PHONETIC AND AUDITORY TRADING RELATIONS BETWEEN ACOUSTIC CUES IN SPEECH PERCEPTION: FURTHER RESULTS*

Bruno H. Repp

Abstract. The series of studies begun by Repp (1981), with the purpose of examining whether trading relations between acoustic cues are obtained within phonetic categories, is continued with three experiments. Despite some unexpected complexities, the results tend to support the hypothesis that the trading relations studied are a consequence of phonetic categorization.

Whenever two or more acoustic cues contribute to the perception of a phonetic distinction, a trading relation among the cues can be demonstrated in categorization, given that the speech stimuli are phonetically ambiguous. That is, a change in one cue can be compensated for by a change in another cue, so as to maintain the same degree of perceptual ambiguity. In a previous paper (Repp, 1981) I asked whether cues would continue to engage in trading relations when the stimuli are phonetically unambiguous. An affirmative answer to this question would mean that the trading relation examined is either psychoacoustic in origin or that it derives from a phonetic mode of processing that extends beyond the mere assignment of category labels. A negative answer, on the other hand, would imply that the trading relation is either tied to phonetic categorization or that it is a psychoacoustic phenomenon specifically limited to the phonetic boundary region. Thus, while these answers do not distinguish between all possible hypotheses, they usefully restrict the set of alternatives. Further arguments and experimental evidence may then be adduced to arrive at the most likely explanation for a given trading relation.

Phonetic classification of unambiguous stimuli evidently does not yield the kind of information sought. In my earlier experiments, I employed instead a fixed-standard same-different discrimination paradigm with stimuli that either straddled a phonetic category boundary or came from within a phonetic category. Four different trading relations were examined. One of them, suspected to be of psychoacoustic origin, held up regardless of phonetic ambiguity; two others, suspected to be byproducts of phonetic categorization, disappeared for within-category stimulus comparisons; the results of the fourth experiment were inconclusive. The three experiments to be reported in the present paper supplement and extend my earlier research using exactly the same methodology.

Acknowledgment. This research was supported by NICHD Grant HD01994 and BRS Grant RR05596. I am grateful to Sharon Manuel for assistance in running the experiments.


121
GENERAL METHOD

A graphic illustration of the paradigm in the form of a geometric analogy is provided in Figure 1. The two acoustic cues whose trade-off is to be investigated are depicted here as the height and width of rectangles. The dimension resulting from the perceptual integration of the two cues, analogous to the phonetic percept (though without any clearly defined category boundary), is the area of the rectangles, a measure of which (in arbitrary units) is given by the numbers in Figure 1. The subjects' task is to discriminate a standard, which occurs first in each stimulus pair, from a limited set of alternative stimuli. A series of practice trials is presented first, with subjects having foreknowledge of the correct responses. Half the stimulus pairs are "same" trials in which the standard is paired with itself; the other half are "different" trials in which the standard is followed by a stimulus that differs in one (the "primary") cue dimension (height in Figure 1) by a fairly large amount. Three blocks of test trials follow. In each of these, there are three types of trials occurring with equal frequency: "same" trials, 1-cue "different" trials in which the difference is only in the primary cue, and 2-cue "different" trials in which the comparison stimulus differs from the standard on both cue dimensions. The difference in the second (the "secondary") cue dimension (width in Figure 1) is fairly small and chosen so as to counteract the difference in the primary cue with respect to the integrated percept; thus, in Figure 1, increased height is coupled with reduced width. The size of the primary cue difference (height) decreases across the three test blocks, whereas the secondary cue difference (width) remains constant.

![COMPARISON STIMULI ON "DIFFERENT" TRIALS](image)

**Figure 1.** Schematic diagram of the experimental paradigm.
If listeners discriminate the stimuli on the basis of an integrated property derived from both cues (area), then the prediction is that, paradoxically, 1-cue differences should be easier to detect than 2-cue differences: In Figure 1, the standard-comparison difference in area is larger on 1-cue than on 2-cue trials. If, however, subjects do not integrate the two cues and instead either focus on a single cue or divide attention between two separable cue dimensions, then there should either be no difference between 1-cue and 2-cue trials (if only the primary cue is attended to), or 2-cue trials should yield higher detection scores than 1-cue trials. In the latter case, a divided-attention strategy may be distinguished from a secondary-cue focus by gauging the extent of the advantage for 2-cue trials and the extent of the decline in 2-cue discrimination performance over test blocks.

Each experiment has two conditions, a between-category (Between) and a within-category (Within) condition. Each condition includes the complete paradigm shown in Figure 1; the difference lies solely in the values chosen for the primary cue dimension. In the Between condition, they are chosen so that the standard stimulus is close to a phonetic boundary and the comparison stimuli tend to fall even closer to, or on the opposite side of, the boundary. This enables listeners to make use of phonetic category distinctions and thus encourages the phonetic strategy of deriving a single integrated percept from the two cue dimensions and of basing same-different judgments on a comparison of these percepts (i.e., categorical perception). This condition should yield the expected phonetic trading relation (revealed as a superiority of 1-cue over 2-cue trials) and thus serves as a control. In the Within condition, the primary cue values are chosen so that all stimuli fall well within a phonetic category. Here, listeners presumably can no longer make phonetic distinctions and have to rely on perceived auditory differences between the stimuli. The critical result is the relative performance on 1-cue and 2-cue trials. If this relation is significantly different from that observed in the Between condition, the conclusion is warranted that a different (presumably nonphonetic) perceptual strategy was used in within-category discrimination. It should be noted that, although the clearest result would be 1-cue superiority in the Between condition and 2-cue superiority in the Within condition, a significant change in the 1-cue versus 2-cue relation across conditions (i.e., a significant Cues by Conditions interaction in an analysis of variance) is sufficient to permit conclusions about differing perceptual strategies. The results may not always be ideal because, as in many other tasks concerned with categorical perception of speech, phonetic and auditory strategies may be used simultaneously in varying degrees, particularly in "between-category" discrimination. (See Repp, 1981, for presumed instances in the present paradigm.)

The experimental setup in the present experiments differed from that of my earlier studies in several minor respects. First, the number of test trials was increased by one-sixth to 84 per block. Second, the number of practice trials was reduced to 28, and instead of following a random sequence, they alternated between "same" and "different." As before, during practice the subjects checked off the correct response printed on the answer sheet. Third, a change in the direction of primary-cue differences was introduced in parts of Experiments 2 and 3 and is described later. Fourth, more extensive identification data were collected than in the earlier studies. These data were always obtained after the discrimination tasks or in a separate session (or from different subjects altogether), to avoid biasing the listeners too
strongly toward use of a phonetic strategy. On the other hand, the Between condition always preceded the more difficult Within condition, to permit subjects to get used to the stimuli and to the task. This, finally, constituted another change from my earlier studies, in which the Within condition was presented twice, both before and after the Between condition. Since there were no significant differences between these two presentations in any of the four previous experiments, the present use of a single run following the Between condition was fully justified, even though the total number of responses obtained was thereby reduced.

EXPERIMENT 1: "SAY"-"STAY"

The purpose of this study was to supplement my earlier Experiment 1, which was concerned with the stop manner distinction in "say" versus "stay." This distinction is of special interest because Best, Morrongiello, and Robson (1981) have reported results that suggest a phonetic basis for the trading relation between the two cues of silent closure duration and first-formant (F1) onset frequency. In my earlier study, I employed stimuli composed of a natural-speech "s" noise followed by a variable amount of silence (the primary cue) and one of two synthetic vocalic portions differing in F1 onset (the secondary cue). The results were encouraging but statistically weak, due to high variability (an aspect of the data that was also encountered in the present experiments, unfortunately). Although the expected trading relation was apparent both in the Between condition (as 1-cue superiority) and in the post-discrimination labeling data, it did not reach significance in either set of data. However, there was a significant 1-cue superiority in the Within condition, and a significant Cue-by-Conditions interaction confirming the reversal. Clearly, then, the phonetic trading relation was absent when the subjects could not draw any category distinctions, which supported the conclusion of Best et al. (1981) that the trading relation may be specific to phonetic perception.

The weakness of the phonetic trading relation in the earlier Between condition may have been due to a mixture of phonetic and auditory strategies in discrimination; however, the similar weakness in the labeling data cannot be so explained. Rather, it suggests that the stimulus materials were not optimal. The original purpose of the present study was to provide a replication with improved stimuli. All-natural stimuli were envisioned for that purpose. Since F1 onset frequency is difficult to manipulate directly in natural speech, it was planned to take vocalic portions from utterances of "say" and "stay," which were thought to contain the required difference in F1. Pilot tests (of a limited nature, to be sure) suggested, however, that the two vocalic portions—the particular tokens used, in any case—had no differential effect in perception and did not generate any trading relation. Although I could have extended my efforts at finding stimuli that "worked," I decided instead to vary a different, but equally relevant, secondary cue: the release burst that occurs immediately following the closure in "stay" but is absent in "say."

Method

The utterances "say" and "stay" were recorded by a female speaker and were digitized at 20 kHz. In order not to bias perception too strongly toward
"stay," the fricative noise portion of "say" was employed in the experimental stimuli. However, to counteract a possible bias in the opposite direction, the final low-amplitude portion was trimmed off, leaving a noise waveform of 157 msec duration. The experimental stimuli were created by following this noise with a variable silent interval and one of two waveforms derived from the 400-msec post-closure portion of the "stay" utterance. Originally, this "day" portion began with a powerful release burst of approximately 25 msec duration, more than sufficient to cue perception of "stay" even when immediately preceded by an "s" noise without closure silence (Repp, 1982). To obtain stimuli that would permit perception of "say" in the same situation, the onset of the "day" portion was cut back by 20 and 29 msec, respectively, resulting in stimuli that, in analogy to Best et al. (1981), may be called strong "day" and weak "day" (relatively speaking). The strong "day" retained the last 4 msec of the release burst, which were of rather low amplitude. In the weak "day," this residual burst was eliminated together with the first 5-msec pitch period, which was of very low amplitude and was overlaid with some aspiration noise. Essentially, then, the strong and weak "day" differed in the presence versus absence of a residual release burst at onset.

In the Between condition, the fixed standard consisted of the "s" noise immediately followed by the strong "day"—a stimulus expected to be perceived as "say." The comparison stimuli in the three test blocks had silent closure intervals of 40, 30, and 20 msec, respectively. In the Within condition, the standard, which again contained the strong "day" portion, had a closure interval of 40 msec (expected to lead to the perception of "stay"), and the comparison stimuli had silences of 100, 80, and 60 msec. A separate identification tape contained ten random sequences of 14 stimuli generated by following the "s" noise with either the strong or the weak "day," separated by silent intervals ranging from 0 to 60 msec in 10-msec steps. The subjects were nine paid volunteers, mostly Yale undergraduates. For details of method not mentioned here, the reader is referred to Repp (1981).

Results and Discussion

Figure 2 displays the average post-discrimination identification results. Percent "stay" responses is shown as a function of silence duration. It is evident that the stimuli containing the strong "day" portion generated an orderly labeling function, with the category boundary at 25 msec of silence. The stimuli that served as standards in the discrimination task, with 0 and 40 msec of silence, received 2 and 91 percent "stay" responses, respectively, which confirms that they had been appropriately chosen as instances of "say" and "stay." The labeling function for the stimuli containing the weak "day," however, was unexpectedly gradual, reaching not even 50 percent "stay" responses at the longest silence. (Only two of the nine subjects reached 100 percent "stay" responses.) This was surprising, for exactly the same stimuli had been used in another study (Repp, 1982) where many more "stay" responses were obtained. The resulting exaggerated trading relation (if it still can be called that) between the silence and release burst cues has implications for the discrimination tasks: On one hand, an especially clear trading relation should emerge in the Between condition; on the other hand, the failure of the weak "day" stimuli to reach 100 percent "stay" responses (presumably even at silences longer than 60 msec, judging from Figure 2) gave subjects an unexpected opportunity to detect phonetic distinctions in the Within condition.
Figure 2. Identification results of Experiment 1.

Figure 3. Discrimination results of Experiment 1.
as well. Here, however, a phonetic strategy should lead to higher scores on 2-cue than on 1-cue trials. (Consider the 40-msec strong "day" standard and the two 60-msec comparison stimuli in Figure 2.) Therefore, a reversal in the relation between 1-cue and 2-cue discrimination scores is predicted on phonetic grounds alone, which complicates (but still permits) an interpretation of the discrimination results.

These results are shown in Figure 3 as d' scores (heavy lines). The pattern is very clear: In the Between condition there is a large advantage for 1-cue trials, \( F(1,8) = 31.7, p < .001 \), while, in the Within condition, there is a strong trend in the opposite direction that, however, failed to reach significance, \( F(1,8) = 3.3 \). The Cue-by-Conditions interaction is highly significant, \( F(1,8) = 25.2, p < .002 \). In addition, performance declined across test blocks, \( F(2,16) = 24.5, p < .001 \), except for blocks 2 and 3 in the Within condition, where scores remained constant.

The results of the Between condition confirm the expected trading relation and bolster the somewhat weak results obtained in the same condition of the earlier "say"-"stay" study (Repp, 1981). The thin lines in Figure 3 indicate the results expected if subjects had relied on phonetic labels alone. These expected d' values were derived after predicting individual hit and false alarm rates according to the classic "Maskin's model" of categorical perception. It can be seen that performance was a good deal better than predicted; this may be attributed to anchoring or contrast effects due to the fixed standard (Repp, Healy, & Crowder, 1979). The smaller gain for 1-cue trials may be attributed to a ceiling effect (d'max = 4.64). Thus, the data are consistent with the hypothesis that, in the Between condition, subjects relied primarily on phonetic labels in discriminating the stimuli. They are also consistent, however, with the alternate hypothesis that a psychoacoustic trading relation localized in the phonetic boundary region is responsible for the effects seen.

The results of the Within condition are less straightforward. Predicted d' values were computed for the last test block and are shown in Figure 3. It can be seen that performance on 1-cue trials was better than predicted (predicted d' was near zero) while performance on 2-cue trials was worse than predicted. As a result, the obtained difference between 1-cue and 2-cue discrimination was smaller than predicted. If the assumption is accepted that subjects used primarily a phonetic strategy even in the Within condition, the depressed scores on 2-cue trials may indicate that a psychoacoustic trading relation favoring 1-cue trials (as in the Between condition) counteracted the trends generated by the phonetic strategy. That purely auditory discrimination played an additional role is clear, at the very least, from the elevated scores in the first test block; note that the predicted scores must be lower in the first than in the last test block, as indicated by the arrow in Figure 3. (This can easily be verified with the aid of Figure 2.)

In the hope of clarifying the situation, the Within-condition results of individual subjects were inspected. All of the five subjects who gave very few "stay" responses to the weak "day" stimuli showed the predicted 2-cue superiority. So did, however, one of the two subjects whose "stay" responses reached 100 percent at or before the 60-msec silence duration (and whose predicted scores were, therefore, zero throughout) and one of two subjects
whose labeling results indicated that 100 percent "stay" responses might have been reached somewhere beyond 60 msec. These results suggest the use of an auditory strategy favoring 2-cue trials, which implies that there was no psychoacoustic trading relation favoring 1-cue trials. On the other hand, one of the two subjects with reasonable labeling scores showed (as the only subject) a substantial advantage for 1-cue trials in the Within condition. The other one of the two subjects with excellent labeling scores performed near chance throughout (as predicted), which suggests that he was a strictly categorical perceiver and failed to make any use of auditory information.

In summary, the results of the present study, while not crystal-clear, do lend some support to the phonetic/localized-psychoacoustic pair of hypotheses; they tend not to favor the generalized-phonetic/psychoacoustic pair. Within the favored pair, the distinction rests on whether the postulated psychoacoustic interaction and its specific location can be supported by independent arguments or evidence. At present, such evidence is in short supply; however, some negative arguments will be presented in the General Discussion.

**EXPERIMENT 2: "SLIT"-"SPLIT"

All the experiments up to now (including the four studies in Repp, 1981, and the present Experiment 1) had in common that the primary cue was temporal in nature, and that the Within condition used longer values on that temporal dimension than the Between condition. This was so out of necessity, since the category boundaries were located at relatively short durations of the temporal cue and did not leave sufficient "room" for a full discrimination paradigm (Figure 1) at the short end of the continuum. Also, to the extent that the boundary coincided with a psychoacoustic threshold of some sort (cf. Miller, Wier, Pastore, Kelly, & Dooling, 1976; Pastore, Ahronon, Baffuto, Friedman, Pulso, & Fink, 1977; Pisoni, 1977), one might have expected discrimination to be at chance below that threshold, i.e., at the very short end of the continuum. Nevertheless, it became increasingly evident that an application of the present paradigm to the short end of a temporal dimension might be a desirable strategy to pursue. After all, few psychoacousticians would be surprised by the finding that an interaction between cues occurring in the vicinity of some hypothesized threshold disappeared at long temporal separations of signal components: Temporal proximity may be a prerequisite for the interactions (be they masking or integration) that are thought to underly a trading relation. If so, however, then the psychoacoustic interaction should become even stronger when temporal separation is further reduced. On the other hand, if the stimuli with these short temporal values all fall in the same phonetic category, then the phonetic hypothesis would predict a disappearance of the trading relation. Moreover, finding that subjects can discriminate these stimuli at all would cast doubt on the hypothesis equating category boundaries with auditory thresholds.

To pursue this possibility, it is necessary to find a stimulus continuum on which the boundary is at somewhat longer durations of a temporal cue. The "slit"-"split" distinction seems to fit the bill. In a recent study by Fitch, Halves, Erickson, and Liberman (1980), the average boundary on a continuum of varying silent closure durations was somewhere between 50 and 80 msec, depending on the precise characteristics of the stimuli. This gives rise to
the hope of obtaining above-chance discrimination scores strictly within the 
"slit" category.

Experiment 2 was conducted in two parts. Part a included the Between 
condition and the Within ("slit") condition just described. Part b included 
the same conditions but with a different choice of standards, as described 
below, plus a second Within ("split") condition using long values of the 
temporal cue dimension.

Method

The stimuli were created in a similar way as those of Experiment 1. A 
female speaker recorded the utterance "split," which was digitized at 20 kHz. 
The pre-closure "a" noise, 141 msec in duration, was separated from the post-
closure "blit" portion, which consisted of an initial 15-msec low-amplitude 
release burst followed by a 230-msec voiced portion, a 137-msec "t" closure, 
and a final "t" release burst. Two versions were derived from this portion by 
waveform editing: a strong "blit" that retained the final 12 msec of the 
release burst, and a weak "blit" that had no release burst left.

In the Between condition of Part a, the standard had a closure silence of 
40 msec preceding the strong "blit." The comparison stimuli in successive 
test blocks had silences of 80, 70, and 60 msec. In the Within ("slit") 
condition of Part a, the standard had no silence preceding the strong "blit," 
while the comparisons had silences of 40, 30, and 20 msec. In the Within 
("split") condition of Part b, the standard had 140 msec of silence preceding 
the strong "blit," while the comparisons had silences of 200, 180, and 160 
msec. The Between and Within ("slit") conditions of Part b essentially 
reversed the standard and comparison stimuli of the corresponding conditions 
in Part a. In the Between condition, the standard initially had 80 msec of 
silence followed by the weak "blit," and the comparisons had 40 msec of 
silence. Over successive test blocks, the silence of the standard decreased 
from 80 to 70 to 60 msec, while that of the comparison remained constant. In 
the Within ("slit") condition, the silence in the standard decreased from 40 
to 30 to 20 msec (followed by the weak "blit"), while that of the comparison 
remained fixed at 0 msec. The reason for these changes will become apparent 
below.

The identification test included ten random sequences of 20 stimuli. 
Silences ranged from 30 to 120 msec in 10-msec steps; stimuli included either 
the weak or the strong "blit." The identification test was taken by nine 
subjects, only four of whom also took Part a of the discrimination tests. 
Eight of the nine paid volunteers in Part a had also been subjects in 
Experiment 1. Seven new subjects were run in Part b. The subjects in Part b 
listened to the Within ("split") tape at the end of the session.

Results and Discussion

The average results of the identification test are shown in Figure 4. 
They proved to be very orderly. The category boundaries were at 49 and 70 
msec for the strong and weak "blit," respectively. Note that the standards 
used in the Within "slit" (Part a) and "split" (Part b) conditions, with
Figure 4. Identification results of Experiment 2.

Figure 5. Discrimination results of Experiment 2, Part a.
silences of 0 and 140 msec, were unambiguous instances of "slit" and "split," respectively, as intended.

The average results of the discrimination tests of Part a are shown in Figure 5. The Within "slit" condition is shown in the left panel and the Between condition is shown in the right panel. In the Between condition, the expected trading relation was initially absent but emerged in the second and third test blocks, $F(2, 16) = 4.6, p < .05$, for the Cues-by-Blocks interaction; $F(1, 8) = 3.3, p > .05$, for the Cues main effect. The reason for this interaction is not known. The Within data are surprising in that they, too, reveal the trading relation in form of a consistent 1-cue superiority, $F(1, 8) = 8.1, p < .05$. The Conditions-by-Cues interaction was not significant, $F(1, 8) = 1.7$, indicating similar patterns of results in the two conditions. The overall advantage for 1-cue trials was significant, $F(1, 8) = 9.9, p < .02$, and so was, of course, the decrease in scores across test blocks, $F(2, 16) = 21.7, p < .001$. The performance level in the Within condition was remarkably high and similar to that in the Between condition of Experiment 1 (Figure 3, left panel), which had employed the same silence durations.

At first blush, these results look exactly like those expected if the trading relation had a purely psychoacoustic basis. However, the high performance level in the Within condition gives rise to suspicion. Indeed, the author's observations as a pilot subject suggest an alternative interpretation: It seems that the consistent presence of the 0-msec standard on every trial may have acted as an anchor that shifted the phonetic boundary toward rather short values, so that tokens with only 40, 30, and even 20 msec of silence began to sound like "split." If so, the trading relation evident in the Within condition may derive from phonetic perception, rather than from a psychoacoustic interaction. It was for this reason that Part b of the experiment was run. By using standards with silences closer to the boundary and different standards in each test block, it was hoped that anchoring effects might be reduced. The Within "split" condition was added to gather additional information comparable to that obtained in Experiment 1.

The results of Part b are shown in Figure 6. The conditions in the two panels on the left correspond to those in Figure 5. The change in standards had a quite dramatic effect. In the Between condition, performance was better than previously and exhibited a clear trading relation, $F(1, 6) = 8.0, p < .05$. Performance in the Within ("slit") condition, on the other hand, was much poorer than previously and showed no significant trading relation, $F(1, 6) = 1.2$. The poor performance suggests that the subjects could no longer rely on a phonetic criterion. Consequently, the absence of any trading relation may be interpreted as supporting the hypothesis that the trading relation in the Between condition had a phonetic, rather than psychoacoustic origin.

One possible objection to that conclusion, however, which cannot be dismissed at present, is that the secondary cue (the brief release burst at the onset of the strong "blit") was effectively masked by the preceding fricative noise in 0-msec silence stimuli. Since all comparison stimuli in the Within ("slit") condition were of that kind, the secondary cue may simply have had no opportunity to produce any perceptual effects, be they phonetic or psychoacoustic. This objection cannot be raised against the results of the
Within ("split") condition (right-hand panel in Figure 6), however, which strongly resemble those of the Within ("split") condition: Again, performance was very poor, and there was no difference at all between 1-cue and 2-cue trials. Thus it appears that subjects did not pay any attention to the secondary cue, unlike the Between condition, where that cue made a large difference (cf. also the labeling data in Figure 4). It seems possible that the lack of any secondary-cue effect in the Within ("split") condition was likewise due to lack of attention, although the possibility of masking remains.

**EXPERIMENT 3: "GA"-"KA"

In the final experiment of this series, another attempt was made to assess within-category discrimination at the short end of a temporal continuum. This time, I chose a voice-onset-time (VOT) continuum for stops with a velar place of articulation, whose phonetic boundary tends to lie at relatively long values of VOT (Lisker & Abramson, 1970). Since the secondary cue was to be the onset frequency of the F1 transition (cf. Lisker, Liberman, Erickson, Dechovitz, & Mandler, 1977; Summerfield & Haggard, 1977), I returned to synthetic stimuli.

**Method**

The stimuli were created on the Haskins Laboratories parallel resonance synthesizer. All stimuli were 250 msec in duration, had a linearly falling fundamental frequency contour and linear 50-msec formant transitions that, in the case of F2 and F3, went from 1764 to 1230 Hz and from 2025 to 2527 Hz, respectively. The primary cue varied was VOT, i.e., the duration of the initial aspiration phase during which F1 was turned off. The secondary cue was the linear F1 transition whose onset frequency, duration, and extent differed between two versions: In short-transition (high F1 onset) stimuli, F1 started at 407 Hz and reached 765 Hz after 50 msec; in long-transition (low F1 onset) stimuli, it started at 279 Hz and reached 765 Hz after 70 msec, given a VOT of 0 msec. At longer VOTs, F1 started at correspondingly higher values. The two F1 trajectories were chosen so as to have the same slope, making the magnitude of the secondary cue difference constant for different values of the primary cue (VOT).

Because Experiment 2 had revealed strong effects of the choice of standard, the present experimental tapes were immediately recorded in two versions. In the Between condition, version A, the standard had 20 msec of aspiration and a short F1 transition, and the comparison stimuli had VOTs of 40, 35, and 30 msec. In version B, the standard had 50 msec of aspiration and a long F1 transition (of which only the last 20 msec remained, of course), and the comparisons had VOTs of 30, 35, and 40 msec. In the Within ("ga") condition, version A, the standard had no aspiration and a short F1 transition, while the comparisons had VOTs of 20, 15, and 10 msec. In version B, the standard had 20 msec of aspiration and a long F1 transition, and the comparisons had VOTs of 0, 5, and 10 msec. Note that the B versions differed from the corresponding conditions in Experiment 2, Part b, in that the standards were held constant through all test blocks, while the comparisons
Figure 6. Discrimination results of Experiment 2, Part b.

Figure 7. Identification results of Experiment 3.
changed from block to block; this resulted in some differences in the precise VOT comparisons used in versions A and B. A Within ("ka") condition could not be included with these stimuli, for the F1 transition did not extend sufficiently into the "ka" category.

A separate identification test included 10 random sequences of long- and short-transition stimuli with VOTs ranging from 0 to 50 msec in 5-msec steps. Ten paid volunteers participated, four of whom had also taken Part b of Experiment 2. Five subjects took version A, and five took version B. The data of one additional subject were discarded because he apparently wrote "same" for "different" (and vice versa) during part of the experiment and responded randomly elsewhere.

Results and Discussion

Figure 7 shows the identification results. The expected trading relation was clearly present, with category boundaries at approximately 23 and 36 msec of VOT for high and low F1 onsets, respectively.

The results of the discrimination tests are shown in Figure 8. They are plotted separately for versions A (top panels) and B (bottom panels) of the tests, not only because the VOT comparisons were slightly different but also because one of the strongest effects in the overall analysis of variance was the Cues by Versions interaction, $F(1,8) = 25.9$, $p < .001$, which suggested that the relationship between scores for 1-cue and 2-cue trials changed across versions. No other interaction with Versions was significant. The overall analysis also revealed a highly significant Conditions by Cues interaction, $F(1,8) = 33.6$, $p < .001$, which indicates that the pattern of results was different for the Within and Between conditions.

Both these effects are evident in Figure 8. Overall, performance was better on 2-cue trials than on 1-cue trials in version A, while the opposite held in version B. Two-cue trials enjoyed a relative advantage in the Within condition, while 1-cue trials were favored in the Between condition. The last-mentioned finding, of course, is the expected phonetic trading relation; because of the strong Cues by Versions interaction, it was small and nonsignificant in version A but large and significant, $F(1,4) = 12.4$, $p < .05$, in version B. In the Within condition, on the other hand, there was a large 2-cue superiority in version A, $F(1,4) = 52.9$, $p < .01$, but no difference whatsoever in version B. Note also the unexpectedly high level of performance in the Within condition in both versions.

These data present some problems for interpretation, but they are quite clear on the main point: There was no sign of any trading relation in the Within condition. When the trading relation was present in the Between condition, it disappeared in the Within condition (version B); when it was absent in the Between condition, a large advantage for 2-cue trials emerged in the Within condition (version A). This pattern of results suggests that the trading relation between F1 onset and VOT is not psychoacoustic in origin (cf. Summerfield, 1982).
Figure 3. Discrimination results of Experiment 3.
One aspect of the present experiment that has not been considered so far is that, in contrast to the previous studies in this series, the primary and secondary cues were not independent. As VOT increased, the effective onset frequency of F1 rose and the F1 transition got shorter. A quick calculation shows that, in all conditions, the differences in F1 onset frequency between the standard and comparison stimuli were larger on 1-cue trials than on 2-cue trials. In fact, the stimuli on 2-cue trials should have been nearly indistinguishable on the basis of F1 onset or duration alone. This contrasts with the large advantage for 2-cue trials in the Within condition, version A, suggesting that these stimuli were discriminated on a basis other than F1 onset. Note also the absence of a decline in 2-cue discrimination scores over test blocks in that condition, which suggests that the secondary cue that caught the subjects' attention was independent of VOT. The only aspect of the secondary cue that was indeed independent of VOT in the short range was its final portion—the point at which F1 reached asymptote relative to the higher formants. This aspect of the stimuli may have been auditorily salient in the Within condition, even though it is apparently not an important factor in phonetic classification (Summerfield & Haggard, 1977). Why it was so much more salient in version A than in version B, where subjects seemed to attend only to the temporal aspect of VOT, is still a mystery. Considering the small number of subjects, however, it may simply have been a difference in listener strategies that was unrelated to the particular arrangement of stimuli.

**GENERAL DISCUSSION**

The present three studies extend the four experiments reported by Repp (1981). Although each experiment in this series has its own individual problems, the cumulative evidence does favor the hypothesis that most trading relations between acoustic cues in phonetic perception are phonetically conditioned. That is, they are a direct consequence of distinguishing between members of phonetic categories that are defined by a multiplicity of acoustic attributes. There is no convincing evidence for any significant psychoacoustic interactions between any of the cues varied, with the sole exception of VOT and aspiration amplitude (Repp, 1981: Exp. 3), which also was the only case in which a trading relation was expected to be psychoacoustic in nature.

To summarize the present findings: Experiment 1 investigated the trading relation between silence duration and presence/absence of release burst as cues to the stop manner contrast. While the trading relation was obtained in the Between condition, it was reversed in the Within condition. Because of the unexpected magnitude of the trading relation in identification, subjects may have applied a phonetic strategy in both conditions. The reversal in the trading relation across conditions was shown to be consistent with that hypothesis. The results are also consistent with the hypothesis that the subjects followed an auditory strategy in the Within condition, different from the phonetic strategy used in the Between condition. However, the results are not consistent with the hypothesis that the same auditory strategy was followed in both conditions, for in this case the pattern of results should have been similar in the two conditions. It may be concluded that the trading relation is either phonetic in origin or, if due to a psychoacoustic interaction, specifically limited to the phonetic boundary region.
Experiment 2, varying similar cues, focused on the within-category region at short values of the primary, temporal cue. At first, a similar trading relation was found in the Between and Within conditions. While this result seemed to lend support to the psychoacoustic hypothesis, it was argued that it may have resulted from a phonetic boundary shift due to anchoring in the Within condition, which thereby became another Between condition. Indeed, a change in stimulus arrangement eliminated the trading relation in the Within condition. An added Within condition using long values of the primary cue likewise yielded no trading relation. These results support the hypothesis that the trading relation is of phonetic origin.

Experiment 3 focused on the trading relation between VOT and F1 onset frequency as cues to the voicing contrast, using short values of VOT for the Within condition. Although the results showed some striking effects of stimulus arrangement, overall the trading relation was obtained in the Between condition but was reversed in the Within condition, thus lending further support to the phonetic hypothesis.

In the Introduction, it was pointed out that the phonetic hypothesis, which maintains that trading relations are a byproduct of phonetic categorization, cannot be clearly distinguished from a version of the psychoacoustic hypothesis that postulates that trading relations are due to auditory interactions occurring only at the phonetic boundary. However, this second hypothesis is weakened by at least two considerations. One emerges from the data of Experiments 2 and 3, which suggest that the trading relations studied disappear not only at relatively long values of the temporal dimension (which may suggest the involvement of a temporal threshold or masking) but also at the shortest values of the same dimension. A psychoacoustic explanation of these findings would have to be quite involved, although it is perhaps not impossible. The second, more serious problem for the boundary-specific psychoacoustic hypothesis is, however, that it rests on the assumption that the placement of the phonetic boundary is itself psychoacoustically conditioned--i.e., that it represents an auditory threshold of some sort (Pisoni, 1977; Pastore et al., 1977; Schouten, 1980). However, there is now ample evidence that linguistic category boundaries, while limited in certain ways by auditory acuity, are placed in accordance with the acoustic-phonetic characteristics of a particular language and, moreover, are flexible under a variety of conditions (Repp & Liberman, Note 1). That is, the location of the boundary is itself phonetically conditioned and therefore cannot be part of a purely psychoacoustic hypothesis.

In conclusion, then, the present data lend support to the classic dual-process view of speech perception (in the laboratory), as proposed by Fujisaki and Kawashima (1969, 1970) and Pisoni (1973) and reaffirmed by such recent authors as Samuel (1977), Soli (in press), and Repp (in press). Within the confines of the auditory perceptual system, these two processes represent the bottom-up and top-down components. (Models of word recognition typically lump both together under the heading of bottom-up.) The phonetic component is top-down because it represents the contribution to perception of the past experience of the individual--of the phonetic category prototypes established through speaking and listening. The auditory, bottom-up component, which includes interactions and nonlinearities of various sorts, merely provides the
raw material on which the interpretive phonetic component operates. Therefore, to say that a specific trading relation is phonetic in origin is quite analogous to saying that the word "apple" refers to the edible object not because of its acoustic (or even phonetic) properties but because the listener knows the word and its meaning. Once this is acknowledged, phonetic trading relations become merely one of many byproducts of categorical perception in the laboratory whose detailed investigation promises few new insights. Rather, the important questions for theoretical and empirical study become the acquisition of phonetic categories and how to conceptualize their internal representation.

REFERENCE NOTE


REFERENCES


**FOOTNOTE**

\(^{1}\)To the best of my knowledge, these were the last stimuli created on that distinguished instrument before it went out of commission in May 1982. A serial synthesizer was avoided because of the amplitude changes consequent upon changes in F1 frequency.