508

Journal of Experimental Psychology:
Human Perception and Performance
1980, Vol. 6, No. 3, 536-563

Information in Speech: Observations on the Perception of [s]-Stop Clusters

Peter J. Bailey and Quentin Summerfield
Haskins Laboratories, New Haven, Connecticut

A series of experiments is reported that investigated the pattern of acoustic information specifying place and manner of stop consonants in medial position after [s]. In both production and perception, information for stop place includes the spectrum of the fricative at offset, the duration of the silent closure interval, the spectral relationship between the frequency of the stop release burst and the following periodically excited formants, and the spectral and temporal characteristics of the first formant transition. Similarly, the information for stop manner includes the duration of silent closure, the frequency of the first formant at the release, the magnitude of the first formant transition, and the proximity of the second and third formants at release. A relationship was shown to exist in perception between the spectral characteristics of the first formant and the duration of the silent closure required to hear a stop. This appears to reciprocate the covariation of these parameters in production across different places of articulation and different vocalic contexts. The existence of perceptual sensitivity to a wide range of the acoustic consequences of production questions the efficacy of accounts of speech perception in terms of the fractionation of the signal into elemental acoustic cues, which are then integrated to yield a phonetic percept. It is argued that it is inappropriate to ascribe a psychological status to cues whose only reality is their operational role as physical parameters whose manipulation can change the phonetic interpretation of a signal. It is suggested that the metric of the information for phonetic perception cannot be that of the cues; rather, a metric should be sought in which acoustic and articulatory dynamics are isomorphic.

Synthesis guided by analysis continues to be an appropriate means by which to devise stimuli for perceptual experiments designed to determine the nature of the information specifying phonetic identity. The strategy was exemplified by Delattre, Liberman, and Cooper (1955), who sought to determine the acoustic specification of the linguistic feature, place of articulation, in syllable-initial stop consonants. Their approach demonstrated that the results of experiments using schematic and constrained stimuli can lead to powerful conclusions. Delattre et al. rationalized the acoustical parameters shown to contribute to perceived place of articulation in terms of the characteristics of speech production. The relationship between acoustical and articulatory variables was corroborated in simulations of vocal tract behavior (Stevens & House, 1956).

These studies were concerned with stop consonants in absolute initial position, which

This research was carried out with the support from National Institute of Child Health and Human Development Grant HD-01994 and National Institutes of Health General Research Support Grant RR-05596 to the Haskins Laboratories, and the authors were funded by North Atlantic Treaty Organization postdoctoral research fellowships from the United Kingdom Science Research Council.

We should like to thank Alvin Liberman for his hospitality and encouragement, and Michael Studdert-Kennedy, Francis Ganong, Bruno Repp, and Mark Haggard for comments on earlier drafts of this article.

Peter J. Bailey is now at the Department of Psychology, University of York, Heslington, York Y01 5DD, England. Quentin Summerfield is now at the Medical Research Council Institute of Hearing Research, University of Nottingham, Nottingham, England.

Requests for reprints should be sent to Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06511.

constrained attention to the acoustical concomitants of consonant release. However, articulatory events unfold over time, and, as their name implies, *stop* consonants involve both a constriction phase, during which the articulators move toward the place of occlusion, and an occlusive phase, during which the vocal tract is literally stopped prior to the release phase realized in consonant–vowel (cv) syllables. Phonetic contexts in which all of these phases are realized acoustically are common in running speech and are useful as experimental stimuli because they provide an opportunity to study the perception of the temporally distributed acoustical consequences of the event of stop consonant production (e.g., Repp, Liberman, Eccardt, & Pesetsky, 1978). The experiments reported here examined a minimally complex example of such a context. We investigated the acoustical specifications of the linguistic features *place* and *manner* of articulation of stop consonants in syllables comprising the sequence [s]-[stop]-[vowel]. (In what follows [s]-[stop], for example, should be read as [s] followed by [stop].)

In a pilot experiment a serial resonance speech synthesizer was used to create a set of 10 steady-state vowels based on the formant frequency data for adult males published by Peterson and Barney (1952). Each vowel was preceded by a period of [s] friction and 100 msec of silence. Some of these syllables were identified as [s]-vowel, others, as [s]-stop-vowel. Overall, the probability of hearing a stop was inversely related to the frequency of the first formant, whereas the place of production of the stop appeared to be determined by the frequency of the second formant, approximately in accordance with the principle of formant loci (Delattre et al., 1955). In general, [s]-[u] was heard as [spu], [s]-[ɛ] as [stɛ], and [s]-[i] as [ski]. No stops were heard in syllables incorporating vowels with high first formants; [s]-[a] gave [sa], for example. Stop percepts were strongest with [spu] and [ski], whereas those with [stɛ] were weak. This contrasts with the results of Delattre et al., who noted that in two-formant CV syllables, the most compelling percept that could be obtained with a steady second formant was of an initial alveolar stop. However, in this pilot experiment, effects attributed to the second formant were confounded not only with the frequency of the first formant but also with variation in the relative levels of the formants across the set of vowels as a result of the serial connection of the synthesizer resonances. In the first two experiments, these variables were controlled by using parallel resonance synthesis to examine the effects of second-formant frequency on place of production of stops perceived in [s]-silence-vowel syllables.

Experiment 1

Method

Stimuli. A continuum of 12 two-formant steady-state vowels was prepared with the parallel resonance synthesizer at the Haskins Laboratories. Throughout the continuum the first-formant frequency was fixed at 260 Hz, and the second-formant frequency increased from 616 Hz to 2307 Hz in steps of approximately 154 Hz. The intensities of the first and second formants were the same; their nominal half-power bandwidths were 60 Hz and 90 Hz, respectively. The amplitude rise and fall times for the vowels were 75 msec and 50 msec, respectively. Each vowel was 250 msec in duration. The endpoints of the continuum were acceptable examples of the English vowels [u] and [i]. The midrange stimuli, on the other hand, were not English vowels but approximated the central vowels [ʉ] and [ɨ], as found, for example, in Swedish and Russian.

A steady-state [s] of 120 msec duration was created with an OVE IIIc serial resonance synthesizer. (In all the experiments reported here, the OVE synthesizer was used to create friction, as its fricative branch permits control over two poles and a zero, in contrast to the single pole available in the Haskins parallel resonance synthesizer.) The fricative formants were set at 3489 Hz and 4532 Hz, and the fricative antiresonance was set to eliminate energy below 2000 Hz. The amplitude rise time of the [s] was 40 msec, and its fall time was 10 msec. The 12 vowels and this [s] were low-pass filtered at 9.7 kHz and digitized with a sampling rate of 20 kHz. Two stimuli were prepared from each vowel by appending it to the [s] with either 10 msec or 100 msec of silence intervening. A randomization of 240 trials was recorded in which each of the 24 stimuli appeared 10 times. The intertrial interval was 3 sec, and the trials were arranged in blocks of 10, with an extra 3-sec pause between blocks.

Subjects. Fourteen undergraduate volunteers served as subjects. They declared themselves to have normal hearing in both ears and to have learned English as their first language in the United States. The two authors (who learned English in the United Kingdom) also took part in the experiment.

Procedure. Stimuli were presented binaurally through Telephonics TDH-39 headphones at a constant peak listening level of 75 dB (SPL). Subjects were instructed that they would hear a randomization of [s]-vowel and [s]-stop-vowel syllables. Their task was to write down the letter V if they heard [s]-vowel, and P, T or K if they heard either [sp]-vowel, [st]-vowel, or [sk]-vowel. They were also permitted the response Q to indicate a stop percept without an identifiable place of production, and the response O to indicate a consonant other than [p], [t], or [k]. Subjects listened first to a 24-trial practice sequence that included an exemplar of each stimulus and then to the 240-trial test sequence.

Results

The data from the 14 volunteer subjects and those from the two authors did not differ qualitatively and were pooled. Each of the 12 stimuli with 10 msec of silence was identified as [s]-vowel on at least 93% of its presentations, whereas no stimulus with 100 msec of silence was identified as [s]-vowel on more than 10% of its presentations. The comparison emphasizes that the concatenation of [s]-silence-vowel is not, per se, sufficient for the percept of a stop; a critical amount of silence is required. One rationalization of the result would be that the vocalic segments could themselves be heard as CV syllables and that a critical duration of silence is required for the initial consonant to evade masking by the [s]. This possibility was tested in a subsidiary experiment. A randomization was recorded that included six instances of each of the 12 vowels and six instances of 12 stimuli derived from these vowels by reducing their amplitude rise time from 75 msec to 5 msec. The randomization was presented to six experienced listeners who were instructed to identify the vowel and to indicate whether or not it was preceded by a stop. Out of 864 identifications there were only four reports of an initial stop. These occurred for stimuli with the more rapid rise in amplitude. We conclude that the vowels used in Experiment 1 did not, of themselves, predispose the percept of a stop.

The results from the 12 stimuli with 100 msec of silence are displayed in Figure 1, in which the percentages of V, P, T, and K responses are plotted against the stimulus number, and hence, the frequency of the second formant in the vowel. For the sake of clarity, responses in the Q and O categories have been combined. There was a slight tendency for O responses to predominate at low second-formant frequencies and for Q responses to predominate at high second-formant frequencies. Variation in the frequency of the second formant produced systematic effects on the place of production of the perceived stop. The probability of [p] percepts decreased and the probability of [k] percepts increased as the second-formant frequency was raised. The fact that [p] predominated when the second formant was below 1695 Hz is consistent with the perception of [p] in the sequence [s]-silence-[lit] (Bastian, Note 1), in which for the [l] the second formant typically onsets below 1000 Hz (Fant, 1960).

As in the pilot experiment, the vowels [u] and [i] gave reasonably strong percepts of [p] and [k], respectively. The [t] category was almost absent here, suggesting that its weakness in the pilot experiment did not result either from an insufficiently low frequency of the first formant for a stop to be heard, or from a relatively low second-formant intensity in midcentral vowels. As we have already noted, the locus principle would predict that [t] would be used as a response to Stimuli 8, 9, and 10, whose second formants span the 1800 Hz alveolar locus. In fact, [t] was made as a response to these stimuli on only 3% of their presentations, whereas the largest proportion of [t] responses made to any stimulus (10%) was made to Stimulus 5, in which the second formant was set to 1232 Hz. The principle of formant loci appears to be insufficient to explain the perception of place of articulation in these stimuli for which the stops were not in absolute initial position.

One reason for the presence of [t] percepts in the pilot experiment and their absence here could be that the vocalic portions of the pilot stimuli incorporated five formants and thus included energy above 2.5 kHz, whereas those in Experiment 1 incorporating only two formants had no energy above 2.5 kHz. This would be consistent with the proven perceptual efficacy of energy around 3 kHz for specifying the release of an alveolar stop (Fant, 1960). Accordingly, Experiment 1 was repeated using stimuli whose vocalic portions included a fixed third formant.

Experiment 2

Method

Procedure. The stimuli in Experiment 2 were derived from those in Experiment 1 by adding a fixed third formant at 3026 Hz to the vocalic portion of each stimulus. The level of the third formant was −15 dB with respect to the first and second formants, and its nominal half-power bandwidth was 120 Hz. Identification tapes were prepared as before and administered to eight volunteer subjects and to the two authors. The instructions and response alternatives were the same as in Experiment 1.

Results

The stimuli with 10 msec of silence were not identified as [s]-vowel as consistently

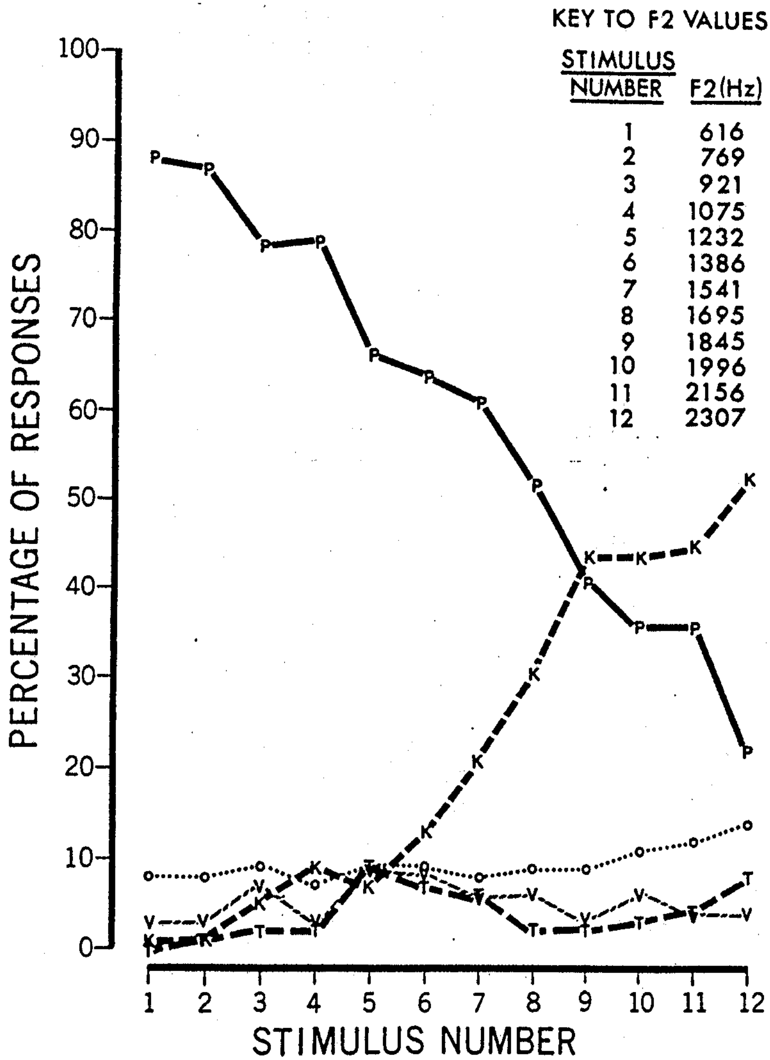

Figure 1. Identification functions from Experiment 1 for stimuli incorporating 100 msec of silence showing patterns of [sp]-vowel (P), [st]-vowel (T), [sk]-vowel (K), [s]-vowel (V), and other (O) responses.

as their counterparts in Experiment 1. On the average they were identified in this way on only 86% of their presentations, due largely to a tendency for stimuli with high second-formant frequency to be identified as [sk]-vowel. This result presages that obtained with the stimuli incorporating 100 msec of silence, whose identification functions are shown in Figure 2. Data from 10 listeners, including the two authors, are displayed. The only systematic difference between the results in Experiments 1 and 2 was an increase in the proportion of [k] responses to stimuli with high second formants in the second experiment. The incorporation of a fixed third formant in the vocalic portions of these stimuli strengthened the velar category, but although the overall proportion of [t] responses increased from 3% to 6%, no clearly defined [t] category emerged.

Discussion

On the basis of perceptual experiments with two-formant patterns, Delattre et al. (1955) deduced fixed second-formant loci of 720 Hz for bilabial stops, 1800 Hz for alveolars, and 3000 Hz for velars before front vowels; a locus at a lower and more variable frequency was found for velars before back vowels. Stevens and House (1956) largely confirmed these results using an electrical analogue of a vocal tract but

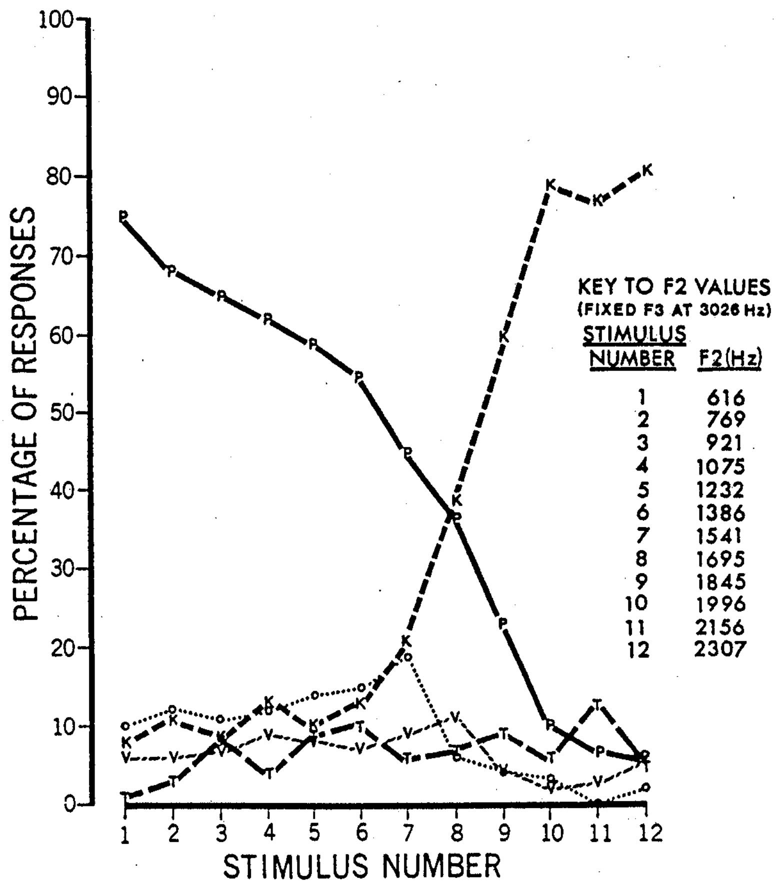

Figure 2. Identification functions from Experiment 2 for stimuli incorporating 100 msec of silence showing patterns of [sp]-vowel (P), [st]-vowel (T), [sk]-vowel (K), [s]-vowel (V), and other (O) responses.

showed that neither the bilabial nor the velar locus is completely fixed. Both loci vary as a function of the extent to which the arrangement of the articulators during closure is permitted to anticipate their organization for the following vowel. With maximum articulatory preconfiguration, the locus for [b] can range from 700 Hz to 1500 Hz, and that for [g] can range from 600 Hz to 2500 Hz. We have already noted the apparent anomaly between our failure to find alveolar percepts in Experiments 1 and 2 and the observation of Delattre et al. (1955) that "of the three stops that are produced when the straight second formants are at the loci, the [d] (at 1800 Hz) is the most compelling" (p. 771). In their stimuli, the information for stop manner was embodied in a rising first-formant transition. We have to explain why alveolar percepts are absent when information for stop manner is embodied in a period of silence. The Stevens and House (1956) simulation demonstrated that the loci are not arbitrary perceptual constructs; they are a direct consequence of stop consonant articulation. The ranges of formant loci computed by Stevens and House could provide a basis for rationalization of the distributions of [p] and [k] responses in Experiments 1 and 2. However, a reliance on formant loci as an explanatory principle would still require an ad hoc account of the absence of the [t] category.

The optimal strategy for explaining the perception of an articulatory event is to consider not just one but all of the acoustic consequences of that event. The present data should be rationalized by determining which natural production is represented most faithfully by the, albeit schematic, stimuli used in Experiments 1 and 2. These stimuli may be characterized as follows: First, there are no formant transitions in the vocalic portions; second, there are no release bursts; third, as a consequence, the spectral energy at and following release is contiguous and continuous. The property of spectral contiguity was identified by Fant (1973) as a general characteristic of initial velar stops: "(for) [k][g] spectral energy is concentrated, strong, and continuously connected, without rapid initial transitions to the formant carrying the main pitch of the vowel" (p. 135). As Fant's spectrograms show, the absence of initial second- and third-formant transitions in velars is strictly characteristic only of front vowels when the second-formant frequency is high. Appreciable second-formant transitions occur before back vowels when the second formant is at a lower frequency. For bilabial stops, Fant (1973) noted that "spectral energy is weak, more spread than in [k][g], and with an emphasis on a lower frequency than [t][d]. Initial transitions are rapid and rising" (p. 135). Thus, in contrast to velar stops, transitions characterizing initial bilabials are minimal before vowels with low second formants, and increase in magnitude before vowels with higher second formants.

Taking these observations in conjunction with the ranges of second-formant loci reported by Stevens and House (1956), we can account for the pattern of results in Experiments 1 and 2. Bilabial percepts would be expected (in the absence of a prominent burst) when the second formant onsets at a low frequency with no formant transition. Velar percepts, on the other hand, would be expected when the second formant onsets at a high frequency without any formant transition. In the stimuli of Experiments 1 and 2, there is inevitable spectral contiguity between energy at the release and that in the following vowel, although there is no release burst as such. The larger number of [k] responses in Experiment 2 following the introduction of a fixed third formant probably resulted both from the presence of higher frequencies in the vocalic onset and from the proximity of the second and third formants when the second formant frequency was high, given that proximity of the second and third formants at onset is a characteristic of the production of velar stops (Fant, 1960; Stevens, 1975).

The absence of alveolar percepts is consistent with Fant's (1973) observation that for [t] and [d] "spectral energy is spread, generally strong, with emphasis on higher frequencies than in [p] and [b] and extending higher than the main [k][g] formant'' (p. 135). As this quotation implies and as Fant's spectrograms show, spectral discontinuity between the release and the periodically excited formants characterizes alveolar stops before all vowels. This discontinuity accompanies, and hence could specify to a perceiver, the sudden increase in the length of the front cavity as the major constriction switches from the alveolar region to a more dorsal position characterizing the vowel. Possibly, the lack of any spectral discontinuity in our stimuli at release is a major contributor to the absence of alveolar percepts.

These observations appear to provide a coherent account of the results of Experiments 1 and 2. The differences between our results and those of Delattre et al. (1955), in which manner information was carried by formant transitions, can be understood in terms of the reciprocal relationship between the importance of bursts and transitions in the perception of stop place (Cooper, Delattre, Liberman, Borst, & Gerstman, 1952). Dorman, Studdert-Kennedy, and Raphael (1977) argued for the functional equivalence of bursts and transitions, demonstrating that whereas the former were perceptually influential, the latter were not, and vice versa.

The major claims of the foregoing account were tested in Experiment 3 by using stimuli that included a spectrally compact, prevocalic release burst. We expected to find that the place of articulation of stops perceived in such stimuli would be determined both by the frequency of the burst and by the relationship between burst frequency and the frequency of the second formant at onset. Specifically, it was predicted that [p] percepts would be obtained when the burst was at a lower frequency than the second formant but might be infrequent because a spectrally compact burst is not typical of the diffuse release spectrum of bilabials; [k] percepts were predicted when the burst frequency and the second-formant onset frequency were contiguous; [t] percepts were predicted when the burst frequency was higher than and not contiguous with the onset frequency of the second formant.

Experiment 3

Method

Stimuli. Twenty-five stimuli were constructed as before by combining a fricative segment synthesized using a serial resonance synthesizer, with a vocalic segment synthesized using a parallel resonance synthesizer. The fricative segments consisted of an [s] of 120 msec duration followed by 100 msec of silence and, in this experiment, a 25-msec release burst. The [s] was spectrally identical to the [s] segments used in Experiments 1 and 2. Five bursts were created by setting the lower fricative pole to 1509 Hz, 1957 Hz, 2466 Hz, 3019 Hz, and 3489 Hz, respectively; in each case the higher fricative pole was set to 6 kHz and was canceled with the antiformant. Bursts synthesized in this way increase in intensity as their frequency rises. Five two-formant steady-state vowels were synthesized. Each was 250 msec in duration, with the first formant set to 260 Hz. The second formant was set to 1386 Hz, 1772 Hz, 2156 Hz, 2540 Hz, and 2910 Hz, respectively in the five patterns. The levels of the two formants were the same, and their amplitude rise time was 20 msec. The five fricative segments and the five vowels were low-pass filtered at 4.9 kHz and digitized with a sampling rate of 10 kHz. Twenty-five test stimuli were constructed by preceding each of the vocalic segments with each of the fricative segments. Thus, each stimulus consisted of the sequence [s] (120 msec)–silence (100 msec)–burst (25 msec)–vowel (250 msec). A 25-trial practice sequence consisting of a single randomization of these 25 stimuli, and a 250-trial test sequence consisting of 10 concatenated randomizations were recorded. The intertrial interval was 3 sec.

Subjects and procedure. The two authors and six experienced listeners who were unaware of the structure of the stimuli listened first to the practice sequence and then to the test sequence. Stimuli were presented binaurally through Telephonics TDH-39 headphones at a constant peak listening level of 75 dB (SPL). The eight listeners were required to identify the stop heard in each syllable as either [p], [t], or [k] but to indicate with a question mark any response of which they were not confident.

Results and Discussion

Of the 2,000 responses only 10 indicated ambiguous percepts, and these will not be distinguished from other responses. The data of the two authors did not differ systematically from those of the other six listeners. Figures 3a–3e display the data pooled over all eight subjects. Each panel shows the percentage of [p], [t], and [k] responses made to the five stimuli with the same burst frequency. In each case the abscissa plots the frequency of the second formant in the vocalic portion of the stimuli, and the burst frequency is indicated by a filled triangle. As predicted, there were few [p] responses. They appeared when the burst frequency was low and below the onset of the second formant, for example, in Figure 3a when F2 was at 1772 Hz, in Figure 3b when F2 was at 2540 Hz, and in Figure 3c when F2 was at 2910 Hz. In general, a [k] percept was most likely when the burst frequency was close to the onset frequency of the second formant, although an exception to this generalization is found in Figure 3a. Here, with the burst at its lowest frequency and hence its lowest intensity, a pattern of results more akin to that of Experiment 1 (in which there were no bursts) would be expected. However, a comparison of Figure 3a with Figure 1 shows that the presence of a burst at this frequency has had the effect of increasing the probability of [k] over [p] percepts. The proportion of [t] responses increased with both the burst frequency and the size of the frequency difference between the burst and the second formant onset. In each of Figures 3b, 3c, and 3d, [t] percepts predominated when the second formant was low, and [k] percepts predominated when the second formant was high. The crossover between alveolar and velar responses occurred at higher second-formant onset frequencies as the frequency of the burst increased.

The results of Experiment 3 are consistent with earlier results (Liberman, Delattre, & Cooper, 1952) and with our rationalization of the data from Experiments 1 and 2. A [t] is perceived medially after [s] and before a vowel in the absence of periodically excited formant transitions if a burst initiates the vowel with a center frequency at least 400 Hz above the main formant in the vowel. The appropriate complement to this result would be the demonstration that [st] percepts do occur in the absence of a burst, provided the vocalic portion of the stimulus incorporates appreciable formant transitions. This was an ancillary finding of Experiments 6 and 7.

The results of the first three experiments confirmed that a complete account of the perception of stop consonants in medial position after [s] depends on an understanding of the acoustic consequences of the underlying articulatory event as a whole. So far, only the release phase of the event has been considered. Before examining the perceptual concomitants of the constriction and occlusion phases, we sought to determine how the acoustic properties of natural productions of [s]-stop-vowel sequences vary according to the place of production of the medial stop.

Production Data

Method

Procedure. The two authors read a randomization of [s]-stop-vowel syllables, in which the stop was one of [p], [t], or [k], and the vowel was one of [i], [ε], [a], or [u]. The syllables were uttered in the phrase, "Now hear [s]-stop-vowel please," at a natural rate cued by a visual metronome. Five tokens of each stop-vowel combination were recorded. The recordings were low-pass filtered at 4.9 kHz and digitized at a sampling rate of 10 kHz. Both the sampled waveform and a hardware spectrum analysis of successive 12.8-msec segments of the signal were displayed. The computer system allows a cursor to be aligned to measure the duration of any segment of the waveform to an accuracy of .2 msec. The spectral section of the 12.8-msec segment containing the cursor is also displayed. Two spectral measures were made in each token; they were the frequencies of the lowest peak in the spectral sections containing (a) the offset of the fricative portion and (b) the release of the stop. (The hardware spectrum analysis is relatively coarse grained and permits only approximate estimates of the frequencies of the spectral peaks corresponding to any particular portion of the waveform.) Three intervals were measured; these were (a) the duration of the fricative portion, (b) the period of silent closure, and (c) the period of aspiration following release prior to the first voicing pulse.

Results

The spectral measurements are summarized in Table 1 in which the average frequencies of the lowest spectral peak at closure and release are tabulated for each syllable. With the exception of the bilabial stops, whose burst frequencies could not be estimated reliably with our measurement procedure, the release burst spectra for medial stops following [s] are in reasonable agreement with measures from stops in initial position (e.g., Fant, 1973), as required by the rationalization of the results of the

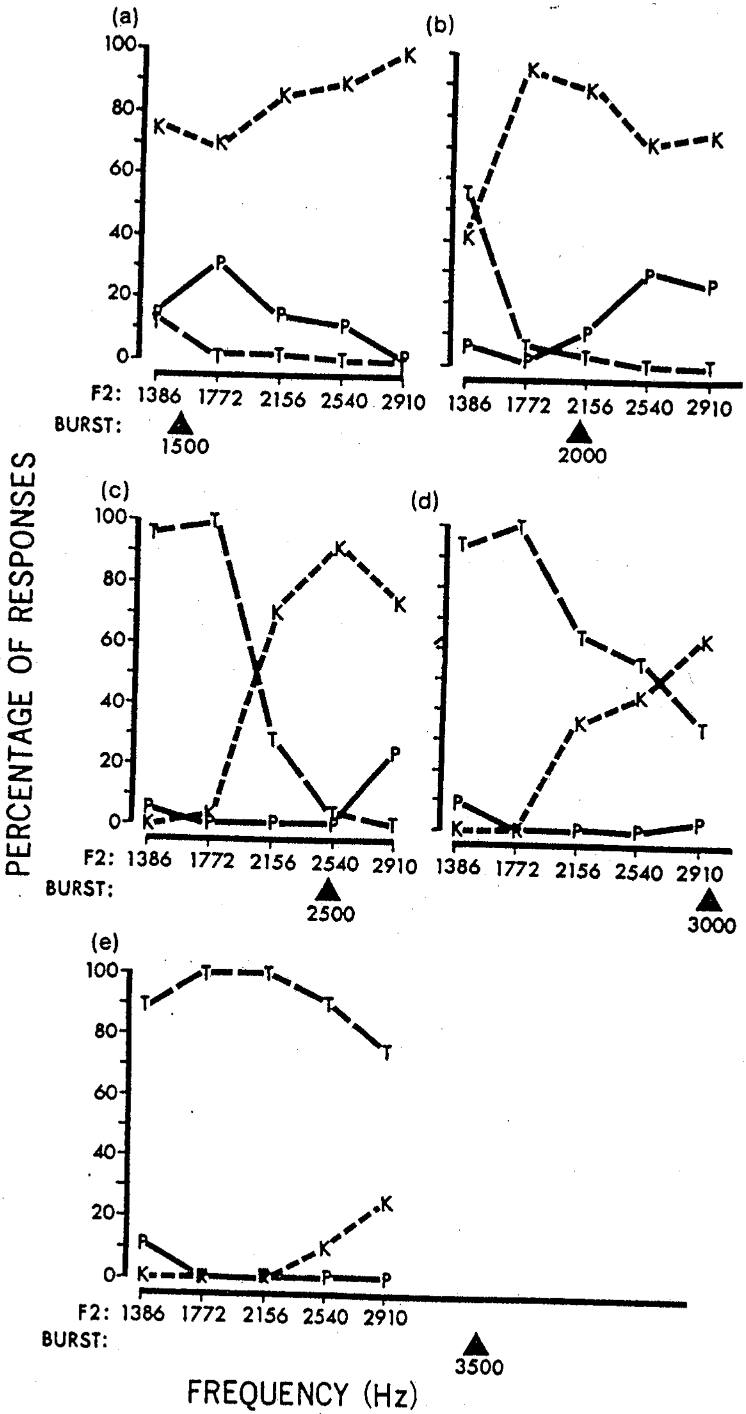

Figure 3. Identification functions from Experiment 3. (Each panel corresponds to stimuli with the burst whose frequency is indicated by a triangle. The percentages of [sp]-vowel (P), [st]-vowel (T), and [sk]-vowel (K) responses are plotted against the frequency of the second formant in the vocalic portion of the stimulus.)

Table 1
Spectral Measurements of the Lowest Frequency Peak of the Offset of the [s] and the Release Burst of the Stop in [s]-[Stop]-[Vowel] Syllables Spoken by Two Speakers of British English

Syllable	Speaker 1: PJB [s] Offset	Speaker 1: PJB Release burst	Speaker 2: AQS [s] Offset	Speaker 2: AQS Release burst
[spi]	3.1	—	3.6	—
[spa]	3.4	—	3.7	—
[spu]	3.1	—	3.8	—
[spɜ]	3.1	—	3.7	—
M	3.2		3.7	
[sti]	3.7	3.1	4.5	3.6
[sta]	3.9	3.4	4.5	3.6
[stu]	3.5	2.7	4.5	3.0
[stɜ]	3.7	3.1	4.6	3.3
M	3.7		4.4	
[ski]	3.8	3.0	4.2	2.9
[ska]	3.8	1.8	3.9	1.4
[sku]	3.7	1.6	3.9	1.4
[skɜ]	3.4	1.8	4.0	1.5
M	3.7		4.0	

Note. — means spectral analysis too coarse to measure burst spectrum. Data are given in kHz.

first three experiments. Fricative offset spectra also show a consistent pattern: Spectral peaks in the fricative offset are at lower frequencies in syllables with bilabial stops than with either alveolars or velars. This reflects the lengthening of the cavity in front of the fricative source as bilabial closure is made and can sometimes be seen in spectrographic displays as a rapid downward transition at the end of the fricative. Given that different spectral changes are concomitants of stops articulated at different places, we should expect to find perceptual sensitivity to the spectral characteristics of the fricative offset in judgments of stop place in [s]-stop clusters. This was investigated later in Experiment 4.

The duration measurements are tabulated in Table 2. The means of the standard deviations of the durations of friction, silence, and aspiration were 17.8 msec, 15.0 msec, and 4.2 msec for Speaker 1, and 10.8 msec, 8.1 msec, and 3.2 msec for Speaker 2. Insofar as they are comparable, these figures are in agreement with the variability in similar duration measurements reported elsewhere (e.g., Klatt, 1974, 1975).

There are three noteworthy aspects of these data. First, both speakers produced longer silent intervals during bilabial closure compared to alveolar or velar closure (cf. Fischer-Jørgensen, 1964). Thus, closure duration appears to be another characteristic of place of production, and Experiments 5 and 6 were designed to measure perceptual sensitivity to this variable. Second, both speakers produced shorter silent intervals before the vowel [a] compared with the other three vowels. To the extent that this difference reflects an inverse correlation in production between the openness of the vowel and the duration of the silent closure, we should expect to find a trading relationship between the magnitude of the first formant transition and the amount of silence required for the perception of a stop in an [s]-stop cluster. This was examined in Experiment 7. The third aspect of the data in Table 2 is parenthetic to our main interest here. It is that the total period of devoicing, that is, the sum of the durations of friction, silence, and aspiration, is less variable across stop place and vowels than is any one of its component durations (cf. Kozhevnikov & Chistovich, 1965). We have noted a similar tendency for total durations of devoicing to be relatively invariant across place in productions of [bəCV] and [bəCrV],

Table 2
Measurements of the Durations of [s] Friction, Silence, and Aspiration Segments of Productions of [s]-Stop-Vowel by Two Speakers of British English

	Speaker 1: PJB				Speaker 2: AQS			
Syllable	Friction	Silence	Aspiration	Total	Friction	Silence	Aspiration	Total
[spi]	174.9	130.7	14.8	320.4	166.5	100.4	12.2	279.1
[spa]	201.1	108.8	15.5	325.4	156.6	96.3	12.9	265.7
[spu]	185.9	142.7	15.6	344.2	160.6	101.1	13.0	274.7
[spɜ]	185.7	127.3	15.1	328.1	166.3	100.3	10.1	276.8
M	186.9	127.4	15.3	329.6	162.5	99.5	12.1	274.0
[sti]	212.1	93.1	33.0	338.2	181.4	67.0	21.2	269.6
[sta]	214.1	66.8	21.1	302.0	179.8	60.6	17.5	257.9
[stu]	222.3	100.2	31.0	353.5	171.3	75.9	20.2	267.4
[stɜ]	218.7	83.0	26.0	327.7	182.8	75.1	19.8	277.8
M	216.8	85.8	27.8	330.4	178.8	69.7	19.7	268.2
[ski]	203.6	86.9	50.9	341.4	186.6	82.9	25.6	295.1
[ska]	212.7	65.2	35.9	317.4	180.7	63.9	23.4	268.0
[sku]	216.0	82.2	44.5	342.7	181.7	87.0	26.6	295.3
[skɜ]	220.5	87.1	44.4	352.0	195.1	77.4	21.4	293.9
M	213.2	80.3	44.4	337.9	186.0	77.8	24.5	288.1
Overall								
M	205.6	97.8	29.0	332.8	175.8	82.3	18.7	276.8
SD	15.7	24.9	13.0	15.3	11.4	14.7	5.5	12.3
SD/M	.076	.255	.448	.046	.065	.179	.294	.044

Note. Data are given in milliseconds.

where C was one of [p], [t], or [k] and V was [i] or [a] (e.g., syllables such as [bəpa] and [bəkri]). Both results suggest that at any particular speaking rate, control in production (at least in stressed syllables) is exercised over the laryngeal event as a whole and not over the temporal microstructure of the sequence of segments that are the acoustic consequences of that event.

It would appear that the spectral properties of stop release after [s] accord with our interpretations of the first three experiments. In addition, the production data have shown that other concomitants of stop place, to which perceptual sensitivity might be demonstrated, exist in both the spectrum of friction offset and the duration of stop closure. They have also suggested a possible articulatory basis for a trading relationship between the duration of stop closure and the characteristics of the first formant at voicing onset in the perception of [s]-stop clusters. These possibilities are explored in the four experiments reported later.

Experiment 4

Experiment 4 was designed to investigate the influence of variation in the spectral properties of fricative offset on the perception of stop place in [s]-silence-vowel syllables. Specifically, we sought to determine whether the relationship between the position of the [p]–[k] boundary and the frequency of F2, shown in Figures 1 and 2, could be changed systematically by varying the spectral properties of the final 35 msec of the [s] friction.

Method

Stimuli. Forty stimuli were created by combining each of four [s] segments with each of 10 vocalic segments. One hundred milliseconds of silence intervened between the two types of segments. Both segments were created with the serial resonance synthesizer. Each [s] segment was 150 msec in duration with amplitude rise and fall times of 50 msec and 15 msec, respectively. Over their first 115 msec, the four [s]s were constant in frequency with the fricative formants set to 3917 Hz and 4932 Hz. The antiformant was set to eliminate energy below 2000 Hz. Different patterns of spectral change distinguished the final 35 msec of the four fricatives. In Pattern S1, the two fricative formants rose linearly to 4936 Hz and 6038 Hz. In Pattern S2 they remained at their steady-state values of 3917 Hz and 4932 Hz. In Pattern S3, the lower fricative formant fell to 3019 Hz. In Pattern S4, the lower fricative formant fell to 1957 Hz. The rise time of the vocalic portions was 30 msec. Over this duration F1 rose linearly from 200 Hz to a steady-state value of 299 Hz. The third formant was constant at 3199 Hz. The 10 vocalic segments were distinguished by the frequencies of their constant second formants, which ranged from 600 Hz to 2400 Hz in steps of approximately 200 Hz. A practice sequence consisting of a single randomization of the 40 stimuli and a test sequence consisting of five concatenated randomizations were recorded.

Subjects and procedure. Eight undergraduates served as subjects. They declared themselves to be phonetically naive, to possess normal hearing in both ears, and to have learned English as their first language in the United States. They were instructed to identify the medial stop in each syllable as either [p], [t], or [k]. They listened first to the practice sequence and then twice to the test sequence. In this way, 10 identifications of each syllable by each subject were collected.

Results

In Figure 4, the data from the eight listeners have been pooled and are displayed in four graphs, one for each of the fricative patterns whose spectral specifications are schematized in the insets. The percentages of [p], [t], and [k] responses are plotted as a function of the stimulus number and hence of the second-formant frequency in the vocalic portion of the stimuli. There is only a minimal difference between the patterns of [p] and [k] identifications of stimuli with fricative S1 and those with fricative S2; the number of [t] responses, already small, decreased slightly with the frequency of the fricative at offset. This trend continued through Patterns S2, S3, and S4, as the fricative offset frequency was further reduced and is significant when assessed in an analysis of variance, $F(3, 21) = 3.40$, $p < .025$. However, the main result of the experiment is that the proportion of [p] responses increased at the expense of [k] responses as the offset of the lower fricative formant was reduced between Patterns S2, S3, and S4. Overall, this effect is significant, $F(3, 21) = 9.26$, $p < .01$. Planned comparisons between adjacent series show that the only significant difference in the proportion of [p] responses is that between S2 and S3, $F(1, 21) = 9.55$, $p < .01$.

In natural productions of [s]-stop clusters, different spectral changes in the offset of the [s] accompany stop closure at different places in the vocal tract (see Table 1). The results of Experiment 4 show that the perceived place of a stop is influenced by the spectral properties of the [s] immediately prior to stop closure (cf. Malecot & Chermak, 1966). However, although consistent, the effect is small. It is manifest as an increase in the region of ambiguity between bilabial and velar responses and not as an increase in the number of stimuli identified unequivocally as bilabial. Nevertheless, these data, taken together with those of the previous experiments, show that just as the event of stop-consonant production occurs over time, so the acoustic information that specifies the identity of a stop for a perceiver is distributed over time.

The experiments that have been described so far have demonstrated perceptual sensitivity to the spectral properties of the segments bounding the period of stop closure. The following two experiments examined the influence of the duration of stop closure itself on the perception of place of articulation of stops in [s]-stop clusters.

Experiment 5

In this experiment the size of the silent interval reflecting stop closure was varied to create four series of syllables, each varying from [s]-vowel to [s]-stop-vowel. The four series were distinguished by different values of second-formant frequency in the vowel. These values were chosen on the basis of the results of Experiment 1 to give a [su-spu] series, a [si-ski] series, and two series intermediate between the two. We wished to determine whether the size of the silent interval necessary for the perception of a particular stop varied as a function of the acoustic specification of the stimulus, its phonetic interpretation, or both of these factors.

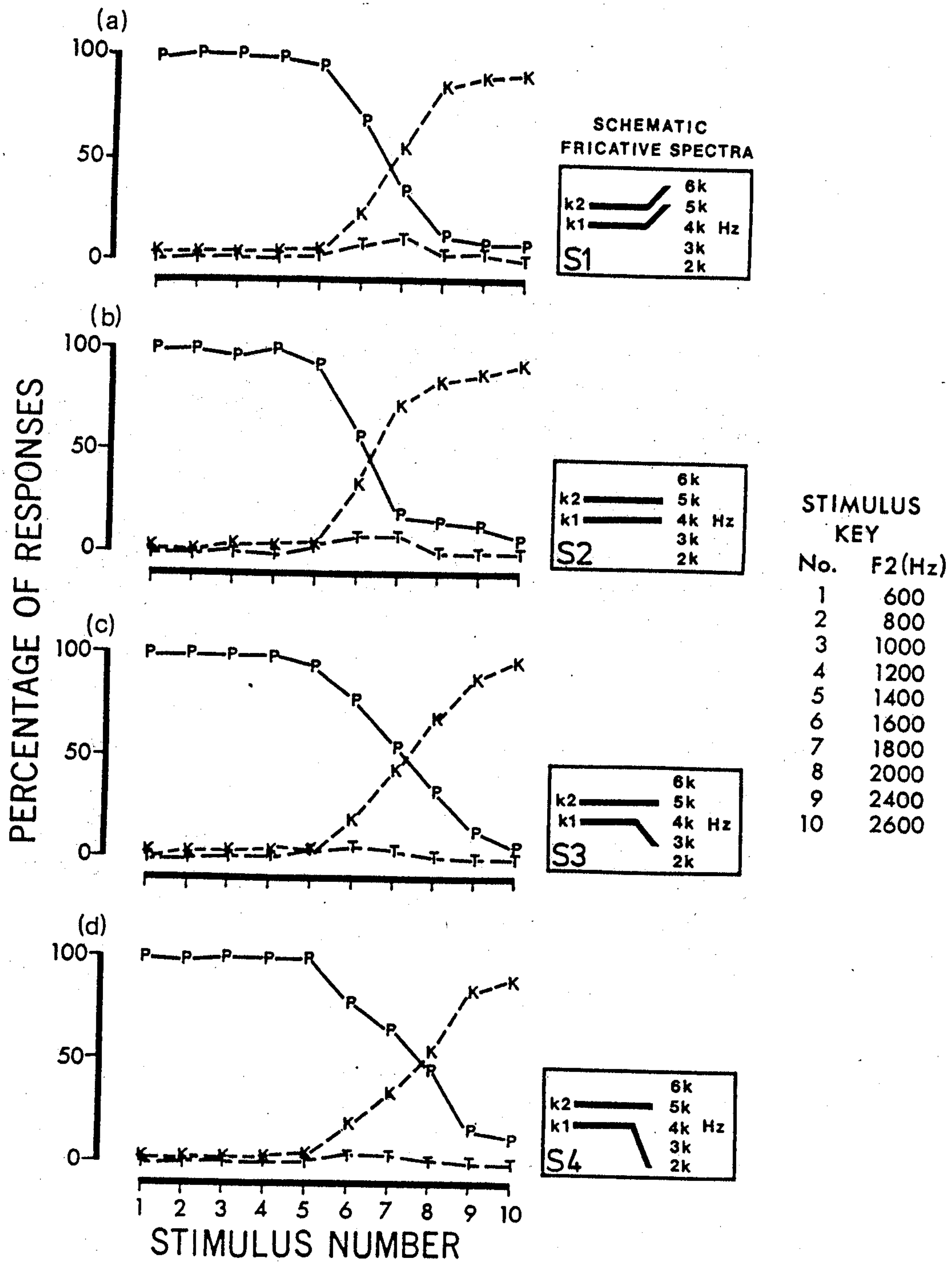

Figure 4. Identification functions from Experiment 4 for [s]-stop-vowel stimuli incorporating four spectrally different [s] segments (shown in insets). (Patterns of [sp]-vowel (P), [st]-vowel (T), and [sk]-vowel (K) responses are plotted as a function of the frequency of the second formant in the vocalic portion of the stimulus.)

Method

Stimuli. Four 11-member [s]-vowel to [s]-stop-vowel series were created by inserting an increasing duration of silence between the [s] friction and the vowel. The two-formant vocalic segments were identical to those in Stimuli 1, 4, 9, and 12 in Experiment 1. Their first formants were set to 260 Hz. Their second formants were set to 616 Hz, 1075 Hz, 1845 Hz, and 2307 Hz, respectively. The [s] friction was the same as that used in Experiment 1. For a given series, the duration of interpolated silence ranged from 0 msec to 100 msec in 10-msec steps. Two sequences were recorded for identification. One was a 44-trial practice sequence, the other was a 440-trial test sequence containing 10 instances of each of the 44 stimuli.

Subjects and procedure. Fifteen subjects, with the same qualifications as those who served in Experiment 4, listened first to the practice sequence and then to the test sequence. The stimuli were presented under the same conditions as in the earlier experiments. Subjects were instructed to identify each stimulus as either [s]-vowel or [s]-stop-vowel. In addition to a response for [s]-vowel, four response alternatives were provided for percepts of [p], [t], [k], and a glottal stop; a fifth alternative was provided for any stop not in these categories.

Results and Discussion

Figures 5a–5d display the data corresponding to each test series pooled over 15 subjects. Each graph plots, as a function of the duration of the silent interval, the percentage of [s]-vowel responses (V), [s]-stop-vowel responses (S), and the breakdown of the stop category into individual functions for [p], [t], and [k], glottal, and other stop responses (Q). Functions are not shown for response categories that received fewer than 10% of the total number of responses.

The production data in Table 2 show that at a fixed rate of speech, stop closures for bilabials are typically longer than are those for alveolars and velars. The present experiment was designed to determine whether velar stops are perceived with shorter silent intervals than bilabials. We can contrast two extreme hypotheses. One, an acoustic hypothesis, suggests that as the frequency of the second formant rises, the probability of hearing any stop at a particular silent interval increases. Since Experiments 1 and 2 have shown that the frequency of the second formant also determines stop place, a longer stop closure would be required to hear a bilabial than a velar. An alternative, phonetic, hypothesis would be that the phonetic decision about stop place is correlated with a criterial duration of silence characterizing stops with that place of articulation.

A comparison of the four graphs in Figure 5 appears to lend some support to the acoustic hypothesis. The crossover between [s]-vowel and [s]-stop-vowel responses in the pooled data occurred at shorter durations of silence as the frequency of the second formant in the vowel was raised. However, this trend does not appear consistently in the data of individual subjects, either when represented as percentages of [s]-vowel responses, or when represented as 50% crossovers on the [s]-vowel identification function estimated by probability unit (PROBIT) analysis (Finney, 1971). Changing the second-formant frequency did not produce significant changes in the duration of silence required to hear a stop. The nonsignificant trend for stops to be heard at shorter durations of silence as the second-formant frequency increased may be related to the rise in discriminability of the duration of silent intervals as the spectral contiguity of their bounding markers is increased (e.g., Divenyi & Danner, 1977). The phonetic hypothesis that the place decision is correlated with a criterial duration of silence can be assessed by comparing the crossover between [s]-vowel and [sp]-vowel responses in Figure 5a with that between [s]-vowel and [sk]-vowel responses in Figure 5d. Although the crossover for the velars occurs at a shorter duration of silence than that for the bilabials, the difference between the positions of the crossovers is not significant. This suggests that there is no simple causative relationship between the phonetic labeling of a stop and the amount of silence required to hear that stop in an [s]-stop cluster.

The stimuli used in this experiment specified information for stop place in a closely controlled fashion and were, therefore, an appropriate starting point for investigating the extreme versions of the acoustic and phonetic hypotheses outlined earlier. However, no correlation between

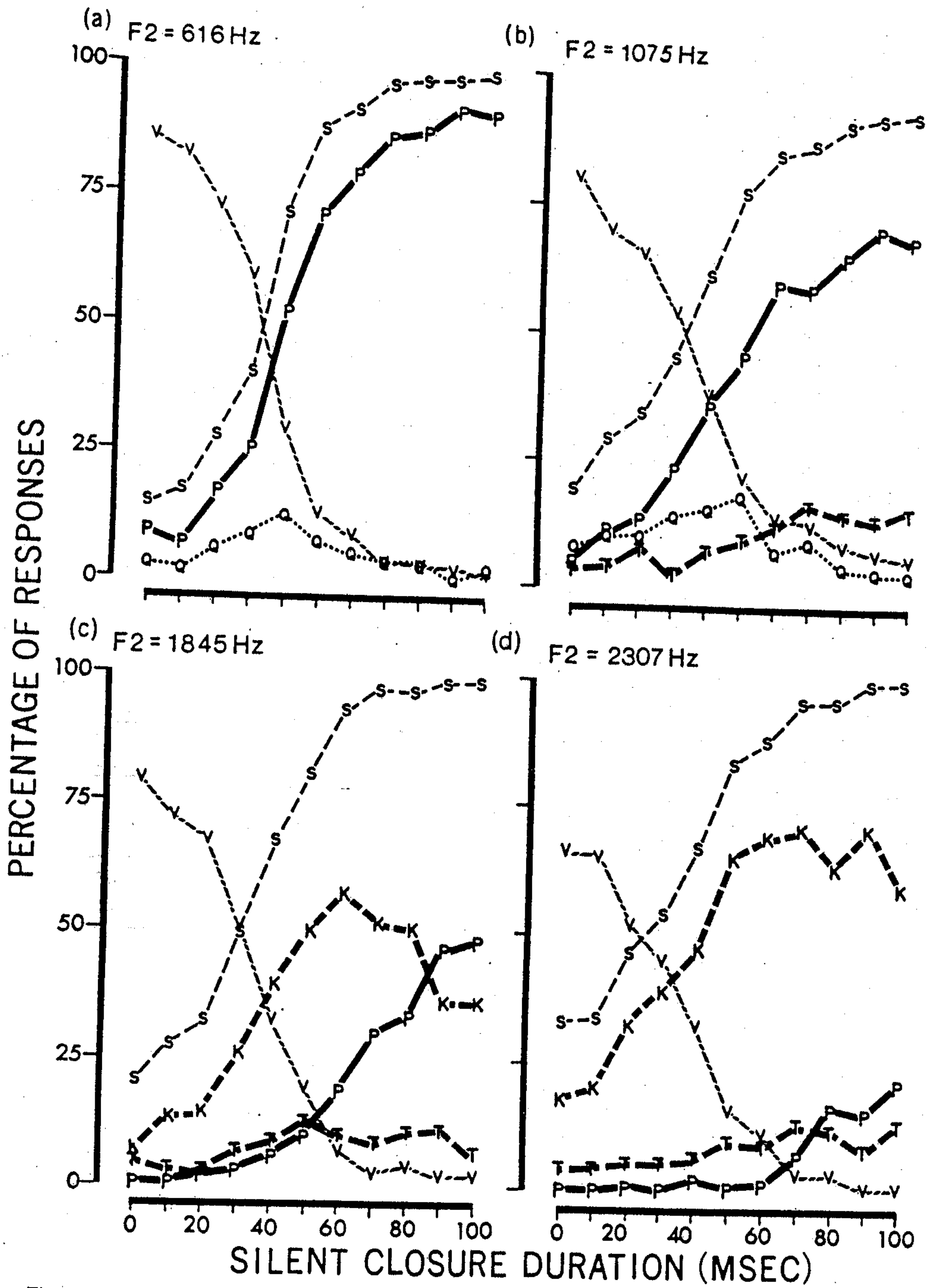

Figure 5. Identification functions from Experiment 5 showing the overall patterns of [s]-vowel (V) and [s]-stop-vowel (S) responses as a function of the duration of the silent closure interval. (Functions for four stimulus series in whose vocalic segments the second formant was fixed at (a) 616 Hz, (b) 1075 Hz, (c) 1845 Hz, and (d) 2307 Hz are displayed. Patterns of individual [s]-stop-vowel responses are shown as [sp]-vowel (P), [st]-vowel (T), [sk]-vowel (K), and [s]-glottal stop-vowel (Q) for those series in which they constituted more than 10% of the responses.)

production and perception emerged, possibly for the very reason that in these schematic stimuli, the acoustic differences between bilabials and velars were reduced to a minimum. A correlation might emerge with stimuli that reflect more fully the acoustic differences that are naturally concomitant with stops produced at different places of articulation. Later, in Experiments 6 and 7, we investigate the effects of manipulating transitions in the second and third formants, and then transitions in the first formant, on the duration of silence required to hear a stop.

The acoustic and phonetic hypotheses outlined earlier do not exhaust the explanatory principles that can be brought to bear on the results of Experiment 5. In making an analytic contrast between two phonetically distinct articulatory events, it is natural to focus first on the most prominent acoustic consequence of articulation that distinguishes the two events. If manipulation of this single parameter leads to different phonetic percepts, the parameter is accorded the status of a cue. However, it is commonly found that if the most potent cues are neutralized by being set to ambiguous values, perceptual sensitivity to less prominent acoustic properties may be demonstrated. Given sufficiently precise control over the acoustic structure of stimuli, it appears to be possible to demonstrate that some cue value attends every acoustic detail that distinguishes two different phonetic events. In the limiting case it becomes clear that every acoustic consequence of an articulatory event is a potential source of information about that event. Thus, according to this rationale, to demonstrate perceptual sensitivity to closure duration as a determinant of stop place, we should neutralize all other cues to place of articulation. This situation was approximated in the stimuli for which identification data are displayed in Figure 5c. Their second-formant frequency was chosen, on the basis of the results of Experiment 1, to give approximately equal numbers of [p] and [k] percepts with a silent interval of 100 msec. Figure 5c shows that both [p] and [k] were perceived. Moreover, other places cues were sufficiently neutral to allow closure duration itself to disambiguate place of articulation: As the production data predict, [k] percepts predominated at short closure durations, whereas [p] percepts predominated at longer closure durations.

The data of Figure 5c suggest more than the simple conclusion that the crossover between [s]-vowel and [sp]-vowel percepts occurred at a longer duration of silence than the crossover between [s]-vowel and [sk]-vowel responses, which would be a direct consequence of the ambiguous place category being resolved in favor of a consistently greater proportion of velars than bilabials. In such a situation, the function corresponding to the less frequent response would always intersect the [s]-vowel function at a longer silent interval than the function corresponding to the more frequent response. Figure 5c shows instead that the ratio of bilabial to velar stops is not fixed, but varies systematically with closure duration, with velars predominating at short closures and bilabials at longer closures. It is unfortunate that the closure durations in the experiment were not extended beyond 100 msec so that the predominance of bilabials at longer closure durations could have been shown more convincingly. However, a more stringent demonstration of the effect can be made. It requires that the identification function corresponding to the less frequently used stop category peak at *shorter* silent intervals than that corresponding to the more frequently used category. This will be shown in Experiment 6.

In summary, the results of Experiment 5 suggest that in the traditional terminology, a silent closure interval is a cue both to stop manner and to stop place. Its latter role is revealed when other cues to place of articulation are neutralized.

Experiment 6

The intention of Experiment 6 was similar to that of the previous experiment in its concern with the duration of the silent closure interval required to hear a stop. It was designed to determine how this duration is influenced by the spectral specification of second- and third-formant transitions introducing the vocalic portion of the stimulus.

Method

Stimuli and procedure. Six consonant-vowel syllables were prepared with the parallel resonance synthesizer. In each syllable the first formant had its onset at 463 Hz and rose linearly to a steady state at 614 Hz. The second and third formants had their steady states at 1845 Hz and 2694 Hz, respectively. The onsets of the second- and third-formant transitions were covaried to produce in informal listening tests two instances each of [b], [d], and [g]. The onset frequencies of the second- and third-formant transitions in these syllables were 1386 Hz and 2525 Hz (B1), 1541 Hz and 2694 Hz (B2), 1695 Hz and 2862 Hz (D1), 1845 Hz and 2862 Hz (D2), 1996 Hz and 2694 Hz (G1), and 2156 Hz and 2525 Hz (G2). The formant transitions were all 40 msec in duration, and the total duration of the CV syllables was 300 msec. Six [s]-vowel to [s]-stop-vowel series were created by combining each of these vocalic portions with [s] followed by silence using the same procedure as in Experiment 5. Each series consisted of 10 stimuli in which the duration of silence ranged from 0 msec to 90 msec in 10-msec steps. Two identification sequences were recorded. A 24-trial practice sequence included two instances of each of the endpoints from the stimulus series. The 300-trial sequence included five presentations of each of the 60 stimuli. Eight subjects with the same qualifications as those who served in Experiment 4 listened first to the practice sequence and then to two presentations of the test sequence. Thus, each subject listened to 10 presentations of each stimulus. The subjects were instructed to identify the syllables as [s]-vowel or [s]-stop-vowel using the same response categories as in Experiment 5.

Results and Discussion

The results of Experiment 6, pooled over subjects, are displayed in Figures 6a–6f. As in Figure 5, percentages of [s]-vowel (V), [s]-stop-vowel (S), and the breakdown of the stop category into individual functions for [p], [t], and [k] are plotted as a function of the duration of the silent interval. Functions are not shown for response categories that received fewer than 10% of the total number of responses. The data will be discussed first in terms of the relationship between [s]-vowel and [s]-stop-vowel responses and then in terms of the particular stop heard.

The closure durations corresponding to the crossovers between [s]-vowel and [s]-stop-vowel responses in the pooled data of Figure 6 are B1—30.0 msec; B2—31.2 msec; D1—29.0 msec; D2—27.4 msec; G1—19.8 msec; and G2—10.1 msec. The significance of any changes in the distributions of [s]-vowel to [s]-stop-vowel responses underlying the difference between these crossovers was assessed in an analysis of variance that examined the proportions of [s]-vowel responses made by each subject to the 10 stimuli in each series combined. Overall, the different stimulus series gave significantly different numbers of [s]-vowel responses, $F(5, 35) = 12.32$, $p < .01$. A posteriori comparisons made according to the criteria recommended by Tukey (Winer, 1971) show that Series G2 received significantly fewer [s]-vowel responses than any other series and that none of Series B1, B2, D1, or D2 differed significantly from one another. These comparisons confirm the finding of Experiment 5 that the duration of silence required to hear a stop is not simply a function of the perceived place of articulation of the stop. Instead, a post hoc examination of the data shows them to correspond closely to an acoustic variable, the frequency separation of the second and third formants at the vocalic onset. As the onset frequencies of F2 and F3 approximate one another, the duration of silence required to hear a stop is reduced. The second and third formants in Pattern G2 are probably close enough at their onsets to fall within one critical band (Scharf, 1970), and the resulting summation of energy could have specified the vocalic onset of this pattern more distinctively than those of the other five patterns. Energy summation may combine with another class of acoustic effect in which the duration of silence required to hear a stop covaries with the amount of spectral change at the vocalic onset. This additional hypothesis is explored later in Experiment 7 for transitions in the first formant. Here, it would predict equivalent outcomes for Patterns B1 and G2. Nonlinear additivity of this effect with that of energy summation could have enhanced the difference between Pattern G2 and the other five patterns.

The results of the previous experiment (Experiment 5) suggested that the relation-

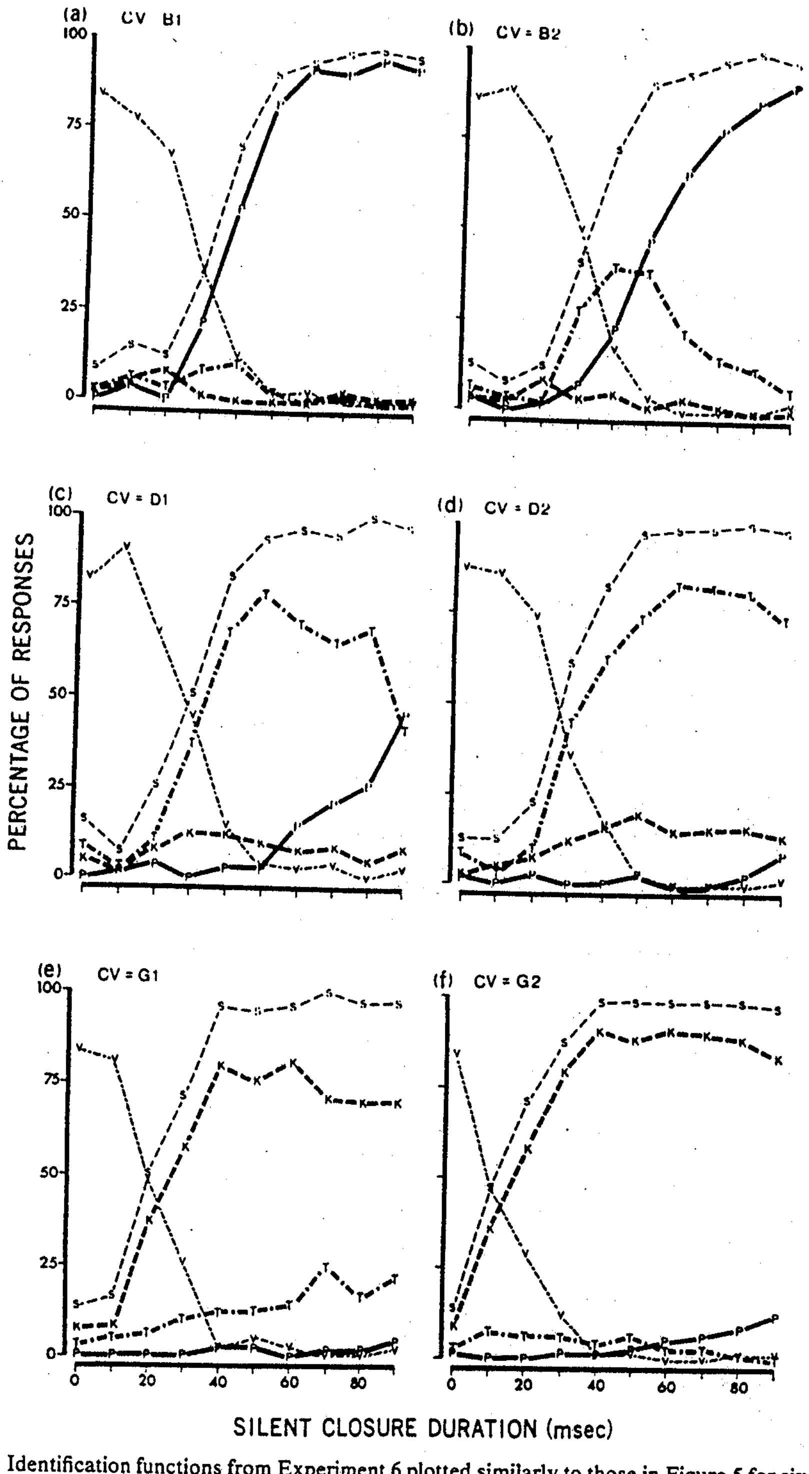

Figure 6. Identification functions from Experiment 6 plotted similarly to those in Figure 5 for six series of stimuli each ranging from [s]-vowel to [s]-stop-vowel incorporating vocalic portions varying in initial F2 and F3 transitions from [bɛ] (B1 and B2) to [dɛ] (D1 and D2) to [gɛ] (G1 and G2).

ship between closure duration and the perceived place of a stop consonant following [s] is most likely to be revealed when other cues to stop place are neutralized. This situation was most closely approximated in the present experiment with Series B2 and D1. Overall, Series B2 received 66.4% [p] responses and 28.7% [t] responses, whereas Series D1 received 18.2% [p] responses and 69.3% [t] responses. The production data in Table 2 show that shorter closure durations characterize alveolar compared to bilabial stops. The patterns of data in Figures 6b and 6c, corresponding to Series B2 and D1, respectively, reflect this relationship: [t] percepts predominate at short closure durations, and [p] percepts predominate at longer closure durations. In Figure 6b, the function for [t] responses intersects the function for [s]-vowel responses at a shorter closure duration than does the function for [p] responses, despite there being a smaller proportion of [t] than [p] responses overall. As was noted in the discussion of Experiment 5, this situation demonstrates that perceived place of stop articulation can depend on stop closure duration. There are two reasons why equivalent effects are not shown in Figures 6d and 6e for series D2 and G1, which straddle the alveolar-velar boundary. First, these patterns specified place of production less ambiguously than did Patterns B2 and D1. Of the [s]-stop responses to series D2, 78.7% were [t], and 90.3% of the [s]-stop responses to series G1 were [k]. Second, in natural productions of [s]-stop clusters, the difference between stop closure durations in alveolars and velars is much smaller than that between either of these categories and bilabials (Table 2).

In summary, the results of Experiment 6 confirm those of Experiment 5. The duration of the stop closure following [s] can serve to disambiguate bilabial from alveolar and velar place categories when other cues to place of articulation are neutralized. On the other hand, the probability of hearing any stop at a particular closure duration following an [s] is largely determined by the acoustic structure of the vocalic portion of the stimulus. (As predicted in the discussion of Experiment 3, [t] percepts are obtained when the vocalic portion contains appropriate formant transitions.) Crossovers between [s]-vowel and [s]-stop-vowel responses fell at about the same closure duration on continua whose vocalic portions exemplified [bɛ] and [dɛ]. Crossovers for continua constructed from vocalic [gɛ] fell at shorter durations. In production, however, alveolars and velars are characterized by similar closure durations, and bilabials are distinguished by longer intervals. If a perceptual trading relationship exists that parallels the differences that occur in production across the different places of articulation, it cannot be completely determined by differences in the spectrotemporal specification of F2 and F3. Accordingly, the final experiment investigated the influence of the characteristics of the first formant on the duration of silence required to hear a stop after [s]. Possibly a trading relationship exists between the characteristics of F1 and the closure duration required to hear a stop. This could compensate both for the differences in closure duration that occur between the different place contexts and for those, noted earlier in relation to the production data, between more and less open vowels.

Experiment 7

Experiment 7 was designed to determine how the onset frequency and the magnitude of the first-formant transition influence the duration of stop closure required to hear a stop after [s].

Method

Stimuli and procedure. Six CV syllables were created with the parallel resonance synthesizer. They had identical second- and third-formant contours. F2 and F3 onset at 1541 Hz and 2862 Hz and fell linearly to steady-state frequencies of 1312 Hz and 2525 Hz, respectively. All transition durations were 35 msec. The duration of each syllable was 300 msec. A typical stimulus is schematized in Figure 7a. The first formant contours fill six cells of a 3 × 3 matrix designated by three values of F1 onset frequency and three values of F1 transition extent as illustrated in Figure 7b. The transition extent of Contours 1 and 2 was 0 Hz, of Contours 3, 4, and 5 was 156 Hz, and of Contour 6 was 309 Hz. The onset frequency of Contours 3 and 6 was 154 Hz; of Contours 1 and 4, 311 Hz; and of Contours 2 and 5, 463 Hz. Six 10-member [s]-vowel to [s]-stop-vowel stimulus series were constructed, as before, by interpolating silent intervals ranging from 0 msec to 90 msec in 10-msec steps between 120 msec of [s] friction and each vocalic segment. Two identification sequences were recorded. A practice sequence of 24 trials included two instances of the endpoint stimuli from each series. A test sequence of 300 trials contained five instances of each of the 60 stimuli.

Eight subjects with the same qualifications as those who had taken part in Experiment 4 listened first to the practice sequence and then to two presentations of the test sequence to yield 10 identifications of each stimulus by each subject. The instructions were identical to those given in the two previous experiments.

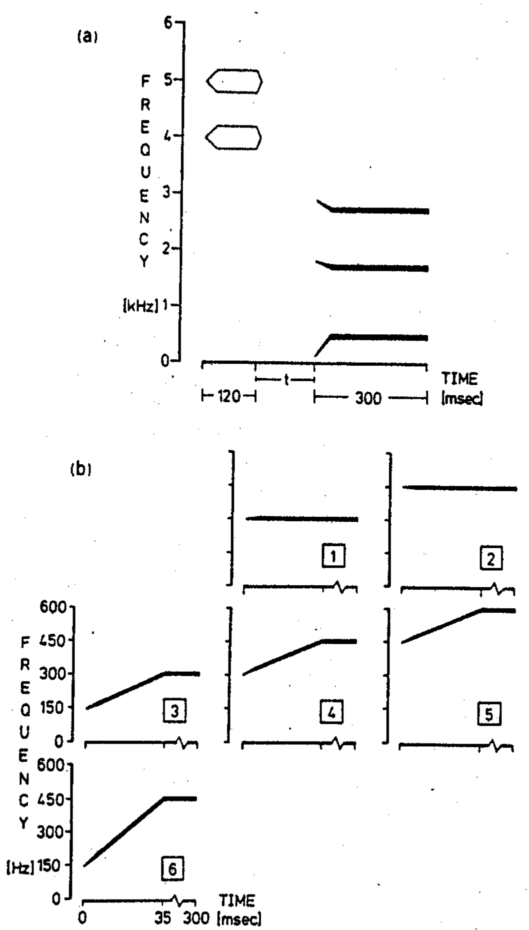

Figure 7a. Schematic spectrogram of a typical stimulus from Experiment 7. (Upper panel.) (Stimulus series were created by varying the interval t from 0 to 90 msec in 10 msec stops.) *Figure 7b.* Schematic representations of the first formant contours in six series of stimuli each ranging from [s]-vowel to [s]-stop-vowel, (Lower panel.)

Results and Discussion

The data from the six stimulus series pooled over the eight subjects are displayed in Figures 8a–8f. As before, each graph displays the percentages of [s]-vowel responses (V), [s]-stop-vowel responses (S), and the breakdown of the stop category into individual functions for [t] and [k] responses, as a function of the duration of the silent interval. Figure 8 is supplemented by Table 3, in which four summary measures are tabulated for each of the six stimulus series. These are (a) the duration of silence at the crossover between [s]-vowel and [s]-stop-vowel responses estimated from the pooled data, (b) the overall percentages of [s]-stop responses, (c) the percentage of [k] responses out of the total number of [s]-stop responses, and (d) the percentage of [t] responses out of the total of [s]-stop responses. Measures c and d were derived from the data of four subjects who made both [k] and [t] responses.

Considering first the relation between [s]-vowel and [s]-stop-vowel responses, two trends are evident: Crossovers between [s]-vowel and [s]-stop-vowel responses occurred at shorter silent intervals both as the onset frequency of F1 was lowered and as the magnitude of the F1 transition was increased (see Table 3a). The statistical significance of these effects could not be assessed directly using the crossover measure because a single crossover could not be estimated for every subject on every stimulus series. As an alternative, directional t tests were performed on the percentages of [s]-stop responses, a measure that could be determined for every subject (see Table 3b). By comparing the pair of percentages in each column of the matrix in Table 3b, an assessment of the effect of increasing the magnitude of the F1 transition by 154 Hz may be made. Series 6 produced more [s]-stop responses than did Series 3, $t(7) = 2.65$, $p < .025$ (one-tailed); significant effects in the same direction were found between Series 4 and 1,

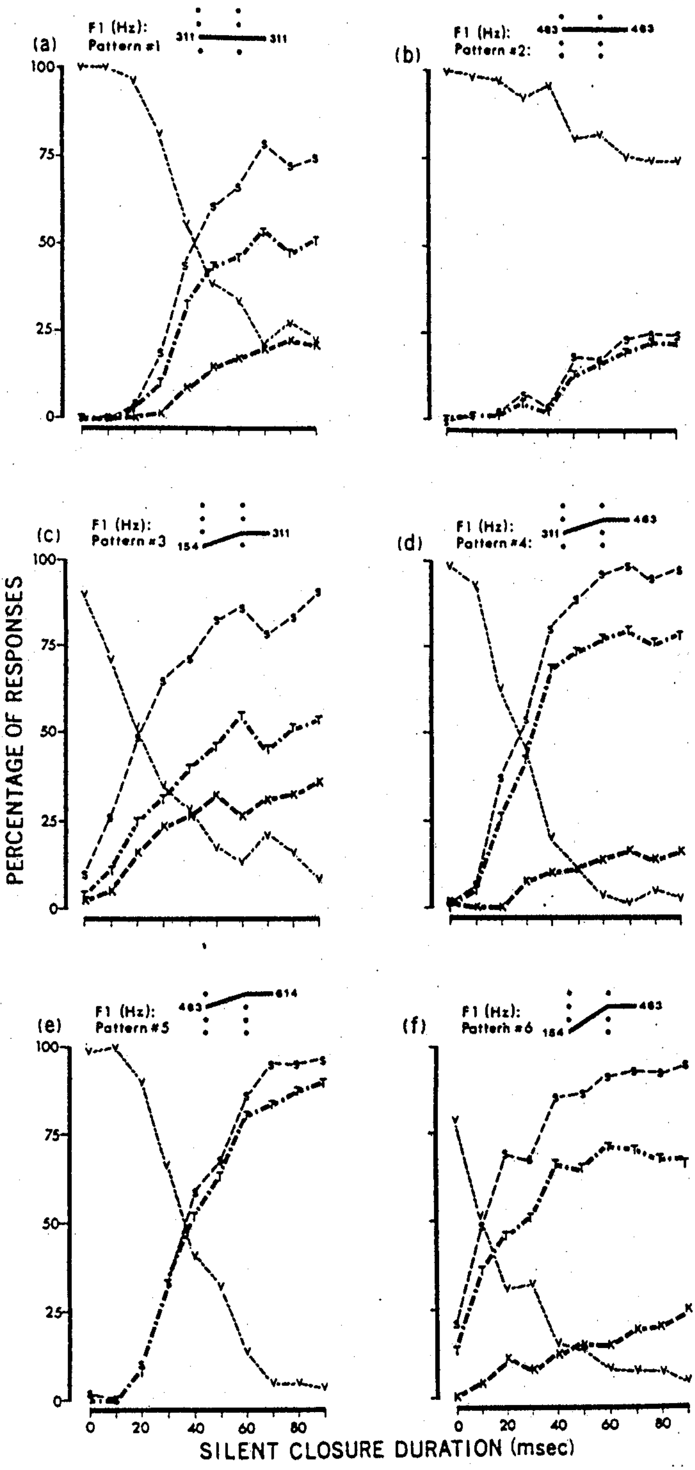

Figure 8. Identification functions from Experiment 7 plotted similarly to those in Figure 5. (Each panel displays data from one stimulus series.)

$t(7) = 5.97$, $p < .01$ (one-tailed), and between Series 5 and 6, $t(7) = 6.60$, $p < .01$ (one-tailed). Thus, in all three cases, a greater magnitude of F1 transition produced a larger percentage of [s]-stop responses.

Similarly, by comparing pairs of percentages in adjacent rows of Table 3b, there are three comparisons that allow an assessment of the effect of lowering the onset frequency of F1 by 154 Hz. Series 3 and 4 did not produce significantly different means, $t(7) = .15$; however, both Series 1 and 2, $t(7) = 3.90$, $p < .01$ (one-tailed), and Series 4 and 5, $t(7) = 2.30$, $p < .05$ (one-tailed), differed significantly. In each case, a lower F1 onset frequency produced a larger percentage of [s]-stop responses. Although one of the changes in F1 onset frequency (from 311 Hz to 154 Hz) failed to produce a significant increase in the percentage of [s]-stop responses, all three were accompanied by a reduction in the duration of silence at the crossover between [s]-vowel and [s]-stop-vowel responses (Table 3a). Overall, it appears that both of the manipulations applied to the first formant in Experiment 7 can produce systematic effects on the duration of silence required to hear a stop after [s].

The response patterns of four of the eight subjects included both [k] and [t] responses, whereas the other four subjects made only [t] responses. The percentage of [k] and [t] responses out of the total number of [s]-stop responses are displayed in Tables 3c and 3d for the four listeners for whom the place category was ambiguous. Each table shows a consistent trend: The percentage of [sk] percepts increased as the onset frequency of F1 was lowered and as the magnitude of its transition was reduced, whereas the inverse pattern applied to [st] percepts. Just as perceptual sensitivity to the covariation of closure duration with place of articulation was demonstrated in Experiments 5 and 6 when other information for place was ambiguous, so here, four subjects have shown perceptual sensitivity to the covariation in production of characteristics of the first-formant transition with place. Spectrograms of natural utterances typically show that this covariation involves longer, slower first-formant transitions for initial velar stops, which may also entail a lower first-formant onset frequency, than alveolars (e.g., Fant, 1973, p. 118).

Table 3
Summary Data Measures From Experiment 7

F1 transition extent (Hz)	F1 onset frequency (Hz): 154	311	463
a. Crossovers between [s]-vowel and [s]-stop-vowel expressed as milliseconds of silent closure interval[a]			
0	—	43.6	90.0[b]
157	20.8	27.7	36.5
309	10.6	—	—
b. Percentages of [s]-stop-vowel responses out of all responses[a]			
0	—	42.0	7.5
157	64.6	65.5	54.6
309	74.9	—	—
c. Percentages of [sk]-vowel responses out of all [s]-stop-vowel responses[c]			
0	—	41.3	26.7
157	55.2	24.3	9.5
309	28.2	—	—
d. Percentages of [st]-vowel responses out of all [s]-stop-vowel responses[c]			
0	—	48.5	64.4
157	32.6	65.9	83.6
309	57.1	—	—

Note. — means combinations of F1 onset frequency and F1 transition extent are not used.
[a] Data are from all eight subjects. [b] Crossover falls beyond 90.0 msec of silence. [c] Data are from four subjects who made both [sk]-vowel and [st]-vowel responses.

Tables 3a and 3b imply the existence of a perceptual trading relationship between the spectral properties of the first-formant transition and the duration of silence required to hear a stop after [s]. Both silence, which is an indicant of a completely constricted vocal tract, and a first-formant rising from a low frequency, which is an indicant of the release of vocal tract constriction, are natural acoustic concomitants of the production of a stop consonant. If it is assumed that a perceptual system for speech exists that is sensitive to both of these attributes and that seeks a criterial amount of information for the presence of a stop, then a trading relationship between the two attributes would be expected. Less silence is required when the spectral attribute is more prominent. (See also, Erickson, Fitch, Halwes, & Liberman, 1977; Summerfield & Haggard, 1977.) The production data in Table 2 endorse the utility of a system organized in this way: The duration of the stop closure is inversely related to the rate at which the oral constriction is released. Thus, longer closures characterize bilabials compared with alveolars and velars, and, for a given place, shorter closures precede the open vowel [a] compared with the more closed vowels [i] and [u].

General Discussion

Summary of Results

The experiments reported here shared a concern for the perception of stop consonants in syllable-initial [s]-stop clusters. The first two experiments showed that the sequence [s]-silence-vowel can give rise to the percept of an initial [s]-stop cluster. The perceived place of the stop was related to the frequency of the second formant: Low second-formant frequencies gave [sp] percepts, and high second formants gave [sk] percepts; few [st] percepts occurred. In contrast, Delattre et al. (1955) reported that in syllable-initial position, where information for stop manner was carried by a rising first formant, a steady second formant at 1800 Hz gave alveolar percepts, substantiating the principle of formant loci.

In Experiments 1 and 2, in which information for stop manner was conveyed by a period of silence simulating stop closure, an explanation for the absence of a significant number of [t] percepts required consideration of a wider range of the consequences of production than are entailed in the principle of formant loci. The steady-state vocalic portions of the stimuli used in Experiments 1 and 2 could only represent natural articulations that give rise either to no release burst or to contiguous energy at and following the release. The hypothesis was developed that alveolar percepts were absent because the stimuli failed to simulate the spectral discontinuity between alveolar release and the following formant pattern.

These notions were tested in Experiment 3, in which the spectral relationship between a release burst and the following second formant was systematically manipulated in [s]-silence-burst-vowel syllables. In accordance with predictions based on the acoustic concomitants of natural productions, the perception of place of articulation in the interpolated stop was determined by the spectral relationship between the burst and the second formant: [st] percepts were reported when the burst was at a higher frequency than, and discontinuous with, F2; [sk] percepts were reported when the burst was spectrally contiguous with F2; [sp] percepts were infrequent with these concentrated bursts but were sometimes reported when the burst frequency was low. We note that the perceptual data in these experiments are rationalized not by identifying a relationship between perceived place and any particular cue dimension, but by determining the articulatory event whose temporally distributed acoustic consequences are most closely approximated by each stimulus pattern.

The relative success of an appeal to the details of articulation as an explanatory principle for perception motivated Experiments 4–7. An analysis of natural productions of syllable-initial [s]-stop clusters showed that both the spectral properties of the offset of [s] friction and the duration of the silent closure interval are characteristic of the place of production of the stop. Experiment 4 showed perceptual sensitivity to the first of these characteristics: Lowering the offset frequency of the fricative predisposes [sp] percepts primarily at the expense of [sk] percepts. Experiments 5 and 6 demonstrated that the duration of the stop closure can determine perceived place of articulation when spectral information for place is ambiguous: Shorter closure intervals predispose [st] and [sk] percepts as opposed to [sp] percepts. In a similar fashion, Experiment 7 showed that for some listeners, the characteristics of the first formant at and following the release can determine perceived place of articulation when other information for place is equivocal. Experiment 7 also demonstrated an interrelationship between the duration of stop closure and the spectral characteristics of the first formant in the perception of stop manner: A shorter duration of silence is required to hear a stop after [s] as the onset frequency of the first formant is lowered and as the magnitude of its transition is increased. This perceptual trading relationship appears to parallel an inverse correlation in production between the duration of stop closure and the openness of the following vowel.

We acknowledge that the stimulus series used in these experiments are not veridically representative of any natural articulatory dimension. Moreover, the schematic vocalic portions of the stimuli were deliberately constrained and, in some cases, involved spectral changes not typical of natural productions. However, the stop percepts in the stimuli were not unnatural, and, when subjects were provided with response categories for ambiguous percepts, they rarely used them. Although interpretations of perceptual data obtained with constrained synthetic stimuli must be tempered by these considerations, the consistency of our results with predictions based on analyses of natural productions endorses the utility of the approach for exploring and accounting for the limits of perceptual sensitivity. The technique of using schematic and geometrically specified stimuli is a powerful tool for demonstrating the gross relationship between perceptual identity and articulatory events. However, its failure to represent the subtleties of natural articulatory dynamics may limit its ability to generate a complete characterization of the information specifying phonetic identity. Nevertheless, the results of the present experiments and those, for example, of Repp et al. (1978) clearly show that in identifying place of articulation of stop consonants, listeners are sensitive to the acoustical concomitants of the articulatory gestures of constriction and occlusion in addition to those of the release phase. Accordingly, an account of perceptual sensitivity to stops should be founded on a characterization of acoustics that includes all three phases.

This conclusion appears to run counter to the emphases of Zue (1976) and of Stevens and Blumstein (1978), who have stressed that specificity to place of production of stop consonants is found in the spectrum sampled at the release, rather than in temporally distributed, time-varying properties. The contradiction is partly superficial. Clearly, spectra sampled at moments adjacent to complete articulatory occlusion possess potentially greater specificity to place of production than do some of the acoustical parameters that have been manipulated here. At moments adjacent to occlusion of the vocal tract, the oral cavity in front of the point of articulatory constriction is well defined. Its resonant qualities, which relate principally to its length and hence provide an indication of the position of the constriction in the mouth (Kuhn, 1975), dominate the spectrum. However, articulatory gestures are continuous and intrinsically time varying. To reciprocate the subtleties of articulatory constraints, phonetic perception should be sensitive to dynamic properties of the spectrum which, it is argued later, are not well specified in a succession of elemental cues.

The Concept of a Cue

The methodology of the present experiments can be traced directly to early work, which sought to specify the acoustic cues of speech (e.g., Liberman et al., 1952). Techniques of analysis and synthesis provided an operational definition of a cue as a physical parameter of a speech signal whose variation could systematically change the phonetic interpretation of the signal. A large body of data attests to the absence of a one-to-one correspondence between a physical specification of those cues and the phonetic percepts that they induce (e.g., Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). The belief that a more nearly invariant relationship exists between phonetic interpretation and the events of articulation motivated a class of perceptual models that sought to specify how articulatory reference might guide the recovery of phonetic identity from the substrate of the acoustic cues. Thus, the cue achieved a functional role in a perceptual system as an element of information used in the construction of a featural description of the signal in articulatory terms. This description was assumed to permit a more direct mapping to phonetic identity (e.g., Mattingly & Liberman, 1969; Stevens & House, 1972). Even models that do not make the reconstruction of articulation explicit allow that articulatory reference can sometimes mediate the acoustic-phonetic translation (e.g., Fourcin, 1974; Pisoni & Sawusch, 1975).

In such models, a particular phonetic feature is detected when a criterial amount of information favoring that feature has been accumulated from the available cues. Trading relationships between cues conveying information for the same feature are inevitable: The greater the amount of information available from one cue, the smaller the amount of information required from another to attain the criterion for deciding that the feature is present.

The notion that information for phonetic identity is conveyed by acoustic cues appeared to possess the attraction of delimiting the critical aspects of the signal and thereby reducing the amount of information that the perceiver must process. Given the rapid rate at which phones are uttered, it is desirable that the amount of information processing required to perceive each phone be minimized. This was seen to be achieved in part by the parallel transmission of cues as an inevitable result of coarticulation. However, in practice, the adequacy of the account would rest on there being in addition a finite and, ideally, small number of cues to be processed for any particular phonetic distinction. The data from the present experiments do not encourage the belief that the set of cues is bounded: For stops after [s], at least, perceptual sensitivity has been demonstrated to each of the acoustic manipulations that we have chosen to investigate.

In the production of a stop consonant, constriction, occlusion, and release of the supralaryngeal vocal tract are aspects of a continuous articulatory event that unfolds over time. In a comparison of stops articulated at two different places of production, the configurations of the articulators, and hence the concomitant acoustic signals, are likely to differ between the two events at every moment within their time span. Given sufficiently precise stimulus control, perceptual sensitivity could probably be demonstrated to every difference between the acoustical consequences of distinct articulatory maneuvers. The set of cues to phonetic categories distinguished by these articulations would then appear to be unbounded. The problem might be resolved by postulating a ranking of the cues in order of importance, such that sufficient information to disambiguate every phone is conveyed by a limited subset of the total number of cues. However, it is a requirement of this solution to the problem that the speech processor ignore the minor cues, at least in the process of speech perception under normal conditions. If this were not the case, part of the perceptual task would be to distinguish major from minor cues and would require, therefore, that all cues be registered. If it is allowed that the minor cues are normally ignored, it is a further requirement that the major cues completely specify the phonetic contrast that they distinguish, since the same cue can play different roles in different contexts (Liberman et al., 1967). However, if the minor cues are ignored under natural conditions of speech perception, it is paradoxical that perceptual sensitivity can be demonstrated to them at all. These general problems are not unique to phonetic perception (see Weisstein, 1973).

It is appropriate to ask whether a list of any number of cues is the best, or even a sufficient, specification for a phonetic category. That it may not be the best specification is suggested by the fact that different, although not necessarily exclusive, sets of cues can specify the same phone in different contexts. That it cannot be a sufficient specification follows from the fact that as the experiments reported here have shown, the information for phonetic perception is distributed over time. A set of cues will only be correctly interpreted if they occur with the proper temporal coordination. It follows that to detect a particular phone, the perceptual system must be constrained to register not only a specific set of cues but also the temporal coordination that articulation imposes on them. Although articulatory events occur over time, and so cues are distributed over time, attention has traditionally been directed toward specifying the cues independently of their temporal coordination. As a complement to this approach, perceptual models have typically included an early stage in which the set of cues for a particular phone, coherently arrayed in the speech stream by the events of articulation, are fractionated out as discrete elements, subsequently to be reintegrated by the mediation of internalized rules of articulatory coherence.

A perceptual system in which the information for phonetic perception was a set of cues would have to incorporate three kinds of knowledge if it were to function successfully. It would have to know, first, which aspects of the acoustic signal are cues and which are not; second, it would need to possess a sensitivity to the pattern of cooccurrence of cues for each phone in its perceptual repertoire; third, it would need to appreciate the proper temporal coordination of the cues within each pattern. There is no reason, in principle, why a device could not be built to perceive phonetic identity from a substrate of acoustic cues, provided it was endowed with an articulatory representation sufficient to embody these three kinds of knowledge. However, we doubt that such a system could evolve in the natural world. For a species to acquire a knowledge of articulatory constraints, it would be necessary first that information specifying those constraints be available for the species, and second that the species possess a prior sensitivity to that information. The knowledge that a particular set of cues combine to indicate the presence of a given phone could be acquired in either of two ways. The identity of the phone could be specified independently of the set of acoustic cues, but this would hardly solve the problem and would preempt the need to evolve a sensitivity to the cues. Alternatively, the signal could specify directly both the identity of the cues and their temporal coordination, but then information in the signal that specified the coherence of its elements would, isomorphically, specify the articulatory event from which that coherence derived. However, the presence of this information about articulation in the signal, and a predisposition to register it on the part of the perceiver, would obviate the need for any internalized articulatory referent to mediate the acoustic-phonetic translation.

These considerations lead us to question the validity of equating the operational and functional definitions of an acoustic cue. A cue was defined operationally as a physical parameter of a speech signal whose manipulation systematically changes the phonetic interpretation of the signal. Although it is clear that perceptual sensitivity must exist to the consequences of manipulating a cue, it is not necessary to suppose that the cue is registered in perception as a discrete functional element.

Conclusion

The cue concept worked well for 25 yr., bringing freedom from nonanalytic approaches to speech; in turn, it has precipitated intransigent theoretical problems. The foregoing argument implies the need for a vocabulary other than that of the traditional acoustic cues to phonetic categories with which to describe the perceived consequences of articulatory events. That vocabulary should preserve the operational potency of traditional cues but not impute to them a functional role in perceptual processing. We suggest that the basis for this revision be sought in those accounts of speech production (e.g., Fowler, Rubin, Remez, & Turvey, 1980) that regard articulatory continuity as the result of speakers so organizing their vocal musculature as to generate for every phoneme an equivalence class of movements, rather than a particular, essentially static, articulatory target configuration. Kinematic details of articulatory maneuvers that distinguish among phonetic classes and that are represented acoustically could provide a vocabulary capturing the temporal dependence of the information for phonetic perception. In that vocabulary the perceptually important details of articulation and acoustics would receive isomorphic descriptions.

The general viability of this orientation is supported by two types of experimental observation. The first is found in demonstrations that phonetic information does not reside immutably in acoustic patterns; percepts may vary when listeners not only hear the talker but can also see his or her face (e.g., MacDonald & McGurk, 1978). This change in phonetic percept in the absence of acoustic change must rest on the possibility of integrating optics and acoustics. Such integration could only occur if a common metric exists in which the perceptually pertinent variables in each modality can be described. The second type of observation is that there exists perceptual sensitivity to higher order, temporally dependent properties of events that is not dependent on an initial fractionation of stimuli into lower order primitives. Velocity may be apprehended directly in vision, for instance, without prior mediation of representations of displacement and time (Lappin, Bell, Harm, & Kottas, 1975). This demonstration allows that the perception of events in general, including articulatory events, may involve the direct apprehension of patterns of change over time and may not, therefore, require the perceptual integration of a succession of discrete cues. Consistent with this suggestion is the fact that vowels uttered in a dynamic consonant-vowel-consonant context are perceived more accurately than are tokens of the same vowels produced in isolation (Shankweiler, Strange, & Verbrugge, 1977; Strange, Verbrugge, Shankweiler, & Edman, 1976). Accordingly, the acoustical specification of a vowel may entail a specification of change over time (Fowler et al., 1980).

These arguments by no means confirm, but do encourage, the belief that the perception of speech could be described as the registration of spectral kinematics isomorphic with the kinematics of their articulatory origins. Experiments of the type reported here would be more fruitful if they were the complement to studies of the relationship between strategies of articulatory control and the concomitant acoustics, so that stimulus patterns could be specified not in the arbitrary metric of Euclidean geometry but in the natural metric of articulatory dynamics. As a result, the endeavor might demonstrate perceptual sensitivity to the kinematics common to articulation, optics, and acoustics, that is, to the information potentially underlying the experience of phonetic perception.

Reference Note

1. Bastian, J. *Silent intervals as closure cues in the perception of stop phonemes*. Haskins Laboratories Quarterly Progress Report, 1959, No. 33, Appendix I.

References

Cooper, F. S., Delattre, P., Liberman, A. M., Borst, J. M., & Gerstman, L. J. Some experiments on the perception of synthetic speech sounds. *Journal of the Acoustical Society of America*, 1952, *24*, 597–606.

Delattre, P. C., Liberman, A. M., & Cooper, F. S. Acoustic loci and transitional cues for consonants. *Journal of the Acoustical Society of America*, 1955, *27*, 769–773.

Divenyi, P. L., & Danner, W. F. Discrimination of time intervals marked by brief acoustic pulses of various intensities and spectra. *Perception & Psychophysics*, 1977, *21*, 125–142.

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.

Erickson, D. M., Fitch, H. L., Halwes, T. G., & Liberman, A. M. Trading relation in perception between silence and spectrum. *Journal of the Acoustical Society of America*, 1977, *61*, S46(A).

Fant, G. *Acoustic theory of speech production*. The Hague, The Netherlands: Mouton, 1960.

Fant, G. *Speech sounds and features*. Cambridge, Mass: M.I.T. Press, 1973.

Finney, D. J. *Probit analysis*. New York: Cambridge University Press, 1971.

Fischer-Jørgensen, E. Sound duration and place of articulation. *Zeitschrift fur Sprachwissenchaft und Kommunicationsforschung*, 1964, *17*, 175–207.

Fourcin, A. J. Speech perception in the absence of speech productive ability. In N. O'Connor (Ed.), *Language, cognitive deficits, and retardation*. London: Butterworth, 1974.

Fowler, C. A., Rubin, P., Remez, R. E., & Turvey, M. T. Implications for speech production of a general theory of action. In B. Butterworth (Ed.), *Language production*. London: Academic Press, 1980.

Klatt, D. H. The duration of [s] in English words. *Journal of Speech and Hearing Research*, 1974, *17*, 51–63.

Klatt, D. H. Voice onset time, frication and aspiration in word-initial consonant clusters. *Journal of Speech and Hearing Research*, 1975, *18*, 686–706.

Kozhevnikov, V. A., & Chistovich, L. A. *Speech, articulation, and perception*. Washington, D.C.: Clearinghouse for Federal Scientific and Technical Information, 1965. (JPRS: 30, 543)

Kuhn, G. M. On the front cavity resonance and its possible role in speech perception. *Journal of the Acoustical Society of America*, 1975, *58*, 428–433.

Lappin, D. S., Bell, H. H., Harm, O. J., & Kottas, B. On the relation between time and space in the visual discrimination of velocity. *Journal of Experimental Psychology: Human Perception and Performance*, 1975, *1*, 383–394.

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.

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

Malecot, A., & Chermak, A. Place cues for /ptk/ in lower cut-off frequency shifts for contiguous /s/. *Language & Speech*, 1966, *9*, 162–169.

Mattingly, I. G., & Liberman, A. M. The speech code and the physiology of language. In K. N. Leibovic (Ed.), *Information processing in the nervous system*. New York: Springer-Verlag, 1969.

Peterson, G. E., & Barney, H. L. Control methods used in a study of the vowels. *Journal of the Acoustical Society of America*, 1952, *24*, 175–184.

Pisoni, D. B., & Sawusch, J. R. Some stages of processing in speech perception. In A. Cohen & S. Nooteboom (Eds.), *Structure and process in speech perception*. Heidelberg, West Germany: Springer-Verlag, 1975.

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.

Shankweiler, D. P., Strange, W., & Verbrugge, R. R. Speech and the problem of perceptual constancy. In R. Shaw & J. Bransford (Eds.), *Perceiving, acting and knowing: Towards an ecological psychology*. Hillsdale, N.J.: Erlbaum, 1977.

Scharf, B. Critical bands. In J. V. Tobias (Ed.), *Foundations of modern auditory theory* (Vol. 1). New York: Academic Press, 1970.

Stevens, K. N. The potential role of property detectors in the perception of consonants. In G. Fant & M. A. A. Tatham (Eds.), *Auditory analysis and the perception of speech*. New York: Academic Press, 1975.

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., & House, A. S. Studies of formant transitions using a vocal tract analogue. *Journal of the Acoustical Society of America*, 1956, *28*, 578–585.

Stevens, K. N., & House, A. S. Speech perception. In J. V. Tobias (Ed.), *Foundations of modern auditory theory* (Vol. 2). New York: Academic Press, 1972.

Strange, W., Verbrugge, R. R., Shankweiler, D. P., & Edman, T. R. Consonantal environment specifies vowel identity. *Journal of the Acoustical Society of America*, 1976, *60*, 213–221.

Summerfield, A. Q., & Haggard, M. P. On the dissociation of spectral and temporal cues to the voicing distinction in initial stops. *Journal of the Acoustical Society of America*, 1977, *62*, 435–453.

Weisstein, N. Beyond the yellow-Volkswagen detector and the grandmother cell: A general strategy for the exploration of operations in human pattern recognition. In R. L. Solso (Ed.), *Contemporary issues in cognitive psychology*. Washington, D. C.: V. H. Winston, 1973.

Winer, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1971.

Zue, V. W. Acoustic characteristics of stop consonants: A controlled study (*Lincoln Laboratory Technical Report No. 523*). Cambridge, Mass.: M.I.T. Press, 1976.

Received May 9, 1979 ■