

Perceptual equivalence of two acoustic cues for stop-consonant manner

HOLLIS L. FITCH, TERRY HALWES, DONNA M. ERICKSON,
and ALVIN M. LIBERMAN

Haskins Laboratories, New Haven, Connecticut 06510

The first experiment examined the effects of orthogonal variation in two acoustic cues that are both produced by the articulation of a stop consonant. One component of the articulation produces a temporal cue, the other a spectral cue. In a phonetic identification task, these quite different cues were found to engage in a trading relation: One cue could, within limits, substitute for the other. In a second experiment, the perceptual equivalence implied by that trading relation was put to a stricter test. If the cues are truly equivalent in perception, then they should have their effects on the same perceptual dimension. That being so, it should be possible to combine them in such a way that, working at cross purposes in the perceptual domain, they effectively cancel each other. Pairs of patterns so produced should, then, be harder to discriminate than pairs produced by either cue alone, and still harder than those produced when the (same) two cues are combined in the opposite way and so cooperate. That expectation was confirmed. We suggest that the equivalence thus demonstrated comes about because the two cues are processed by a system specialized to take account of their common origin in speech production. So interpreted, the equivalence may be viewed as an instance of distinctively phonetic perception.

In speech, the many-to-one relationship between stimulus and percept has two aspects: Several phonetic contrasts can be produced by the same acoustic cue; conversely, several acoustic cues can produce the same phonetic contrast. Examining the first aspect, one finds that the effect pervades all three phonetic dimensions. Thus, with all else constant, duration of (intersyllabic) silence, for example, can cue contrasts in manner (Haskins Laboratories, Note 1; Kuipers, Note 2), voicing (Lisker, 1957), and place (Port, 1976). As for the other aspect, the various acoustic cues for a particular contrast can be radically different. For example, an intervocalic voicing contrast in disyllables with the trochaic stress (*rapid* vs. *rabid*) can be cued, all else constant, by the duration of the intersyllabic silence or, alternatively, by the formant transitions at the end of the first syllable and the beginning of the next (Lisker, Note 3).

It is with the second aspect of the many-to-one relationship that this paper is concerned. Specifically, it examines two of the cues for the manner contrast

exemplified by the words *slit* and *split*. One cue is temporal (the duration of silence between the noise associated with the initial [s] and the vocalic portion of the syllable); the other is spectral (the presence or absence of appropriate formant transitions at the onset of the vocalic portion of the syllable). Our aim is to see whether two such different acoustic cues engage in a trading relation, and also to discover how they might be combined so as to cause their effects to summate or to cancel each other. Taken together, the results may reflect an equivalence that implies phonetic (as distinct from auditory) perception.

EXPERIMENT 1

The importance of silence as a cue for the perception of stop-consonant manner was shown in early studies by Bastian, Delattre, and Liberman (1959) and Bastian (Note 4). Starting with magnetic-tape recordings of the real speech utterance *sag*, the former investigators inserted snippets of blank tape (hence silence) between the noise associated with the initial fricative and the vocalic portion of the syllable. Listeners perceived *sag* or *stag*, depending on the duration of the (intrasyllabic) silent interval.

More recent experiments on silence as a cue for manner have been designed to test the hypothesis that its effects in speech are instances of distinctively phonetic perception. In the phonetic domain, silence would lead to perception of a stop consonant, not only because the ear hears, but also, and crucially, be-

This work was supported by NICHD Grant HD01994. Special thanks are due to A. Quentin Summerfield for valuable suggestions, including especially one that led to the procedure used in Experiment 2. We also thank Bruno Repp and David Isenberg for helpful comments on an earlier draft of this paper. Some of the work reported here was presented at the Spring 1977 meeting of the Acoustical Society of America, Pennsylvania State University, University Park, Pennsylvania, June 6-10, 1977. H. L. Fitch is also at University of Connecticut; A. M. Liberman is also at University of Connecticut and Yale University.

cause the silence specifies to an appropriately specialized perceptual system that the speaker closed his vocal tract, as he must when he produces a stop. Among the data relevant to that hypothesis are some that imply a trading relation—hence an equivalence in perception—between silence and various aspects of sound that are also related to the closing (and opening) gestures. An especially telling set of such data is owing to Bailey and Summerfield (Note 5). These investigators found that the amount of silence necessary to produce the stop consonant in fricative-stop-vowel syllables varied with another acoustic correlate of closure—namely, the onset frequency of the first formant. As the onset was lower, less silence was needed to hear the stop.

In the case just described—and in other cases that do not involve the silence cue—diverse acoustic events appear to sound alike. (For a review, see Liberman & Studdert-Kennedy, 1977.) One asks why.

A purely auditory explanation of this equivalence would rely on a description of the way the auditory system processes sound, regardless of whether or not the sound is perceived as speech. One such auditory process is the kind of energy integration that leads to time-intensity tradeoffs in the perception of very brief tones and light flashes. In processes of this kind, a longer light flash of lower intensity, for example, is indistinguishable from a shorter light flash that has just the increment of intensity required to keep the total amount of energy constant. This energy integration is presumably due to limits on the resolving power of the peripheral sensory apparatus. It is unlikely, however, that such energy integration could account for the equivalence of two components of signal which, as in the cases of speech perception we are concerned with, are themselves separately resolvable.

Another auditory process that might explain the perceptual equivalence of two different acoustic dimensions would be some kind of functional interaction, presumably at “higher” levels of the system. To the extent that such interactions might apply to all patterns, and not just to those that can be perceived as speech, they would be of an auditory, not a phonetic, sort. However, nothing presently known about the auditory system provides a basis in principle for the many interactions that must be assumed if we are to account for the various trading relations among the speech cues. It would appear, then, that we could only multiply the assumed interactions ad hoc, with the result that we would, in the end, have as many assumed auditory interactions as there are trading relations.

An alternative view of the trading relations, which we find more appealing, differs from any auditory account in that it is not neutral with regard to the events by which the acoustic cues are produced. The advantage of this view is that it provides a principle

that can be seen to underlie many different trading relations: Cues that are the common but distributed products of the same linguistically significant act will tend to trade. Consider again, for example, the equivalence between silence and the starting frequency of the first formant in the experiment by Bailey and Summerfield. As those investigators point out, the silence occurs as a consequence of the vocal-tract closure necessary for the stop, and the low-frequency onset of the first formant as a consequence of the subsequent opening. If we assume a perceiver sensitive to these cues as information about the source of phonetically significant acts, then silence and the starting frequency of the first formant might lead to the same percept because they specify the same phonetic act. On that interpretation, the trading relations we have described would hold only for sounds that were being processed as speech, and they would, accordingly, be reflections of phonetic perception.

The first of the two experiments to be reported in this paper is similar to that of Bailey and Summerfield, already described, in that it will examine a trading relation between silence and spectrum in the perception of stop-consonant manner in an initial fricative-stop cluster. Our aim is to carry out a further study of this kind of trading relation, but also—indeed, especially—to provide a basis for a second experiment in which we will find the results of combining the two cues in different ways, and thus achieve a stricter test of the perceptual equivalence that the trading relation implies.

Method

Our aim in constructing the stimuli was to have two cues to the perceived contrast *sllt* vs. *spllt*: temporal variation in silence and spectral variation in sound. To that end, we synthesized stimuli that consisted of an s-like noise, followed, after a variable amount of silence, by either of two vocalic syllables.

To synthesize the vocalic syllables, we first obtained approximate parameter values by making relevant measurements of the utterances [lit] and [pllt] as spoken by a male. Using these to set the parameter values of a serial resonance synthesizer (OVE III), we synthesized the two different (vocalic) syllables. We then modified these syllables so as to meet the following requirements: that they be the same except at their onsets, where there would be a spectral difference (in formant transitions), and that the spectral difference not be so great as to override the effects of the variable silent interval in cueing *sllt* vs. *spllt*. (It was also important, of course, that other cues to the contrast *sllt* vs. *spllt*, such as amplitude rise-time, be effectively neutralized so as not to override the effects of the two cues, silence and spectrum, we wished to study.) The syllables are shown in Figure 1. One sees that the frequency contours of the formant onsets were relatively flat for one of the patterns and somewhat rising for the other. These shapes closely approximated the natural tokens of [lit] and [pllt], respectively. Each vocalic section was 170 msec long.

To determine if the spectral difference alone would successfully bias those syllables towards [lit] and [pllt], we tested them, without any preceding s-like noise, in forced-choice identification trials. Each of five subjects heard a randomized series containing 30 instances of each of the two syllables, labeling each token as [lit] or [pllt]. The stimulus with the relatively flat formants was

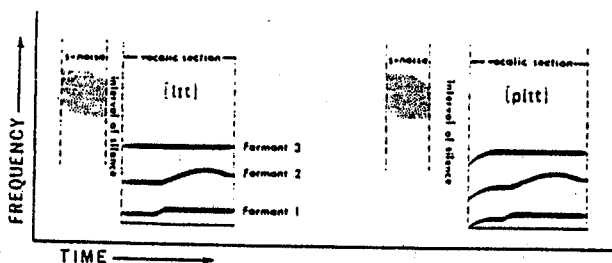


Figure 1. Schematic spectrograms of the stimulus patterns, showing two of the settings of the silent interval and both settings of the formant transitions at the onset of the vocalic portion.

heard as [plit] 57% of the time; the stimulus with the rising formants was heard as [plit] 97% of the time. For convenience, we will call the former syllable [lit], the latter [plit].²

To produce the full stimulus patterns, we generated a 96-msec patch of band-limited noise appropriate for [s] (referred to hereafter as 's'), placed it in front of the [lit] and [plit] vocalic syllables and varied the interval of silence between the 's' and the vocalic portion from 8 to 160 msec in steps of 8 msec, making a total of 20 stimuli in each series.

To produce a set of test stimuli that would enable us to determine the location of the phonetic boundary in each series—and thus to see the trading relation, if any, between duration of silence and formant transitions as cues for the distinction between *slit* and *split*—we recorded the stimuli onto a tape appropriate for presentation to listeners. That tape contained six randomizations of the full set of 40 stimuli, 20 from the 's' + [lit] series and 20 from the 's' + [plit] series. There was a 3-sec pause between items and a 10-sec pause between randomizations.

The experimental tape was presented once to each subject, with instructions to identify each stimulus as "slit" or "split" and to guess if necessary. It was played over a loudspeaker at a comfortable listening level. The subjects were 12 college students. All were native American English speakers and claimed to have good hearing in both ears.

Results

The variation in silent duration was effective in producing a perceived contrast between [slit] and [split], as is apparent in Figure 2. For both series ('s' + [lit] and 's' + [plit]), judgments shifted from [slit] to [split] as the silent interval increased.

More interesting is the displacement of the perceptual boundary between [slit] and [split] in the two series, for that reflects the trading relation between the temporal and the spectral cues that is the subject of this investigation. For the series 's' + [plit], the phonetic boundary (here defined as the point where the interpolated function crosses the 50% level) is at about 55 msec of silence, while for the series 's' + [lit] the boundary is at about 80 msec. Thus, it appears that about 25 msec less silence was required, on the average, to hear [split] when the formant transitions appropriate for [p] were present ('s' + [plit] series) than when they were absent ('s' + [lit] series). That finding defines a trading relation between the temporal cue and the spectral cue. Within the limits of that relation, these two very different acoustic cues appear to have equivalent effects in perception.

The results we have so far discussed were averaged across subjects. It is appropriate, therefore, to note that, although individual listeners differed in the absolute position of the phonetic boundary, every one of the 12 showed a boundary shift and in the same direction, though not necessarily by the same amount. The smallest shift shown by any subject was 8 msec; the largest was 40 msec.

We should note, further, that the magnitude of the phonetic boundary shift, whether for individuals or for the group, is presumably not fixed; settings of the formant transitions other than the two we selected would likely produce a smaller or larger shift, as would different settings of other relevant cues. These latter include, for example, the offset characteristics of the 's' (Bailey, Summerfield, & Dorman, Note 6). Moreover, as indicated in the method section, there are settings of the cues that cause one or more of them to override the others and thus preclude a trading relation. For our purposes, however, the important fact is that, within limits, the trading relation reported here does exist.

EXPERIMENT 2

The trading relation found in Experiment 1 implies a perceptual equivalence between two acoustic cues for the presence (or absence) of stop-consonant manner in the distinction between *slit* and *split*: One cue is silence, and its dimension of variation is temporal; the other is sound, and its dimension of variation is spectral. This perceptual equivalence is provocative, because the two cues that participate in it appear to have little in common from an acoustic or auditory point of view. As in the cases of equivalence referred to in the introduction to Experiment 1,

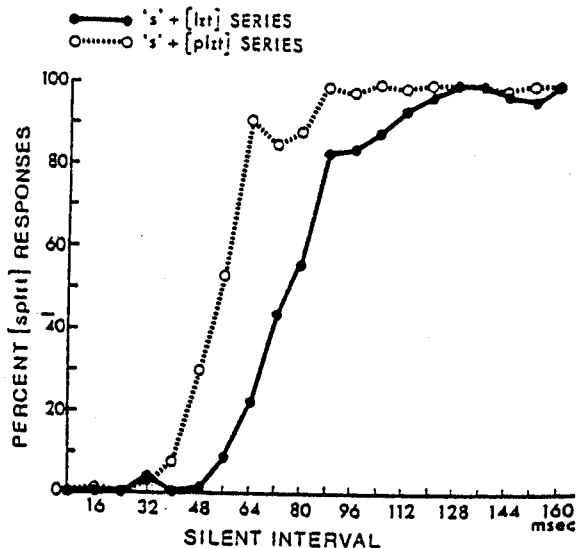


Figure 2. Effect of silent interval on the identification of the experimental syllables for each of the two stimulus series.

however, they are the common products of the same phonetically significant gesture. Perhaps, then, the equivalence is owing to distinctively phonetic processes sensitive not to duration of silence or direction of formant excursions per se, but rather to these as information that reflects what the speaker did. If so, the equivalence might stand as an instance of phonetic perception.

Perhaps the simplest and most direct meaning of the perceptual equivalence we are now considering is that, as diagrammed in the top half of Figure 3, the contrast between *slit* and *split* should sound the same whether it is made by the temporal cue (Pair 1) or the spectral cue (Pair 2). But there was, in Experiment 1, no provision for direct comparison of such pairs. Results from the phonetic labeling task alone do not preclude the possibility that the apparent equivalence is due to the fact that the listeners were limited to a forced choice between *slit* and *split*. A simple test of the discriminability of the two types of contrasts did not seem methodologically wise, however, because it would depend on producing an absolute negative result: inability to discriminate.

The second experiment, therefore, used a test of equivalence that not only provides for a reasonably direct assessment, as in the comparison just described, but also rests on an outcome that depends on differences in relative discriminability across several stimulus conditions rather than on utter indiscriminability in one. Basic to the development of such a test are the following considerations. If the silence

cue and the spectral cue are truly equivalent in phonetic perception, then they should have their effects on the same (perceived) phonetic dimension. In that case, it should be possible to arrange the two cues so that, as in Pair 4 of Figure 3, they effectively "neutralize" each other, producing two syllables that sound more alike than those that are produced, as in Pair 1 or Pair 2, by either cue alone. It should also be possible to arrange the same cues, as in Pair 3, so that they "cooperate," producing two syllables that are more discriminable than the syllables produced by either cue alone and, a fortiori, still more discriminable than the syllables produced when the cues are so arranged as to neutralize each other. Thus, the assumption of equivalence leads us to expect that the temporal and spectral cues can be added together so as to make pairs of stimuli that are either more or less discriminable than pairs that differ by one of the cues alone. The purpose of Experiment 2 is to provide a test of that expectation.

Method

Taking advantage of a particular procedure devised for us by Quentin Summerfield, we undertook to compare the discriminability of the stimuli of Experiment 1 in the three conditions of cue combination just described in the introduction. In one condition, the members of each pair of to-be-discriminated stimuli differed by only one of the two cues. We chose to make that the spectral cue. In the other two conditions, the members of each pair differed by both cues, temporal as well as spectral. For one of these latter conditions, the two cues were combined so as to "cooperate": One member of each pair had both cues biased toward *split*; the other had both cues biased toward *slit*. The anticipated effect of such cooperating cues would be to place

	DESCRIPTION OF STIMULI		PERCEPT	CHARACTERIZATION OF CUES			
	SILENT INTERVAL	VOCALIC PORTION		TEMPORAL	SPECTRAL	TEMPORAL	SPECTRAL
PAIR 1	"s"	short-- [ɪt] long -- [ɪt]	<i>slit</i> <i>split</i>	-p +p	-p -p	different	same
PAIR 2	"s"	short-- [ɪt] short-- [pɪt]	<i>slit</i> <i>split</i>	-p -p	-p +p	same	different
PAIR 3	"s"	short-- [ɪt] long -- [pɪt]	<i>slit</i> <i>split</i>	-p +p	-p +p	different	different
PAIR 4	"s"	short-- [pɪt] long -- [ɪt]	<i>split</i> <i>split</i>	-p +p	+p -p	different	different

Figure 3. Diagrams illustrating the phonetically equivalent effects of spectral and temporal cues and the phonetically different effects of combining these cues in two ways.

the percepts more clearly on opposite sides of the phonetic boundary and thus, given any region of uncertainty at that boundary, to increase discriminability by comparison with the pairs that differ by one cue alone. For the remaining condition in which the pairs differed by both cues, the arrangement of the cues was such as to put them in conflict: One member of each pair had one cue biased toward *split*, the other toward *slit*; in the other member of the pair, the cues (and their biases) were reversed. If these cues have their effect on the same perceived dimension, then this combination ought, by a kind of neutralization, to decrease discriminability by comparison with the condition in which the two cues cooperate, and even, indeed, by comparison with the condition in which the pairs differ by only one cue.

For the condition in which the pairs differed by only the spectral cue—let us call this the “one-cue condition”—the procedure was simple enough: The members of each pair had (of course) the same setting (duration) of the temporal cue, but the setting of that cue varied among the pairs. Thus, at one end of the range there was a pair comprising the stimuli (*'s' - 8 msec silence - [lit]*) vs. (*'s' - 8 msec silence - [plit]*); at the other end there was the pair (*'s' - 144 msec silence - [lit]*) vs. (*'s' - 144 msec silence - [plit]*). Between these two extremes were similar pairs for all intermediate settings of the temporal cue. (Since the stimuli with more than 144 msec of silence had all been perceived as *split*, they were, in the interest of economy, omitted.)

In both the other two conditions, where the members of each pair differed always by both cues, the procedure was somewhat more complex. Consider, first, the condition in which the cues were arranged so as to cooperate. Let us call this the “two-cooperating-cues condition.” As in the one-cue condition, each pair had one member made from [lit] and one member made from [plit], but in the two-cooperating-cues condition the [plit] member of each pair had a silent interval 24 msec longer than the [lit] member. Thus, at the one end of the continuum of silent intervals, the stimulus (*'s' - 8 msec silence - [lit]*) was paired with (*'s' - 32 msec silence - [plit]*), and similar pairings were arranged through the entire range of silent intervals up to the pair (*'s' - 120 msec silence - [lit]*) vs. (*'s' - 144 msec silence - [plit]*).

Consider, finally, the other condition in which both cues differed, but now in such a way as to be in conflict. Let us call this the “two-conflicting-cues condition.” Now the pairwise arrangement of the to-be-discriminated stimuli was here exactly the same as that for the two-cooperating-cues condition, except that, to produce the conflict, each [plit] member had 24 msec less silence than its [lit] companion. Thus, at the one end of the continuum of silent intervals, (*'s' - 32 msec silence - [lit]*) was paired with (*'s' - 8 msec silence - [plit]*), and similar pairings were made at all increasing values of the silent interval through the pair (*'s' - 144 msec silence - [lit]*) vs. (*'s' - 120 msec silence - [plit]*).

For the two-cooperating-cues and the two-conflicting-cues conditions, choosing the amount of silence by which the members of each pair differ is, of course, critical, if the experiment is to reveal most sensitively such differences in discriminability as there may be between the two conditions. That amount of silence would be equal to the amount by which the two perceptual identification functions—the one for stimuli made of [lit], the other for stimuli made of [plit]—are displaced (as in Figure 2 of Experiment 1), since that is the amount of silence that, according to identification judgments, just compensates for bilabial transitions. Ideally, the amount of silence would be adjusted appropriately for each subject. For experimental convenience, we did not make the adjustment for each subject, but rather used for all a single value, 24 msec, which is close to the average displacement obtained in Experiment 1.

An oddity test was used to measure discriminability of the members of these pairs. On each trial, one member of a pair was presented twice and the other member once. The listener was to determine which of the three stimuli was the odd one. For each pair of stimuli to be tested, six different oddity triplets—that is, all possible permutations—were generated. Each of these triplets occurred three times, yielding 24 presentations per pair for the subject to judge. There was 1 sec between successive syllables in

a trial, and 4 sec between trials. The full test was a random ordering of all the triplets in the experiment. The discrimination test was administered in four 1-h sessions.

At each of these four discrimination-testing sessions, identification functions like those in Experiment 1 were obtained. The stimuli were presented one at a time and in random order for judgment as “slit” or “split,” half at the beginning and half at the end of each session. The purpose of this part of the procedure was twofold: to see if the subjects in this experiment showed the same trading relation as found in Experiment 1, and to see if the trading relation was stable across the experimental sessions. The latter purpose was especially important, since it would not be proper to combine the discrimination data across the four experimental sessions if, for any reason, the trading relation that those data are supposed to test had itself undergone some change.

Five American-English-speaking college students with no known hearing deficit served as subjects.

Results

Part of this experiment—the identification of the stimuli as [slit] or [split]—was identical with Experiment 1, though with more repetitions and with different subjects. Figure 4 shows that the same result was obtained: Less silence was necessary to hear [split] when the transitions appropriate for [plit] were present. In the first experiment, the average difference was 25 msec; here it was 28 msec. As in Experiment 1, every subject showed a shift in the phoneme boundary and in the same direction. The smallest difference between *'s' + [lit]* and *'s' + [plit]* crossover points for any subject was 20 msec; the largest difference was 33 msec. No significant changes in the positions of these phonetic boundaries were noted across the four days of testing.

The results of the discrimination tests can be seen in Figure 5. Consider first the one-cue condition—that is, the condition in which the pairs of stimuli to be discriminated differed only in the spectral cue at the

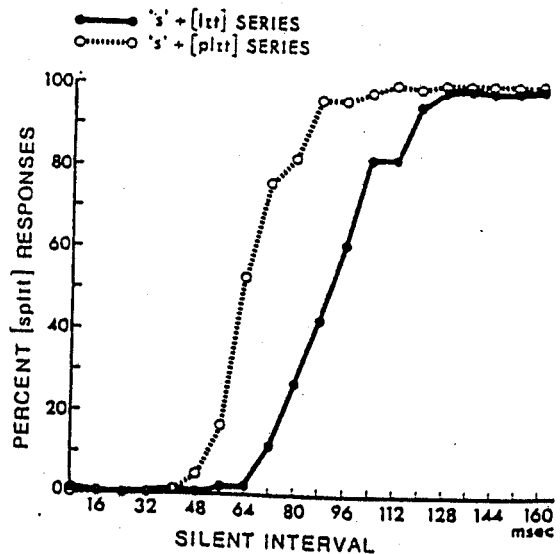


Figure 4. Effect of silent interval on the identification of the experimental syllables for each of the two stimulus series in Experiment 2.

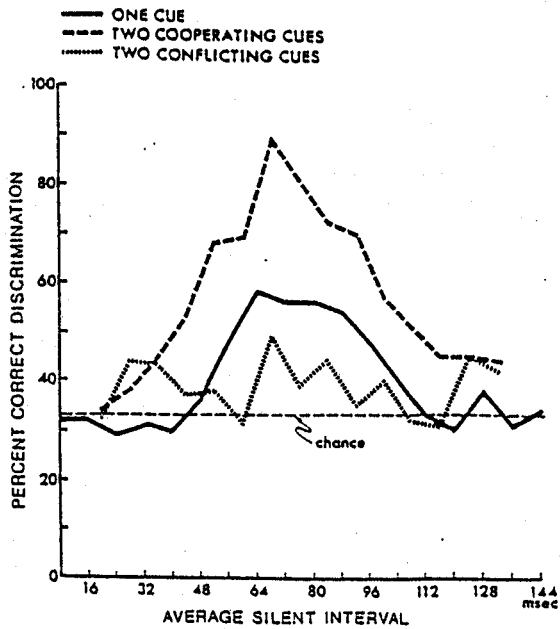


Figure 5. Percent correct discrimination for three types of stimulus pairs.

beginning of the vocalic section. Examining the appropriate data, which are shown as the solid line, it can be seen that discrimination is relatively low at the extremes and relatively high in the region of the phonetic boundary. Given the outcome of the identification test, this is what would have been expected: At very short intervals of silence, patterns made of 's' plus either [lɪt] or [plɪt] should have sounded like [slɪt]; at long intervals of silence both should have sounded like [splɪt]; and only in the region of the phonetic boundary should there have been a relatively high level of discrimination, for there the 's' plus [plɪt] would have begun to sound like [splɪt], but the 's' plus [lɪt] would not.

Taking the one-cue condition as the baseline, consider next the results obtained in the two conditions in which the patterns to be discriminated differed by two cues. It will be remembered that in one of these two-cue conditions—the one called two cooperating cues—the spectral and temporal cues were so arranged that, based on the identification data of Experiments 1 and 2, it was expected that they would reinforce each other, resulting in enhanced discriminability by comparison with the one-cue condition.³ The discriminability data for this two-cooperating-cues condition are shown as the dashed line in Figure 5. As in the one-cue condition, discrimination rises to a peak in the region of the phonetic boundary, but now the peak reaches a greater height. Thus, adding the temporal cue to the spectral cue in this condition made the pairs easier to discriminate.

In the other two-cue condition—the one called two conflicting cues—the spectral and temporal cues were

arranged so that, based on the identification data of Experiments 1 and 2, it was expected that they would neutralize each other and thus make discrimination difficult. The dotted line of Figure 5, which represents the relevant data, indicates that discrimination was, in fact, difficult—more difficult, apparently, than in the one-cue condition, and more difficult by far than in the other two-cue condition.

Thus, taking as the base the condition in which the stimuli to be discriminated differ only by the spectral cue, we find that it is possible to add a fixed difference in the temporal cue so as to increase discriminability or to decrease it.

Having seen the average results for the group, examine now the relative discriminability of the pairs in the one-cue, two-cooperating-cues, and two-conflicting-cues conditions for the individual subjects. Since the differences in these three conditions occurred primarily at the phoneme boundary (as we would expect), and since, although the position of the phoneme boundary varied slightly from subject to subject, it was always located near the middle of the range, the results from the middle third of the range are averaged. Those averages are presented in Table 1. For every subject, the order of difficulty, from easiest discrimination to most difficult, is: (1) the condition with two cooperating cues, (2) the condition with one cue, and (3) the condition with two conflicting cues. Thus the group results, plotted in Figure 5, accurately reflect the performance of each one of the subjects.

To compute the probability that this result could have been obtained by chance, we may take advantage of the fact that a particular ordering of the three conditions was expected. Since there are six possible orderings of three conditions, and only one of these was expected, then for any subject the exact likelihood of this particular ordering of results in the three conditions is one in six. The likelihood of obtaining this same ordering for each of five subjects is $(1/6)^5$, or once in 7,776 experiments.

Having seen the relative order of difficulty in discriminating the different types of syllable pairs, we should recall what this has to do with the claim that the spectral and temporal cues are equivalent when they converge on the perception of [p] in [splɪt]. For that purpose, consider the two-conflicting-cues condition. If the two cues are truly equivalent in phonetic perception, then it should be possible to arrange

Table 1
Percent Correct Discrimination Averaged Over the Middle Third of the Stimulus Series for Each Subject

Condition	Subject				
	1	2	3	4	5
Two Cooperating Cues	.77	.75	.81	.77	.70
One Cue	.56	.55	.56	.50	.46
Two Conflicting Cues	.40	.43	.42	.34	.38

Note—Chance is 33%.

them so that they effectively neutralize each other, producing two syllables—for example, [splɪt] and [spɪlt]—that sound exactly alike. But while that should have been possible in principle, in practice it would have required that a precise adjustment of the cues be made separately for each subject to take account of the individual differences in the exact value of the tradeoff between the cues. (Moreover, it would have required that each subject be perfectly consistent in the value for the tradeoff.) That adjustment was not attempted. Rather, a setting of the cues was used that was appropriate only on average. A result is that the patterns could not be expected to be perfectly indiscriminable, only less or more discriminable, depending on the arrangement and number of the cues.

It is possible, even likely, that the absolute level of discrimination in the two-conflicting-cues condition could have been raised by using a different kind of discrimination test or more highly practiced subjects. However, there is reason to believe that changes in procedure would not have altered the *ordering* of discriminability among the conditions. Indeed, the same ordering of results has now been obtained in a study (Best, 1979) similar to this one but using different syllables (*say* vs. *stay*) and a different kind of discrimination test (same-different). And it is this *relative ordering* among the conditions, rather than the absolute level of discriminability in any one condition, that is important to the inferences we would draw from our results.⁴

DISCUSSION

As pointed out in earlier sections, the several cues for a phonetic contrast are typically found, in perceptual studies, to engage in a trading relation: Given a phonetic contrast for which each of two cues is relevant, the effects of varying one cue can, within limits, be compensated for by appropriate variations in the other. Such trading relations are of interest if only because they imply an equivalence among aspects of stimulation that are often quite different from an acoustic point of view. They are the more interesting if we are right in supposing that the equivalence reflects a sensitivity to the common origin in articulation of the different acoustic cues—that is, to processes that are distinctively phonetic.

The trading relation observed in Experiment 1 is novel only in that it is another token of a type. Surely, this token is a striking one, for the contrast between the cues that trade is very great indeed: One of them is silence, the other is sound. But even that contrast is not entirely new, as the reader can see in the paper by Bailey and Summerfield (Note 5), cited earlier. There one can find experiments on trading relations different from ours and also considerably more comprehensive. But we need note here only that, like us,

Bailey and Summerfield studied the contrast between fricative-vowel and fricative-stop-vowel syllables; they employed temporal variations in silence and spectral variations in sound, just as we did; they found trading relations not different in principle from ours; and they offered an interpretation that is, in some important respects, similar to the one we favor.

But trading relations among acoustic cues, with the phonetic equivalence they imply, have been based on perceptual tests that only require a subject to attach phonetic labels. As mentioned in the introduction to Experiment 2, this leaves open the possibility that, for want of an alternative, subjects might sometimes have attached the same phonetic label to stimuli that were, in some peculiar way, as different as two stimuli to which they had found it possible to assign different labels. What is novel about this experiment is that we have subjected the claim of perceptual equivalence to a more rigorous test. The results of that test help to justify the inference about equivalence that had been made from the way the subjects assigned phonetic labels.

Now we would remark a possibly interesting by-product of Experiment 2: It may provide a test of phonetic perception that can be applied to nonhuman animals. Consider, again, the result with adult humans, which was that the two very different cues (silence and sound) could be so combined as to “neutralize” each other and thus produce pairs of syllables that are hard to discriminate—harder than pairs in which the same two cues are made to augment each other, and harder even than pairs that are distinguished by one of the two cues alone. That the cues have such effects can reasonably be taken to mean that, in the proper phonetic context, they are actually equivalent—that is, they do sound alike. But if we are right in supposing that the syllables with the different cues sound alike only to animals specialized to perceive their phonetic import, then to nonhuman animals they should sound quite different—as different, presumably, as they would to us when heard in a nonphonetic context. In that case, we should suppose that the two cues could not be made to neutralize each other. The result for the animals would then be that both of the pairs differing by the two cues would be easier than the pairs that differ by either cue alone. In any case, the outcome of the appropriate experiments could be straightforward and telling. What we should have to look for is only a difference in the order of discriminability. Moreover, the result to be expected is that the pair that is relatively the hardest for the human beings to discriminate would be one of the two easiest for the nonhuman animals. Such a result, if obtained, could not then be attributed to inattention or lack of motivation.

We realize that such an experiment might prove to be difficult in practice, if only because it requires a demonstration, before the critical test, that the ani-

mals are able to discriminate each of the two cues taken singly. Simpler syllables, such as the *say* vs. *stay* syllables used by Best (1979), may be easier than these to use with animals. In any case, all cues that engage in trading relations are candidates.

If, in the event, animals do show the different order of relative discriminability that we rather expect, then it would, of course, be of interest to apply the same test to human infants. Obviously, the considerations that make it a good test for animals would apply equally to human infants, though, just as obviously, it might be difficult with human infants to find cues that are discriminable yet appropriate. But if such cues can be found, then the kind of test we have proposed might provide a useful way to reveal (and study) an important biological predisposition to language.

At all events, the experiments reported here have isolated two very different acoustic cues that engage in a trading relation and are, in an important sense, perceptually equivalent. We have suggested that this equivalence is due to the fact that the cues are the common products of a single phonetically significant act, and are perceived by a system specialized to take account of that fact. If that is so, then a test for perceptual equivalence of two such different cues may provide an interesting basis for comparative studies among adult humans, preverbal infants, and nonhuman animals.

REFERENCE NOTES

1. Haskins Laboratories. *Affricates: Duration cues for the perception of c in intervocalic position* (Quarterly Progress Report No. 12). New York: Haskins Laboratories, 1954.
2. [Kuipers, A.]. *Affricates in intervocalic position* (Quarterly Progress Report No. 15, Appendix 6). New York: Haskins Laboratories, 1955.
3. LISKER, L. *Closure duration, first-formant transitions, and the voiced-voiceless contrast of intervocalic stops* (Quarterly Progress Report No. 23, Appendix 1). New York: Haskins Laboratories, 1957.
4. Bastian, J. *Silent intervals as closure cues in the perception of stop phonemes* (Quarterly Progress Report No. 33, Appendix 1). New York: Haskins Laboratories, 1959.
5. BAILEY, P. J., & SUMMERFIELD, A. Q. *Some observations on the perception of [s] + stop clusters* (Haskins Laboratories Status Report on Speech Research, SR-53, 2, 25-60). New Haven: Haskins Laboratories, 1978.
6. Bailey, P. J., Summerfield, A. Q., & Dorman, M. F. *Personal communication*, 1977.
7. Best, C. T. *Personal communication*, 1979.

REFERENCES

- BASTIAN, J., DELATRE, P., & LIBERMAN, A. M. Silent interval as a cue for the distinction between stops and semi-vowels in medial position. *Journal of the Acoustical Society of America*, 1959, 31, 1568 (A).

- BEST, C. T., MORRONGIELLO, B., & ROBSON, R. Perceptual equivalence of cues for a phonetic trading relation: Primacy of phonetic over psychoacoustic effects. *Journal of the Acoustical Society of America*, 1979, 66, S50 (A).
- LIBERMAN, A. M., & STUDDERT-KENNEDY, M. Phonetic perception. In R. Held, H. Leibowitz, & H.-L. Teuber (Eds.), *Handbook of sensory physiology* (Vol. VIII). Heidelberg: Springer-Verlag, 1977.
- LISKER, L. Closure duration and the intervocalic voiced-voiceless distinction in English. *Language*, 1957, 33, 42-49.
- PORT, R. Influence of tempo on the closure interval cue to the voicing and place of intervocalic stops. *Journal of the Acoustical Society of America*, 1976, 59, S41-42 (A).

NOTES

1. In addition to the intended difference in the frequency patterns of the formant transitions, there was a slight amplitude difference between the two syllables due to the particular characteristics of the OVE III synthesizer. The [plit] syllable was .8 dB less than the [lit] syllable on the first pitch pulse, and they were of equal amplitude from the second pitch pulse on. (If the difference in rise-time had an effect, it would be to bias the stimuli *against* the trading relation we expected to find.)
2. Although this experiment concerns cues for a manner contrast and not for place, one may nevertheless wonder why the listeners heard *split* rather than *sklit* or *stlit*. Perhaps the most important reason is that the acoustic cues in the vocalic section are reasonably appropriate for [p], even in the pattern that was minimally biased in that direction. Of course, we cannot exclude the possibility that listeners might also have been biased toward *split* because, unlike *stlit*, it is phonotactically regular and, unlike either *stlit* or *sklit*, it is a word.
3. On the assumption that perception of the speech patterns was categorical, or nearly so, one may ask how discriminability could be increased by adding another acoustic difference. The answer lies in the fact that the identification functions have sufficiently shallow slopes and are sufficiently close together that the members of the one-cue condition pairs never fail unambiguously on opposite sides of the phonetic boundary. Thus, as can be seen in Figure 4, at 64 msec of silence the [lit] token is identified 100% of the time as [silit], but the [plit] token is identified as [split] only 62% of the time. Likewise, at 112 msec of silence, when the [plit] token has reached 100% identification as [split], the [lit] token is identified as [silit] only 19% of the time. It follows that, even if perception of the patterns were categorical, the one-cue condition would not be expected, for any of the pairs used in the experiment, to produce discrimination at the level of 100%. In the two-cooperating-cues condition, on the other hand, some of the pairs in the middle of the series comprised stimuli that the subject had consistently put into different phonetic categories; for such pairs, we should expect that discrimination would be enhanced.
4. There is also some evidence that the ordering of discriminability among the conditions *does* change, depending on whether the stimuli are heard as speech. In an experiment employing sine-wave analogs to [se] and [ste], listeners who, in fact, heard those syllables (whether spontaneously or through instruction) produced the "phonetic" ordering; listeners who heard the stimuli as non-speech noises did not (Best, Note 7).