*Perception & Psychophysics*
*1986, 39 (2), 129-142*

# Trading relations in speech and nonspeech

ELLEN M. PARKER, RANDY L. DIEHL, and KEITH R. KLUENDER
*University of Texas at Austin, Austin, Texas*

Two acoustic variables that correlate with the distinction between intervocalic [b] and [p] are closure duration and presence or absence of low-frequency glottal pulsing during the closure interval. These variables may be considered to exhibit a trading relation (Repp, 1982), to the extent that a longer closure is required to perceive the consonant as voiceless when glottal pulsing is present than when it is not. Such trading relations have been interpreted as reflecting a special speech mode of perception. In the present experiments, we demonstrated a trading relation between closure duration and closure pulsing for a set of [aba]-[apa] stimuli. Next we showed that a similar effect could be obtained with square-wave analogue stimuli that mimicked the segment durations and peak amplitudes of the speech stimuli but that were not phonetically categorizable. This nonspeech trading relation depended on the degree of spectral continuity between the low-frequency pulsing and the adjacent portions of the square wave. The implications of these results for the speech mode hypothesis are discussed.

A persistent and influential claim is that the perception of speech sounds requires a specialized mode of phonetic processing above and beyond the general auditory and cognitive capabilities used in detecting, discriminating, and recognizing nonspeech acoustic patterns (Liberman, 1970; Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Liberman & Studdert-Kennedy, 1978; Repp, 1982). In recent years, the speech mode hypothesis has been proffered as a special case of Fodor's (1983) modularity proposal (Liberman, 1982; Liberman & Mattingly, 1985; Mattingly & Liberman, 1985). An input system is taken to a modular if, among other things, it is vertically organized, that is, if it operates on a limited stimulus domain (such as speech) with special-purpose mechanisms that are largely innate. Horizontal systems, by contrast, are more general-purpose mechanisms that operate across stimulus domains. According to more recent versions of the speech mode hypothesis, speech perception is accomplished by vertical, rather than horizontal, processing mechanisms.

Until about 1975, the most favorable evidence for the speech mode hypothesis came from studies of categorical perception. Listeners are said to perceive a dimension categorically if they divide the items along the dimension into discrete labeling categories and discriminate only between items from separate categories. Speech dimensions that were found in early studies to be perceived in a nearly categorical manner included place of articulation of word-initial stop consonants (Liberman, Harris, Hoffman, & Griffith, 1957; Mattingly, Liberman, Syrdal, & Halwes, 1971), voicing of initial and medial stops (Abramson & Lisker, 1970; Liberman, Harris, Eimas, Lisker, & Bastian, 1961; Lisker & Abramson, 1970), [r] versus [l] (Miyawaki et al., 1975), and others.

In many of these early studies, both speech and nonspeech dimensions were investigated, and almost invariably only the speech dimensions yielded categorical results.[1] For example, Miyawaki et al. (1975) tested subjects using both a three-formant stimulus series ranging from [ra] to [la] and a corresponding nonspeech series consisting of the third formant alone of each item in the [ra]-[la] series. (The third formant was the critical cue distinguishing [ra] from [la], because it alone varied across the [ra]-[la] series.) Whereas the [ra]-[la] stimuli were categorically perceived by English-speaking listeners, the nonspeech control stimuli were well discriminated across the entire series. The finding that categorical perception was apparently unique to speech encouraged the belief in a speech mode of processing.

More recently, however, the speech mode hypothesis has been challenged on the grounds that categorical perception is neither unique to speech sounds nor specific to human listeners. Although certain nonspeech stimuli, such as isolated formants or formant transitions, are not perceived categorically (Mattingly et al., 1971; Miyawaki et al., 1975), more complex nonspeech patterns, particularly those that have abstract acoustic properties analogous to speech, *have* been shown to yield categorical perception (Jusczyk, Pisoni, Walley, & Murray, 1980; Miller, Wier, Pastore, Kelly, & Dooling, 1976; Pisoni, 1976, 1977; see Pastore et al., 1976, for a discussion of potential problems associated with constructing suitable nonspeech analogue stimuli). One speech dimension that is perceived in a strongly categorical manner is voice onset time (VOT), the interval between the release burst of a stop consonant and the onset of waveform periodicity associated with voicing or vocal-fold vibration (Abramson & Lisker, 1970; Lisker & Abramson, 1970). Differ-

This work was supported by National Institutes of Health Grant HD 18060. We thank Gerald Lame for creating the waveform synthesis program used in generating the nonspeech stimulus patterns. Requests for reprints should be sent to Randy L. Diehl, Department of Psychology, 330 Mezes, University of Texas, Austin, Texas 78712.

ences in VOT are sufficient to signal the distinction between voiced and voiceless stops ([b] versus [$p^h$]; [d] versus [$t^h$]; [g] versus [$k^h$]) in word-initial position. The studies by Pisoni (1977) and Jusczyk et al. (1980) used stimuli consisting of two steady-state tones varying in relative temporal onset. These stimuli, which are abstractly analogous to VOT stimuli but which are not perceived as speechlike, were discriminated (by adults and infants) and labeled (by adults) in a manner comparable to VOT items. Such findings are significant for two reasons. They show that categorical perception is not unique to speech, and they suggest that, at least in some instances, categorical perception of speech dimensions may derive largely from psychoacoustic factors that are not speech specific.

Perhaps the most damaging evidence against the speech mode hypothesis comes from studies of speech perception in the chinchilla (Kuhl, 1981; Kuhl & Miller, 1975, 1978). The animals were trained to respond differently to two endpoint stimuli of a synthetic VOT series ([da], 0 msec VOT; [$t^h$a], 80 msec VOT) and then were tested with stimuli at intermediate values. Identification performance corresponded almost exactly to that of adult English-speaking listeners. Further generalization tests with bilabial ([ba]-[$p^h$a]) and velar ([ga]-[$k^h$a]) VOT stimuli, as well as tests of VOT discriminability, also showed agreement with English-speaking listeners. Additional evidence of categorical perception has been obtained recently with macaque monkeys (Kuhl & Padden, 1982). These results suggest that categorical perception of dimensions such as VOT may reflect general constraints on auditory processing among mammals and may have little to do with any species-specific speech mode of processing.

Miller et al. (1976) hypothesized that the enhanced discriminability at the voiced-voiceless boundary occurs because that region of the VOT dimension corresponds to a natural psychophysical boundary. A VOT smaller than the boundary value is below the listener's threshold for judging successive events (in this case, the release burst and voicing onset) as nonsimultaneous (see Hirsh, 1959; Hirsh & Sherrick, 1961). Above the boundary value, VOT discriminability decreases according to Weber's law. This model correctly predicts that nonspeech stimuli that mimic the temporal properties of VOT stimuli will also be perceived categorically (Miller et al., 1976; Pisoni, 1977; Stevens & Klatt, 1974); it also accounts for human adult and infant perceptual data, as well as for the animal results.

Given these challenges to the classical version of the speech mode hypothesis, some investigators have abandoned the hypothesis in favor of a general psychophysical approach to speech perception (e.g., Howell & Rosen, 1984; MacMillan, in press; Miller et al., 1976; Pastore, 1976, 1981; Pastore et al., 1977; Schouten, 1980), whereas others have sought new evidence for the speech mode (e.g., Liberman, 1982; Repp, 1982). Among the various recent findings that are taken to support the speech mode hypothesis, Repp (1982) has distinguished two important classes of phenomena: phonetic trading relations and context effects. For any given phonetic segment, there are, in general, many relevant acoustic cues (see Diehl & Kluender, in press). Voicing in initial stops, for example, is signaled by several acoustic variables, including the following: VOT (Lisker & Abramson, 1970; Lisker, Liberman, Erickson, Dechovitz, & Mandler, 1977), first-formant onset frequency (Lisker, 1975; Summerfield & Haggard, 1977), duration of voiced formant transitions (Lisker et al., 1977; Stevens & Klatt, 1974; Summerfield & Haggard, 1974), intensity of aspiration (Repp, 1979), and direction of fundamental frequency change at voicing onset (Haggard, Ambler, & Callow, 1970; Haggard, Summerfield, & Roberts, 1981). A phonetic trading relation is said to exist when a change in the value of one cue can be offset by an opposing change in another cue so that phonetic quality is preserved. Thus, for example, as VOT value is decreased, first-formant onset frequency may be increased to maintain a voiceless percept. Repp (1982) argued that the diverse phonetic trading relations that have been observed are not readily explained in terms of general auditory psychophysics, but instead seem to require the existence of a speech mode of perception. Listeners appear to have tacit knowledge of the full range of normal acoustic correlates of phonetic categories or of the underlying articulatory correlates. It is this phonetic knowledge that, according to Repp, accounts for the trading relations.

Repp (1982) made much the same argument concerning various context effects. They are analogous to trading relations, except that the interaction between cues occurs over a greater distance, say, between adjacent segments rather than within a segment. For example, Mann and Repp (1980) found an effect of the following vowel on the identification of a noise segment that was intermediate between [s] and [ʃ]. This segment was more likely to be labeled "s" before [u] than before [a]. Such a context effect can apparently be explained in terms of the listener's tacit knowledge of fricative-vowel coarticulation and/or its acoustic effects. Lip rounding appropriate for [u] occurs during the preceding fricative segment, causing the fricative noise to be lowered in frequency. In compensating perceptually for this, listeners may more readily accept a lower frequency noise as [s] (rather than [ʃ]) before a rounded vowel.

Although particular trading relations and context effects may plausibly be explained in terms of the listener's speech-specific knowledge, it is necessary to show that such effects do not arise for acoustically analogous nonspeech stimuli. Best, Morrongiello, and Robson (1981) synthesized speech patterns ranging from [sei] to [stei] and demonstrated a trading relation between silent gap duration (following [s]) and first-formant onset frequency. In a parallel experiment, they used nonspeech analogue stimuli with sine waves replicating the vocalic formant center-frequency patterns of the [sei] and [stei] items. Subjects who were able to hear the sine-wave analogues as [sei] and [stei] showed a trading relation between silent

gap duration and onset frequency of the lowest frequency sine wave. However, subjects who heard them as nonspeech either failed to show a trading relation or showed a trading relation that was considerably larger than that shown for the corresponding speech stimuli.

Other evidence was provided by Summerfield (1982), who compared listeners' identification performance for VOT stimuli ([ga]-[k$^h$a]) with that for two nonspeech stimulus sets that simulated the temporal properties of the VOT items. The stimuli in one set consisted of coterminous tones of varying relative temporal onset; in the other set, the items consisted of a higher frequency noise band that varied in onset relative to a 100-Hz pulse train that was band limited by the same first-formant filter that was used to generate the speech stimuli. For the speech stimuli, there was a clear trading relation between VOT and first-formant onset frequency (see also Lisker, 1975; Summerfield & Haggard, 1977); smaller and less reliable effects were observed for the nonspeech stimulus sets under analogous spectral manipulations.

To date, there have been at least two successful attempts to replicate speech trading relations or context effects with nonspeech analogue stimuli. Miller and Liberman (1979) had listeners identify sets of synthetic speech patterns that varied in formant-transition duration and ranged perceptually from [ba] (short transitions) to [wa] (long transitions). When the vowel was lengthened, the [b]/[w] perceptual boundary shifted toward longer transition durations. The authors argued that this effect represented an appropriate perceptual normalization for rate variation in speech. However, Pisoni, Carrell, and Gans (1983) demonstrated the same type of context effect using sine-wave analogues of the [ba]-[wa] stimuli and concluded that the effect for both speech and nonspeech conditions was attributable to general auditory factors, rather than to a speech mode of perception.

More recently, Hillenbrand (1984) showed that the observed trading relation between VOT and formant-transition duration in the perception of voicing contrasts (e.g., Lisker et al., 1977) could be duplicated with sine-wave analogue stimuli.

A potential problem in interpreting the positive results obtained with sine-wave analogues is that such stimuli have been shown to be *phonetically* categorizable (Bailey, Summerfield, & Dorman, 1977; Best et al., 1981; Remez, Rubin, Pisoni, & Carrell, 1981). It is possible, therefore, that any trading relations observed with sine-wave stimuli may reflect a speech mode of perception.[2] A more convincing test of the speech-mode explanation of trading relations requires a set of nonspeech analogue stimuli that are not phonetically categorizable. The present experiments were designed to satisfy this requirement.

We chose to investigate a very robust trading relation that characterizes the perception of voicing in word-medial stop consonants. Among the many articulatory/acoustic variables that distinguish a word such as *rapid* from its medially voiced counterpart *rabid*, two are known to be especially significant: stop closure duration and presence or absence of glottal pulsing during the closure interval. In general, medial voiceless stops are produced with considerably longer closure intervals than are corresponding voiced stops (Lisker, 1957), and voiceless closure intervals are free of the glottal pulsing (a low-frequency buzz) that characterizes voiced closure intervals (Lisker, Abramson, Cooper, & Schvey, 1969). Not surprisingly, each of these variables has a large effect on the perception of medial voicing, and Lisker (1978a) has demonstrated a reliable trading relation between them.

In two experiments, we examined the perceptual effects of closure duration and closure glottal pulsing in both speech and nonspeech stimulus sets. The nonspeech stimuli were constructed so as to preserve the segment durations and peak amplitudes of the speech stimuli without being phonetically categorizable. Our aim was to assess whether this trading relation might depend on general auditory constraints and processes or whether it is instead unique to speech processing.

## EXPERIMENT 1

### Method

**Stimuli.** Two speech stimulus series, both ranging perceptually from [aba] to [apa], were created by varying the closure interval of the medial stop. The two series differed only with respect to the presence or absence of glottal pulsing during the closure interval. Both stimulus sets were generated by digitally editing a token of [apa] produced by a male talker (R.L.D.). An 8-msec portion of the [p] burst was first eliminated to reduce somewhat the voiceless quality of the medial segment. For the *no-pulsing* stimuli, silent intervals varying in 10-msec steps from 20 msec to 120 msec were inserted between the preclosure and postclosure portions of the disyllable, the original medial closure segments having been deleted. The *pulsing* stimulus series was identical to the no-pulsing series, except that the corresponding closure intervals contained glottal pulsing extracted from the medial closure interval of an [aba] (spoken by the same male talker that produced the [apa]). Glottal pulsing started at the beginning of the closure interval and lasted 60 msec or until the end of the closure interval, whichever came first. (Thus, for example, the stimulus with a 120-msec closure interval contained pulsing over only the first half of that interval.) The upper limit on pulsing duration was chosen in light of Lisker's (1978a) finding that listeners are unable to hear medial stops as voiceless, even with very long closure intervals, when those intervals are completely filled with glottal pulsing.

The pre- and postclosure segments of the disyllables were 184 msec and 188 msec, respectively, and the peak amplitude of the preclosure segment was about 3 dB greater than that of the postclosure segment. Fundamental frequency remained approximately 96 Hz throughout both segments. The glottal buzz used in the closure interval of the pulsing stimuli had a nearly constant fundamental frequency of about 88 Hz and an amplitude approximately 12 dB less than the peak of the preclosure segment.

Two nonspeech stimulus sets, each mimicking the temporal and (to some extent) the amplitude properties of the speech pulsing and no-pulsing stimuli, were prepared by means of a waveform synthesis program on a PDP 11/34 computer. In the first set, every item consisted of two steady-state square-wave segments equal in duration to the pre- and postclosure segments, respectively, of the speech stimuli.[3] One nonspeech stimulus series in this set had silent intervals of varying durations separating the square-wave segments. These intervals corresponded exactly in duration to the silent closure intervals of the speech no-pulsing items. The other

nonspeech series in this set was identical to the first, except that the intervals separating the square-wave portions contained the same segments of glottal buzz (or buzz plus silence) that were used in the speech pulsing stimuli. The fundamental frequency of the square-wave segments remained constant at 256 Hz, and the steady-state amplitudes of the two segments matched the peak amplitudes of the pre- and postclosure speech segments, respectively. Each square-wave segment (both before and after the medial gap) had linear rise and decay times of 10 msec. No attempt was made to model in the square-wave stimuli the onset and offset amplitude contours of the corresponding speech stimuli.

The second set of nonspeech stimuli was identical to the first, except that fundamental frequency decreased linearly from 256 Hz to 175 Hz over the last 40 msec of the initial square-wave segment and increased linearly from 175 Hz to 256 Hz over the first 40 msec of the final square-wave segment. (The medial segments of silence or glottal buzz were unchanged.) The rationale for including this second set of nonspeech stimuli is that frequency-ramped square waves are abstractly more analogous to [VbV] or [VpV] stimuli than are steady-state square waves. For labial stops in medial position, formant frequencies generally fall near the onset of closure and rise after the release of closure (Fant, 1960). Thus, for both the speech and frequency-ramped nonspeech stimuli, there was a considerable degree of spectral continuity between the glottal pulsing segment and the low-frequency regions of the adjacent stimulus segments.

Sine-wave analogues schematically replicate actual formant patterns (thus accounting for their perceived phonetic character), whereas square-wave analogues replicate neither the harmonic nor the formant structure of speech and are very unlikely to be phonetically categorizable. Even the frequency-ramped square waves were judged by all three experimenters to be highly nonspeechlike, and, in particular, to bear little perceptual resemblance to any vowel-consonant-vowel (VCV) disyllables.[4]

Figures 1 and 2 show schematized spectrograms of tokens of the square-wave stimuli.

**Subjects and Procedure.** Fifty-two undergraduate students enrolled in an introductory psychology course participated in the study in partial fulfillment of course requirements. All were native English speakers and reported having no hearing defects.

Subjects were randomly assigned to one of the three stimulus conditions (one speech and two nonspeech conditions). Every listener identified a pulsing and a corresponding no-pulsing stimulus series, each presented in 20-min sessions separated by a 5-min break. Before identifying an entire series, subjects were presented with a random sequence of the two endpoint stimuli (40 trials each) and were required to learn (by means of feedback lights) which of two response buttons corresponded to each endpoint stimulus. The correspondence between the buttons and the stimuli was counterbalanced across subjects within each condition, but remained constant for each subject for both the pulsing and the no-pulsing stimulus series. The order of presentation of the pulsing and no-pulsing series was also counterbalanced across subjects within each condition.

After training with the series-endpoint stimuli, subjects identified 10 randomized blocks of the full 11-item series, with 2 sec between stimuli and 3 sec between blocks. Listeners were instructed to "press the button corresponding to the endpoint that each stimulus sounds more like."

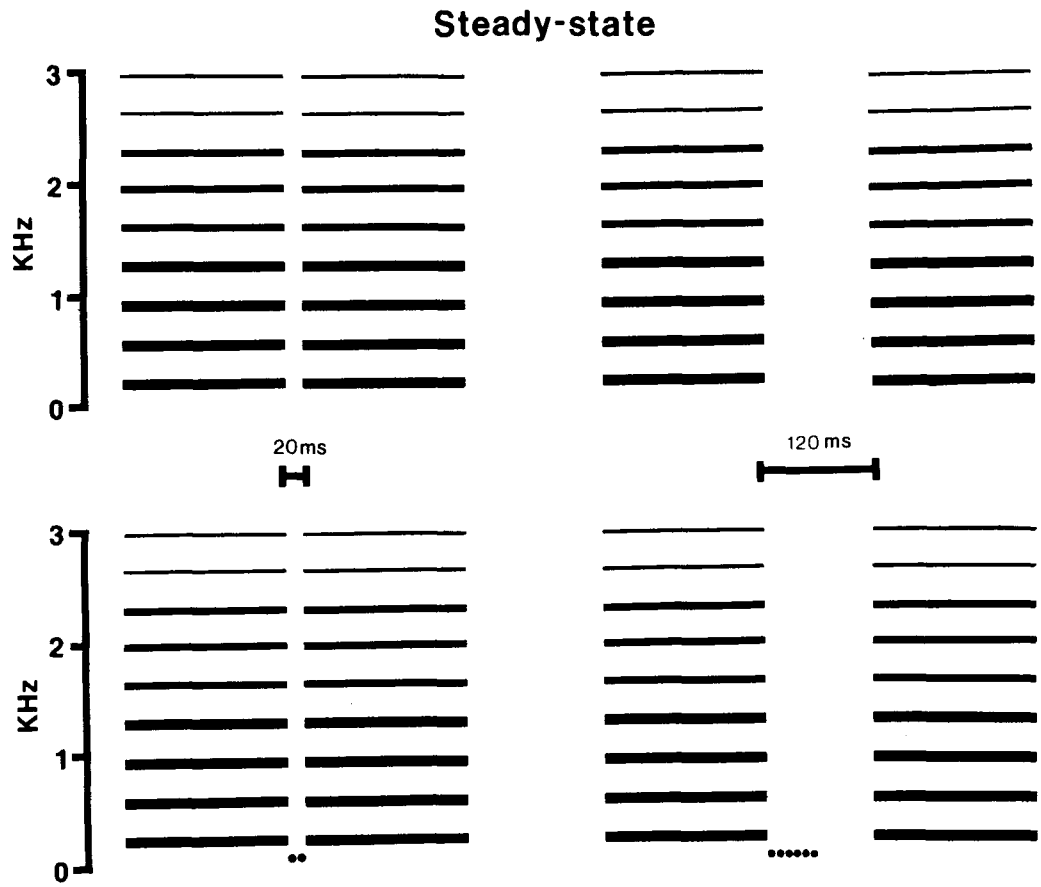


**Figure 1. Schematized spectrograms of the steady-state square-wave stimuli used in Experiment 1. Stimuli with medial pulsing are displayed at the bottom.**

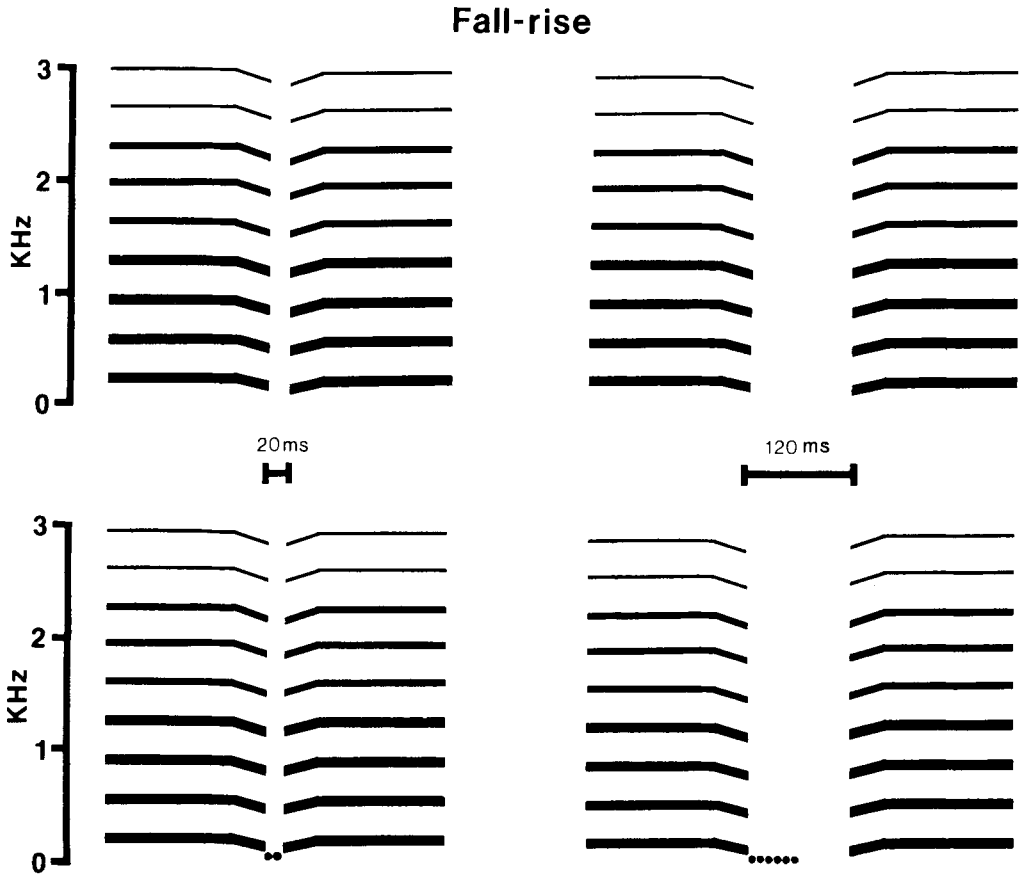


**Figure 2. Schematized spectrograms of the fall-rise square-wave stimuli used in Experiment 1. Stimuli with medial pulsing are displayed at the bottom.**

The stimuli, which had been stored on a computer disk, were digital-to-analog converted and lowpass filtered (4.9 kHz cutoff frequency), and were presented to subjects via TDH-49 earphones at a comfortable level (approximately 80 dB SPL). From 2 to 4 subjects, assigned to separate response stations in a soundproof chamber (Industrial Acoustics Corporation), participated in each experimental session.

## Results and Discussion

We included in the analysis only data from subjects who satisfied a predetermined criterion of 90% correct identification on both series-endpoint stimuli for the pulsing and no-pulsing conditions. Sessions were conducted until 10 subjects in each condition had met this criterion (there were 11 subjects in the speech condition, 19 in the steady-state square-wave condition, and 22 in the frequency-ramped square-wave condition).

Figures 3, 4, and 5 display the identification results for the speech and the two nonspeech conditions. Each data point represents the mean percentage of times the subjects labeled the test item as more like the endpoint stimulus with the 20-msec closure interval. We refer to these as *short-gap* responses. Probit analyses (Finney, 1971) were performed on each subject's identification functions to estimate the location of the boundary between the two response categories. All statistical comparisons were made with respect to these boundary values.

For the [aba]-[apa] stimuli (see Figure 3) the presence of glottal pulsing during the closure interval produced a substantial boundary shift in the direction of more short-gap responses [$t(9) = 5.53$, $p < .01$], replicating the findings of Lisker (1978a). Glottal pulsing failed to have a significant effect on the identification of the steady-state square-wave analogue stimuli [$t(9) = 1.54$] (see Figure 4). However, pulsing did produce a reliable increase in short-gap responses to the frequency-ramped (fall-rise) square-wave analogues [$t(9) = 2.44$, $p < .05$] (see Figure 5).

Although the boundary shift observed for the fall-rise square-wave stimuli was approximately one-third the size of that observed for the [aba]-[apa] items, the two effects were in the same direction. It is reasonable, therefore, to conclude that the trading relation between closure duration and closure pulsing is at least partially attributable to psychoacoustic factors that are not specific to speech. This conclusion is supported by subjects' descriptions of the two nonspeech stimulus sets. After the experimental session, each listener in the nonspeech conditions was asked to characterize the perceptual quality of the stimuli.

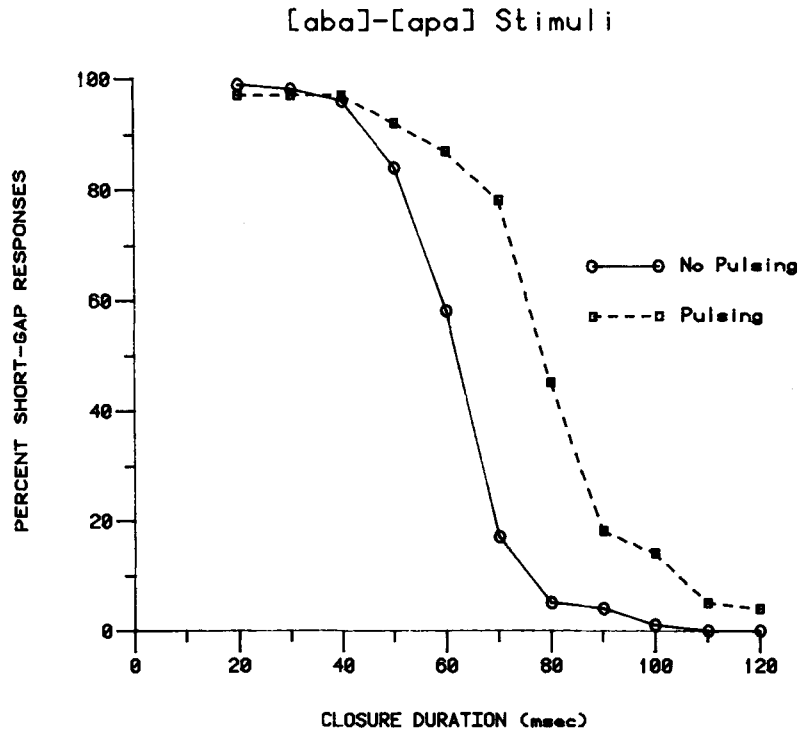


Figure 3. Mean percentage of short-gap responses to the [aba]-[apa] stimuli in Experiment 1.

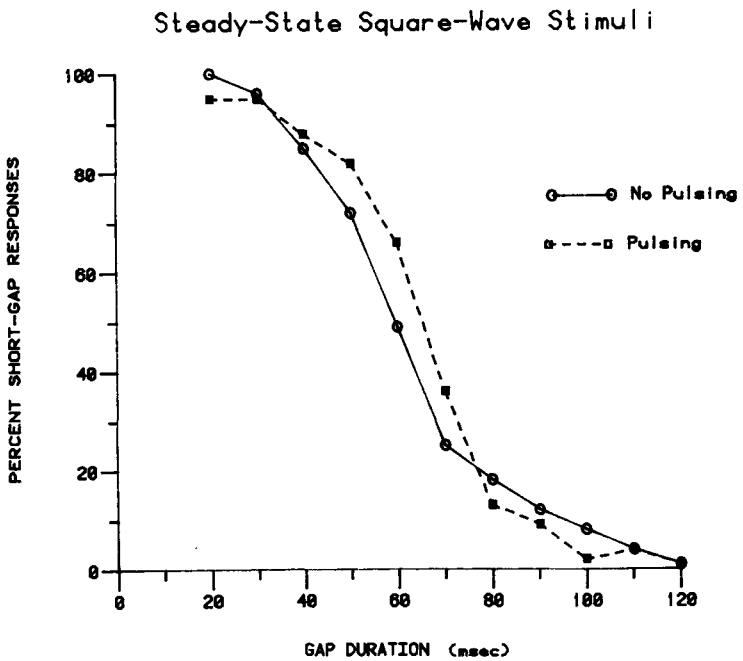


Figure 4. Mean percentage of short-gap responses to the steady-state square-wave stimuli in Experiment 1.

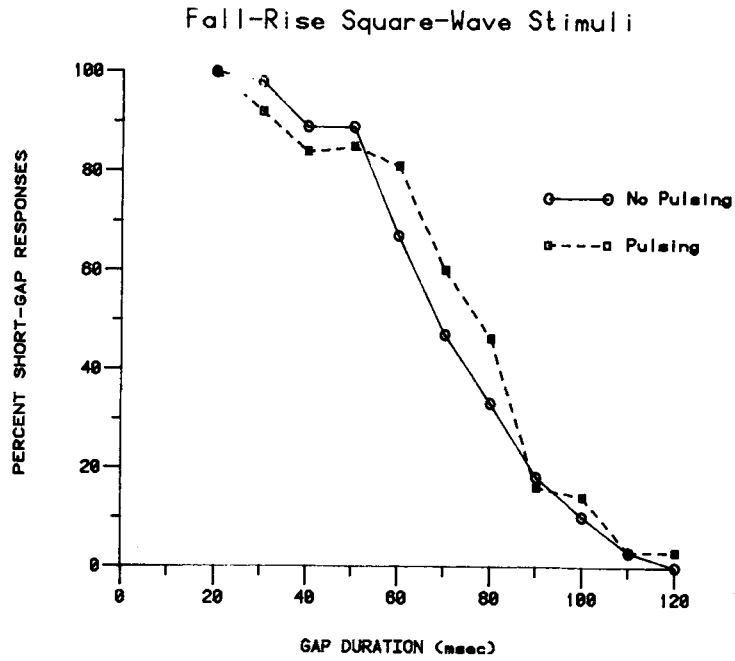


**Figure 5. Mean percentage of short-gap responses to the fall-rise square-wave stimuli in Experiment 1.**

Most of the descriptions were variations of "machine noises," "computer sounds," or "electronic music." Not a single subject in either nonspeech condition volunteered any type of phonetic or speechlike characterization. When specifically asked if the sounds could be heard as speechlike, the subjects uniformly said no, and many expressed surprise that the question was even posed. On the basis of these reports, it appears very unlikely that the nonspeech trading relation we obtained might derive from explicit or implicit phonetic categorization by the listeners.

What, then, is the basis of the boundary shift for the fall-rise square-wave stimuli? One possibility is simply that the presence of glottal pulsing during the medial gap may effectively shorten the perceived duration of the gap, biasing the listener toward short-gap responses. Such an effect would presumably be common to both speech and nonspeech conditions. However, in order to explain the absence of a significant trading relation for the steady-state square-wave stimuli, an additional assumption is required. Apparently, the glottal pulsing reduces the perceived duration of the gap only when it is spectrally continuous (or nearly so) with frequency components in adjacent portions of the stimulus (see, e.g., Steiger & Bregman, 1981).

The frequency components of the steady-state square wave were highly discontinuous with those of the glottal pulsing. In contrast, the fundamental frequency of the fall-rise square-wave stimulus dropped to 175 Hz (from 256 Hz) at the onset of glottal pulsing, where most of the energy was concentrated at the fundamental frequency (88 Hz). Thus, there was a greater degree of spectral continuity between higher energy components of the fall-rise square wave and the glottal pulsing. This was true as well of the [aba]-[apa] stimuli in which the first-formant frequency dropped near the onset of closure to a value approximating the frequency range of the pulsing.

An alternative to our spectral continuity hypothesis is that the trading relation between closure duration and closure pulsing depends merely on the presence of a frequency transition near the onset (or offset) of the closure interval. Both the [aba]-[apa] stimuli and the fall-rise square-wave stimuli contained such a transition, whereas the steady-state square-wave stimuli obviously did not. One purpose of Experiment 2 was to test this alternative hypothesis. Specifically, we designed a new nonspeech condition in which the square-wave stimuli were ramped *upward* in frequency in the vicinity of the medial gap, such that there was a large spectral discontinuity between the square wave and the closure pulsing.

Another potential problem of interpretation concerns the type of responses elicited from subjects in Experiment 1. Ordinarily, trading relations in speech perception are investigated by having listeners provide actual phonetic labels for the stimulus items. In order to make the speech and nonspeech conditions in Experiment 1 more comparable, we opted to have listeners label each speech stimulus not as [aba] or [apa] but as "more similar" to one series-endpoint stimulus or the other. We assumed that this procedure would yield identification functions much like those obtained with the conventional phonetic labeling procedure. It is possible, however, that our procedure may have distorted the underlying trading relation by biasing listeners against a response shift. Be-

cause the range and distributional properties of the stimuli were identical between the pulsing and no-pulsing conditions, listeners may have tended simply to divide both stimulus series at the same location, for example, at a point roughly equidistant from the series endpoints (see Parducci, 1965, 1974). This tendency would plainly reduce the size of any underlying trading relation. The conventional phonetic labeling procedure may be less susceptible to such biasing effects, if only because the judgments are less explicitly linked to properties of the stimulus range, such as the endpoints. We therefore decided to repeat the [aba]-[apa] condition in Experiment 2, this time allowing subjects to apply phonetic labels to the stimulus items.

## EXPERIMENT 2

### Method

**Stimuli.** The nonspeech stimulus sets used in Experiment 2 were identical to the frequency-ramped square-wave stimuli of Experiment 1, except that the direction of frequency change was reversed. Specifically, the fundamental frequency rose linearly from 256 Hz to 337 Hz over the last 40 msec of the initial square-wave segment and fell linearly from 337 Hz to 256 Hz over the first 40 msec of the final square-wave segment. (The remaining portions of the square wave had a fixed fundamental frequency of 256 Hz.) These rise-fall square-wave stimuli were designed to contain spectral changes comparable to those of the fall-rise stimuli of Experiment 1, but without spectral continuity between the square-wave segments and the glottal pulsing.

The [aba]-[apa] stimuli used in the present experiment were identical to those of Experiment 1.

**Subjects and Procedure.** The 35 subjects were undergraduate psychology students drawn from the same population as those who participated in Experiment 1. All were native English speakers and reported having no hearing defects.

Eighteen subjects served in the speech condition and 17 in the nonspeech condition. For the nonspeech condition, all procedures were identical to those used in Experiment 1. The [aba]-[apa] condition was exactly like that of Experiment 1, except that, following training, listeners were instructed to judge whether the consonant in each stimulus sounded more like [b] or [p] and to press the appropriate button.

### Results and Discussion

Again, we included in the analysis only data from subjects who satisfied a criterion of 90% correct identification on both series-endpoint stimuli for the pulsing and no-pulsing conditions. Sessions were conducted until 10 subjects from each condition met this criterion.

Figures 6 and 7 show the identification functions for the speech and nonspeech conditions of Experiment 2. Each data point represents the average percentage of [b] responses (Figure 6) or short-gap responses (Figure 7) to a given stimulus. For the [aba]-[apa] condition, the presence of glottal pulsing produced a sizable boundary shift toward more [b] responses [$t(9) = 4.15, p < .01$], an effect that was slightly smaller than that of the corresponding speech condition in Experiment 1. Apparently, the use of explicit phonetic labels is no more conducive to demonstrating the underlying trading relation than is the similarity-to-endpoint procedure used in Experiment 1. The parallel results of the two procedures sug-

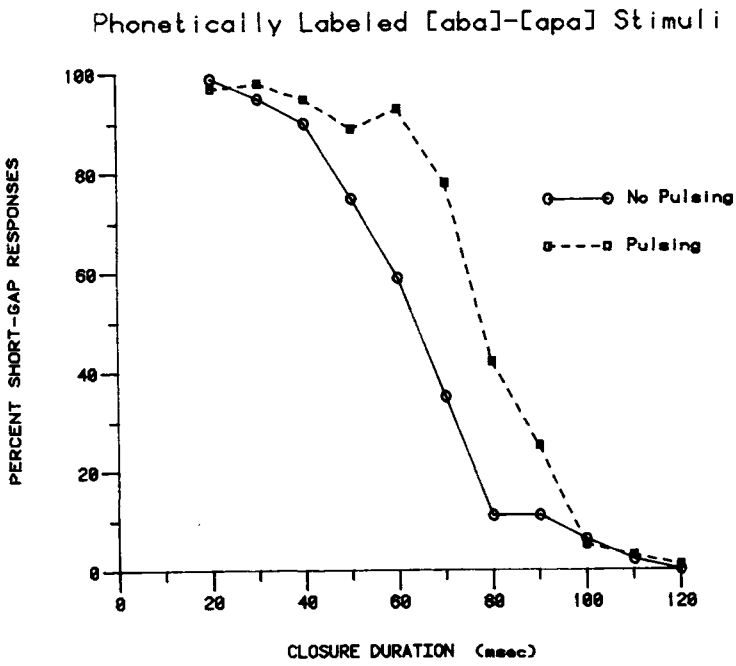


**Figure 6. Mean percentage of [b] responses to the [aba]-[apa] stimuli in Experiment 2.**

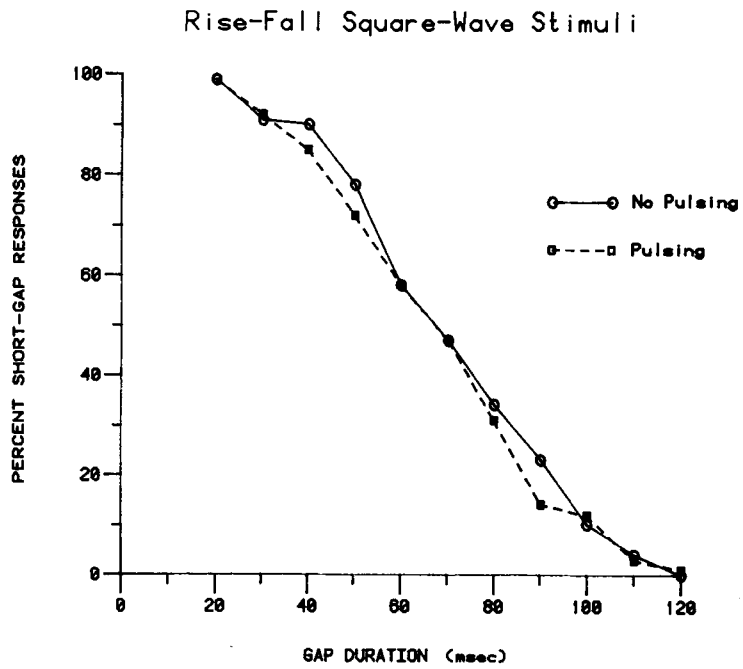


**Figure 7. Mean percentage of short-gap responses to the rise-fall square-wave stimuli in Experiment 2.**

gest that similarity-to-endpoint judgments provide a reasonably valid measure of category boundaries in both the speech and nonspeech conditions.

For the rise-fall square-wave condition, glottal pulsing had no significant effect on the boundary location [$t(9) = -0.95$]. There was, in fact, a slight response shift in the direction opposite to that obtained with the [aba]-[apa] stimuli and the fall-rise square-wave stimuli of Experiment 1. Frequency transitions per se are not sufficient to yield the normal trading relation between closure duration and closure pulsing. What appears to be required is some degree of spectral continuity between the glottal pulsing and the adjacent portions of the stimulus.

As in the first experiment, we asked the subjects to characterize the square-wave items after the experimental session. The responses were quite similar to those we had obtained earlier with respect to the other square-wave stimulus sets, and included no phonetic labels of any kind.

The boundary shifts from the five conditions of Experiments 1 and 2 were submitted to a single-factor analysis of variance, and Newman-Keuls post hoc comparisons were performed on all possible pairs of shifts. The size of the shift in the [aba]-[apa] condition of Experiment 1 was significantly larger than that of the rise-fall square-wave condition of Experiment 2 [$q(45) = 5.73, p < 01$]. The same was true for the speech condition of Experiment 2 [$q(45) = 5.57, p < .01$]. However, the boundary shift for the fall-rise square-wave condition of Experiment 1, although reliable, did not differ significantly from those of either the steady-state square-wave condition of Experiment 1 [$q(45) = 0.43$] or the rise-fall square-wave condition of Experiment 2 [$q(45) = 3.04$]. Finally, the shift in the fall-rise square-wave condition was not reliably different from the speech conditions of Experiment 1 [$q(45) = 2.68$] or Experiment 2 [$q(45) = 2.53$].

## GENERAL DISCUSSION

The results of the present experiments support two conclusions. First, the trading relation between closure duration and closure pulsing observed for speech sounds such as [aba] and [apa] can be at least partially duplicated with nonspeech analogue stimuli that are not phonetically categorizable. Second, a necessary condition for achieving this nonspeech effect is that there be some degree of spectral continuity between the glottal pulsing and the acoustic signal adjacent to the medial gap. We note that this condition almost invariably obtains in natural speech (Fant, 1960), and may well be a general constraint on this particular trading relation for both speech and nonspeech signals.

Various investigators (e.g., Best et al., 1981; Fitch, Halwes, Erickson, & Liberman, 1980; Repp, 1982) have argued that perceptual trading relations are most naturally explained in terms of the listener's tacit knowledge of underlying articulatory/phonatory acts. Thus, for example, Fitch et al. wrote:

> Nothing presently known about the auditory system provides a basis in principle for the many interactions that must be assumed if we are to account for the various trading relations among the speech cues.... An alternative view of the trading relations, which we find more appealing, differs from the auditory account in that it is not neutral with regard to

> the events by which the acoustic cues are produced. The advantage of this view is that it provides a principle that can be seen to underlie many different trading relations: Cues that are common but distributed products of the same linguistically significant act will tend to trade.... On that interpretation, the trading relations we have described would hold only for sounds that were being processed as speech, and they would, accordingly, be reflections of phonetic perception. (p. 344)

The present results indicate, however, that the mere presence of a perceptual trading relation among speech cues should not be construed as evidence for a speech mode of perception (any more than, say, categorical perception should be so construed). If an analogous result can be demonstrated with nonspeech stimuli, considerations of parsimony suggest that a general auditory explanation should be sought. Earlier we proposed an auditory account of the trading relation between closure duration and closure pulsing: The presence of pulsing may effectively shorten the perceived duration of a medial gap, especially when there is some degree of spectral continuity between the pulsing and the adjacent segments. According to this account, pulsing and closure duration are jointly regulated by talkers to signal the medial voicing distinction, in part because they have mutually reinforcing auditory effects. Their natural covariation is not simply a necessary consequence of the biomechanics and aerodynamics of speech production; it is, rather (at least to some extent), an exploitation of existing auditory constraints and interactions in the service of reliable speech communication. Our argument is fully compatible with the views of Miller et al. (1976) and Kuhl and Miller (1978), who suggested that boundaries between phonetic categories may be located so as to exploit auditory discontinuities along certain acoustic dimensions. Where possible, we would extend this auditory hypothesis to encompass natural variations in phonetic boundary placement due to the presence or absence of other acoustic properties (i.e., trading relations and context effects).

Several potential objections to the auditory hypothesis need to be addressed at this point. First, what do we make of the discrepancy between the magnitudes of the trading relations in the speech and nonspeech conditions? The boundary shift for the fall-rise square-wave stimuli was approximately one-third the size of the [aba]-[apa] shift (although in neither experiment was this difference reliable). One interpretation of the difference is that there is a speech-specific component of the trading relation, in addition to the general auditory component. Perhaps the speech trading relation is a conventionalization or exaggeration of an underlying auditory bias. Such conventionalizations appear to be a common feature of phonological systems (e.g., Hyman, 1975, 1984), although the underlying biases have typically been assumed to derive more from the physics and physiology of speech production than from auditory factors (see, e.g., Lisker, 1974).

On the basis of the present results, we certainly cannot rule out the possibility of a speech-specific component of the trading relation, but nor are we obligated to posit such a component. There are at least two alternative explanations of the difference in magnitude between the speech and nonspeech results. First, as we demonstrated in the three nonspeech conditions of the present two experiments, the trading relation between medial gap duration and gap pulsing depends on the degree of spectral continuity between the pulsing and the adjacent portions of the stimulus. Even for the fall-rise square-wave stimuli, there was a substantial discontinuity in fundamental frequency between the pulsing (88 Hz) and the adjacent square wave (175 Hz). In contrast, for the speech stimuli, the fundamental frequency just prior to the closure interval was approximately 96 Hz, yielding a greater degree of spectral continuity in the low-frequency region.

A second, and compatible, explanation of the difference between the speech and nonspeech results derives from considerations of the probable effects of practice and level of uncertainty. Watson and his colleagues (Spiegel & Watson, 1981; Watson & Kelly, 1981; Watson, Kelly, & Wroton, 1976; Watson, Wroton, Kelly, & Benbassat, 1975) investigated listeners' discrimination of complex word-length patterns consisting of ten 40-msec tonal components. Not surprisingly, performance in most conditions improved substantially with practice. Of more interest were the effects of uncertainty on discriminability. When the temporal location of the target tone was fixed across trials, and the frequencies and amplitudes of the remaining tones in the pattern were also held constant (minimal uncertainty condition), frequency discrimination was as great as when the target tones were presented in isolation. However, when there was trial-to-trial variation in target location and in properties of the context (high uncertainty condition), performance degraded substantially.

The results obtained by Watson and his colleagues bear importantly on the interpretation of experiments that compare performance on speech and nonspeech stimuli. Adult subjects obviously have vast experience in listening to speech sounds, and we can assume that they know where to listen for critical information when the stimulus variation involves a known phonetic distinction. By contrast, the square-wave patterns are quite novel stimuli, and although they incorporate the same type of physical variation as the speech stimuli, listeners may have difficulty in knowing what to attend to in order to detect that variation, especially early in the experimental session. (See Pisoni, 1977, for a discussion of the role of stimulus familiarity and attention in the perception of nonspeech analogues.) In other words, relative to nonspeech analogue stimuli, speech sounds present the listener with a low-uncertainty situation. It may, therefore, be unreasonable to expect exact convergence between speech and nonspeech results.

A second possible objection to our general auditory hypothesis is that it is based on too small a sample (i.e, one) of trading relations found in speech. The observed similarity between the speech and nonspeech results may be purely accidental, an interpretation apparently sup-

ported by recent purported failures to demonstrate trading relations with nonspeech analogue stimuli (Best et al., 1981; Summerfield, 1982).

We agree that evaluation of the auditory hypothesis requires consideration of a diverse set of trading relations, and we are now in the process of conducting this kind of extended investigation. So far we have successfully demonstrated two additional cases in which trading relations for nonspeech stimuli parallel those for speech sounds, and we have found no instances to the contrary. Kluender, Diehl, and Wright (1985) showed that, for sets of [aba] and [apa] stimuli, there is a reliable trading relation between initial syllable duration and medial closure duration, with respect to the voicing contrast. With longer syllable durations, longer closures are required to achieve a voiceless percept. Moreover, a trading relation of virtually the same magnitude was also observed for square-wave analogues of the speech stimuli, even though the nonspeech patterns were not phonetically categorizable by any of our subjects.

One plausible auditory account of this trading relation is that the medial gap duration is judged relative to the duration of the initial segment: longer initial segments make a given gap duration seem shorter by contrast, so the gap must be lengthened to achieve a comparable perceptual result. We offer this auditory hypothesis as one explanation of the nearly universal tendency among languages for vowels to be longer before voiced than before voiceless stops. Although various articulatory accounts of the vowel length distinction have been proposed, Lisker (1974) argued convincingly that each of these accounts has serious flaws. In view of his arguments and our own data, the alternative auditory hypothesis seems quite compelling.

Parker (1985) recently showed in five separate experiments that the ability of listeners to judge that two tones have asynchronous onsets depends on the frequency of the lower tone. As this frequency is reduced (while the upper-frequency component is held constant), a greater onset asynchrony is required in order for listeners to judge the tones as nonsimultaneous. This trading relation closely parallels that found between VOT and first-formant onset frequency in voicing identification (Lisker, 1975; Summerfield, 1982; Summerfield & Haggard, 1977).

Parker's (1985) results appear to be at odds with those of Summerfield (1982), who concluded that there was no systematic interaction in nonspeech between perceived onset asynchrony and the frequency of the lower component. In our view, Summerfield's results do not warrant this conclusion. In his first experiment, there was a significant ($p < .01$) monotonic trend, such that the just-noticeable onset asynchrony increased as the lower-tone frequency decreased, thus duplicating the direction, if not the magnitude, of the effect found by Parker. Only in Summerfield's second experiment, in which subjects were more practiced and several methodological changes were introduced, did the monotonic trend disappear.

Given the relative lack of experience listeners have in categorizing nonspeech stimuli, the absence of a systematic effect in Summerfield's second experiment may reflect either the failure of individual subjects to adopt a consistent strategy in responding to these stimuli or variability in the response strategies adopted by different subjects. In an effort to account for the discrepancy between Summerfield's results and her own, Parker investigated a variety of methodological factors that might account for the divergent outcomes. Although none of the procedural variables examined appeared to be responsible for the effects obtained by Parker, it was noted that the magnitude of the effects obtained varied substantially among subjects. If this was also the case in Summerfield's study, such variability may have obscured the effect, because of the small number of subjects used in that study. In any event, it appears that most listeners have an auditory bias that is consistent with the speech trading relation between VOT and first-formant onset frequency.

We turn now to the results of Best et al. (1981), which seem to pose a challenge to our general auditory account of speech trading relations. Recall that Best et al. obtained different perceptual outcomes for sine-wave analogues of [sei] and [stei], depending on whether or not the listeners heard them as speech. When the stimulus items were perceived phonetically, there was a reliable trading relation between silent gap duration and onset frequency of the lowest frequency sine wave (mirroring a similar trading relation for speech stimuli). However, among subjects who failed to hear the stimuli as speechlike, some showed no trading relation at all, and others showed a trading relation even larger than that of the phonetic perceivers. Best et al. concluded that psychoacoustic factors cannot explain this pattern of results and that the speech trading relation arises "specifically from perception of phonetic information" (p. 191).

The problem with positing a special speech mode to explain the performance of the phonetic perceivers is that, by parity of reasoning, one should also posit two distinct nonspeech modes to account for the divergent results of the subgroups that failed to perceive the sine-wave analogues as speech. To the extent that different modes are understood to represent distinct specialized mechanisms, such a proliferation of modes to explain the results of one experiment seems unparsimonious. A reasonable alternative is to assume that different subgroups of listeners, including the phonetic perceivers, were attending to somewhat different aspects of the complex signal. If, for example, the stimuli were perceived as speechlike, an appropriate attentional strategy would have been to focus on those properties of the signal that were most informative with respect to the phonological contrast in question. As we suggested earlier, knowing that a given set of stimuli are speech sounds involves knowing where to listen for critical information. If this was all that was intended by Best et al. (1981) when they claimed that the speech trading relation arises "specifically from the perception

of phonetic information" (p. 191), then we would not disagree. If, on the other hand, they were positing a specialized perceptual mode or mechanism for speech sounds, their case is less than convincing.

One additional point should be made about the results of Best et al. (1981). For half of the subjects who failed to perceive the sine-wave analogues as speechlike, there was a very large trading relation between silent gap duration and onset frequency of the lowest frequency sine wave. Best et al. appeared to imply that this trading relation had nothing to do with the somewhat smaller trading relation in the same direction obtained for the phonetic perceivers. As we argued above, differences in the magnitude of speech and nonspeech trading relations should not necessarily be taken to imply distinct perceptual mechanisms. In the present case, we think there may well be a common auditory basis for the observed trading relations, although the effects of these auditory factors may be mitigated or enhanced by various attentional strategies.

A final objection to our auditory hypothesis concerns the sheer number and diversity of cues (and hence trading relations) that can be demonstrated for any particular phonological contrast. For example, judgments about the voicing category of initial stops have been shown to depend on at least 9 different acoustic parameters (Diehl & Kluender, in press), and perceived voicing in medial stops is influenced by at least 15 parameters (Lisker, 1978b). One may argue that it is simply implausible to suggest that these cues, which are the normal acoustic concomitants of the phonetic categories in question, are all the products of an articulatory strategy to exploit general auditory constraints. Is it not safer to assume that many of these cues are just necessary physical byproducts of the way that phones are produced, rather than the intended result of an articulatory strategy? Liberman and Mattingly (1985), for example, call it "implausible" to suppose that "the articulators are always able to behave so as to produce just those sounds that conform to the manifold and complex requirements that the auditory interactions impose" (p. 19).

Our reply to this objection is to concede that it is indeed unlikely that every cue and every trading relation in speech is the specific intended result of an articulatory strategy to exploit general audition capabilities. In some cases, no doubt, a given acoustic property of a phone is simply an unavoidable consequence of the physics or physiology of speech production, but to the extent that it correlates reliably with the phonetic category, it will have informational value. According to Diehl and Kluender (in press), in order to recognize speech, the listener must learn a great many facts about the acoustic correlates of phonetic categories and the manner in which they are affected by variation in, for example, phonetic context, utterance rate, stress level, and talker characteristics. However, to acknowledge the important role of perceptual learning in speech perception is by no means to embrace the speech mode hypothesis. We assume that the principles of perceptual learning that apply to speech recognition are precisely those that apply to other stimulus categories, both auditory and nonauditory. Identification of speech sounds certainly involves the acquisition and use of domain-specific rules, but the same can obviously be said about the identification of music, traffic noises, faces, or wines. No one would seriously suggest that each of these domains requires its own specialized perceptual module. By the same token, the use of domain-specific knowledge in categorizing speech sounds is not a sufficient reason to posit a speech module.

A model of speech perception should include a specification of the general auditory transfer function and a compendium of all the speech-specific facts, tacitly known by the listener, that are relevant to phonetic classification. To the extent that a speech trading relation or context effect can be directly explained by properties of the auditory transfer function, it need not be assigned to the listener's store of tacit knowledge, and the overall model is thereby simplified. An appeal to speech-specific tacit knowledge should always be, in other words, an explanation of last resort.

Earlier we referred to Fodor's (1983) distinction between vertical, or modular, processing mechanisms that are largely innate and entirely domain specific, and horizontal mechanisms that operate across stimulus domains (see Liberman, 1982; Liberman & Mattingly, 1985; Mattingly & Liberman, 1985). The burden of the present paper has been to suggest that it is unnecessary to posit a vertical or modular device to account for speech recognition. What are needed, rather, are processes and strategies involved in general audition, attention, and perceptual learning.

## REFERENCES


ABRAMSON, A. S., & LISKER, L. (1970). Discriminability along the voicing continuum: Cross-language tests. *Proceedings of the 6th International Congress of Phonetic Sciences, Prague, 1967* (pp. 569-573). Prague: Academia.

BAILEY, P. J., SUMMERFIELD, Q., & DORMAN, M. (1977). *On the identification of sine-wave analogues of certain speech sounds* (Status Report on Speech Research, SR-51/52, 1-25). New Haven, CT: Haskins Laboratories. (ERIC Document Reproduction Service No. ED 147 892)

BEST, C. T., MORRONGIELLO, B., & ROBSON, R. (1981). Perceptual equivalence of acoustic cues in speech and nonspeech perception. *Perception & Psychophysics*, **29**, 191-211.

DIEHL, R. L., & KLUENDER, K. R. (in press). On the categorization of speech sounds. In S. Harnad (Ed.), *Categorical perception*. New York: Oxford University Press.

FANT, G. (1960). *Acoustic theory of speech production*. The Hague: Mouton.

FINNEY, D. J. (1971). *Probit analysis*. New York: Cambridge University Press.

FITCH, H. L., HALWES, T. G., ERICKSON, D. M., & LIBERMAN, A. M. (1980). Perceptual equivalence of two acoustic cues for stop consonant manner. *Perception & Psychophysics*, **27**, 343-350.

FODOR, J. A. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.

HAGGARD, M., AMBLER, S., & CALLOW, M. (1970). Pitch as a voicing cue. *Journal of the Acoustical Society of America*, **47**, 613-617.

HAGGARD, M. P., SUMMERFIELD, A. Q., & ROBERTS, M. (1981). Psychoacoustical and cultural determinants of phoneme boundaries: Evi-

dence from trading FO cues in the voiced-voiceless distinction. *Journal of Phonetics*, **9**, 49-62.

HILLENBRAND, J. (1984). Perception of sine-wave analogs of voice onset time stimuli. *Journal of the Acoustical Society of America*, **75**, 231-240.

HIRSH, I. J. (1959). Auditory perception of temporal order. *Journal of the Acoustical Society of America*, **31**, 757-767.

HIRSH, I. J., & SHERRICK, C. E. (1961). Perceived order in different sense modalities. *Journal of Experimental Psychology*, **62**, 423-432.

HOWELL, P., & ROSEN, S. (1984). Natural auditory sensitivities as universal determiners of phonemic contrasts. In B. Butterworth, B. Comrie, & O. Dahl (Eds.), *Explanations for linguistic universals* (pp. 205-235). The Hague: Mouton.

HYMAN, L. M. (1975). *Phonology: Theory and analysis*. New York: Holt, Rinehart & Winston.

HYMAN, L. M. (1984). Form and substance in language universals. In B. Butterworth, B. Comrie, & O. Dahl (Eds.), *Explanations for linguistic universals* (pp. 67-85). The Hague: Mouton.

JUSCZYK, P. W., PISONI, D. B., WALLEY, A., & MURRAY, J. (1980). Discrimination of relative onset time of two-component tones by infants. *Journal of the Acoustical Society of America*, **67**, 262-270.

KLUENDER, K. R., DIEHL, R. L., & WRIGHT, B. A. (1985). Perception of duration of medial silent intervals in speech and nonspeech signals. *Journal of the Acoustical Society of America*, **77**, S27 (Abstract).

Kuhl, P. K. (1981). Discrimination of speech by nonhuman animals: Basic auditory sensitivities conducive to the perception of speech-sound categories. *Journal of the Acoustical Society of America*, **70**, 340-349.

KUHL, P. K., & MILLER, J. D. (1975). Speech perception by the chinchilla: The voiced-voiceless distinction in alveolar plosive consonants. *Science*, **190**, 69-72.

KUHL, P. K., & MILLER, J. D. (1978). Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli. *Journal of the Acoustical Society of America*, **63**, 905-917.

KUHL, P. K., & PADDEN, D. M. (1982). Enhanced discriminability at the phonetic boundaries for the voicing feature in macaques. *Perception & Psychophysics*, **32**, 542-550.

LIBERMAN, A. M. (1970). Some characteristics of perception in the speech mode. In D. A. Hamburg et al. (Eds.), *Perception and its disorders* (ARNMD Research Publications Series, Vol. 48, pp. 238-254). New York: Raven Press.

LIBERMAN, A. M. (1982). On finding that speech is special. *American Psychologist*, **37**, 148-167.

LIBERMAN, A. M., COOPER, F. S., SHANKWEILER, D. P., & STUDDERT-KENNEDY, M. (1967). Perception of the speech code. *Psychological Review*, **74**, 431-461.

LIBERMAN, A. M., HARRIS, K. S., EIMAS, P. D., LISKER, L., & BASTIAN, J. (1961). An effect of learning on speech perception: The discrimination of durations of silence with and without phonemic significance. *Language & Speech*, **4**, 175-195.

LIBERMAN, A. M., HARRIS, K. S., HOFFMAN, H. S., & GRIFFITH, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, **54**, 358-368.

LIBERMAN, A. M., & MATTINGLY, I. G. (1985). The motor theory of speech perception revised. *Cognition*, **21**, 1-36.

LIBERMAN, A. M., & STUDDERT-KENNEDY, M. (1978). Phonetic perception. In R. Held, H. W. Leibowitz, & H. L. Teuber (Eds.), *Handbook of sensory physiology: Vol. 8. Perception* (pp. 143-178). New York: Springer-Verlag.

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

LISKER, L. (1974). On "explaining" vowel duration variation. *Glossa*, **8**, 233-246.

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

LISKER, L. (1978a). *On buzzing the English /b/* (Status Report on Speech Research, SR-55/56, 181-188). New Haven, CT: Haskins Laboratories. New Haven, CT: Haskins Laboratories. (ERIC Document Reproduction Service No. ED 166 757)

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

Lisker, L., & Abramson, A. S. (1970). The voicing dimension: Some experiments in comparative phonetics. In *Proceedings of the 6th International Congress of Phonetic Sciences, Prague, 1967* (pp. 563-567). Prague: Academia.

LISKER, L., ABRAMSON, A. S., COOPER, F. S., & SCHVEY, M. H. (1969). Transillumination of the larynx in running speech. *Journal of the Acoustical Society of America*, **45**, 1544-1546.

LISKER, L., LIBERMAN, A. M., ERICKSON, D. M., DECHOVITZ, D., & MANDLER, R. (1977). On pushing the voice-onset-time (VOT) boundary about. *Language & Speech*, **20**, 209-216.

MACMILLAN, N. A. (in press). Beyond the categorical/continuous distinction: A psychophysical approach to processing modes. In S. Harnad (Ed.), *Categorical perception*. New York: Oxford University Press.

MANN, V. A., & REPP, B. H. (1980). Influence of vocalic context on perception of the /ʃ/-/s/ distinction. *Perception & Psychophysics*, **28**, 213-228.

MATTINGLY, I. G., & LIBERMAN, A. M. (1985). Verticality unparalleled. *The Behavioral & Brain Sciences*, **8**, 24-26.

MATTINGLY, I. G., LIBERMAN, A. M., SYRDAL, A. K., & HALWES, T. (1971). Discrimination in speech and nonspeech modes. *Cognitive Psychology*, **2**, 131-157.

MILLER, J. D., WIER, C. C., PASTORE, R. E., KELLY, W. J., & DOOLING, R. J. (1976). Discrimination and labeling of noise-buzz sequences with varying noise-lead times: An example of categorical perception. *Journal of the Acoustical Society of America*, **60**, 410-417.

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

MIYAWAKI, K., STRANGE, W., VERBRUGGE, R., LIBERMAN, A. M., JENKINS, J. J., & FUJIMURA, O. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception & Psychophysics*, **18**, 331-340.

PARDUCCI, A. (1965). Category judgment: A range-frequency model. *Psychological Review*, **72**, 407-418.

PARDUCCI, A. (1974). Contextual effects: A range-frequency analysis. In E. C. Carterette & M. P. Friedman (Eds.), *Handbook of perception: Vol. 2. Psychophysical judgment and measurement* (pp. 127-141). New York: Academic Press.

PARKER, E. M., (1985). *Auditory constraints on phonetic categorization: Trading relations in speech and nonspeech.* Unpublished doctoral dissertation, University of Texas, Austin, TX.

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

PASTORE, R. E. (1981). Possible psychoacoustic factors in speech perception. In P. Eimas & J. Miller (Eds.), *Perspectives on the study of speech* (pp. 165-205). Hillsdale, NJ: Erlbaum.

PASTORE, R. E., AHROON, W. A., BAFFUTO, K. J., FRIEDMAN, C., PULEO, J. S., & FINK, E. A. (1977). Common-factor model of categorical perception. *Journal of Experimental Psychology: Human Perception & Performance*, **3**, 686-696.

PASTORE, R. E., AHROON, W. A., PULEO, J. S., CRIMMINS, D. B., GOLOWNER, L., & BERGER, R. S. (1976). Processing interaction between two dimensions of nonphonetic auditory signals. *Journal of Experimental Psychology: Human Perception & Performance*, **2**, 267-276.

PISONI, D. B. (1976). *Discrimination of brief frequency glissandoes* (Research on Speech Perception, Progress Report No. 3). Bloomington: Indiana University, Department of Psychology.

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

PISONI, D. B., CARRELL, T. D., & GANS, S. J. (1983). Perception of the duration of rapid spectrum changes in speech and nonspeech signals. *Perception & Psychophysics*, **34**, 314-322.

REMEZ, R. E., RUBIN, P. E., PISONI, D. B., & CARRELL, T. D. (1981).

Speech perception without traditional speech cues. *Science*, **212**, 947-950.

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

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

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

SPIEGEL, M. F., & WATSON, C. S. (1981). Factors in the discrimination of tonal patterns: III. Frequency discrimination with components of well-learned patterns. *Journal of the Acoustical Society of America*, **69**, 223-230.

STEIGER, H., & BREGMAN, A. S. (1981). Capturing frequency components of glided tones: Frequency separation, orientation, and alignment. *Perception & Psychophysics*, **30**, 425-435.

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

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

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

SUMMERFIELD, Q., & HAGGARD, M. (1977). On the dissociation of spectral and temporal cues to the voicing distinction in initial stop consonants. *Journal of the Acoustical Society of America*, **62**, 435-448.

WATSON, C. S., & KELLY, W. J. (1981). The role of stimulus uncertainty in the discrimination of auditory patterns. In D. J. Getty & J. N. Howard (Eds.), *Auditory and visual pattern recognition* (pp. 37-59). Hillsdale, NJ: Erlbaum.

WATSON, C. S., KELLY, W. J., & WROTON, H. W. (1976). Factors in the discrimination of tonal patterns: II. Selective attention and learning under various levels of stimulus uncertainty. *Journal of the Acoustical Society of America*, **60**, 1176-1186.

WATSON, C. S., WROTON, H. W., KELLY, W. J., & BENBASSAT, C. A. (1975). Factors in the discrimination of tonal patterns: I. Component frequency, temporal position, and silent intervals. *Journal of the Acoustical Society of America*, **57**, 1175-1185.

## NOTES

1. In all of these early comparisons between speech and nonspeech, only discrimination tests were administered in the nonspeech conditions, whereas in the speech conditions both discrimination and identification tests were given.

2. It should be pointed out that in the study by Pisoni et al. (1983), posttest questionnaires indicated that none of the subjects perceived the sine-wave analogues as speech or speechlike.

3. Owing to certain properties of the waveform synthesis program, the nonspeech stimuli contained some energy at the even-numbered harmonics, so they deviated slightly from true square waves.

4. We think it is likely that frequency components of a complex waveform will be interpretable by the listener as formant analogues only when (1) they have frequencies roughly similar to the formant-center frequencies of natural speech and (2) they do not form a harmonic series. Presumably, the components of a harmonic series are most simply interpreted as harmonics rather than formants. Although our square-wave stimuli may conceivably satisfy requirement 1, they fail to satisfy requirement 2.

(Manuscript received June 10, 1985;
revision accepted for publication February 17, 1986.)