Psychological Bulletin
1982, Vol. 92, No. 1, 81-110

Phonetic Trading Relations and Context Effects: New Experimental Evidence for a Speech Mode of Perception

Bruno H. Repp
Haskins Laboratories, New Haven, Connecticut

This article reviews a variety of recent experimental findings that show that the perception of phonetic distinctions relies on the integration of multiple acoustic cues and is sensitive to the surrounding context in very specific ways. Most of these effects have correspondences in speech production and are readily explained by the assumption that listeners make continuous use of their tacit knowledge of speech patterns. A general auditory theory that does not make reference to the specific origin and characteristics of speech can, at best, handle only a small portion of the phenomena reviewed here. Special emphasis is placed on studies that obtained different patterns of results depending on whether the same stimuli were perceived as speech or as nonspeech. These findings provide strong empirical evidence for the existence of a speech-specific mode of perception.

Speech is a specifically human capacity. Just as humans are uniquely enabled to produce the complex stream of sound called *speech*, one might suppose that they make use of special perceptual mechanisms to decode it. Because speech is so remarkably different from all other environmental sounds, it is perhaps a truism that some perceptual and cognitive processes occur only when speech is the input. Otherwise, speech would not be perceived as what it is. To make sense, the question of whether speech perception is different from other forms of perception is best restricted to those aspects of speech that are not obviously unique (e.g., to speech as an acoustic signal that can be described in the same physical terms as other environmental sounds). Then one can ask whether the perceptual translation of this acoustic signal into the sequence of discrete linguistic units that we experience (i.e., phonetic perception) requires special mechanisms, or whether it can be reduced to a combination of auditory processes involved also in perceiving and interpreting nonspeech sounds. The second alternative presupposes that phonetic categories, though appropriate only for speech, are not essentially unique but rather can be viewed as labels applied to specific auditory patterns. This may be wrong, but it must be granted if the argument is to be taken seriously.

The precise nature of the processes and mechanisms that support phonetic perception has been much discussed. One view is that speech perception is special in that it takes account of the origin of the signal in the action of a speaker's articulatory system. This view underlies the motor theory of speech perception (e.g., Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967) as well as the theory of analysis-by-synthesis (Halle & Stevens, 1959). More recently, it has been augmented by ideas derived from Gibson's (1966) theory of event perception (see, e.g., Bailey & Summerfield, 1980; Neisser, 1976; Summerfield, 1979), which postulates that all perception is directed toward the source of stimulation. Although in the Gibsonian framework speech perception is not seen as basically different from the perception of other auditory (and visual) events, the special nature of the source (the human vocal tract) is acknowledged and em-

Preparation of this article was supported by National Institute of Child Health and Human Development Grant HD01994 and BRS Grant RR05596 to Haskins Laboratories.

Valuable comments on an earlier draft were obtained from Carol Fowler, Alvin Liberman, Michael Studdert-Kennedy, and Janet Werker. A special contribution was made by Peter Jusczyk, whose comments were exceptionally thorough and constructive. My intellectual debt to Alvin Liberman must be evident throughout; nevertheless, I am solely responsible for the shortcomings of this article.

Requests for reprints should be sent to Bruno H. Repp, Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06510.

phasized. In this view, speech perception is special because the source of speech is special.

Other researchers would concur with only the second half of the preceding statement. Although they recognize the constraints introduced by the articulatory system on speech production, they pursue the hypothesis that the processes involved in speech perception are essentially the same as those that support the auditory perception of nonspeech sounds. In this view, the specific complexity of speech perception results merely from the diversity and the number of elementary auditory processes required to deal with an intricately structured input (see, e.g., Divenyi, 1979; Kuhl & Miller, 1978; Pastore, 1981; Schouten, 1980).

These two contrasting views are perhaps most clearly distinguished by their different orientations to the evolution of speech perception: Whereas, according to the first view, special perceptual processes evolved hand in hand with articulatory capabilities to handle the complex output of a speaker's vocal tract, the second view assumes that the vocal productions of early hominids were fitted into a preexisting mold created by the sensitivities and limitations of their auditory systems.

Which of these two views is correct is, in part, an empirical question that rests on many possible sources of evidence, including the reactions to speech of animal and infant subjects, traditional laboratory experiments, and electrophysiological and clinical observations. In this review, I focus on the most recent attempts to demonstrate the peculiarities of speech perception in the laboratory with normal adult human subjects. This kind of evidence has been, and continues to be, central to the argument, as it is easier to obtain, permits a variety of approaches, and is perhaps more readily interpreted than some of the other research. This is not to deny that some of the most crucial results will come from infant and animal experiments; nevertheless, this research characteristically lags one step behind the standard laboratory findings, and studies that extend the latest findings on college students' perception to other subject populations are just getting under way as this review is being written.

Less than a decade ago, a rich set of experimental data suggested the existence of a special speech mode of perception that was distinct from other kinds of auditory perception. Within a few years, however, that support seems to have all but evaporated. The history of these events is summarized and commented on in the first part of the present article. Insofar as the main purpose of that section is to set the stage for the following review, my treatment of what are complex and often controversial issues is necessarily sketchy and betrays my biases. In the second part, new evidence—most of it collected over the last few years—is reviewed and discussed. It leads to the conclusion that we have, once again, strong experimental support for a phonetic mode of perception.

The Old Evidence

In a well-known article, Wood (1975) listed six laboratory phenomena that, at that time, seemed to provide strong converging evidence for the existence of processes that are unique to speech perception. One phenomenon is the "phoneme boundary effect," which is commonly subsumed under the more general term, *categorical perception.* It is the finding that two speech stimuli are easier to distinguish when they can be assigned to different linguistic categories than when, though separated by an equivalent physical difference, they are perceived as belonging to the same category. A second phenomenon is selective adaptation, the shift of the category boundary on a synthetic speech continuum following repeated presentation of one endpoint stimulus. Three other phenomena have to do with hemispheric specialization: the dichotic right-ear advantage, the right-ear advantage in temporal-order judgments of speech stimuli, and differences in evoked potentials from the two hemispheres in response to speech stimuli. A sixth phenomenon concerned asymmetric interference between auditory and phonetic stimulus dimensions in a speeded classification task. Many of the findings that Wood referred to under these headings have been excellently reviewed by Studdert-Kennedy (1976).

At the time the Wood (1975) and Studdert-Kennedy (1976) articles were written, all of the above-named phenomena seemed to be specific to speech; that is, they were apparently not obtained with nonspeech stimuli. A few years later, however, the picture had changed considerably. Using Wood's enumeration of findings as their starting point, both Cutting (1978) and Schouten (1980) reviewed more recent research using the various paradigms and concluded independently that there was no evidence for a special phonetic mode of perception. After that statement, the views of these two authors diverge: Cutting, a vigorous proponent of the Gibsonian view, argues for considering speech perception as merely one instance of auditory event perception (i.e., the perception of auditory events other than speech may be as—or nearly as—complex and special as speech perception). On the other hand, Schouten, who represents a more narrowly psychophysical orientation, states rather bluntly that "speech and non-speech auditory stimuli are probably perceived in the same way" (p. 71), implying that all perception through the ears rests on the same elementary auditory processes.

The conclusions of both authors reflect their disillusion over the failure of a number of experimental techniques to produce results that are specific to speech. Because the relevant evidence has been competently reviewed by them and by others, I deal with it only briefly, focusing primarily on its interpretation.

Categorical Perception

The "phoneme boundary effect" singled out by Wood (1975)—the enhanced discriminability across the phonetic boundary on a synthetic speech continuum—is merely one aspect of the complex phenomenon termed *categorical perception*, which includes sharp category boundaries, reduced context sensitivity, and predictability of discrimination performance from identification scores (Studdert-Kennedy, Liberman, Harris, & Cooper, 1970). The speech specificity of the phoneme boundary effect has been challenged on the ground that analogous effects have been demonstrated for a variety of nonspeech continua: noise–buzz sequences (Miller, Wier, Pastore, Kelly, & Dooling, 1976), tone onset time (Pisoni, 1977), tone amplitude in the presence of a reference signal and visual flicker (Pastore et al., 1977), musical intervals (Burns & Ward, 1978), and amplitude rise time (Cutting & Rosner, 1974). The results for the rise time ("pluck"–"bow") continuum, which have been widely cited and followed up and on which Cutting (1978) rested his whole argument, have recently been claimed to be artifacts due to faulty stimulus construction (Rosen & Howell, 1981), but the other findings appear to be solid. Nevertheless, some of them are not very surprising. If a psychophysical continuum is chosen on which some kind of threshold is *known* to exist—such as the critical flicker-fusion threshold—it is obvious that two stimuli from opposite sides of the threshold will be more discriminable than two stimuli from the same side. It does not necessarily follow that, therefore, the phoneme boundary effect on a speech continuum is also caused by a psychophysical boundary that happens to coincide with the phoneme boundary. The problem is that, in most cases, we have no good idea of what the psychophysical boundary ought to be on a given speech continuum. Moreover, there is evidence that a phoneme boundary effect may be caused by the application of labels in the absence of any psychophysical threshold (see Ades's, 1977, distinction between two types of boundary effects). There are several reasons why the nonspeech studies referred to above have made only slow progress toward clarifying the issue.

First of all, only results obtained with nonspeech stimuli that have something in common with speech (so-called nonspeech analogs) are directly relevant to the question of whether a specific phoneme boundary falls on top of a psychoacoustic threshold. For example, the observations on the flicker-fusion threshold (Pastore et al., 1977) cannot have any direct implications for speech perception because flicker fusion does not occur in speech.

Second, just how much certain nonspeech stimuli have in common with speech stimuli they are intended to emulate is a matter of debate. It is doubtful, for example, whether the relative onset time of two sinusoids (Pisoni, 1977) successfully simulates the distinction between a voiced and a voiceless stop consonant (see Pastore, Harris & Kaplan, 1982; Pisoni, 1980; Summerfield, 1982), or whether amplitude rise time has much to do with the fricative–affricate distinction (Remez, Cutting, & Studdert-Kennedy, 1980).

Third, even those nonspeech continua (such as noise–buzz sequences) that appear to copy a speech cue more or less faithfully yield results that, on closer inspection, are not in agreement with speech perception. For example, individual listeners in the Miller et al. (1976) study showed boundaries as short as 4 msec on a noize–buzz continuum, which is much shorter than any boundaries for English-speaking listeners on the supposedly analogous voice onset time dimension (see, e.g., Zlatin, 1974). Note also that auditory discrimination thresholds generally improve with practice, whereas linguistic boundaries ordinarily remain stable; this creates a problem for comparing the locations of the two.

Fourth, the various comparisons of categorical perception of speech and supposedly analogous nonspeech stimuli generally have not taken into account the fact that there are multiple cues for each phonetic contrast and that perception of one cue, as it were, is not independent of the settings of other relevant cues. This issue, which has received particular attention only in the last few years, is central to the second part of this article.

Fifth, there are a variety of other factors that influence the locations of phonetic boundaries: language experience, speaking rate, stress, phonetic context, semantic factors, and so on. It remains to be shown that psychophysical thresholds are sensitive to all, or even some, of these variables (or their psychoacoustic analogs, if they exist).

Finally, we note that there are examples of category boundary effects on nonspeech continua that have no obvious psychophysical boundaries, namely, for musical intervals (Burns & Ward, 1978; Siegel & Siegel, 1977) or chords (Blechner, 1977; Zatorre & Halpern, 1979), which suggests that well-established categories of nonpsychophysical origin may dominate perception.

In view of these arguments, one plausible account of the phoneme boundary effect remains that it arises from the use of category labels in discrimination. The best support for this hypothesis comes from studies that show a change in speech sound discriminability as a consequence of a redefinition of linguistic categories for the same stimuli and the same listeners (e.g., Carden, Levitt, Jusczyk, & Walley, 1981). Nevertheless, a reliance on category labels in discrimination is not unique to speech. The difference between speech and nonspeech in the discrimination paradigm probably rests on the nature of the categories: Phonetic categories are not only more deeply engrained than other categories, but they also bear a special relation to the acoustic signal. As Studdert-Kennedy (1976) put it, speech sounds "name themselves." Therefore, linguistic categories will dominate perception in a discrimination task to a larger extent than will nonspeech categories that often are not very familiar to the subjects and sometimes merely serve to bisect an arbitrarily selected stimulus range. In addition, the acoustic distinctions underlying a phonetic category contrast may be rather fine and also are habitually ignored by listeners in a natural situation; therefore, they are less accessible in the context of a discrimination task than are acoustic differences on a nonspeech continuum.

The strongest evidence for the alternative hypothesis, that categorical perception of speech rests on nonlinguistic auditory discontinuities in perception, comes from research on human infants (for recent summaries, see Jusczyk, 1981; Morse, 1979; Walley, Pisoni, & Aslin, 1981) and on nonhuman animals, particularly chinchillas (Kuhl, 1981; Kuhl & Miller, 1978). Allowing for the inevitable methodological differences and limitations, infants and (so far) chinchillas appear to perceive synthetic speech stimuli essentially the same way adults do, including superior discrimination of stimuli from different (adult) categories than of stimuli from the same category. These effects obviously reflect some "natural" boundaries, but it is not entirely clear whether these boundaries are strictly psychoacoustic in nature or whether they perhaps reflect some innate or acquired sensitivity to articulatory patterns. Even if they were psychoacoustic (this being the received interpretation of the infant and chinchilla findings), it is not certain that linguistic categories in fact depend on them. (See, however, Aslin & Pisoni, 1980, for a different view.) For example, children in the early stages of language use often are not able to make the perceptual distinctions infants seem to be capable of (Barton, 1980). There are still many open questions here. A fair assessment of the situation may be that the evidence on phoneme boundary effects neither supports nor disconfirms the existence of a special speech mode of perception.

Selective Adaptation

The shifting of phoneme boundaries on a continuum by repeated presentation of stimuli from one category has been a favorite pastime of many speech perception researchers ever since Eimas and Corbit (1973) discovered the technique. (See Diehl, 1981, for a recent critical review.) In hindsight, this effort seems not to have been worthwhile. Because various kinds of nonspeech dimensions show selective adaptation effects, it was to be expected that auditory dimensions of speech can be adapted as well. On the whole, this is what a score of studies show. The technique was considered interesting because it was thought to reveal the existence of "phonetic feature detectors" (Eimas & Corbit, 1973). Nevertheless, the evidence for specifically phonetic effects in selective adaptation is scant, and what there is can probably be explained as shifts in response criteria or as effects of remote auditory similarity. Recent experiments by Sawusch and Jusczyk (1981) and particularly by Roberts and Summerfield (1981) strongly suggest that there is no phonetic component in selective adaptation at all and that the effect takes place exclusively at a relatively early stage in auditory processing.

The concept of phonetic feature detectors is useless not only for the explanation of selective adaptation results (see Remez, 1979) but also from a wider theoretical perspective. No one expresses this better than Studdert-Kennedy (1982) when he says that "we are dealing with tautology, not explanation. . . . The error lies in offering to explain phonetic capacity by making a substantive physiological mechanism out of a descriptive property of language" (p. 225). For, ". . . the perceived feature is an attribute, not a constituent, of the percept, and we are absolved from positing specialized mechanisms for its extraction" (p. 227). For these reasons, selective adaptation results cannot have any implications for or against the existence of a specia! speech mode.

Hemispheric Specialization

The empirical results relating to the hemispheric asymmetry for speech and language are rich and complex. Although left-hemisphere advantages have been reported for certain kinds of nonspeech sounds, the evidence that speech processes are lateralized to the left hemisphere in the large majority of individuals is unassailable. It has been claimed, however, that precisely because certain nonspeech stimuli show similar effects, the lateralization of speech should be explained by a more general principle, for example, by a specialization of the left hemisphere for auditory properties that are characteristic of speech (Cutting, 1978; Schouten, 1980) or by an analytic–holistic distinction between the two hemispheres (e.g., Bradshaw & Nettleton, 1981). In commenting on the last-named article, Studdert-Kennedy (1981) argued that the analytic–holistic hypothesis, although descriptively adequate, is ill conceived from a phylogenetic viewpoint. Rather, insofar as lateralization presumably evolved to support some behavior important to the species, it seems more likely that lateralization of motor control preceded or caused lateralization of speech processes, which in turn may be responsible for the superior analytic capabilities of the left hemisphere. The apparent specialization of the left hemisphere for certain auditory characteristics of speech may just as well be the consequence as the cause of the lateralization of linguistic functions. Thus, the existing evidence on hemispheric specialization can be interpreted in an alternative way that is more compatible with a biological viewpoint and that recognizes the special status of speech.

Other Laboratory Phenomena

Various other findings have been cited as evidence for or against a speech mode of perception. Thus, Wood (1975) mentioned the phenomenon of asymmetric interference between auditory and linguistic dimensions in a speeded classification task. Although this finding (whose methodological details need not concern us here) may reveal something about the auditory processing of speech, its implications for the existence of a special speech mode of perception are limited. Similar patterns of results have been obtained with nonspeech auditory stimuli (Blechner, Day, & Cutting, 1976; Pastore et al., 1976), suggesting that the asymmetry has a nonphonetic basis.

Schouten (1980) added to Wood's (1975) list two findings that seem to have even less bearing on the question of a phonetic mode of perception: a difference in the stimulus duration needed for correct order judgments with sequences of speech or nonspeech sounds (Warren, Obusek, Farmer, & Warren, 1969) and an asymmetry in the perception of truncated consonant-vowel and vowel-consonant syllables (Pols & Schouten, 1978). The first finding probably reflects the fact that speech stimuli are more readily categorized than nonspeech stimuli, whereas the second finding seems altogether irrelevant, having most likely a psychoacoustic explanation. It is a mistake to believe (as Schouten apparently does) that the "case against a speech mode of perception" is strengthened by various findings of auditory (nonphonetic) effects in speech perception experiments. Such effects are likely to occur, for after all, speech enters through the ears. The thesis of the present article is, however, that these effects are relatively inconsequential for the linguistic processing of speech.

By focusing primarily on the experimental paradigms listed in Wood's (1975) article, Cutting (1978) and Schouten (1980) neglected a variety of other observations that suggest the existence of a speech mode of perception. Liberman et al. (1967) reviewed many properties that are peculiar to speech and that seem to require special perceptual skills. Foremost among these properties is the invariance of phonetic perception over substantial changes in the acoustic information; consider the well-known [di]-[du] example, which shows that the "d" percept can be cued by radically different transitions of the second formant (in the absence of any other, more invariant, cues). To achieve the same classification without reference to the articulatory gesture common to [di] and [du], an exceedingly complex "auditory decoder" would be required.

Liberman et al. (1967) also noted that the formant transitions distinguishing [di] and [du] sound quite different from each other when they are presented in isolation and do not engage the speech mode. In fact, when second- or third-formant transitions are removed from a synthetic syllable and presented to one ear while the rest of the speech pattern is presented to the other ear, the transitions are found to do double duty: They are perceived as whistles or chirps in one ear, but they also fuse with the remainder of the syllable in the other ear to produce a percept that is equivalent to the original syllable (Cutting, 1976; Rand, 1974). This "duplex perception" demonstrates the simultaneous use of speech and nonspeech modes of perception and has recently been further explored in experiments that are reviewed later in this article.

Other authors have noted striking differences in subjects' responses depending on whether identical or similar stimuli were perceived as speech or nonspeech. For example, House, Stevens, Sandel, and Arnold (1962) found that an ensemble of speech stimuli was easier to learn than various ensembles of speechlike stimuli that, however, were not perceived as speech by the subjects (see also Grunke & Pisoni, Note 1). Several studies of categorical perception have shown that speech stimuli from a synthetic continuum are discriminated well across a phonetic category boundary, whereas nonspeech analogs or components of the same stimuli are discriminated poorly or at chance (e.g., Liberman, Harris, Eimas, Lisker, & Bastian, 1961; Liberman, Harris, Kinney, & Lane, 1961; Mattingly, Liberman, Syrdal, & Halwes, 1971). Although none of these studies is without its specific methodological problems (some of which have been overcome in more recent work, discussed below), taken together they suggest strongly that speech and nonspeech stimuli are responded to in qualitatively different ways. As long ago as two decades, House et al. concluded that

> an understanding of speech perception cannot be achieved through experiments that study classical psychophysical responses to complex acoustic stimuli. . . . Although speech stimuli are accepted by the peripheral auditory mechanism, their interpretation as linguistic events transfers their processing to some nonperipheral center where the detailed characteristics of the peripheral analysis are irrelevant. (p. 142)

This conclusion is still valid, as the remainder of this article attempts to show.

Summary

Of the various paradigms reviewed by Cutting (1978) and Schouten (1980), some failed to support the existence of a speech mode of perception because the paradigms were irrelevant to begin with. As far as categorical perception and hemispheric specialization are concerned, some of the evidence may have been misinterpreted. The fact that categorical perception and left-hemisphere superiority can be obtained for certain nonspeech stimuli does away with earlier claims that these phenomena are speech-specific. It does not necessarily imply, however, that similar patterns of results occur for the same reason in speech and nonspeech, and if they do, it is not necessarily true that the processes involved in the perception of nonspeech are more basic than, or are the prerequisites for, those supporting speech perception. We have seen that there are other findings, not considered by Cutting and Schouten, that suggest that speech perception differs from nonspeech auditory perception. It must be acknowledged, however, that the empirical results are complex, and although they hardly argue against the existence of a speech mode, they do not provide an overwhelming amount of positive evidence either.

Certainly, the argument that speech perception is special would be strengthened if new, less controversial, results could be brought to bear on the issue. The second part of this article focuses on a set of findings that add a new dimension to the argument. Because these results are recent and have not been reviewed previously, they are treated in more detail. They may be grouped into three categories: phonetic trading relations, context effects, and other perceptual integration phenomena. What is common to all of them is that they deal with *integration* (over frequency, time, or space) in phonetic perception.

The New Evidence

The Distinction Between Trading Relations and Context Effects

It is known from many previous studies that virtually every phonetic contrast is cued by several distinct acoustic properties of the speech signal. It follows that, within limits set by the relative perceptual weights and by the ranges of effectiveness of these cues, a change in the setting of one cue (which, by itself, would have led to a change in the phonetic percept) can be offset by an opposed change in the setting of another cue so as to maintain the original phonetic percept. This is a phonetic trading relation. According to Fitch, Halwes, Erickson, and Liberman (1980), there is a *phonetic equivalence* between two cues that trade with each other. I prefer to use this term in a slightly different way, for neither cue is perceived in isolation; rather, they are perceived together and integrated into a unitary phonetic percept. Therefore, the equivalence holds not so much between (a − b) units of Cue 1 and (c − d) units of Cue 2 but rather between the phonetic percept caused by setting a of Cue 1 and setting d of Cue 2 and the phonetic percept caused by setting b of Cue 1 and setting c of Cue 2. These two percepts are phonetically equivalent in the sense that they yield exactly the same distribution of identification responses and are difficult to discriminate (see below).

Trading relations occur among different cues for the same phonetic contrast. On the other hand, when the perception of a phonetic distinction is affected by a preceding or following context that is not part of the set of direct cues for the distinction (as illustrated in the next paragraph), we speak of a context effect. The context may be "close"; that is, it may constitute portions of the same coherent speech signal. Or it may be "remote", referring to the relation between separate stimuli in a sequence or between a precursor and a test stimulus. (Of course, the distinction between close and remote context is, to some extent, arbitrary.)

Effects of close context, which are of special interest to us, are similar to trading relations in that they can be canceled by an appropriate change in one or another cue relevant to the critical phonetic distinction. Conversely, a trading relation could be described (inappropriately) as a context effect, with one cue (the context) affecting the perception of another (the target). Formally, trading relations and context effects are quite similar, but it is useful to distinguish them on theoretical grounds. The distinction is best illustrated with an example.

Mann and Repp (1980) presented listeners with fricative noises from a synthetic [ʃ]-[s] continuum, immediately followed by one of four periodic stimuli. The periodic stimuli derived from natural utterances of [ʃa], [sa], [ʃu], and [su], from which the fricative noise portion had been removed; thus, they contained formant transitions appropriate for either [ʃ] or [s], and the identity of the vowel was either [a] or [u]. The results showed that, for a given ambiguous noise stimulus, listeners reported more instances of "s" when the following formant transitions were appropriate for [s] rather than for [ʃ], and they also reported more instances of "s" when [u] rather than [a] followed. The first effect is a trading relation, the second a context effect. The effect of formant transitions on perception of the [ʃ]-[s] distinction is a trading relation because the transitions are a cue to fricative place of articulation. They are also a direct consequence of fricative production, and this is obviously the reason why they are a cue to fricative perception. Note that the transitions are integrated with the fricative noise cue into a unitary phonetic percept; listeners do not perceive a noise plus transitions, or a fricative consonant followed by a stop consonant, although a stop would be perceived if the fricative noise were removed or if silence were inserted between it and the periodic portion (Cole & Scott, 1973; Mann & Repp, 1980). The effect of vowel identity on fricative perception is different. Whether the vowel is [a] or [u] is not a consequence of fricative production, and vowel quality therefore does not constitute a direct cue for fricative perception. The vowel is not perceptually integrated with the noise cue—it remains audible as a separate phonetic segment. It is appropriate here to say that the perceived vowel quality modifies the perception or interpretation of the fricative cues. This is a context effect, as distinct from a trading relation.[1]

As we will see below, trading relations and context effects have distinct (though related) explanations in a theory of phonetic perception, and it is that theoretical view that underlies the distinction in the first place. Before we turn to the issue of explanation, however, a brief review of empirical findings is presented.

Phonetic Trading Relations

Overview

The fact that there are multiple cues for most phonetic contrasts has been known for a long time. Much of this early knowledge derives from the extensive explorations at Haskins Laboratories since the late 1940s. For example, Delattre, Liberman, Cooper, and Gerstman (1952) showed that the first two formants are important cues to vowel quality; Harris, Hoffman, Liberman, Delattre, and Cooper (1958) demonstrated that both second- and third-formant transitions contribute to the place-of-articulation distinction in stop consonants; and Gerstman (1957) found that both frication duration and rise time are relevant to the fricative–affricate distinction. Lisker (1978b), drawing on observations collected over a number of years, listed no less than 16 distinguishable cues to the /b/-/p/ distinction in intervocalic position.

From these and many other studies, a nearly complete list of cues has been accu-

[1] A rule of thumb for distinguishing a trading relation from a context effect is that the phonetic equivalence resulting from a trading relation is strong in the sense that two phonetically equivalent stimuli (syllables or words) are difficult to tell apart (Fitch et al., 1980), whereas the phonetic equivalence produced by trading a critical cue against some contextual influence is restricted to the target segment, as it always involves a readily detectable change in one or more contextual segments. To the extent that a change in context (e.g., vowel quality) also modifies critical cues (e.g., formant transitions), context effects may sometimes include disguised trading relations.

mulated over the years. Nevertheless, the data were typically collected by varying one cue at a time, although there are some exceptions, such as Hoffman's (1958) heroic study that varied three cues to stop place of articulation simultaneously. Restrictions on the size of stimulus ensembles were imposed by the limited technology of the time, which made stimulus synthesis and test randomization very cumbersome. With the advent of modern computer-controlled synthesis and randomization routines, however, orthogonal variation of several cues in a single experiment became an easy task, and the limit to the number of stimuli was set by the patience of the listener rather than by that of the investigator. The new technology led to a resurgence of interest in the way in which multiple cues cooperate in signaling a phonetic distinction. Because, for one reason or another, many of the early Haskins studies had remained unpublished, certain results that had been known for years by word of mouth or from preliminary reports only recently found their way into the literature, after having been replicated with contemporary methods.

A word is in order about the definition of cues. The traditional approach, exemplified especially by the Haskins work (including my own), has been to dissect a spectrographic representation of the speech signal, following essentially visual Gestalt principles. A cue, then, is a portion of the signal that can be isolated visually, that can be manipulated independently in a speech synthesizer constructed for that purpose, and that can be shown to have some perceptual effect.

The above way of defining cues has been challenged in two ways: (a) The spectrogram is not the only, and not necessarily the best, representation of the speech signal. For example, the well-known work of Stevens and Blumstein (1978; Blumstein & Stevens, 1979, 1980) pursues the hypothesis that the shape of the total short-term spectrum at certain critical points in the signal constitutes a perceptual cue; thus, the individual formants and adjacent noise bursts are not treated as separate cues. Such a redefinition of cues is justified as long as it does not bypass the legitimate empirical issue of whether the elementary, spectrographically defined signal components are indeed integrated by the auditory system in this way (as they may be in the case of individual formants but probably not in the case of other, more disparate types of cues). Although definitions of such complex cues effectively combine distributed information on one dimension (e.g., in the spectral domain), they typically sacrifice information on other dimensions (e.g., in the temporal domain). Thus, the onset spectra examined by Stevens and Blumstein are static and do not easily permit the description of dynamic change over time. The issue revolves, in large part, around the question of how the perceptually salient information in the signal is best characterized—a question that, of course, lies at the heart of the present article as well. The essential problem is that the totality of the cues for a given phonetic contrast apparently cannot be captured in a fully integrated fashion as long as purely physical (rather than articulatory or linguistic) terms are used.[2] (b) Another criticism of a more far-reaching sort denies altogether the usefulness of fractionating the speech signal into cues (see, e.g., Bailey & Summerfield, 1980). This view, which rests on the precepts of Gibsonian theory (Gibson, 1966), is taken up in the concluding comments of this article.

I do not attempt to review in detail all recent studies of phonetic trading relations, of which there are quite a few. A brief and selective overview shall suffice. Most studies had the purpose of clarifying the roles and surveying the effectiveness of different cues to various phonetic distinctions. Some studies that depart from this standard pattern are considered later in more detail. Whereas the large majority of studies have used synthetic speech, some obtained similar information by cross-splicing components of natural utterances or by combining such

[2] The attempt to define integrated cues must be distinguished from independent efforts to represent the speech signal in a way that takes into account peripheral auditory transformations (Searle, Jacobson & Rayment, 1979; Zwicker, Terhardt, & Paulus, 1979). Such representations are, of course, very useful and may lead to the redefinition of some cues; nevertheless, they do not by themselves solve the problem of cue definition.

components with synthetic stimulus portions. Not all authors describe their findings as *trading relations* (a term used primarily by the Haskins group), but such relations are implied by the pattern of results.

Voicing cues. Many studies have investigated multiple cues to the voiced–voiceless distinction. For stop consonants in initial position, both voice onset time (VOT) and the first-formant (F1) transition contribute to the distinction (Lisker, Liberman, Erickson, Dechovitz, & Mandler, 1977; Stevens & Klatt, 1974). The critical feature of the F1 transition, which can be traded against VOT, is its onset frequency: If the onset frequency is lowered in a phonetically ambiguous stimulus, the VOT must be increased for a phonetically equivalent percept to obtain (Lisker, 1975; Summerfield & Haggard, 1977). Another cue that can be traded for VOT is the amplitude of the aspiration noise preceding the onset of voicing: If the amplitude of the noise is increased, its duration (i.e., the VOT) must be decreased to maintain phonetic equivalence (Repp, 1979). The fundamental frequency (FO) at the onset of the voiced stimulus portion is another relevant cue (Haggard, Ambler, & Callow, 1970) that can be traded against VOT (Haggard, Summerfield, & Roberts, 1981).

For stop consonants in intervocalic position, Lisker (1978b) catalogued all of the different aspects of the acoustic signal that contribute to the voicing distinction. They include the duration and offset characteristics of the preceding vocalic portion, the duration of the closure interval, the amplitude of voicing during the closure, and the onset characteristics of the following vocalic portion. Lisker's catalogue is based on a large number of studies, not all of which have been published; nevertheless, see Lisker (1957, 1978a, 1978c), Lisker and Price (1979), and Price and Lisker (1979). Trading relations between voicing cues for intervocalic stops have also been studied in French (Serniclaes, 1974, Notes 2 & 3) and in German (Kohler, 1979).

The voicing distinction for stop consonants in final position has also been intensively studied. Here, the duration of the vocalic portion is important (especially if no release burst is present) and so are its offset characteristics, the properties of the release burst, and the duration of the preceding closure. Trading relations among these cues have been investigated by Raphael (1972, 1981), Wolf (1978), and Hogan and Rozsypal (1980), among others.

The voicing distinction for fricatives in initial position has been studied by Massaro and Cohen (1976, 1977), who focused on the trading relation between fricative noise duration and F0 at the onset of periodicity. In a similar fashion, Derr and Massaro (1980) and Soli (1982) studied the trading relations among duration of the periodic ("vowel") portion, duration of fricative noise, and F0 as cues to fricative voicing in utterance-final position. Earlier studies of these cues include Denes (1955) and Raphael (1972).

Place of articulation cues. Trading relations among place of articulation cues for stop consonants in initial position—F2 and F3 transitions, burst frequency and burst amplitude—were studied long ago by Harris et al. (1958) and Hoffman (1958) and more recently by Dorman, Studdert-Kennedy, and Raphael (1977) and Mattingly and Levitt (1980). For stop consonants in intervocalic position, Repp (1978) found a trading relation between the formant transitions in and out of the closure, and Dorman and Raphael (1980) reported additional effects of closure duration and release burst frequency. Bailey and Summerfield (1980), in a series of painstaking experiments, investigated place cues for stops in fricative-stop-vowel syllables; these cues included the offset spectrum of the fricative noise, the duration of the closure period, and the formant frequencies at the onset of the vocalic portion. Repp and Mann (1981b) recently demonstrated a trading relation between fricative noise offset spectrum and vocalic formant transitions in similar stimuli. Fricative noise spectrum and vocalic formant transitions as joint cues to fricative place of articulation were investigated by Whalen (1981), Mann and Repp (1980), and Carden et al. (1981).

Manner cues. Cues to stop manner of articulation (i.e., to presence vs. absence of a stop consonant) following a fricative and preceding a vowel were investigated by Bailey and Summerfield (1980), Fitch et al. (1980), and Best, Morrongiello, and Robson

(1981). In each case, the trading relation studied was that between closure duration and formant onset frequencies in the vocalic portion. The two last-named studies are discussed in more detail below. Summerfield, Bailey, Seton, and Dorman (1981) showed that duration and amplitude contour of the fricative noise preceding the silent closure also contribute to the perception of stop manner.

Several cues to the fricative–affricate distinction in initial position (rise time, noise duration) were investigated by Gerstman (1957); see also van Heuven (1979). In a more recent set of experiments, Repp, Liberman, Eccardt, and Pesetsky (1978) traded vocalic offset spectrum, closure duration, and fricative noise duration as cues to a four-way distinction between vowel-fricative, vowel-stop-fricative, vowel-affricate, and vowel-stop-affricate. Trading relations among cues to the fricative–affricate distinction in final position were reported by Dorman, Raphael, and Liberman (1979, Exp. 5) and Dorman, Raphael, and Isenberg (1980).

Phonetic Equivalence

It is obvious that, whenever two or more cues contribute to a given phonetic distinction, they can be traded against each other, within certain limits. What is not so obvious is that two different stimuli with equal response distributions are truly equivalent in perception. Insofar as most data on trading relations were collected in identification tasks with a restricted set of response categories, subjects may have had no opportunity to report that certain stimuli sounded like neither of the alternatives. At a more subtle level, it may be that phonetically equivalent stimuli, even though they are labeled similarly, sound different in some way that subjects cannot easily explain in words. One way to assess this possibility is by means of a discrimination task.[3]

This task was undertaken by Fitch et al. (1980) for the trading relation between silent closure duration and vocalic formant transition onsets as cues to stop manner in the "slit"–"split" distinction and by Best et al. (1981) for the similar trading relation between silent closure duration and F1 transition onset in the "say"–"stay" contrast. First, these authors determined in an identification task how much silence was needed to compensate for a certain difference in formant onset frequency. Then they devised a discrimination task containing three different types of trials: On *single-cue* trials, the stimuli to be discriminated differed only in the spectral cue (formant onset frequency); they had the same setting of the temporal cue (silence). On *cooperating-cues* trials, the stimuli differed in both cues, such that the stimulus with the lower formant onsets (which favor "split" or "stay" percepts) also had the longer silence (which favors the same percepts). On *conflicting-cues* trials, the stimuli again differed in both cues, but now the stimulus with the lower formant onsets had the shorter silence, so one cue favored "split" ("stay") and the other "slit" ("say"). Because the silence difference chosen was the one found to compensate exactly for the spectral difference in the identification task, the stimuli in the conflicting-cues condition were (on the average) phonetically equivalent.[4]

[3] In essence, this kind of study investigates whether multidimensionally varying speech stimuli are perceived categorically. Traditional studies of categorical perception have been exclusively concerned with stimuli varying on a single dimension or varying on several dimensions in a perfectly correlated fashion. Note that, in these studies, physically different stimuli from the region of the category boundary are not phonetically equivalent—they have different response distributions. As soon as two or more cues are varied, however, pairs of phonetically equivalent stimuli can be found for any given response distribution. Thus, the influence of phonetic categorization on discrimination judgments can be factored out, at least in principle (see Footnote 4).

[4] To produce precise (rather than just average) phonetic equivalence, it would not only be necessary to take into account the fact that individual listeners show trading relations of varying magnitude but also that (covert) labeling responses may change in the context of a discrimination task (Repp, Healy, & Crowder, 1979). Thus, the stimulus parameters would have to be adjusted separately for each listener, based on labeling data collected with the stimulus sequences of the discrimination task. This procedure would optimize the opportunity to verify the prediction that stimuli in the conflicting-cues condition are more difficult to discriminate than those in the cooperating-cues condition, with the single-cue condition in between. This order of difficulty, however, is likely to obtain also when the choices of parameters are less than optimal.

The results of these experiments showed a clear difference among the three conditions: Subjects' discrimination performance in the category boundary region was best in the cooperating-cues condition, worst in the conflicting-cues condition, and intermediate in the single-cue condition. Thus, it is true that (approximately) phonetically equivalent stimuli—namely, those in the conflicting-cues condition—are difficult to discriminate; they "sound the same," whereas stimuli in the cooperating-cues condition sound different, even though they exhibit the same physical differences on the two relevant dimensions. The pattern of discrimination results follows that predicted from identification data, showing that stimuli that differ on two auditory dimensions simultaneously are still categorically perceived (given that perception is categorical when each of these dimensions is varied separately). It is likely that listeners could be trained to become more sensitive to the physical differences that do exist between phonetically equivalent stimuli, and the interesting question arises whether discrimination on cooperating-cues trials would continue to be superior to that on conflicting-cues trials. So far, no study has taken this approach. Nevertheless, preliminary results from a related series of experiments (Repp, 1981a) indicate that some trading relations disappear when listeners try to discriminate stimuli that unambiguously belong to the same phonetic category (i.e., phonetically equivalent stimuli that are not from the boundary region), suggesting that these trading relations operate only when the stimuli are phonetically ambiguous. This leads us to the question of the origin of trading relations.

Explanation of Trading Relations: Phonetic or Auditory?

The large number of trading relations surveyed above poses formidable problems for anyone who would like to explain speech perception in purely auditory terms. Why should cues as diverse as, say, VOT and F1 onset, or silence and fricative noise duration, trade in the way they do? Auditory theory has only two avenues open: Either the cues are integrated into a unitary auditory percept at an early stage in perception (the *auditory integration hypothesis*), or selective attention is directed to one of the cues (which then must be postulated to be the essential cue for the relevant phonetic contrast), and the perception of that cue is affected by the settings of other cues (the *auditory interaction hypothesis*).

The auditory integration hypothesis is implicit in the work of Stevens and Blumstein (1978; Blumstein & Stevens, 1979, 1980). To account for the fact that release burst spectrum and formant transition onset frequencies are joint cues to place of articulation of syllable-initial stop consonants, Stevens and Blumstein assume that the perceptually relevant variable is the integrated spectrum of the first 25 msec or so of a stimulus. In other words, the burst (which is usually shorter than 25 msec) and the onsets of the several formant transitions are considered an integral auditory variable. Insofar as both cues are spectral in nature and occur within a short interval, this is not an unreasonable hypothesis, notwithstanding the different sources of excitation (noise vs. periodic) of the two sets of cues in voiced stops. In fact, Ganong (1978) found support for the perceptual integrality of burst and formant transition cues in an ingenious experiment involving interaural transfer of selective adaptation effects. Stevens and Blumstein, however, had only limited success with automatic classification of stop consonants according to onset spectrum alone, and Kewley-Port (1981) recently demonstrated that automatic stop consonant identification can be improved by incorporating a measure of spectral change while retaining the notion of auditory integration. Thus, even though onset spectrum may be an important cue, it does not contain all of the relevant information in the signal.

The main problem with the auditory integration hypothesis seems to be that it applies only when the relevant cues are all spectral in nature, are of short duration, and occur simultaneously or in close succession. The cues, however, are often spread out over a considerable stretch of time. For example, an explanation of the fact that both the formant transitions into and out of a stop closure contribute to the perceived place of articulation of a stop in medial position (Dorman & Raphael, 1980; Repp, 1978; Repp & Mann, 1981b) would require integration of spectra across a closure (i.e., over as much as 100 msec). Such a long integration period seems unlikely; certainly, it is much longer than that envisioned by Stevens and Blumstein (1978). Trading relations that involve spectral and temporal cues (e.g., F1 onset and VOT for stop voicing in initial position) cannot be easily translated into purely spectral terms; trading relations between purely temporal cues (e.g., silent closure duration and fricative noise duration for the fricative-affricate distinction in medial position—Repp et al., 1978) require a different explanation altogether.

To be sure, there are some trading relations that do suggest auditory integration, such as that between VOT (i.e., aspiration noise duration) and aspiration noise amplitude (Repp, 1979), which is reminiscent of certain time-intensity reciprocities at the auditory threshold. In fact, preliminary data (Repp, 1981a) support this suggestion by showing that this trading relation operates independently of whether a listener is making phonetic or auditory judgments of speech stimuli. In most other cases, however, the cues that participate in a trading relation are simply too diverse or too widely spread out to make auditory integration seem plausible. Or, to put it somewhat differently, whereas any such trading relation could be *described* as resulting from "auditory integration," this integration would no longer seem to be motivated by general principles of auditory perception; thus, it would have to be considered a speech-specific process.

The auditory interaction hypothesis, which postulates that trading relations occur because perception of a primary cue is affected by other cues, has even less concrete evidence in its favor, in part because most of the relevant studies remain to be done. In particular, it is not clear whether auditory interactions (masking, contrast, etc.) of the kind and extent required to explain certain trading relations are at all plausible. For example, to explain the trading relation between VOT and F1 onset frequency as cues to stop consonant voicing, it would have to be the case that a noise-filled interval (VOT) sounds subjectively longer when followed by a periodic stimulus with a relatively low onset frequency. At present, there are no psychoacoustic data to support this hypothesis. Auditory psychophysics involving nonspeech stimuli of speechlike complexity is still in its infancy (see Pastore, 1981). Perhaps as more is learned about the perception of complex sounds and sound sequences, some auditory explanations of what now appear to be phonetic phenomena will be forthcoming.[5]

One serious problem that has vexed researchers since the time of the early Haskins research is that of finding appropriate nonspeech analogs for speech stimuli. If the analogs are too similar to speech, they may be perceived as speech and thereby cease to be good analogs and become bad speech. If they are too different from speech, the generalizability of the findings to speech may be questioned. There is a way out of this dilemma: If stimuli could be constructed that are sufficiently like speech to be perceived as speech by some listeners but not by others (perhaps prompted by different instructions), or even by the same listeners on different occasions, and if a trading relation were obtained when speech is heard but not when the percept is nonspeech, this would then be proof of specialized perceptual processes serving speech perception.

It is from this perspective that a recent study by Best et al. (1981) receives special importance. These authors investigated the trading relation between silent closure duration and F1 transition onset frequency as cues to stop manner in the "say"-"stay" contrast. After replicating the trading relation previously obtained with the similar "slit"-"split" contrast by Fitch et al. (1980),

[5] What is most interesting is that the only completed attempt (so far) to establish a trading relation in human infants (Miller & Eimas, Note 4) has yielded a positive result: The boundary on a VOT continuum was significantly affected by the duration of the formant transitions, a variable that is confounded with F1 onset frequency (see Summerfield & Haggard, 1977). Kuhl and Miller (1978) obtained a similar result with chinchillas (shifts in the VOT boundary with place of articulation) that may reflect the same trading relation. This trading relation, then, may be of auditory origin, even though the principle involved is not yet clear. It seems likely, though, that not all trading relations will follow this pattern.

they proceeded to test for the presence of an analogous trading relation in "sine-wave analogs" of their stimuli. Sine-wave analogs are obtained by imitating the formant trajectories of (voiced) speech signals with pure tones. Such analogs of simple consonant–vowel syllables have been used previously by Cutting (1974) and by Bailey, Summerfield, and Dorman (1977), whose work is discussed below; recently, Remez, Rubin, Pisoni, and Carrell (1981) successfully synthesized whole English sentences in that way. The interesting thing about these stimuli is that they are heard as nonspeech whistles by the majority of naive listeners, but they may be heard as speech when instructions point out their speechlikeness or spontaneously after prolonged listening. Once heard as speech, it is difficult (if not impossible) to hear them as pure whistles again, although the speech that is heard retains a highly artificial quality (Remez et al., 1981). This phenomenon was exploited by Best et al. in their main experiment.

Best et al. (1981) constructed sine-wave analogs of a "say"–"stay" continuum by following a noise resembling [s]-frication with varying periods of silence and a sine-wave portion whose component tones imitated the first three formants of the periodic portion of the speech stimuli. There were two versions of the sine-wave portion, one with a low onset of the tone simulating F1 and one with a high onset. (In the speech stimuli, less silence was needed to change "say" to "stay" when F1 had a low onset than when it had a high onset.) The sine-wave stimuli were presented to listeners in an AXB format, where the critical X stimulus had to be designated as being more similar to either the A or the B stimulus, which were analogs of a clear "say" (no silence, high F1 onset) and a clear "stay" (long silence, low F1 onset), respectively. Some of the subjects were told that the stimuli were intended to sound like "say" or "stay", whereas others were only told that the stimuli were computer sounds.

After the experiment, the subjects were divided into two groups: those who reported that they had heard the stimuli as "say"–"stay", either spontaneously or after instructions, and those who reported various auditory impressions or inappropriate speech percepts. Only members of the first group, who—according to their self-reports—used a phonetic mode of perception, showed a trading relation between silence and F1 onset frequency, and this trading relation closely resembled that obtained with synthetic speech stimuli. None of the subjects in the other group showed this pattern of results. These other subjects could be further subdivided into two groups: those who reported that the stimuli differed in the amount of separation between the two stimulus portions (noise and sine waves), and those who reported that the stimuli differed in the quality of the onset of the second portion ("water dripping", "thud", etc.). The AXB results substantiated these reports: The results of the first group indicated that the subjects paid attention only to the silence cue, whereas the second group seemed to make their judgments primarily on the basis of the spectral cue (F1-analog onset frequency). The response patterns of the two groups were radically different from each other, but neither resembled that of the group who heard the stimuli as speech.

It seems reasonable to conclude that the subjects who heard nonspeech used an auditory mode of perception. Being in this mode, they were unable to integrate the two cues into a unitary percept and instead focused on one or the other cue separately, thereby disconfirming the auditory integration hypothesis for this set of cues.[6] There was some evidence of an auditory interaction in that those listeners who paid attention to the spectral cue were influenced also by the setting of the temporal cue. Nevertheless, this effect was not sufficiently strong to account for the trading relation observed in speech-mode listeners; moreover, those subjects who focused on the silence cue (which is the primary cue for stop manner) were not

[6] That the subjects focused on one cue only was a strategy furthered by the AXB classification task of Best et al. (1981). In a different paradigm, the subjects may pay attention to both cues at the same time (see Repp, 1981a). The important point is that, in the auditory mode, the cues are not integrated into a unitary percept, so listeners may choose between selective-attention and divided-attention strategies—a choice they are not free to make in the speech mode.

affected at all by the setting of the spectral cue.

The results of Best et al. (1981) provide the strongest evidence we have so far that a trading relation is specific to phonetic perception: When listeners are not in the speech mode, the trading relation disappears, and selective attention to individual acoustic cues becomes possible. The data argue against any general auditory explanation of the trading relation at hand, and they support the existence of a phonetic mode of perception that is characterized by specialized ways of stimulus processing. Recently, Repp (1981a) further confirmed the phonetic nature of the trading relation between silence and F1 onset for the "say"–"stay" distinction by showing that it is obtained only in the phonetic boundary region of the speech continuum (i.e., when listeners can make a phonetic distinction) but not within the "stay" category (i.e., when listeners cannot make a phonetic distinction and must rely on auditory criteria for discrimination). We may suspect that many other trading relations will behave similarly. This is already indicated for the trading relation between closure duration and fricative noise duration in the "say shop"–"say chop" distinction (Repp, 1981a) and for that between fricative noise spectrum and formant transitions in the [ʃ]–[s] distinction (Repp, 1981b, discussed in the next section).

How, then, are trading relations to be explained, if not in terms of auditory interactions or integration? The proposed answer is this: Speech is produced by a vocal tract, and the production of a phonetic segment (assuming that such segments exist at some abstract level in the articulatory plan) has complex and temporally distributed acoustic consequences. Therefore, the information supporting the perception of the same phonetic segment is acoustically diverse and spread out over time. The perceiver recovers the abstract units of speech by integrating the multiple cues that result from their production. The basis for that perceptual integration may be conceptualized in two ways. One is to assume that listeners know from experience what a given phonetic segment "ought to sound like" in a given context. Insofar as phonetic contrasts almost always involve more than one acoustic property, trading relations among these properties must result when the stimulus is ambiguous because, in this view, it is being evaluated with reference to idealized representations or "prototypes" that differ on all of these dimensions simultaneously: A change in one dimension can be offset by a change in another dimension so that the perceptual distances from the prototypes remain constant. The other possibility is that perceptual integration does not require specific knowledge of speech patterns (whose form of memory storage is difficult to imagine) but is predicated directly on the articulatory information in the signal. In other words, trading relations may occur because listeners perceive speech in terms of the underlying articulation and resolve inconsistencies in the acoustic information by perceiving the most plausible articulatory act. This explanation requires that the listener have at least a general model of human vocal tracts and of their ways of action. The question remains: How much must an organism know about speech to exhibit a phonetic trading relation? An important issue for future research will be the question whether phonetic trading relations are obtained in human infants, and if not, how and when they begin to develop.[7]

Context Effects

Effects due to Immediate Phonetic Context

Like phonetic trading relations, certain kinds of phonetic context effects have been known for a long time. The most familiar example is, perhaps, the dependence of stop release burst perception on the following vowel. Liberman, Delattre, and Cooper (1952) showed that, when noise bursts of varying frequencies are followed by different steady-state periodic stimuli, the stop con-

[7] In that connection, a study by Simon and Fourcin (1978) might be mentioned that showed that the trading relation between VOT and F1 transition trajectory as cues to stop consonant voicing emerged at age 4 in British children but was absent in 2- and 3-year olds. Recently, however, Miller and Eimas (Note 4) found a related trading relation (between VOT and transition duration) in American infants. This conflict needs to be resolved.

sonant categories reported by listeners may depend on the quality of the vowel. For example, if a noise burst centered at 1600 Hz is followed by steady states appropriate for [i] or [u], listeners report "p", but if [a] follows, they report "k".

A similar effect has been reported by Summerfield (1975), who found that the nature of the vowel influences the location of the category boundary on a continuum of stop-consonant-vowel syllables varying in VOT. This context effect may actually be a trading relation because it probably reflects the influence of F1 onset (rather than vowel quality *per se*) on the voicing decision, that is, a trading relation between F1 onset and VOT (Summerfield & Haggard, 1974, 1977). Recently, Summerfield (1982) conducted an important series of experiments in which he tested whether this effect has an auditory basis. He used speech stimuli varying in VOT and in the F1 frequency of the following steady-state vocalic portion, and he compared their perception with that of two kinds of nonspeech analogs. One was a tone onset time (TOT) continuum (Pisoni, 1977) that varied the relative onset time of two pure tones, matched in frequency and amplitude to the first two formants of the speech stimuli. The other set of nonspeech stimuli formed a noise onset time (NOT) continuum (cf. Miller et al., 1976) that varied the lead time of a noise-excited steady-state F2 relative to a periodically excited steady-state F1. The stimuli were presented for identification as "g" or "k" (speech) or as "simultaneous onset" versus "successive onset" (nonspeech). Although the VOT boundary exhibited the expected sensitivity to F1 onset frequency, neither nonspeech continuum evinced any reliable influence of F1(-analog) frequency on listeners' judgments. Pastore et al. (1981) recently reported mixed results in comparing the effects of different secondary variables (duration, rise time, and trailing stimuli) on VOT and TOT category boundaries. These results suggest that the context effect obtained in speech does not have an auditory basis but is specific to the phonetic mode. (Nevertheless, see Footnote 5 for some potentially conflicting evidence.)

An effect of vocalic context on the perception of stop consonant place of articulation was investigated by Bailey et al. (1977). These authors constructed two synthetic speech continua ranging from [b] + vowel to [d] + vowel by varying the transition onset frequencies of F2 and F3. The two continua differed in the terminal (steady-state) frequency of F2, which was high in one and low in the other. On each continuum, the transition onsets were arranged so that the center stimulus had completely flat F2 and F3 and both transitions rose in one endpoint stimulus to the same degree as they fell in the other endpoint stimulus. When these stimuli were presented to subjects for classification in an AXB task, it turned out that the category boundaries were at different locations on the two continua, neither being exactly in the center: One (on the continuum with the low-F2 vowel) was displaced toward the [d] end, and the other boundary was displaced toward the [b] end.

Bailey et al. (1977) wished to test whether this difference (a kind of context effect, especially when direction of transition is considered the relevant cue, rather than absolute transition onset frequency, which varied with context) has a psychoacoustic basis. They pioneered in using sine-wave analogs for that purpose. The sine-wave stimuli were presented in the same AXB paradigm to a group of subjects that was subdivided afterwards according to self-reports as to whether or not the stimuli sounded like speech. It turned out that those listeners who claimed to hear [b] and [d] had their category boundaries on the two continua at different locations that corresponded to those found with speech stimuli. The other listeners, however, who reported only nonspeech impressions, had their boundaries close to the centers of both continua, as one might predict on simple psychophysical grounds. This experiment provided evidence that phonetic categorization is based on principles that are different from those of auditory psychophysics. Although this was not shown directly by Bailey et al., the asymmetrical boundaries obtained with speech stimuli were obviously in accord with the acoustical characteristics of stop consonants in these particular vocalic contexts.

Let us turn now to other context effects that are of special interest because they involve segments that are not as obviously interdependent as stop consonants and following vowels. One effect concerns the influence of vocalic context on fricative perception. If a noise portion that is ambiguous between [ʃ] and [s] is followed by a periodic portion that is appropriate for a rounded vowel such as [u], listeners are more likely to report "s" than when the vowel is unrounded, for example, [a] (Mann & Repp, 1980; Whalen, 1981; Kunisaki & Fujisaki, Note 5). A preceding vowel has a similar but smaller effect (Hasegawa, 1976). In addition to roundedness, other features of the vowel (such as the front–back dimension) also seem to play a role (Whalen, 1981). Repp and Mann (1981b) also discovered a small but reliable effect of a following stop consonant on fricative perception: Listeners are more likely to report "s" when the formant transitions in the following vocalic portion (separated from the noise by a silent closure interval) are appropriate for [k] than when they are appropriate for [t].

Several additional effects of context on the perception of stop consonants have been discovered in recent experiments. Mann and Repp (1981) found that, in fricative-stop-vowel stimuli, listeners are more likely to report "k" when vocalic stimuli with formant transitions that are ambiguous between [t] and [k] are preceded by an [s]-noise plus silence than when they are preceded by an [ʃ]-noise plus silence. They showed that the effect has two components, one due to the spectral characteristics of the fricative noise (perhaps an auditory effect), the other due to the category label assigned to the fricative (which must be a phonetic effect). Subsequently, Repp and Mann (1981b) showed the context effect to be independent of the effect of direct cues to stop place of articulation in the fricative noise offset spectrum (which proves that it is a true context effect and not a trading relation), and they also ruled out simple response bias as a possible cause. In another experiment, Mann (1980) found that, when stimuli that are ambiguous between [da] and [ga] were preceded by either [al] or [ar], listeners reported many more "g" percepts after [al] than after [ar]. In experiments with vowel-stop-stop-vowel stimuli, Repp (1978, 1980a, 1980b) found various perceptual interdependencies between the two stops cued by the formant transitions on either side of the closure interval; in particular, perception of the first stop was influenced strongly by the second.

How are all of these effects to be explained? Auditory explanations would have to be formulated in the manner of the interaction hypothesis for trading relations: The perception of the relevant acoustic cues must be somehow affected by the context. As in the case of trading relations, however, no plausible mechanisms that might create such effects have been suggested, and no similar effects with nonspeech analogs have been reported so far. On the other hand, reference to speech production provides a straightforward explanation of most, if not all, context effects. Just as trading relations reflect the dynamic nature of articulation (of a given phonetic segment), so are context effects accounted for by *coarticulation* (of different phonetic segments). The articulatory movements characteristic of a given phonetic segment exhibit contextual variations that may be either part of the articulatory plan (allophonic variation or anticipatory coarticulation) or due to the inertia of the articulators (perseverative coarticulation). Presumably, human listeners possess implicit knowledge of this coarticulatory variation.

Coarticulatory effects corresponding to the perceptual phenomena just cited have been observed in most cases. Thus, it is well known that the release burst spectrum of stop consonants varies with the following vowel (Zue, Note 6) in a manner that is quite parallel to the perceptual findings of Liberman et al. (1952). Fricative noises exhibit a downward shift in spectrum when they precede or follow a rounded vowel, due to anticipatory or carry-over lip rounding (Fujisaki & Kunisaki, 1978; Hasegawa, 1976; Mann & Repp, 1980), which explains the effect of vocalic context on fricative perception. The formant transitions of stop consonants vary with preceding fricatives (Repp & Mann, 1981a, 1981b) and liquids (Mann, 1980) in a manner that is consistent with the corresponding perceptual effects. Thus, the available evidence suggests that most perceptual context effects are parallel to coarticulatory effects. The implication is, then, that listeners "expect" coarticulation to occur (by referring to prototypes of allophonic variants or to some more general articulatory model) and compensate for its absence in experimental stimuli by shifting their response criteria accordingly. For example, if an [ʃ]-like noise followed by [u] is not sufficiently low on the spectral scale (as it should be because of anticipatory lip rounding), it might be perceived as an "s". Similarly, if it is true that [d] is articulated with a more forward place of articulation following [l] than following [r] (Mann, 1980), then ambiguous formant transitions following [l] may be interpreted as [g] because they do not signal the fronting expected for a [d]. Thus, the evidence is highly persuasive that context effects, just like trading relations, reflect the listeners' intrinsic knowledge of articulatory dynamics.

A critical test of the auditory versus phonetic explanations of context effects can again be performed with appropriate nonspeech analogs or with stimuli that can be perceived as either speech or nonspeech. Two such studies (Bailey et al., 1977; Summerfield, 1982) were discussed above. In a recent experiment, I took an alternative approach (Repp, 1981b): Rather than using nonspeech stimuli that can be perceived as speech, I used speech stimuli, a portion of which can be fairly readily perceived as nonspeech. Although it is usually difficult to abandon the phonetic mode when listening to speech, except in cases where the speech is strongly distorted or poorly synthesized, fricative-vowel syllables offer an opportunity to do so because they contain a sizable segment of fairly steady-state noise whose auditory properties ("pitch," length, loudness) are relatively accessible.

In my study, the fricative noise spectrum was varied along a continuum from [ʃ]-like to [s]-like, and the vowel was either [a] or [u]. It was known from earlier experiments (Mann & Repp, 1980) that listeners are more likely to label ambiguous fricatives "s" in the context of [u] than in the context of [a]. A secondary cue to the [ʃ]–[s] distinction was deliberately confounded with the context effect: The [a] vocalic portion contained formant transitions that are appropriate for [ʃ], and the [u] portion contained transitions that are appropriate for [s]; this increased the differential effect of the two vocalic contexts on fricative identification. (Thus, this experiment tested a context effect and a trading relation at the same time.) The stimuli were presented in a same-different discrimination task, where the difference to be detected was in the spectrum of the noise portion, and the vowels were either the same or different but irrelevant in any case.

The majority of naive subjects perceived the stimuli fairly categorically: Their discrimination performance was poor; the pattern of responses suggested that they relied on category labels; and there were pronounced effects of vocalic context, just as in previous labeling tasks. Two subjects, however, performed much better than the others. Their data resembled those of three experienced listeners who also participated in the experiment. Comments and introspections of these subjects suggested that they were able to bypass or ignore phonetic categorization and to focus instead on the spectral properties (the "pitch") of the fricative noise. The crucial result was not that these listeners performed much better than the rest (although this supported the hypothesis that they used an auditory mode of perception) but that they did not show any effect of vocalic context. These results were confirmed in a follow-up study in which naive listeners were induced (with partial success) to adopt an auditory listening strategy. Thus, vocalic context affected the perceived phonetic category of the fricative but not the perceived pitch quality of the noise. Therefore, the context effect due to the quality of the vowel, as well as the cue integration underlying the contribution of the vocalic formant transitions to fricative identification, must be phonetic in nature.

Speaker Normalization Effects

A phenomenon that is related to the context effects just discussed is that of speaker normalization. In an experimental demonstration of this effect, the perception of a critical phonetic segment is influenced not by a phonetic change in an adjacent segment but by an acoustic change such as might result from a change in speaker. For example, a (roughly proportional) upward shift of vowel formants on the frequency scale signifies that the speech signal originated in a smaller vocal tract. (How listeners "decide" that the same vowel has been produced by a smaller vocal tract, rather than a different vowel by the same vocal tract, is an unresolved issue.) Such a change may influence the perception of phonetic segments in the vicinity as long as the listener perceives the whole test utterance as coming from a single speaker.

Although speaker normalization is a well-recognized problem in speech recognition research, there have been relatively few experimental studies. Rand (1971) constructed three stop consonant continua ranging from [b] to [d] to [g] by varying the onset of the F2 transition of synthetic two-formant stimuli whose vocalic portion was intended to represent, respectively, an [æ] produced by a large vocal tract, an [æ] produced by a small vocal tract (differing from the former only in F2 frequency), and an [ɛ] produced by a large vocal tract (differing from the former only in F1 frequency). The results showed similar category boundaries (expressed in terms of absolute F2 onset frequency) for the two stimulus continua associated with large vocal tracts but a shift toward higher frequencies on the continuum associated with a small vocal tract. Rand interpreted his findings as evidence for perceptual normalization, although this may not be the only possible explanation.

In a more recent study, May (1976) followed fricative noises from a synthetic [ʃ]–[s] continuum with one of two synthetic periodic portions intended to represent the same vowel produced by two differently sized vocal tracts. The [ʃ]–[s] boundary shifted as expected: Listeners reported more "s" percepts in the context of the larger vocal tract. Subsequently, Mann and Repp (1980) conducted a similar experiment in which synthetic fricative noises were followed by vocalic portions derived from natural utterances produced by a male or a female speaker. The results replicated those by May. These findings are consistent with the fact that smaller vocal tracts (i.e., women) produce fricative noises that are of higher average frequency than those produced by large vocal tracts (i.e., men; Schwartz, 1968).

To these results must be added the evidence from studies that have shown speaker normalization effects due to "remote" context, that is, due to other stimuli in a sequence or to precursor stimuli or phrases (e.g., Ladefoged & Broadbent, 1957; Strange, Verbrugge, Shankweiler, & Edman, 1976; Summerfield & Haggard, 1975). They all demonstrate the same point: Listeners interpret the speech signal in accordance with the perceived (or expected) dimensions of the vocal tract that produced it. Information about vocal tract size is picked up in parallel with information about articulator movements; these are, respectively, the static and dynamic (or structural and functional) aspects of articulatory information. Speaker normalization effects are difficult to explain in terms of a general auditory theory that does not make reference to the mechanisms of speech production. Although some effects could, in principle, result from auditory contrast, interactions of similar complexity have not yet been demonstrated in nonspeech contexts.

Rate Normalization Effects

The somewhat larger literature on perceptual effects of speaking rate has recently been thoroughly reviewed by Miller (1981). Rate normalization, like speaker normalization, is a kind of context effect, and it can be produced by either close or remote context. Rate normalization is said to occur when the perception of a phonetic distinction signaled by a temporal cue (i.e., by the duration of a stimulus portion or by the rate of change in some acoustic parameter) is modified after a temporal change is introduced in portions of the context that are not themselves cues for the perception of the target segment.

Only a few representative findings are mentioned here. Miller and Liberman (1979) examined the stop–semivowel distinction ([ba]–[wa]), cued by the duration and rate of the initial formant transitions, and found that the category boundary shifted systematically with the duration of the vocalic portion (i.e., of the whole stimulus). A corresponding shift of the discrimination peak in an oddity task was reported by Miller (1980). This effect may have an auditory basis, for it has been found not only in human infants (Eimas & Miller, 1980) but also with sinewave analogs that were apparently perceived as nonspeech (Carrell, Pisoni, & Gans, Note 7). Simple durational variation may not be sufficient to create variations in perceived speaking rate, and therefore Miller's results may represent rate normalization of a basic psychophysical sort.

Fitch (1981) recently attempted to dissociate information about speaking rate from phonetically distinctive durational variation. The phonetic distinction studied was that between [dabi] and [dapi], as cued by the duration of the first stimulus portion ([dab] or [dap]). By manipulating the duration of natural utterances produced at different rates, she was able to show that speaking rate had a perceptual effect that was separate from that of physical duration. Thus, the information about speaking rate seems to be carried, in part, by more complex structural variables, such as the rate of spectral change in the signal. Soli (1982) recently obtained similar results in a thorough investigation of cues to the [jus]–[juz] distinction. These findings are considerably more difficult to explain by psychoacoustic principles.

The most convincing instances of rate normalization derive from studies that varied remote context. The perception of a variety of phonetic distinctions is sensitive to the perceived rate of articulation of a carrier sentence (e.g., Miller & Grosjean, 1981; Pickett & Decker, 1960; Summerfield, 1981). Miller and Grosjean (1981) showed that the articulation rate of the carrier sentence was more important than its pause rate, even though the critical phonetic contrast ("rabid"–"rapid") was cued primarily by the perceived duration of a silent interval. Findings such as these suggest that speaking rate is a rather abstract property whose perception requires an appreciation of articulatory and linguistic variables (see also Grosjean & Lane, 1976). Summerfield (1981) showed that the rate of a nonspeech carrier (a melody) does not affect speech perception, confirming that the listener's rate estimate must derive from speech to be relevant.

These findings are just a sampling of a much larger literature on perceptual adjustments for speaking rate (see Miller, 1981). Whether or not there are corresponding contextual effects in auditory judgments is not known (except for the above-cited study by Carrell et al., Note 7), although there is some plausibility in the hypothesis that the durations of adjacent or corresponding auditory intervals are judged relative to each other. Perhaps because this hypothesis seems more plausible than do possible auditory explanations of other context effects in speech, there have been few attempts so far to simulate speaking rate effects using nonspeech analog stimuli. Nevertheless, there is some evidence that even simple durational changes may be interpreted differently in speech and nonspeech modes. Smith (1978) presented two identical syllables in succession and varied their relative durations. Listeners had to judge either which syllable was more stressed (a linguistic judgment) or which syllable was longer in duration (an auditory judgment). The two kinds of judgment diverged: Stress judgments exhibited a tendency for the first syllable to be judged as stressed, whereas duration judgments showed no such bias. These results indicate that the linguistic function of acoustic segment duration cannot be directly predicted from auditory judgments of that duration. Presumably, in speech perception, acoustic segment duration is interpreted, as are all other cues, within a framework of tacitly known acoustic and/or articulatory patterns, such as the well-known lengthening of a final syllable (Klatt, 1976).

Sequential (Remote) Context Effects

Context effects due to preceding and following stimuli in a test sequence are a ubiquitous phenomenon and are well known also in auditory psychophysics. They include effects of neighboring stimuli (preceding and/or following a target stimulus) as well as effects due to a whole series of preceding stimuli, referred to variously as selective adaptation, anchoring, range, or frequency effects. Even though these effects are clearly not in any way specific to speech—and speech stimuli are by no means immune to them, as was once believed with regard to anchoring (Sawusch & Pisoni, 1973; Sawusch, Pisoni, & Cutting, 1974)—the pattern of the data obtained for speech may nevertheless exhibit peculiarities that are not observed with nonspeech stimuli. The most striking of these is, of course, the relative stability of phonetic boundaries. Although all boundaries can be shifted to some extent by contextual influences, most boundaries do not change very much. (Isolated vowels are a significant exception—see below.) Presumably, this is so because listeners have internal criteria based on their long experience with speech and especially with their native tongue. It might be argued that phonetic boundaries are stable because they coincide with natural auditory boundaries of some sort. However, the evidence for such a coincidence is not convincing (see my earlier discussion of categorical perception), and nonhuman subjects seem to exhibit much larger range-contingent boundary shifts for speech stimuli than do adult human subjects (Waters & Wilson, 1976).

Another example of an interesting discrepancy between speech and nonspeech is provided by the pattern of vowel context effects. Repp et al. (1979) found not only that isolated synthetic vowel stimuli presented in pairs exhibit large contextual effects (as shown earlier by Fry, Abramson, Eimas, & Liberman, 1962; Lindner, 1966; Thompson & Hollien, 1970; and others) but also that backward contrast (the influence of the second stimulus on perception of the first) was stronger than forward contrast (the influence of the first stimulus on perception of the second). These results are interesting in the light of other findings that show that nonspeech stimuli exhibit smaller contrast effects and no (or the opposite) difference between forward and backward contrast (Eimas, 1963; Fujisaki & Shigeno, 1979; Healy & Repp, 1982; Shigeno & Fujisaki, Note 8). Although it seems possible that an auditory explanation of these differences will eventually be found, the peculiar flexibility of vowel perception may also be grounded in the special status of vowels as nuclear elements in the speech message. Perhaps the modifiability of vowel perception corresponds to the remarkable contextual variability that vowels exhibit in the speech signal.

Other Perceptual Integration Effects

A discussion of evidence for a phonetic mode of perception would not be complete without mention of two strands of research that make a particularly important contribution. They both deal with the integration of cues separated not in time but in space or even occurring in different modalities.

Duplex Perception

Duplex perception is the newly coined (Liberman, 1979) name for a phenomenon originally discovered by Rand (1974) and described earlier in this article: An isolated formant transition presented to one ear simultaneously with the "base" (a synthetic syllable bereft of that formant transition) in the other ear is perceived as a lateralized nonspeech "chirp," although at the same time, it contributes (presumably by some process of central integration) to the perception of the syllable in the other ear. The phenomenon by itself shows that the same input may be perceived in auditory and phonetic modes at the same time: The transition is segregated at the auditory level yet phonetically integrated with the base. Several recent studies show that various experimental variables affect either the auditory or the phonetic part of the duplex percept but not both.

Thus, Isenberg and Liberman (1978) varied the intensity of the isolated transition. The subjects perceived changes in the loudness of the chirp, but they could not detect any change in the loudness of the syllable in the other ear, even though they perceived the phonetic segment specified by the transition. Liberman, Isenberg, and Rakerd (1981) immediately preceded the base with a fricative noise that was appropriate for [s], which (in the absence of any intervening silence) inhibited the perception of the stop consonant ([p] or [t]) that the base otherwise would have generated in conjunction with the transition in the other ear. Listeners found it difficult to discriminate [s] + [pa] and [s] + [ta] as long as they attended to the side on which the speech was heard, for both stimuli sounded like [sa]. Nevertheless, their discrimination of [p]-chirps from [t]-chirps in the other ear was highly accurate. Recently, Mann, Madden, Russell, and Liberman (1981) used the duplex perception paradigm to examine further the effect (discovered by Mann, 1980) of a preceding liquid on stop consonant perception. When the syllables [al] or [ar] preceded the base of a stimulus from a [da]-[ga] continuum, the context effect was obtained in phonetic perception (more [ga] percepts following [al]) while the perception of the isolated transition in the other ear was unaltered.

Effects similar to duplex perception have been reported, where some nonspeech stimulus in one ear affected phonetic perception in the other ear while retaining its nonspeech quality. For example, Pastore (1978) found that when the syllable [pa] in one ear was accompanied by a burst of noise in the other ear, the phonetic percept changed to [ta]. Apparently, the noise—even though it did not have the appropriate timing, duration, and envelope—was interpreted by listeners as a [t]-release burst and was integrated with the syllable in the other ear. There is no doubt, however, that listeners continued to hear the noise burst as a nonspeech sound. The finding by Repp (1976) that the pitch of an isolated vowel in one ear affected the perception of the voiced–voiceless distinction for stop-consonant-vowel syllables in the other ear may be taken as another instance of duplex perception. Presumably, listeners could have accurately judged the pitch of the isolated vowel without destroying its phonetic effect.

Duplex perception phenomena provide evidence for the distinction between auditory and phonetic modes of perception. They show that, in the duplex situation, the auditory mode can gain access to the input from individual ears, whereas the phonetic mode operates on the combined input from both ears. The "phonological fusion" discovered by Day (1968)—two dichotic utterances such as "banket" and "lanket" yield the percept "blanket"—is yet another example of the abstract, nonauditory level of integration that characterizes the phonetic mode.

Audiovisual Integration

A most important recent discovery is the finding of an influence of visual articulatory information on phonetic perception (MacDonald & McGurk, 1978; McGurk & MacDonald, 1976; Summerfield, 1979). Of course, it has been known for a long time that lip reading aids speech perception, especially for the hard of hearing, but only recently has it become clear how tight audiovisual integration can be. McGurk and MacDonald (1976) presented a video display of a person's face saying simple syllables in synchrony with acoustic recordings of syllables from the same set. When the visual and auditory information disagreed, the visual information exerted a strong influence on the subjects' percepts, primarily due to the readily perceived presence versus absence of visible lip closure. Thus, when a visual [da] or [ga] was paired with an auditory [ba], subjects usually reported "da".[8]

The interpretation of this finding is straightforward and of great theoretical significance. Clearly, subjects somehow combine the articulatory information gained from the visual display with that gained from the acoustic signal. In Summerfield's (1979) words, "optical and acoustic displays are coperceived in a common metric closely related to that of articulatory dynamics" (p. 314). Audiovisual integration provides some of the strongest evidence we have for the existence of a speech-specific mode of perception that makes use of articulatory, as

[8] I have experienced this effect myself (together with a number of my colleagues at Haskins) and can confirm that it is a true perceptual phenomenon and not some kind of inference or bias in the face of conflicting information. The observer really believes that he or she hears what, in fact, he or she only sees on the screen; there is little or no awareness of anything odd happening. Nevertheless, the effect is not always that strong; its presence and strength depend on the particular combination of syllables in a way that can also, in part, be explained by reference to articulation. It is strongest when the visual information makes the auditory information impossible in articulatory terms. The details of the effect and of the relevant variables remain to be investigated.

opposed to general auditory, information. The common metric of visual and auditory speech input represents a modality-independent, presumably articulation-based level of abstraction that is the likely site of the integration and context effects reviewed above. Phonetic perception in the auditory modality (when speech enters through the ears) is likely to be in every sense as abstract as it is in the visual modality (when articulatory movements are observed directly).

In a recent ingenious study, Roberts and Summerfield (1981) used the audiovisual technique to demonstrate that selective adaptation of phonetic judgments is a purely auditory effect. Although conflicting visual information changed the listeners' phonetic interpretation of an adapting stimulus, it had no effect whatsoever on the direction or magnitude of the adaptation effect. Besides its implications for the selective adaptation paradigm (see also Sawusch & Jusczyk, 1981), this elegant study provides further evidence for the autonomy of phonetic perception.

Disruption of Perceptual Integration

As was pointed out in the discussion of speaker normalization effects, a simulated change in vocal tract size (or in any other speaker characteristic, such as fundamental frequency) must not disrupt the perceptual coherence of an utterance if a normalization effect is to be observed. In the case of formant transitions leading into a vocalic stimulus portion, or of an aperiodic portion (fricative noise) being followed by a periodic portion, perceptual coherence is easily maintained when the formant frequencies of the vowel are changed. However, when two periodic signal portions that are appropriate to different vocal tracts are juxtaposed, a change in speaker may be perceived, and this may lead to the disruption of whatever perceptual interactions (trading relations or context effects) may have taken place between the two periodic signal portions. There are several examples of this phenomenon in the recent literature.

For example, Darwin and Bethell-Fox (1977) showed that, by changing fundamental frequency abruptly at points of transition, a speech stimulus originally perceived as a smooth alternation of a liquid consonant (or semivowel) and a vowel could be changed into a train of stop-vowel syllables perceived as being produced in alternation by two different speakers. The manipulation of fundamental frequency signaled a change in source and thus "split" the formant transitions into portions that effectively became new cues, signaling stop consonants rather than liquids or semivowels.

Dorman et al. (1979, Exp. 6) studied a situation in which the perception of a syllable-final stop consonant depends on whether or not there is a sufficient period of (near) silence to indicate closure. An utterance such as [babda] is generally perceived as [bada] if the stop closure interval is removed. Dorman et al. found, however, that when the first syllable, [bab], is produced by a male speaker and the second syllable, [da], by a female speaker, the syllable-final stop in [bab] is clearly perceived. Because of the perceived change in speakers, listeners no longer recognize the absence of a closure interval; the critical syllable-final stop is now in utterance-final position. It is interesting that two subjects who reported that they did not notice a change in speaker also failed to perceive the syllable-final stop consonant in the absence of closure.

Conversely, an interval of silence in an utterance may lose its perceptual value when a change of speaker is perceived to occur across it (Dorman et al., 1979, Exp. 7): When silence is inserted into the utterance "say shop" immediately preceding the fricative noise, listeners report "say chop". Nevertheless, when "say" is spoken by a male voice and "shop" by a female voice, this effect no longer occurs; the silence loses its phonetic significance, and the second syllable remains "shop".[9]

This effect was further investigated by Dechovitz, Rakerd, and Verbrugge (1980),

[9] These experiments concern the disruption of perceptual integration of cues. Context effects, however, can presumably be similarly blocked by a change in apparent source. Diehl, Souther, and Convis (1980) recently reported a study in which a rate normalization effect (of a precursor on the /ga/-/ka/ distinction) was eliminated by a change of voice. Unfortunately, their data were not entirely consistent and call for replication.

who varied the perceived continuity of the test utterance "Let's go shop (chop)" by having speakers produce either the whole phrase or just "Let's go". Silence inserted (or removed from) between the "go" and the "shop (chop)" of a continuous utterance had the expected effect on phonetic perception: "Shop" was perceived as "chop" when silence was present, and "chop" was perceived as "shop" when there was no silence. Nevertheless, when the "Let's go" with phrase-final intonation was followed by either "shop" or "chop" from a different production, there were no such effects: "Shop (chop)" remained "shop (chop)". A change from a female speaker to a male speaker between "Let's go" and "shop (chop)" did *not* disrupt perceptual integration as long as the "Let's go" derived from a continuous utterance of "Let's go shop (chop)". This finding is in apparent contradiction to that described in the preceding paragraph. Dechovitz et al. interpreted it as showing that dynamic information for utterance continuity may override a perceived change in source (despite the concomitant auditory discontinuities). If this interpretation is correct, it may point to another instance where purely auditory principles fail to explain phonetic perception. Although some of the variables that determine the perceived continuity of an utterance are likely to be auditory (see Bregman, 1978), there may also be speech-specific factors that reflect what listeners consider plausible and possible in the dynamic context of natural utterances.

Conclusions

The findings reviewed above provide a wealth of results that, in large measure, cannot be accounted for by our current knowledge of auditory psychophysics. Although there remains much to be learned about the perception of complex auditory stimuli, some trading relations and context effects seem a priori unlikely to reflect an auditory level of interaction, and at least one phenomenon—audiovisual integration—simply cannot derive from that level. Although efforts to delineate the role of general auditory processes in speech perception should certainly continue, it may be predicted that this role will be restricted largely to the perception of nonphonetic stimulus attributes.

This is not to say that auditory properties of the signal are not the basic carrier of the linguistic message. However, auditory psychophysics gains knowledge about the perception of these properties mainly from listeners' judgments in psychophysical experiments, and these judgments are made in a different frame of reference from the judgments of speech. Auditory variables, but *not* auditory judgments, are the basis of phonetic perception. Limitations of detectability and resolution imposed by the auditory system may not play any important role in phonetic distinctions. For instance, there is no reason why phonetic category boundaries could not be placed at suprathreshold auditory parameter settings that seem arbitrary from a psychophysical viewpoint but are well motivated by the articulatory and acoustic patterns that characterize a given language. Furthermore, even though phonetic and auditory boundaries may sometimes coincide, there is the more fundamental question whether such "laboratory boundaries" play any role in the perception of natural speech, considering the fact that natural speech is different in a number of ways from the artificial stimuli used in speech discrimination tasks. Although the objection of ecologically invalid stimuli extends to most of the studies reviewed in this article, the present emphasis has been on *processes of perceptual integration* that promise to be more general than static concepts such as boundary locations.

Two possible criticisms of the research reviewed here should be mentioned. One is that nearly all of the studies demonstrated perceptual integration in situations of high uncertainty produced by ambiguous settings of the primary cue(s) for a given phonetic distinction. The perceptual integration observed may have been motivated by that ambiguity. In that case, it may be that perceptual integration does not occur to the same extent in natural situations, where the primary cues are often sufficient for accurate phonetic perception.

The other criticism is that, although the trading relations and context effects reviewed here have been described as complex interactions between separate cues, it may well be that these cues do not function as perceptual entities that are "extracted" and then recombined into a unitary phonetic percept (see Bailey & Summerfield, 1980). In that view, cues serve only descriptive purposes; the perceptual interactions between them can be understood as resulting from the listeners' apprehension of the articulatory events they convey. Although cues (i.e., acoustic segments) are indispensable for describing how the articulatory information is represented in the signal, we need not postulate special perceptual processes that construct or derive the articulatory information from these elementary pieces. Rather, the articulatory information may be said to be directly available (Gibson, 1966; Neisser, 1976). This is an attractive proposal; we should not forget, however, that there are real questions to be answered about the *mechanisms* that accomplish phonetic perception and that we know very little about at present. If cues and their interactions have no place in a description of these mechanisms, we face the more fundamental problem of finding the proper ingredients for a model of speech perception.

To understand how our perceptual systems work, we need to understand how a complex biological system (our brain) integrates and differentiates information, how it is modified by experience, and how the structure of the input (i.e., the environment) gets to be represented in the system. These are complex questions whose solution will not come easily. The computer analogies that underlie most current models of perception are largely tautological and distract from the fundamental biological and philosophical problems that lie at the heart of the problem of perception (see, e.g., Hayek, 1952; Piaget, 1967; Studdert-Kennedy, 1982, Note 9). In a particularly enlightening discussion, Fodor (Note 10) recently argued for the modularity of the speech (and language) system (i.e., for its specificity and relative isolation from other perceptual and cognitive systems). He also pointed out that it is precisely such modular systems that we have some hope of understanding, whereas explanations of perception in terms of general principles are probably doomed to failure. Thus, we should not be surprised to find that speech perception is accomplished by means that are entirely particular to that mode. The problem of how to investigate and describe those means will keep us busy for some time to come.

Reference Notes

1. Grunke, M. E., & Pisoni, D. B. Some experiments on perceptual learning of mirror-image acoustic patterns. *Research on Speech Perception* (Department of Psychology, Indiana University), 1979, *5*, 147-182.
2. Serniclaes, W. La simultanéité des indices dans la perception du voisement des occlusives. *Rapport d'Activités de l'Institut de Phonétique* (Bruxelles: Université Libre), 1973, *7*(2), 59-67.
3. Serniclaes, W. Traitement indépendant ou interaction dans le processus de structuration perceptive des indices de voisement? *Rapport d'Activités de l'Institut de Phonétique* (Bruxelles: Université Libre), 1975, *9*(2), 47-57.
4. Miller, J. L., & Eimas, P. D. *Contextual perception of voicing by infants*. Paper presented at the meeting of the Society for Research in Child Development, Boston, April 1981.
5. 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.
6. Zue, V. W. *Acoustic characteristics of stop consonants: A controlled study* (Tech. Rep. 523). Lexington, Mass.: Massachusetts Institute of Technology, Lincoln Laboratory, May 1976.
7. 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* (Department of Psychology, Indiana University), 1980, *6*, 421-436.
8. 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.
9. Studdert-Kennedy, M. *Are utterances prepared and perceived in parts? Perhaps.* Paper presented at the First International Conference on Event Perception, University of Connecticut, Storrs, June 1981.
10. Fodor, J. A. *The modularity of mind*. Unpublished manuscript, Massachusetts Institute of Technology, 1981.

References

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

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. Perception*. New York: Academic Press, 1980.

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. (ERIC Document Reproduction Service No. ED 147 892)

Barton, D. Phonemic perception in children. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child phonology: Vol. 2. Perception*. New York: Academic Press, 1980.

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.

Blechner, M. J., Day, R. S., & Cutting, J. E. Processing two dimensions of nonspeech stimuli: The auditory-phonetic distinction reconsidered. *Journal of Experimental Psychology: Human Perception and Performance*, 1976, *2*, 257-266.

Blumstein, S. E., & Stevens, K. N. Acoustic invariance in speech production. *Journal of the Acoustical Society of America*, 1979, *66*, 1001-1017.

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.

Bradshaw, J. L., & Nettleton, N. C. The nature of hemispheric specialization in man. *Behavioral and Brain Sciences*, 1981, *4*, 51-63.

Bregman, A. S. The formation of auditory streams. In J. Requin (Ed.), *Attention and performance VII*. Hillsdale, N.J.: Erlbaum, 1978.

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, *63*, 456-468.

Carden, G., Levitt, A. G., 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.

Cole, R. A., & Scott, B. Perception of temporal order in speech: The role of vowel transitions. *Canadian Journal of Psychology*, 1973, *27*, 441-449.

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

Cutting, J. E. Auditory and linguistic processes in speech perception: Inferences from six fusions in dichotic listening. *Psychological Review*, 1976, *83*, 114-140.

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.

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

Darwin, C. J., & Bethell-Fox, C. E. Pitch continuity and speech source attribution. *Journal of Experimental Psychology: Human Perception and Performance*, 1977, *3*, 665-672.

Day, R. S. *Fusion in dichotic listening*. Unpublished doctoral dissertation, Stanford University, 1968.

Dechovitz, D. R., Rakerd, B., & Verbrugge, R. R. Effects of utterance continuity on phonetic judgments. *Haskins Laboratories Status Report on Speech Research*, 1980, *SR-62*, 101-116. (ERIC Document Reproduction Service No. ED 196 099)

Delattre, P. C., Liberman, A. M., Cooper, F. S., & Gerstman, L. J. An experimental study of the acoustic determinants of vowel color: Observations on one- and two-formant vowels synthesized from spectrographic patterns. *Word*, 1952, *8*, 195-210.

Denes, P. Effect of duration on the perception of voicing. *Journal of the Acoustical Society of America*, 1955, *27*, 761-764.

Derr, M. A., & Massaro, D. W. The contribution of vowel duration, F_0 contour, and frication duration as cues to the /juz/-/jus/ distinction. *Perception & Psychophysics*, 1980, *27*, 51-59.

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

Diehl, R. L., Souther, A. F., & Convis, C. L. Conditions on rate normalization in speech perception. *Perception & Psychophysics*, 1980, *27*, 435-443.

Divenyi, P. L. Some psychoacoustic factors in phonetic analysis. *Proceedings of the Ninth International Congress of Phonetic Sciences* (Vol. 2). Copenhagen: University of Copenhagen, 1979.

Dorman, M. F., & Raphael, L. J. Distribution of acoustic cues for stop consonant place of articulation in VCV syllables. *Journal of the Acoustical Society of America*, 1980, *67*, 1333-1335.

Dorman, M. F., Raphael, L. J., & Isenberg, D. Acoustic cues for a fricative-affricate contrast in word-final position. *Journal of Phonetics*, 1980, *8*, 397-405.

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

Dorman, M. F., Studdert-Kennedy, M., & Raphael, L. J. Stop-consonant recognition: Release bursts and formant transitions as functionally equivalent, context-dependent cues. *Perception & Psychophysics*, 1977, *22*, 109-122.

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

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

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

Fitch, H. L. Distinguishing temporal information for speaking rate from temporal information for intervocalic stop consonant voicing. *Haskins Laboratories Status Report on Speech Research*, 1981, *SR-65*, 1-32. (ERIC Document Reproduction Service No. ED 201 022)

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.

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

Fujisaki, H., & Shigeno, S. Context effects in the categorization of speech and nonspeech stimuli. In J. J. Wolf & D. H. Klatt (Eds.), *Speech communication papers presented at the 97th Meeting of the Acoustical Society of America.* New York: Acoustical Society of America, 1979.

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

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, Mass.: Houghton-Mifflin, 1966.

Grosjean, F., & Lane, H. How the listener integrates the components of speaking rate. *Journal of Experimental Psychology: Human Perception and Performance,* 1976, *2,* 538–543.

Haggard, M. P., Ambler, S., & Callow, M. Pitch as a voicing cue. *Journal of the Acoustical Society of America,* 1970, *47,* 613–617.

Haggard, M., Summerfield, Q., & Roberts, M. Psychoacoustical and cultural determinants of phoneme boundaries: Evidence from trading F_0 cues in the voiced–voiceless distinction. *Journal of Phonetics,* 1981, *9,* 49–62.

Halle, M., & Stevens, K. N. Analysis by synthesis. In W. Wathen-Dunn & L. E. Woods (Eds.), *Proceedings of the seminar on speech compression and processing* (Vol. 2). AFCRC-TR-59-198, USAF Cambridge Research Center, 1959.

Harris, K. S., Hoffman, H. S., Liberman, A. M., Delattre, P. C., & Cooper, F. S. Effect of third-formant transitions on the perception of the voiced stop consonants. *Journal of the Acoustical Society of America,* 1958, *30,* 122–126.

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

Hayek, F. A. *The sensory order.* Chicago: University of Chicago Press, 1952.

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

Hoffman, H. S. Study of some cues in the perception of the voiced stop consonants. *Journal of the Acoustical Society of America,* 1958, *30,* 1035–1041.

Hogan, J. T., & Rozsypal, A. J. Evaluation of vowel duration as a cue for the voicing distinction in the following word-final consonant. *Journal of the Acoustical Society of America,* 1980, *67,* 1764–1771.

House, A. S., Stevens, K. N., Sandel, T. T., & Arnold, J. B. On the learning of speechlike vocabularies. *Journal of Verbal Learning and Verbal Behavior,* 1962, *1,* 133–143.

Isenberg, D., & Liberman, A. M. Speech and nonspeech percepts from the same sound. *Journal of the Acoustical Society of America,* 1978, *64* (Supplement No. 1), S20. (Abstract)

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

Kewley-Port, D. *Representations of spectral change as cues to place of articulation of stop consonants.* Unpublished doctoral dissertation, City University of New York, 1981.

Kohler, K. J. Dimensions in the perception of fortis and lenis plosives. *Phonetica,* 1979, *36,* 332–343.

Klatt, D. H. Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America,* 1976, *59,* 1208–1221.

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: Identification functions for synthetic VOT stimuli. *Journal of the Acoustical Society of America,* 1978, *63,* 905–917.

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

Liberman, A. M. Duplex perception and integration of cues: Evidence that speech is different from nonspeech and similar to language. *Proceedings of the Ninth International Congress of Phonetic Sciences* (Vol. 2). Copenhagen: University of Copenhagen, 1979.

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

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, *4,* 175–195.

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., Isenberg, D., & Rakerd, B. Duplex perception of cues for stop consonants: Evidence for a phonetic mode. *Perception & Psychophysics,* 1981, *30,* 133–143.

Lindner, G. Veränderung der Beurteilung synthetischer Vokale unter dem Einfluss des Sukzessivkontrastes. *Zeitschrift für Phonetik, Sprachwissenschaft, und Kommunikationsforschung,* 1966, *19,* 287–307.

Lisker, L. Closure duration and the intervocalic voiced-voiceless distinction in English. *Language,* 1957, *33,* 42–49.

Lisker, L. Is it VOT or a first-formant transition detector? *Journal of the Acoustical Society of America*, 1975, *57*, 1547–1551.

Lisker, L. Closure hiatus: Cue to voicing, manner and place of consonant occlusion. *Haskins Laboratories Status Report on Speech Research*, 1978, *SR-53* (Vol. 1), 79–86. (ERIC Document Reproduction Service No. ED 155 760) (a)

Lisker, L. *Rapid* vs. *rabid*: A catalogue of acoustic features that may cue the distinction. *Haskins Laboratories Status Report on Speech Research*, 1978, *SR-54*, 127–132. (ERIC Document Reproduction Service No. ED 161 096) (b)

Lisker, L. On buzzing the English /b/. *Haskins Laboratories Status Report on Speech Research*, 1978, *SR-55/56*, 181–188. (ERIC Document Reproduction Service No. ED 166 757) (c)

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.

Lisker, L., & Price, P. J. Context-determined effects of varying closure duration. In J. J. Wolf & D. H. Klatt (Eds.), *Speech communication papers presented at the 97th meeting of the Acoustical Society of America*. New York: Acoustical Society of America, 1979.

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

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

Mann, V. A., Madden, J., Russell, J. M., & Liberman, A. M. Further investigation into the influence of preceding liquids on stop consonant perception. *Journal of the Acoustical Society of America*, 1981, *69* (Supplement No. 1), S91. (Abstract)

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

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

Massaro, D. W., & Cohen, M. M. The contribution of fundamental frequency and voice onset time to the /zi/–/si/ distinction. *Journal of the Acoustical Society of America*, 1976, *60*, 704–717.

Massaro, D. W., & Cohen, M. M. Voice onset time and fundamental frequency as cues to the /zi/–/si/ distinction. *Perception & Psychophysics*, 1977, *22*, 373–382.

Mattingly, I. G., & Levitt, A. G. Perception of stop consonants before low unrounded vowels. *Haskins Laboratories Status Report on Speech Research*, 1980, *SR-61*, 167–174. (ERIC Document Reproduction Service No. ED 185 636)

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. Vocal tract normalization for /s/ and /š/. *Haskins Laboratories Status Report on Speech Research*, 1976, *SR-48*, 67–73. (ERIC Document Reproduction Service No. ED 135 028)

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

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

Miller, J. L. 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, N.J.: Erlbaum, 1981.

Miller, J. L., & Grosjean, F. How the components of speaking rate influence perception of phonetic segments. *Journal of Experimental Psychology: Human Perception and Performance*, 1981, *7*, 208–215.

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.

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

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

Pastore, R. E. Contralateral cueing effects in the perception of aspirated stop consonants. *Journal of the Acoustical Society of America*, 1978, *64* (Supplement No. 1), S17. (Abstract)

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

Pastore, R. E. et al. Processing interactions between two dimensions of nonphonetic auditory signals. *Journal of Experimental Psychology: Human Perception and Performance*, 1976, *2*, 267–276.

Pastore, R. E. et al. Common factor model of categorical perception. *Journal of Experimental Psychology: Human Perception and Performance*, 1977, *3*, 686–696.

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

Piaget, J. *Biology and knowledge*. Chicago: University of Chicago Press, 1967.

Pickett, J. M., & Decker, L. R. Time factors in perception of a double consonant. *Language and Speech*, 1960, *3*, 11–17.

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.

Pols, L. C. W., & Schouten, M. E. H. Identification of deleted consonants. *Journal of the Acoustical Society of America*, 1978, *64*, 1333–1337.

Price, P. J., & Lisker, L. (/b/→/p/) but ~(/p/→/b/). In J. J. Wolf & D. H. Klatt (Eds.), *Speech communication papers presented at the 97th* *meeting of the Acoustical Society of America*. New York: Acoustical Society of America, 1979.

Rand, T. C. Vocal tract size normalization in the perception of stop consonants. *Haskins Laboratories Status Report on Speech Research*, 1971, *SR-25/26*, 141-146. (ERIC Document Reproduction Service No. ED 056 560)

Rand, T. C. Dichotic release from masking for speech. *Journal of the Acoustical Society of America*, 1974, *55*, 678-680.

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.

Remez, R. E. Adaptation of the category boundary between speech and nonspeech: A case against feature detectors. *Cognitive Psychology*, 1979, *11*, 38-57.

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

Remez, R. E., Rubin, P. E., Pisoni, D. B., & Carrell, T. D. Speech perception without traditional speech cues. *Science*, 1981, *212*, 947-950.

Repp, B. H. Dichotic "masking" of voice onset time. *Journal of the Acoustical Society of America*, 1976, *59*, 183-194.

Repp, B. H. Perceptual integration and differentiation of spectral cues for intervocalic stop consonants. *Perception & Psychophysics*, 1978, *24*, 471-485.

Repp, B. H. Relative amplitude of aspiration noise as a voicing cue for syllable-initial stop consonants. *Language and Speech*, 1979, *22*, 173-189.

Repp, B. H. Bidirectional contrast effects in the perception of VC-CV sequences. *Haskins Laboratories Status Report on Speech Research*, 1980, *SR-63/64*, 157-176. (ERIC Document Reproduction Service No. ED 197 416) (a)

Repp, B. H. Perception and production of two-stop-consonant sequences. *Haskins Laboratories Status Report on Speech Research*, 1980, *SR-63/64*, 177-194. (ERIC Document Reproduction Service No. ED 197 416) (b)

Repp, B. H. Auditory and phonetic trading relations between acoustic cues in speech perception: Preliminary results. *Haskins Laboratories Status Report on Speech Research*, 1981, *SR-67/68*, 165-190. (ERIC Document Reproduction Service [No. not yet assigned]) (a)

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

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 acoustic cues for stop, fricative and affricate manner. *Journal of Experimental Psychology: Human Perception and Performance*, 1978, *4*, 621-637.

Repp, B. H., & Mann, V. A. Fricative-stop coarticulation: Acoustic and perceptual evidence. *Haskins Laboratories Status Report on Speech Research*, 1981, *SR-67/68*, 255-268. (ERIC Document Reproduction Service [No. not yet assigned]) (a)

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

Roberts, M., & Summerfield, Q. Audio-visual adaptation in speech perception. *Perception & Psychophysics*, 1981, *30*, 309-314.

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

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., & Pisoni, D. B. Category boundaries for speech and nonspeech sounds. *Journal of the Acoustical Society of America*, 1973, *54*, 76. (Abstract)

Sawusch, J. R., Pisoni, D. B., & Cutting, J. E. Category boundaries for linguistic and non-linguistic dimensions of the same stimuli. *Journal of the Acoustical Society of America*, 1974, *55* (Supplement No. 1), S55. (Abstract)

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

Schwartz, M. F. Identification of speaker sex from isolated voiceless fricatives. *Journal of the Acoustical Society of America*, 1968, *43*, 1178-1179.

Searle, C. L., Jacobson, J. Z., & Rayment, S. G. Stop consonant discrimination based on human audition. *Journal of the Acoustical Society of America*, 1979, *65*, 799-809.

Serniclaes, W. Perceptual processing of acoustic correlates of the voicing feature. *Proceedings of the Speech Communication Seminar* (Stockholm), 1974, 87-93.

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

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

Smith, M. R. Perception of word stress and syllable length. *Journal of the Acoustical Society of America*, 1978, *63* (Supplement No. 1), S55. (Abstract)

Soli, S. D. Structure and duration of vowels together specify fricative voicing. *Journal of the Acoustical Society of America*, 1982, *71*, in press.

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.

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

Studdert-Kennedy, M. Speech perception. In N. J. Lass (Ed.), *Contemporary issues in experimental phonetics*. New York: Academic Press, 1976.

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

Studdert-Kennedy, M. A note on the biology of speech perception. In J. Mehler, M. Garrett, & E. Walker (Eds.), *Perspectives in mental representation*. Hillsdale, N.J.: Erlbaum, 1982.

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

Summerfield, A. Q. *Information-processing analyses of perceptual adjustments to source and context variables in speech*. Unpublished doctoral dissertation, Queen's University of Belfast, 1975.

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

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

Summerfield, Q. Does VOT equal TOT or NOT? Examination of a possible auditory basis for the perception of voicing in initial stops. *Journal of the Acoustical Society of America*, 1982, *71*, in press.

Summerfield, Q., Bailey, P. J., Seton, J., & Dorman, M. F. Fricative envelope parameters and silent intervals in distinguishing "slit" and "split". *Phonetica*, 1981, *38*, 181-192.

Summerfield, A. Q., & Haggard, M. P. Perceptual processing of multiple cues and contexts: Effects of following vowel upon stop consonant voicing. *Journal of Phonetics*, 1974, *2*, 279-295.

Summerfield, A. Q., & Haggard, M. P. Vocal tract normalization as demonstrated by reaction times. In G. Fant & M. A. A. Tatham (Eds.), *Auditory analysis and perception of speech*. London: Academic Press, 1975.

Summerfield, Q., & Haggard, M. P. 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*, 435-448.

Thompson, C. L., & Hollien, H. Some contextual effects on the perception of synthetic vowels. *Language and Speech*, 1970, *13*, 1-13.

van Heuven, V. J. The relative contribution of rise time, steady time, and overall duration of noise bursts to the affricate-fricative distinction in English: A reanalysis of old data. In J. J. Wolf & D. H. Klatt (Eds.), *Speech communication papers presented at the 97th meeting of the Acoustical Society of America*. New York: Acoustical Society of America, 1979.

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. Petersen (Eds.), *Sensory and perceptual development*. New York: Academic Press, 1981.

Warren, R. M., Obusek, C. J., Farmer, R. M., & Warren, R. P. Auditory sequence: Confusion of patterns other than speech or music. *Science*, 1969, *164*, 586-587.

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.

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

Wolf, C. G. Voicing cues in English final stops. *Journal of Phonetics*, 1978, *6*, 299-309.

Wood, C. C. Auditory and phonetic levels of processing in speech perception: Neurophysiological and information-processing analyses. *Journal of Experimental Psychology: Human Perception and Performance*, 1975, *104*, 3-20.

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

Zlatin, M. A. Voicing contrast: Perceptual and productive voice onset time characteristics of adults. *Journal of the Acoustical Society of America*, 1974, *56*, 981-994.

Zwicker, E., Terhardt, E., & Paulus, E. Automatic speech recognition using psychoacoustic models. *Journal of the Acoustical Society of America*, 1979, *65*, 487-498.

Received September 24, 1981 ■