psychophysics (see, Petzold, 1981) and is compatible with Braida and Durlach's (1972) two-factor theory of perceptual coding (see Chap-

ter 2 by Macmillan, this volume). Thus, Petzold (1981) has found that preceding stimuli exert a contrastive effect whereas preceding responses exert an assimilative effect. On the other hand, Shigeno and Fujisaki (1980) have proposed a two-factor model for sequential effects in speech and nonspeech that predicts precisely the opposite. The limited data available suggest, on the contrary, that for speech *both* components of sequential effects are contrastive in nature.

Global sequential (range-frequency) effects

Shifts in phonetic-category boundaries may occur as a consequence of variations in the overall composition of a stimulus sequence – that is, the range of stimuli used and the frequency of occurrence of the individual stimuli. In general, if the stimulus range is shifted or expanded in a certain direction, the boundary will shift in the same direction; and if one stimulus (typically one of the endpoints, the "anchor") occurs more frequently than other stimuli, the boundary will shift toward it. In other words, the effects are contrastive in nature, and in the case of speech sounds, they exhibit variations in magnitude similar to those observed for simple sequential effects: For stop consonants varying in place or voicing, the effects are small (Brady & Darwin, 1978; Rosen, 1979), whereas for isolated vowels (Sawusch & Nusbaum, 1979), certain other consonantal contrasts (Repp, 1980), and even for stop consonants in Polish (Keating, Mikos, & Ganong, 1981), they may be quite large.

An interesting asymmetry has been observed in the anchoring paradigm for isolated vowels (Sawusch, Nusbaum, & Schwab, 1980): An analysis of anchoring effects on an /i/-/I/ continuum has suggested that the effect of the /i/ anchor was due to sensory adaptation while that of the /I/ anchor represented a change in response criterion. In a recent and similar study in which the anchor always came first in a stimulus pair and only the second stimulus required a response, Crowder and Repp (1984) found an effect of /i/ but not of /I/. The explanation for this asymmetry may be found in the acoustics of the stimuli; alternatively, it may be owing to the special status of /i/ as one of the corners of the vowel space.

We should note, perhaps, that although range-frequency effects are usually considered to derive from stimulus context beyond the immediate local environment, they are often confounded with sequential probabilities: If a given endpoint stimulus (the anchor) occurs more often than other stimuli, the probability that a given stimulus is immediately preceded by the anchor will be increased relative to an equal-frequency (or a different anchoring) condition. Similarly, if the stimulus range is shifted or expanded in one direction, the likelihood that certain critical stimuli are preceded by other stimuli from that part of the continuum is increased. Therefore, range-frequency effects may in many cases be just local sequential effects in disguise. The extent to which nonlocal stimulus context makes any additional contribution has, to our knowledge, not been ascertained experimentally for speech stimuli. It is possible, however, that the frequent occurrence of a single ·

stimulus has an additional adapting influence not evident in regular balanced stimulus sequences. In that sense, the anchoring paradigm approximates the selective-adaptation paradigm.

Selective adaptation

In selective-adaptation experiments, an adapting stimulus (frequently one or the other endpoint stimulus of a speech continuum) is presented repeatedly many times before responses to a few test stimuli are collected. The original motivation for using this paradigm in speech research was the assumption that the effects of the adapting stimulus might reveal the existence and nature of "phonetic feature detectors" (Eimas & Corbit, 1973; see Chapter 6 by Remez and Chapter 7 by Diehl & Kluender, this volume). Apart from the difficulty of conceiving that phonetic features (place, manner, voicing) could possibly be perceived by detectors that respond to such simple features as the auditory analogs of edges and angles in vision (see, e.g., Diehl, 1981; Studdert-Kennedy, 1981), a large number of experiments suggest that the effect of selective adaptation takes place primarily at the auditory, not the phonetic (judgmental) level. (However, see Elman, 1979.)

The most striking demonstrations of the auditory (as opposed to the phonetic) nature of selective adaptation were provided in two recent studies. In one of these, Roberts and Summerfield (1981) presented audiovisual adapting stimuli that, due to the overriding influence of a conflicting visual display, were never classified into the category normally associated with the auditory stimulus. Nevertheless, the audiovisual adaptors had exactly the same influence on the identification of auditory test stimuli as did purely auditory adaptors. Thus, the phonetic category assigned to the adaptors seemed to play no role in selective adaptation. A similar result was obtained by Sawusch and Jusczyk (1981), who used adaptors of the form /spa/, in which the stop consonant was classified phonetically as *p* but was acoustically identical with the initial *b* in /ba/. The adapting effects of /spa/ and /ba/ did not differ.[3] These studies, together with several earlier attempts to dissociate acoustic and phonetic stimulus properties (Blumstein, Stevens, & Nigro, 1977; Sawusch & Pisoni, 1976), suggest that selective adaptation with speech is an exclusively auditory phenomenon. Even though studies of interaural transfer of adaptation effects suggest more than one site at which adaptation takes place (Ganong, 1978; Sawusch, 1977), both of these sites appear to be auditory (nonphonetic) in nature.

There are two types of evidence, however, that do indicate some involvement of phonetic processing in selective adaptation. One has to do with the influence of the listeners' native language. The relevant finding is that selective-adaptation effects on the same stimulus continuum are different for American and Thai listeners, as independently demonstrated by Donald (1976) and Foreit (1977). The continuum was one of stop consonants varying in voice-onset time (VOT), ranging from prevoiced (voicing lead) to devoiced (0-ms VOT) to aspirated (voicing lag). For

American listeners, who do not distinguish prevoiced and devoiced stops, a −60-ms VOT and a 0-ms VOT adaptor had the same effect on the category boundary. For Thai listeners, on the other hand, who have three separate categories on the continuum, only the 0-ms adaptor affected the devoiced-aspirated boundary while the −60-ms adaptor was ineffective. This finding agrees with earlier results of Cooper (1974) showing that, on a place-of-articulation continuum divided into three categories, adapting stimuli affected only the adjacent but not the remote category boundary.

The other piece of evidence for a role of phonetic categorization in selective adaptation comes from studies that have revealed differences in the effectiveness of adaptors as a function of their distance from the category boundary. In general, the effectiveness of an adaptor increases with its distance from the boundary (Ainsworth, 1977; Cole & Cooper, 1977; Miller, 1977b) unless it crosses another phonetic boundary (Cooper, 1974; Donald, 1976; Foreit, 1977). Of course, this may be just another instance of the well-confirmed fact that the spectral similarity of adaptor and test stimuli is the major determinant of the size of the adaptation effect. In other words, the distance effect may have a purely auditory explanation. In a recent study, however, Miller, Connine, Schermer, and Kluender (1983) demonstrated that even if no other phonetic boundary intervenes, the adaptation effect does not increase indefinitely as the adaptor moves away from the boundary, but instead reaches a maximum and then declines (or, for some subjects, remains on a plateau). The adaptor that produces the maximum effect has characteristics that may reasonably be assumed to be optimal for its category, which led Miller et al. to conjecture that the size of the adaptation effect is related to the adaptor's distance from the listener's internal category prototype. Preliminary support for this hypothesis was obtained by Miller et al. in a condition in which the category boundary on a /ba/-/wa/ continuum, and with it the presumable location of the /wa/ prototype (see Miller & Baer, 1983), was made to shift by reducing the duration of the syllables. The peak in the function relating the size of the adaptation effect to the location of the adaptor on the continuum shifted accordingly, as predicted.

Even stronger support for a role of "category goodness" in selective adaptation comes from a study by Samuel (1982). He first asked his subjects to locate the optimal /ga/ on a /ga/-/ka/ VOT continuum. The subjects were then divided into two groups — those with short-VOT and those with long-VOT prototypes. Two adapting stimuli matching the two average prototypes were then selected. For each group of subjects, the adaptor matching the group's prototype produced the larger boundary shift. Because exactly the same adaptors were used for both groups, the listeners' internal category prototype seemed to be responsible for the magnitude of the adaptation obtained.

These recent results lead to the tentative conclusion that selective adaptation takes place at an auditory level that is *phonetically relevant*. Perhaps this should not come as a surprise. The adapting stimuli, after all, are speech, and therefore *are* phonet-

ically relevant auditory patterns. Conversely, the internal standards or category prototypes against which listeners presumably compare stimuli in the process of categorization must entail detailed auditory specifications; otherwise, in the absence of a common metric, the comparison would be impossible. Selective adaptation may then be viewed as a temporary modification of the prototype itself – a weakening of the critical specifications that is proportional to the degree to which the auditory input meets those specifications. With this interpretation, the results reviewed in this section can be reconciled with the numerous earlier demonstrations of "purely auditory" effects in selective adaptation.

From this vantage point, the various "low-level" effects reviewed so far – sequential contrast, range-frequency effects, and selective adaptation – are relevant to the topic of our chapter, the flexibility of phonetic boundaries. In essence, the data seem to show that not even a psychophysical procedure like selective adaptation has its effects exclusively at a "general auditory level" of processing; rather, as long as the adapting stimuli are speech, their effects reflect the extent to which they engage the speech-processing apparatus. Because speech stimuli ordinarily engage the mechanisms of phonetic categorization (even in the absence of an overt or covert response), selective adaptation with speech is properly viewed as a speech-specific phenomenon – a modification of the frame of reference within which speech stimuli are interpreted. The same is true for range-frequency and sequential-contrast effects, except that overt responses to contextual stimuli may have additional effects at a judgmental level. In other words, although speech must pass through the auditory nerve, there may be no "general auditory" level of representation beyond the peripheral transduction. Speech perception takes place within a preestablished frame of reference, and the auditory representation of speech cannot be separated from the (equally "auditory") internal structures, due to cumulative experience in conjunction with biological predispositions, through which the incoming information is filtered. (See also Chapter 2 by Macmillan and Chapter 13 by Wilson, this volume.)

Stimulus-structure effects

Under this heading we consider perceptual dependencies that arise among different components of a single coherent speech stimulus. That stimulus may be as short as a single syllable or as long as a whole sentence. Stimulus-structure effects, even though they are most easily revealed in the laboratory, are closer to the real-life situation than the stimulus-sequence effects discussed in the preceding section, which represent or exploit artifacts of test-sequence construction. Although the experimental induction of selective adaptation or sequential contrast may be useful for the purpose of probing perceptual mechanisms, there is no reason to believe that these phenomena (as distinct from the mechanism they reveal) play any significant role in the perception of coherent speech. The various effects discussed in the

present section, on the other hand, have more direct implications for normal speech perception, as they reflect the perceptual functions of integration and normalization that make speech perception so effortless and efficient.[4]

Cue-integration effects

Distinctions among phonetic segments are known to rest on a multiplicity of acoustic cues in the speech signal. Typically, these many cues are acoustically diverse, relatively widely distributed in time, and overlap with cues for other segments. Yet the perceiver somehow integrates these diverse and distributed aspects of the speech signal to recover the phonetic structure of the message (Liberman & Studdert-Kennedy, 1978; Repp, Liberman, Eccardt, & Pesetsky, 1978). Exactly how the individual acoustic cues are characterized depends to some extent on the methods of analysis and experimental manipulation and on the descriptive framework chosen by the investigator. From a purely acoustic point of view, however, they seem in most cases to be incoherent. From an articulatory point of view, on the other hand, they make sense – that is, they reflect a unitary event in the domain of articulatory planning.[5]

The statement that there are multiple cues for each phonetic contrast must be qualified by the fact that some cues are more important than others. That is, some cues are easily overriden by others. Listeners' sensitivity to the weaker cues can be demonstrated in the laboratory by eliminating the stronger ones or by setting them at ambiguous values. From the existing evidence it can indeed be concluded that, given the opportunity, listeners will make use of *any* cue for a given phonetic distinction (Bailey & Summerfield, 1980). This general observation suggests that, as Bailey and Summerfield have pointed out, the concept of cue has limited theoretical relevance. As a practical matter it is useful, even essential, in dealing with the acoustic basis of speech perception. But the sensitivity to the many and various cues for a phonetic segment suggests, as we have already implied, that listeners are perceiving just what all the cues have in common – namely, some economical representation of the coherent process underlying the peripheral articulation.

The relevance of cue integration to the topic of our chapter is evident when we consider that a phonetic category boundary is usually determined on a continuum of stimuli varying in only one important cue dimension. The flexibility of that phonetic boundary may then be assessed by introducing other, usually less important, cues that favor either one or the other response alternative. That boundaries are indeed flexible in this particular sense has been demonstrated in numerous studies. (For a recent review, see Repp, 1982.) By definition, phonetic boundaries are located at the point of maximal ambiguity, where weaker cues have their strongest effect. The perceptual-cue integration, or phonetic "trading relation," revealed by the boundary shift generally takes place without the listener's awareness. (The term "trading relation" refers to the fact that the physical settings of several cues to the same

phonetic contrast can be played off against each other, so as to yield the same phonetic identification probabilities.) Perception tends to remain categorical even in the presence of multiple acoustic differences among stimuli (see, e.g., Fitch, Halwes, Erickson, & Liberman, 1980).

The ubiquity of trading relations among acoustically diverse cues provides one of the strongest arguments against theories that predict fixed boundary locations on any acoustic speech continuum. In many cases, cues are so disparate as to be extremely unlikely to engage in any direct psychoacoustic interaction. Rather, what seems to unite them is that they are common consequences of the articulatory gestures that differentiate phonetic segments; at the same time, they are members of the set of structural acoustic differences that characterize a particular phonetic contrast. To cite only one specific example, the primary cue for the prevocalic /s/-/ʃ/ distinction is the spectrum of the fricative noise, but a secondary cue is provided by the voiced formant transitions following the noise. The phonetic boundary on an /s/-/ʃ/ continuum, obtained by varying the spectral properties of the fricative noise, is at different locations depending on whether the formant transitions are appropriate for /s/ or for /ʃ/ (Mann & Repp, 1980). Considering that the fricative noise is of relatively long duration, produced by a different source, and of a spectral composition quite different from that of the following signal, there is little reason to expect any direct effect of the formant transitions on the auditory representation of the fricative noise. Indeed, when listeners are led to focus on the "pitch" of the fricative noise (rather than on the phonetic fricative category), there seems to be no influence of the following formant transitions on their judgments (Repp, 1981). Thus, the perceptual integration of the cues provided by fricative-noise-spectrum and formant transitions seems to be phonetically motivated; this is related to the fact that both cues are consistently correlated with different places of fricative production. Similar arguments may be applied to other phonetic trading relations, even including those that could, in principle, result from some psycho-acoustic interaction.

Feature-integration effects

The trading relations discussed in the preceding section (and reviewed by Repp, 1982) take place among cues to a single phonetic feature – for example, voicing or place of articulation. This is a consequence of the fact that the phonetic categories constituting the endpoints of a speech continuum nearly always differ only in a single feature. Here we consider a related class of effects that reveals perceptual dependencies among cues to different features of the same phonetic segment. The main reason for considering these effects separately is that they give the theorist an additional degree of freedom: Feature interactions may be hypothesized to occur after a process of "feature extraction" but before assembly of the features into a phonetic segment (see, e.g., Miller, 1977a; Sawusch & Pisoni, 1974). For theorists

who instead postulate either direct psychoacoustic interactions among the cues or reference to phoneme- or syllable-sized prototypes, the effects considered here are further instances of cue integration (see Oden & Massaro, 1978; Massaro, Chapter 8, this volume).

The literature on genuine feature-integration effects is rather small, because it is difficult to vary cues for different features in a strictly orthogonal fashion. A well-known finding is that the voicing boundary on a VOT continuum is at increasingly larger voicing lags for labial, alveolar, and velar stop consonants (Lisker & Abramson, 1970). In most studies, however, the duration of the first-formant transition, which itself constitutes a voicing cue (as well as a weak cue for place of articulation) has covaried with place of articulation, so that the boundary shifts may be considered as arising from a simple trading relation among voicing cues. In one experiment, however, the $F1$ transition was held constant (with only the $F2$ and $F3$ transitions varying so as to cue differences in place of articulation), and a small but reliable voicing boundary shift as a function of place of articulation was obtained (Miller, 1977a). (See, however, Massaro & Oden, 1980, for a failure to replicate this result.) Subsequently, Miller (1977a) showed that the boundary on a labial-alveolar place of articulation continuum shifted depending on whether the stop consonants were synthesized as nasal, voiced, or voiceless. She interpreted these results as revealing processing dependencies among phonetic features. An alternative interpretation has been proposed in a model that builds feature dependencies into prespecified critical feature values and so avoids any processing interactions after the feature extraction stage (Oden & Massaro, 1978; Massaro & Oden, 1980). Because of the built-in dependencies, however, the model rests on the assumption of phoneme- or syllable-sized prototypes and merely pays lip service to phonetic features.

Feature interactions of the kind observed by Miller (1977a) presumably reflect the inherent nonorthogonality of articulatory features and their acoustic correlates. Clearly, the binary-feature matrix devised by phonologists is inadequate from a phonetic viewpoint. Because of their longer VOTs, initial velar stops, for example, simply are relatively "more voiceless" than labial stops. The possibility of psychoacoustic interactions among signal components must be considered, but there is no well-supported psychoacoustic explanation for the observed feature interactions.

One case in which a psychoacoustic interaction between feature dimensions can definitely be ruled out is the finding (Carden, Levitt, Jusczyk, & Walley, 1981) that, given a single continuum of formant transitions, listeners place the phonetic boundary at different locations depending on whether they are instructed to hear the stimuli as stops ([ba], [da]) or as fricatives ([fa], [θa]). This can only be accounted for as an adjustment – and apparently a perfectly automatic one – for the fact that the places of production are somewhat different for the two stops from what they are for the fricatives. Hence, it becomes yet another example of the rule that phonetic categorization is guided by internal criteria that reflect the prototypical acoustic and articulatory characteristics of speech.

Segmental-context effects

A third class of perceptual interactions taking place within a single utterance concerns perceptual dependencies among cues for different phonetic segments. Although the conceptual distinction from the two classes we have discussed (integration of cues to the same feature, or to different features of the same segment) is straightforward, practical distinctions are somewhat fuzzy because acoustic cues generally cannot be apportioned exclusively to one or the other phonetic segment. However, an experimental dissociation is usually possible between those signal aspects that provide weak (coarticulatory) cues to one segment and those that are strong and sufficient cues for a different segment, even when both very nearly coincide in time.

For example, take the effect of a following vowel on fricative perception, investigated – among others – by Mann and Repp (1980). The periodic signal portion following a fricative noise necessarily has formant transitions characteristic of the fricative's place of production, which contribute to the fricative percept, particularly when the fricative-noise spectrum carries little distinctive information (Carden et al., 1981; Mann & Repp, 1980). This effect therefore belongs under the heading of cue integration. The identity of the vowel itself, however, is quite independent of the preceding fricative and therefore cannot provide any direct cues to fricative place of production. Nevertheless, as Mann and Repp (1980) and others (Kunisaki & Fujisaki, 1977; Whalen, 1981) have shown, the vowel also exerts an influence on fricative perception: When the fricative noise is ambiguous between /s/ and /ʃ/, listeners report more instances of /s/ when the following vowel is rounded (/u/) than when it is not (/a/), resulting in a quite substantial boundary shift on an /s/-/ʃ/ fricative-noise continuum.

A number of other effects of this kind have been found in recent research. For example, a preceding fricative noise (/s/ versus /ʃ/) affects the perception of a following stop consonant (/t/ versus /k/): The /t/-/k/ boundary shifts in favor of /k/ when the precursor is /s/ (Mann & Repp, 1981). The effect is independent of coarticulatory cues to stop place of articulation in the fricative noise, and it occurs also when the fricative appears to belong to a preceding syllable (Repp & Mann, 1981). Yet another effect operating across a syllable boundary has been obtained by Mann (1980): The boundary on a /da/-/ga/ continuum shifts in favor of /g/ when the preceding syllable is /al/ rather than /ar/.

How are such segmental context effects to be accounted for? Psychoacoustic interactions between adjacent signal portions, although not impossible, become rather implausible. For example, there is little reason to expect that a fricative noise would "sound" different before different vowels. Indeed, when listeners are required to judge the "pitch" of the noise rather than the phonetic category of the fricative, effects of the following vowel disappear (Repp, 1981). The most plausible hypothesis is that segmental-context effects represent a perceptual compensation for coarticulatory interactions in speech production. For example, anticipatory lip

rounding for rounded vowels is known to affect the noise spectrum of preceding fricatives (Fujisaki & Kunisaki, 1978; Mann & Repp, 1980), and there are indications that the formant transitions of stop consonants shift with the place of articulation of preceding fricatives (Repp & Mann, 1982) and liquids (Mann, 1980). The ability of listeners to compensate for these coarticulatory effects implies an internal representation of these dependencies, which may be conceptualized in dynamic or static terms.

Segmental-context effects have been demonstrated even among nonadjacent segments. Thus, shifts in the place-of-articulation boundaries for initial stop consonants have been found to occur as a function of the place of articulation of the final stop consonant in the same syllable (Alfonso, 1981). Perceptual interdependencies between two vowels separated by a consonant have also been reported (Kanamori, Kasuya, Arai, & Kido, 1971). These effects may reflect perceptual compensation for coarticulatory dependencies operating over wider time spans (see Martin & Bunnell, 1981, 1982; Öhman, 1966).

Speaking-rate effects

The perception of phonetic distinctions that rest on temporal cues may be affected by the temporal structure of surrounding signal portions. As these effects have been thoroughly reviewed by Miller (1981), we can be brief here.

It is useful to distinguish between experimental manipulations of the duration of selected (steady-state) acoustic segments and of time-varying spectral changes connected with actual (or simulated) changes in articulatory rate. Both temporal and spectrotemporal manipulations have been shown to affect the perception of certain temporal cues, but it is not clear whether their effects take place at the same level.

Some experiments on effects of "speaking rate" concern trading relations among cues for the same phonetic segment. When two temporal cues contribute to the same distinction, a change in one will necessarily require a compensatory change in the other to maintain perceptual constancy. An example of such a trading relation is that between (preceding) silence duration and fricative-noise duration as joint cues to the fricative-affricate distinction (Repp et al., 1978). Affricate percepts are favored by both long silences and short noises, so an increase in silence duration can be compensated for, within limits, by an increase in noise duration. But when this trading relation was examined in the context of a true rate manipulation – the critical cues were embedded in sentence frames produced at a fast or at a slow rate – relatively *more* silence was needed in the fast sentence frame to maintain the same level of affricate responses. One possible interpretation of this reliable effect (cf. Dorman, Raphael, & Liberman, 1979) is that, in the rapidly articulated context, the (constant) fricative noise sounded relatively longer and hence more fricative-like, so that a longer silence was required to restore the same level of affricate responses. This assumes that the perception of the silence cue was less affected by the rate manipulation. Why this should be so is not clear at present. We should also remark that the speaking-rate effect was probably mediated primarily by the immediately

adjacent signal portions – the durations of the vocalic segments preceding the silence and following the fricative noise. If so, the speaking-rate effects observed may have been a special instance of a segmental-context effect or even a trading relation.

A good example of another "speaking-rate effect" that could just as well be put in the preceding section on segmental-context effects is the influence of the duration of a following vowel on the perception of the /b/-/w/ distinction cued by varying formant transition duration (Miller & Liberman, 1979): The longer the vowel, the longer the formant transition duration at the /b/-/w/ boundary. This finding was interpreted as a speaking-rate effect, and it is indeed consistent with observed changes in /w/ transition duration at different speeds of articulation (Miller & Baer, 1983). However, the effect has also been obtained with infants (Eimas & Miller, 1980) and with nonspeech stimuli (Pisoni, Carrell, & Gans, 1983), which suggests a possible psychoacoustic origin – in other words, a temporal normalization early in the perceptual process. It is indeed questionable whether changes in the duration of a (steady-state) synthetic vowel are sufficient to convey anything like "speaking rate." Within the context of cue-trading relations, both Fitch (1981) and Soli (1982) have been able to separate perceptual effects of vowel duration from effects due to vowel "structure," which are more complex spectral changes taking place over time. It is the latter that are more properly viewed as the carriers of information about rate of articulation.

The examples given in this section illustrate that true "speaking-rate effects" are not easy to distinguish from simpler temporal trading relations and local-context effects. Moreover, if speaking rate is varied, those changes that occur closest to the target segment will affect its perception most (Summerfield, 1981). In addition, Miller, Aibel, and Green (1984) have recently demonstrated that listeners' overt judgments of speaking rate do not predict the perceptual effects of rate manipulations. On the other hand, considering the extensive speech knowledge that listeners must possess, it seems reasonable to assume that they also have intrinsic knowledge of the acoustic changes that accompany changes in speaking rate and that they "know" how to apply this knowledge in perception. An example of this was also provided by Miller and Liberman (1979) in their study of the /b/-/w/ distinction. When the following vowel was extended by a nonstationary portion containing transitions appropriate for a syllable-final /d/, the effect on the /b/-/w/ boundary was equivalent to that of *shortening* the steady-state vowel. This paradoxical finding presumably reflects an increase in the perceived rate of articulation caused by the additional phonetic segment in the syllable.

Speaker-normalization effects

Phonetic boundaries along a spectral cue dimension may shift in accordance with the size of the vocal tract that is perceived to be the source of the utterance – that is the hypothesis, at least. As with speaking-rate effects, genuine speaker-normalization effects are not easy to distinguish from local-context effects and spectral trading

relations. Moreover, a demonstration of true speaker normalization requires that the test utterance representing different sources be perceived as coming from a single source (speaker), which is possible only with target segments that are relatively ambiguous as to their source. For these reasons, there are few convincing demonstrations of speaker-normalization effects in the literature.

One of the earliest demonstrations was provided by Ladefoged and Broadbent (1957), who showed that synthetic vowel targets were perceived differently in sentence carriers simulating different speakers. This result was replicated with natural speech by Dechovitz (1977). More recently, May (1976) with synthetic speech and Mann and Repp (1980) with natural speech found a shift in the /ʃ/-/s/ boundary when the same fricative noises occurred in the context of vowels produced by different-sized vocal tracts. More experiments along these lines are needed to establish firmly listeners' sensitivity to the static aspects of the perceived speech source.

Semantic and syntactic effects

It is a commonplace observation that listeners tend to hear what they expect to hear. Effects of semantic context are ubiquitous in speech perception (Bagley, 1900–01; Cole & Rudnicky, 1983). However, these effects are generally obtained only when some acoustic information is missing and needs to be "filled in." Apparently, semantic factors can also influence the phonetic boundary on an acoustic continuum characterized by ambiguous (rather than missing) cues.

That such factors can influence the category boundary on a VOT continuum was demonstrated by Ganong (1980). He found that the boundary shifted in favor of word responses when one of the alternatives was a word and the other a nonword, even though the phonetic distinction was in the initial consonant. The pattern of the data suggested that the effect was not merely a response bias; rather, lexical status seemed to influence phonetic categorization directly. But this kind of direct interaction between "top-down" and "bottom-up" processes is a controversial notion (see, e.g., Swinney, 1982), and we do not wish to enter into a discussion of the matter here. Suffice it to point out that phonetic boundaries may be shifted by semantic biases. Such biases can be manipulated not only by changing the lexical status of the target word but also by inducing expectations through preceding sentence context (Garnes & Bond, 1977; Miller, Green, & Schermer, 1982). However, the phonetic-boundary shift obtained in that case may be eliminated by selective attention to the target word (Miller et al., 1982), suggesting that semantic processing can be consciously avoided in certain conditions (e.g., when the same materials are repeated over and over). Interestingly, the same study by Miller et al. (1982) also revealed that effects on segmental perception due to the speaking rate of a carrier sentence could not be voluntarily disengaged.

Effects of syntactic boundaries on certain phonetic distinctions have also been reported (Dechovitz, 1979; Price and Levitt, 1983): If the critical cue for the distinction is silence duration (as in the fricative-affricate contrast), more silence is

needed if a syntactic boundary is made to coincide with the silence. Although claims have been advanced that this effect can be produced by syntactic structure *per se* (Dechovitz, 1979), no convincing evidence for such "pure syntax" effects exists so far. Rather, the effects of syntactic boundaries seem to be mediated by the prosodic changes that accompany them. The fricative-affricate boundary may shift depending on whether the preceding word has clause-final intonation and lengthening (Price & Levitt, 1983; see also Rakerd, Dechovitz, & Verbrugge, 1982). To what extent these effects should be considered merely local-context effects or temporal trading relations remains to be seen. In either case, they seem genuinely phonetic rather than psychoacoustic.

Cross-language effects

For the purpose of ruling out psychoacoustic factors and establishing that the location of a phonetic boundary is largely determined by factors internal to the listener, cross-language comparisons are most instructive. Languages do differ in their articulatory-acoustic patterns, frequently even for phonetic categories that seem phonemically identical (see Ladefoged, 1983). To the extent that these cross-linguistic differences are captured by a single acoustic speech continuum (and this is not always the case), we should want to know whether the phonetic boundaries in fact differ for speakers of different languages.

Unfortunately, cross-linguistic studies using the same stimuli and procedures are not very numerous. Among those that do exist, most have dealt with the voicing dimension, as cued by VOT, taking advantage of the fact that languages such as English, French, and Thai make their voicing contrasts in phonetically different ways. Whereas English distinguishes voiced (either prevoiced or voiceless unaspirated) and voiceless aspirated stops, French, Spanish, and Polish contrast prevoiced with voiceless unaspirated stops and Thai makes both distinctions. The single voicing boundary for English listeners is located in the short-lag values of VOT, between roughly 20 and 40 ms, depending on place of articulation (Lisker & Abramson, 1970). The single boundary for French, Spanish, and Polish listeners, on the other hand, is generally located at shorter lag times, close to zero, and is considerably more variable (Caramazza, Yeni-Komshian, Zurif, & Carbone, 1973; Keating et al., 1981; Williams, 1977). Thai listeners have two boundaries, one in the voicing-lead region (where none of the other languages mentioned exhibits any boundary), and the other at voicing lags somewhat longer than in English (Lisker & Abramson, 1970; Foreit, 1977). Thus, native language does seem to influence the location of comparable phonetic boundaries on a VOT continuum, and it certainly determines whether a boundary exists at all.

There is ample evidence that discrimination performance is best in the vicinity of a phonetic boundary. Thus, discrimination peaks shift with the phonetic boundaries across languages. Speakers of a language such as Thai have a discrimination peak in the voicing-lead region where English listeners' ability to detect differences is extremely poor (Abramson & Lisker, 1970). Another well-known example of such

a cross-language difference is provided by the /r/-/l/ contrast, which is easily discriminated by English listeners but nearly indistinguishable for speakers of Japanese, a language that does not contain these phonetic segments (Miyawaki et al., 1975). For a review of these and related data, see Strange and Jenkins (1978) and Repp (1984).

In view of the flexibility of phonetic boundaries, demonstrations of a coincidence of category boundaries obtained for chinchillas or monkeys with those of English-speaking humans lose some of their impact. To the extent that these animal boundaries are stable at all (see Waters & Wilson, 1976, for a demonstration of large range effects), they may reveal certain psychoacoustic sensitivities that, however, seem to exert only a weak constraint on the possible locations of human boundaries. (See also Chapter 11 by Snowden and Chapter 12 by Kuhl, this volume.)

It is likely, of course, that the locations of phonetic boundaries in the languages of the world are not totally arbitrary. The structure of the speech-production apparatus imposes universal constraints on articulation that may be reflected in a limited number of preferred boundary locations. The hypothesis that human infants may possess some innate sensitivity to these universal *potential* phonetic boundaries (see Aslin & Pisoni, 1980) has recently gained momentum through the remarkable findings of Werker and Tees (1984), who showed that prelinguistic American infants are capable of distinguishing phonetic categories foreign to English, but lose that ability around ten months of age. It has not been conclusively established, however, that these prelinguistic category distinctions are truly phonetic, rather than psychoacoustic, in nature. Exposure to the phonetic distinctions of the native language may merely induce a "speech mode" of listening in the one-year-old infant and thereby lead it to ignore irrelevant acoustic detail. Similarly, several demonstrations of adults' ability to discriminate foreign phonetic categories in certain laboratory situations (MacKain, Best, & Strange, 1981; Pisoni, Aslin, Perey, & Hennessy, 1982) may, at least in part, reflect skills of deploying a nonphonetic mode of processing, and not the acquisition of a new phonetic distinction that can be generalized beyond the laboratory. On the other hand, mastery of a new language does imply the establishment of new phonetic categories, and it is primarily a matter of implementing all the necessary controls to permit the conclusion that this is indeed what has happened in any given laboratory experiment. Rigorous investigations of the process of phonetic learning, which may be a good deal slower than the time span of the typical speech experiment, are just beginning (e.g., Flege & Port, 1981).

Conclusion

Evidence from a variety of experiments on speech perception establishes that phonetic category boundaries are flexible in response to each of two quite different sets of conditions. One set is commonly created by the way utterances are arranged in experiments that require the presentation of sequences of test stimuli. Most of the

effects of such conditions are found with nonspeech sounds as well, though, for reasons that are not yet clear, some may be peculiar to speech. The other conditions are the more interesting, at least for our purposes, because they seem to be integral parts of the processes by which utterances are perceived in any test sequence and so, presumably, in the real-life situation. Their effects are of several superficially different kinds, but, common to all, there is a (more or less) apparent correspondence between the shift in the perceived category boundary and the acoustic effects of an articulatory or coarticulatory maneuver. Thus, these boundary shifts imply a link between speech perception and speech production, much as if perception were constrained by tacit "knowledge" of what a vocal tract does when it makes linguistically significant gestures. Considerations of this kind, roughly similar to those that led originally to the (so-called) "motor theory of speech perception" (Liberman, Delattre, & Cooper, 1952), lead us to suppose that such boundary shifts as these are peculiar to speech.

Notes

1. We are uncertain where to place in the present framework another important class of hypotheses, that of acoustic invariance (Stevens & Blumstein, 1978; Kewley-Port, 1983; Lahiri, Gewirth, & Blumstein, 1984). Sometimes invariant properties are described in terms that suggest a boundary-oriented approach – for example, when a spectral shape is considered to be either rising or falling. On the other hand, the use of optimal "templates" (Stevens & Blumstein, 1978) suggests a prototype-oriented approach. Because the invariance hypothesis postulates invariant acoustic correlates for linguistically distinctive features, it would seem to permit little flexibility in category boundaries, particularly if the boundaries themselves are taken to be the invariant correlates.

2. Not all the studies we will cite actually examined boundary shifts. Some studies showed only that the perception of a single ambiguous stimulus could be influenced in one or the other direction. It is safe to infer, however, that had that stimulus been part of an acoustic continuum, the category boundary on that continuum would have shifted in precisely the same direction.

3. In the same study, however, sequential contrast was found to be contingent on the perceived phonetic category; in other words, the effect of /spa/ differed from that of /ba/ (Sawusch & Jusczyk, 1981). It is worth noting that in the selective-adaptation paradigm the adaptors are typically presented at a fast rate that may discourage even covert categorization. Phonetic (judgmental) effects may be contingent upon overt or covert labeling of contextual stimuli.

4. We call them perceptual *functions*, rather than perceptual *processes*, because we believe that these accomplishments of the perceptual system should not be viewed in process terms. In any case, whatever neural or cognitive processes may underlie these functions is totally unknown at present.

5. Although there have been persistent attempts to conceptualize single "invariant" acoustic properties for distinctive features in speech (see Stevens & Blumstein, 1978; Kewley-Port, 1983; Lahiri et al., 1984), these properties never fully capture the phonetically distinctive information. It seems to be a fact to be accepted that what may be a unitary event at the level of linguistic structure or articulatory planning emerges in a fractionated form at the level of acoustic description.

References

Abramson, A. S., & Lisker, L. (1970). Discriminability along the voicing continuum: Cross-language tests. *Proceedings of the Sixth International Congress of Phonetic Sciences* (pp. 569–573). Prague: Academia.

Ainsworth, W. A. (1977). Mechanisms of selective feature adaptation. *Perception & Psychophysics, 21,* 365–370.

Alfonso, P. (1981). Context effects on the perception of place of articulation. *Journal of the Acoustical Society of America, 69* (Suppl. 1), S93 (Abstract).

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

Bagley, W. C. (1900) The apperception of the spoken sentence: A study in the psychology of language. *American Journal of Psychology, 12,* 80–130.

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

Blumstein, S. E., Stevens, K. N., & Nigro, G. N. (1977). Property detectors for bursts and transitions in speech perception. *Journal of the Acoustical Society of America, 61,* 1301–1313.

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

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

Caramazza, A., Yeni-Komshian, G. H., Zurif, E. B., & Carbone, E. (1973). The acquisition of a new phonological contrast: The case of stop consonants in French-English bilinguals. *Journal of the Acoustical Society of America, 54,* 421–428.

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

Cole, R. A., & Cooper, W. E. (1977). Properties of friction analyzers for /j/. *Journal of the Acoustical Society of America, 62,* 177–182.

Cole, R. A., & Rudnicky, A. I. (1983). What's new in speech perception? The research and ideas of William Chandler Bagley, 1874–1946. *Psychological Review, 90,* 94–101.

Cooper, W. E. (1974). Adaptation of phonetic feature analyzers for place of articulation. *Journal of the Acoustical Society of America, 56,* 617–627.

Crowder, R. G. (1981). 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.

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

Crowder, R. G., & Repp, B. H. (1984). Single formant contrast in vowel identification. *Perception & Psychophysics, 35,* 372–378.

Dechovitz, D. (1977). Information conveyed by vowels: A confirmation. *Haskins Laboratories Status Report on Speech Research,* SR-51/52, 213–219.

Dechovitz, D. (1979). Effects of syntax on the perceptual integration of segmental features. In J. J. Wolf & D. H. Klatt (Eds.), *Speech communication papers presented at the 97th Meeting of the Acoustical Society of America* (pp. 319–322). New York: Acoustical Society of America.

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

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

Donald, L. (1976). The effects of selective adaptation on voicing in Thai and English. *Haskins Laboratories Status Report on Speech Research,* SR-47, 129–136.

Dorman, M. F., Raphael, L. J., & Liberman, A. M. (1979). Some experiments on the sound of silence in phonetic perception. *Journal of the Acoustical Society of America, 65,* 1518–1532.

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

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

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

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

Fitch, H. L. (1981). Distinguishing temporal information for speaking rate from temporal information for intervocalic stop consonant voicing. *Haskins Laboratories Status Report on Speech Research*, SR-65, 1–32.

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

Flege, J. E., & Port, R. Cross-language phonetic interference: Arabic to English. (1981). *Language and Speech, 24*, 125–146.

Foreit, K. G. (1977). Linguistic relativism and selective adaptation for speech: A comparative study of English and Thai. *Perception & Psychophyics, 21*, 347–351.

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

Fujisaki, H., & Kunisaki, O. (1978). Analysis, recognition, and perception of voiceless fricative consonants in Japanese. *IEEE Transactions (ASSP), 26*, 21–27.

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

Ganong, W. F. III. (1978). The selective adaptation effects of burst-cued stops. *Perception & Psychophysics, 24*, 71–83.

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

Ganong, W. F. III., & Zatorre, R. J. (1980). Measuring phoneme boundaries four ways. *Journal of the Acoustical Society of America, 68*, 431–439.

Games, S., & Bond, Z. S. (1977). The influence of semantics on speech perception. *Journal of the Acoustical Society of America, 61* (Suppl. 1), S65 (Abstract).

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

Kanamori, Y., Kasuya, H., Arai, S., & Kido, K. (1971). Effect of context on vowel perception. *Proceedings of the Seventh International Congress on Acoustics*, Budapest: 37–40.

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

Kewley-Port, D. (1983). Time-varying features as correlates of place of articulation in stop consonants. *Journal of the Acoustical Society of America. 73*, 322–335.

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

Ladefoged. P. (1983). Cross-linguistic studies of speech production. In P. F. MacNeilage (Ed.), *The production of speech* (pp. 177–188). New York: Springer.

Ladefoged, P., & Broadbent, D. E. (1957). Information conveyed by vowels. *Journal of the Acoustical Society of America, 29*, 98–104.

Lahiri, A., Gewirth, L., & Blumstein, S. E. (1984). A reconsideration of acoustic invariance for place of articulation in stop consonants: Evidence from a cross-language study. *Journal of the Acoustical Society of America, 76*, 391–404.

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

Liberman, A. M., Delattre, P. C., & Cooper, F. S. (1952). The role of selected stimulus-variables in the perception of the unvoiced stop consonants. *American Journal of Psychology, 65*, 497–516.

Liberman, A. M., & Studdert-Kennedy, M. (1978). Phonetic perception. In R. Held, H. W. Leibowitz,

& H.-L. Teuber (Eds.), *Handbook of sensory physiology, Vol. VIII: Perception* (pp. 143–178). New York: Springer.

Lisker, L., & Abramson, A. S. (1970). The voicing dimension: Some experiments in comparative phonetics. *Proceedings of the 6th International Congress of Phonetic Sciences* (pp. 563–567). Prague: Academia.

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

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

Mann, V. A., & Repp, B. H. (1980). Influence of vocalic context on perception of the [ʃ]-[s] distinction. *Perception & Psychophysics, 28*, 213–228.

Mann, V. A., & Repp, B. H. (1981). Influence of preceding fricative on stop consonant perception. *Journal of the Acoustical Society of America, 69*, 548–558.

Martin, J. G., & Bunnell, H. T. (1981). Perception of anticipatory coarticulation effects. *Journal of the Acoustical Society of America, 69*, 559–567.

Martin, J. G., & Bunnell, H. T. (1982). Perception of anticipatory coarticulation effects in vowel-stop consonant-vowel sequences. *Journal of Experimental Psychology: Human Perception and Performance, 8*, 473–488.

May, J. Vocal tract normalization for /s/ and /ʃ/. (1976). *Haskins Laboratories Status Report on Speech Research, SR-48*, 67–73.

Massaro, D. W., & Oden, G. C. (1980). Evaluation and integration of acoustic features in speech perception. *Journal of the Acoustical Society of America, 67*, 996–1013.

Miller, J. L. (1977a). Nonindependence of feature processing in initial consonants. *Journal of Speech and Hearing Research, 20*, 519–528.

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

Miller, J. L. (1981). The effect of speaking rate on segmental distinctions: Acoustic variation and perceptual compensation. In P. D. Eimas & J. L. Miller (Eds.), *Perspectives on the study of speech*. Hillsdale, NJ: Erlbaum.

Miller, J. L., Aibel, I. L., & Green, K. (1984). On the nature of rate-dependent processing during phonetic perception. *Perception & Psychophysics, 35*, 5–15.

Miller, J. L., & Baer, T. (1983). Some effects of speaking rate on the production of /b/ and /w/. *Journal of the Acoustical Society of America, 73*, 1751–1755.

Miller, J. L., Connine, C. M., Schermer, T. M., & Kluender, K. R. (1983). A possible auditory basis for internal structure of phonetic categories. *Journal of the Acoustical Society of America, 73*, 2124–2133.

Miller, J. L., Green, K., & Schermer, T. (1982). On the distinction between prosodic and semantic factors in word identification. *Journal of the Acoustical Society of America, 71* (Suppl. 1), S95 (Abstract).

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

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

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

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

Petzold, P. (1981). Distance effects on sequential dependencies in categorical judgment. *Journal of Experimental Psychology: Human Perception and Performance, 7*, 1371–1385.

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

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

Price, P. J., & Levitt, A. G. (1983). The relative roles of syntax and prosody in the perception of the /s-č/ distinction. *Language and Speech, 26,* 291–304.

Rakerd, B., Dechovitz, D. R., & Verbrugge, R. R. (1982). An effect of sentence finality on the phonetic significance of silence. *Language and Speech, 25,* 267–282.

Repp, B. H. (1980). A range-frequency effect on perception of silence in speech. *Haskins Laboratories Status Report on Speech Research, SR-61,* 151–166.

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

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

Repp, B. H. (1984). Categorical perception: Issues, methods, findings. In N. J. Lass (Ed.), *Speech and language: Advances in basic research and practice* (Vol. 10). New York: Academic Press.

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

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

Repp, B. H., & Mann, V. A. (1981). Perceptual assessment of fricative-stop coarticulation. *Journal of the Acoustical Society of America, 69,* 1154–1163.

Repp, B. H., & Mann, V. A. (1982). Fricative-stop coarticulation: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America, 71,* 1562–1567.

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

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

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

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

Sawusch, J. R. (1977). Peripheral and central processing in speech perception. *Journal of the Acoustical Society of America, 62,* 738–750.

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

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

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

Sawusch, J. R., & Pisoni, D. B. (1974). On the identification of place and voicing features in synthetic stop consonants. *Journal of Phonetics, 2,* 181–194.

Sawusch, J. R., & Pisoni, D. B. (1976). Response organization and selective adaptation to speech sounds. *Perception & Psychophysics, 20,* 413–418.

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

Simon, H. J., & Studdert-Kennedy, M. (1978). Selective anchoring and adaptation of phonetic and nonphonetic continua. *Journal of the Acoustical Society of America, 64,* 1338–1357.

Soli, S. D. (1982). Structure and duration of vowels together specify fricative voicing. *Journal of the Acoustical Society of America, 72,* 366–378.

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

Strange, W., & Jenkins, J. J. (1978). Role of linguistic experience in the perception of speech. In R. D.

Walk & H. L. Pick, Jr. (Eds.), *Perception and experience* (pp. 125–169). New York: Plenum Press.

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

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

Summerfield, Q. (1981). Articulatory rate and perceptual constancy in phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance, 7*, 1074–1095.

Swinney, D. A. (1982). The structure and time-course of information interaction during speech comprehension, lexical segmentation, access, and interpretation. In J. Mehler, E. C. T. Walker, & M. Garrett (Eds.), *Perspectives on mental representation* (pp. 151–167). Hillsdale, NJ: Erlbaum.

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

Werker, J. F. & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 7*, 49–63.

Whalen, D. H. (1981). Effects of vocalic formant transitions and vowel quality on the English [s]-[š] boundary. *Journal of the Acoustical Society of America, 69*, 275–282.

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