435

REPRINTED FROM:

SPEECH COMMUNICATION

Volume 2, No. 4, December 1983

TRADING RELATIONS AMONG ACOUSTIC CUES IN SPEECH PERCEPTION ARE LARGELY A RESULT OF PHONETIC CATEGORIZATION

Bruno H. REPP

Haskins Laboratories, 270 Crown Street, New Haven, CT 06511, USA

pp. 341–362

NORTH-HOLLAND PUBLISHING COMPANY – AMSTERDAM

Speech Communication 2 (1983) 341–361
North-Holland

TRADING RELATIONS AMONG ACOUSTIC CUES IN SPEECH PERCEPTION ARE LARGELY A RESULT OF PHONETIC CATEGORIZATION

Bruno H. REPP

Haskins Laboratories, 270 Crown Street, New Haven, CT 06511, USA

Received 5 April 1983
Revised 28 June 1983

Abstract. When several different acoustic cues contribute to the perception of a phonetic distinction, a trading relation among the cues can be demonstrated in an identification task as long as the speech stimuli are phonetically ambiguous. The present experiments address the question of whether the cues trade also in unambiguous stimuli, using a fixed-standard AX discrimination task with stimuli either from the vicinity of the phonetic category boundary or from within a phonetic category. The results suggest that four of the five trading relations examined are tied to the perception of phonetic contrasts; they disappear or reverse within categories. The one predicted exception represents a trading relation presumed to originate at a psychoacoustic level. The data severely restrict psychoacoustic explanations for these effects and also suggest that within-category discrimination is not achieved in a phonetic mode of perception, thus affirming a dual-process view of speech discrimination.

Zusammenfassung. Wenn mehrere verschiedene akustische Eigenschaften zur Wahrnehmung eines phonetischen Kontrastes beitragen, dann kann eine reziproke Beziehung zwischen zwei Eigenschaften in einem Identifikationstest demonstriert werden, solange die Stimuli phonetisch zweideutig sind. Hier wird die Frage gestellt, ob diese Beziehungen auch in phonetisch eindeutigen Stimuli existieren. Mit Hilfe eines standardisierten AX-Diskriminationstests untersuchen die hier beschriebenen Experimente fünf verschiedene phonetische Kontraste; für jeden stammen die Stimuli dabei entweder aus der Nähe einer Kategoriengrenze oder aus dem Feld innerhalb einer phonetischen Kategorie. Die Ergebnisse zeigen, dass vier der fünf reziproken Beziehungen mit der phonetischen Klassifizierung verbunden sind; innerhalb einer Kategorie verschwinden sie oder kehren sich um. Die einzige, vorhergesagte Ausnahme war eine Beziehung, für die ein psychoakustischer Ursprung angenommen wurde. Diese Daten schränken psychoakustische Erklärungen für reziproke Beziehungen stark ein und legen ausserdem nahe, dass Unterscheidungen innerhalb einer phonetischen Kategorie nicht in einem phonetischen Modus der Wahrnehmung erfolgen, was für ein duales Modell der Sprachwahrnehmung spricht.

Résumé. Quand plusieurs caractéristiques acoustiques différentes contribuent à la perception d'une distinction phonétique, on peut montrer l'existence d'une relation compensatoire lors de l'identification de stimuli de parole lorsque ceux-ci sont phonétiquement ambigus. Les expériences présentes essayent d'observer si ces relations compensatoires existent également avec des stimuli non-ambigus, au moyen d'une tâche de discrimination AX avec des stimuli, soit proches de la limite entre deux catégories phonétiques, soit à l'intérieur d'une catégorie. Les résultats suggèrent que parmi les cinq relations compensatoires, quatre sont liées à la perception des contrastes phonétiques; elles disparaissent ou s'inversent à l'intérieur des catégories. L'unique exception prévue représente une relation compensatoire supposée tirer son origine du niveau psychoacoustique. Ces données limitent strictement les explications psychoacoustiques pour les relations compensatoires et suggèrent également que la discrimination à l'intérieur des catégories n'est pas réalisée dans un mode phonétique de perception, affirmant ainsi que le discrimination de la parole peut être traitée de deux façons; phonétique et psychoacoustique.

Keywords. Trading relations, categorical perception, speech perception.

1. Introduction

Virtually every phonetic distinction has multiple correlates in the acoustic speech signal. That is, the articulatory adjustments required to change from one phonetic category to another (other things equal) cause acoustic changes along several separable physical dimensions—spectrum, amplitude, time. While a listener typically perceives only a single change—viz., one of phonetic category—the physical changes that led to this unitary percept can only be described in form of a list with multiple entries. When the signal properties thus listed are manipulated individually in an experiment, it is generally found that they all have perceptual cue value for the relevant phonetic distinction, although they may differ in their relative importance. If one cue in such an ensemble is changed to favor category ***B***, another cue can be modified to favor category ***A***, so that the phonetic percept remains unchanged. This is called a trading relation. Presumably, any two cues for the same phonetic distinction can be traded off against each other within limits set by their acceptable range of values and by their relative perceptual weights. Numerous recent studies of trading relations have been reviewed by Repp [24].

The mechanisms by which a listener's brain combines a number of diverse cues into a single phonetic percept are not known, but there are two contrasting views on that issue. One view (e.g., [13,26]) holds that the perceptual integration of acoustic cues is motivated by their common origin in the production of a phonetic contrast; that is, listeners are assumed to possess and apply detailed tacit knowledge of articulatory-acoustic speech patterns. The other view (best spelled out by Pastore [18]) maintains that integration of, and trading relations between, acoustic cues might arise from integration or interactions (such as masking or contrast) at a purely auditory (psychoacoustic) level of processing, without reference to the articulatory origin of the cues. The evidence so far (summarized by Repp [24]) is almost exclusively in favor of the first view. However, this conclusion is not as firm as it seems, as little is known about the auditory perception of stimuli as complex as speech. It is conceivable that, as more is learned about auditory mechanisms, certain trading relations between acoustic cues will find auditory explanations.

One way of investigating the issue is to contrast adult human perception of speech with that of human infants or nonhuman animals. Few such studies have been completed so far, but the existing data show similar trading relations in infants and chinchillas as in adult human subjects, suggesting that the particular effects studied may have an auditory basis [3,11,12,17]. In addition, Delgutte [2] has reported some interesting interactions in the firing patterns of cats' auditory neurons in response to speech stimuli, which might underly certain trading relations. These findings demonstrate that the question of the origin of trading relations cannot be considered settled. Indeed, it may be that different trading relations require different explanations.

In studies using adult human subjects, three methods have been applied to address the question of the origin of trading relations. One compares the perception of speech stimuli with that of 'nonspeech analogs' which contain the critical cues under investigation but are different from speech in other respects [31,10]. A second approach holds the acoustic stimulus structure constant by using nonspeech analogs that, although they are perceived as nonspeech by naive subjects, are sufficiently speechlike to be perceived as speech by more experienced or specially instructed subjects. The technique of imitating the speech formants with pure tones has served this purpose well [1]. The third method is to use speech stimuli and to lead listeners, through special instructions and practice, to perceive them analytically—to segregate them into their auditory components, as it were. This is a difficult task, and it is possible only with certain special stimuli, e.g., with fricative-vowel syllables [23]. In all of these studies, subjects' response patterns were radically different for stimuli heard as speech and for stimuli heard as nonspeech; in particular, the trading relations under investigation were observed only in the speech mode. However, as noted above, this result may not hold for all trading relations.

The present experiments took yet another approach, employing a simplified version of a procedure first used by Fitch, Halwes, Erickson, and Liberman [5]. Fitch et al. were concerned with a trading relation between a temporal and a spectral cue for the 'slit-split' contrast. The temporal cue was the amount of silence between the fricative noise and the periodic stimulus portion, and the spectral cue was the presence or absence of formant transitions (appropriate for a labial stop) at the onset of the periodic portion. In an identification task, less silence was needed to change 'slit' to 'split' when formant transitions were present than when they were absent. In a subsequent oddity discrimination task, Fitch et al. compared performance on three types of trials: (1) spectral difference only ('one-cue condition'); (2) spectral and temporal difference, where the stimulus with the formant transitions always had the longer silence ('two-cooperating-cues condition'); and (3) spectral and temporal difference, where the stimulus with the formant transitions had the shorter silence ('two-conflicting-cues condition'). Subjects were considerably more accurate in the second than in the third condition, with performance in the first condition lying in between. This ordering of conditions was predicted from the way the stimuli were labeled by the subjects. In essence, these results amount to a demonstration of categorical perception for speech stimuli varying on two dimensions: The listeners appeared to base their discrimination judgments primarily on the phonetic labels of the stimuli, and thus the trading relation between the two cues was exhibited in discrimination as well as in labeling responses.

The question explored by the present experiments is: What would happen if subjects could not rely on phonetic labels? Such a situation would arise if the stimuli to be discriminated are perceived as belonging to the same phonetic category. We know from many earlier studies of categorical perception that such discriminations are difficult to make, but subjects typically perform at a level better than chance, and their performance may be enhanced by increasing physical stimulus differences and/or by using a paradigm that reduces stimulus uncertainty (see Repp [25] for a review). If subjects cannot rely on phonetic labels, they presumably must make their discriminations on the basis of the auditory properties of the stimuli. If some of these properties interact at the auditory level of perception and thereby generate a trading relation, then this trading relation might be observed regardless of whether or not listeners can make phonetic category distinctions. On the other hand, if a trading relation is phonetic in origin, then the unavailability of phonetic contrasts might lead to a disappearance of the trading relation. Since, in this case, the cues may be independent at the auditory level, a difference in two cues should be at least as easy to discriminate as a difference in one cue (cf. [4]), regardless of whether the cue values are paired in the cooperating or the conflicting manner (in the fashion of [5]).

This is the rationale underlying the present experiments. To simplify the design, the cooperating-cues condition was omitted. The critical comparison was between 1-cue and 2-cue (conflicting-cues) trials in two discrimination conditions: *Between* phonetic categories and *Within* a single phonetic category. A trading relation in the Between condition (where stimuli contrasted phonetically on most trials) should show up as poorer performance on 2-cue than on 1-cue trials. The same pattern in the Within condition would suggest that the trading relation is auditory in origin. On the other hand, equal or better performance on 2-cue than on 1-cue trials in the Within condition would indicate that the trading relation is absent and, therefore, that its occurrence in the Between condition has a phonetic basis.

This contrast between an 'auditory' and a 'phonetic' account of trading relations is actually somewhat oversimplified. In truth, each account has a 'companion hypothesis' which represents a modified version of the opposing account. Thus, data favoring the 'auditory' hypothesis may also be taken to favor a generalized mode of phonetic processing that applies to speech stimuli regardless of whether phonetic contrasts are perceived, and that enables listeners to make fine within-category discriminations. According to this hypothesis, trading relations characterize this general speech mode rather than merely the act of phonetic categorization. Conversely, data favoring the 'phonetic' hypothesis could also be interpreted as reflecting a psychoacoustic trading relation that is specifically restricted to the category boundary region. Thus, the present studies are really trying to decide among two pairs of hypotheses. The relative merits of the two surviving hypotheses will be considered further in the General Discussion.

2.3. Analysis

Individual subject scores in each test block were converted into d' values, taking the proportions of 'different' responses on 1-cue and 2-cue trials, respectively, as separate hit rates, and the proportion of 'different' responses on trials of identical stimuli as the single false-alarm rate. Proportions of 0 and 1 were treated as 0.01 and 0.99, respectively, thus limiting d' to a maximum value of 4.66.

Analyses of variance were conducted on subjects' d' scores in each experiment. The main analysis had three factors: Conditions (Within vs. Between), Cues (1-cue vs. 2-cue), and Blocks (three level of increasing difficulty). The two repetitions of the Within condition in Exps. 1–2 were combined in this analysis. The critical effect was the Conditions by Cues interaction, which was expected to reveal whether or not the same trading relation (or other response pattern) held in the two conditions. Separate analyses of variance were also conducted on the Between and Within conditions to determine the significance of the Cues main effects. There was generally a significant decrease in performance across test blocks, resulting in a significant main effect of Blocks, which will not be mentioned further below. Repetitions was an additional factor in the separate analysis of the Within conditions in Experiments 1–2. There were no significant effects involving this factor, so repetitions were combined in the data reported below.

3. Experiment 1: 'say'-'stay'

The trading relation studied here concerned, as the primary cue, the amount of silence following the initial fricative noise and, as the secondary cue, the onset frequency of the first formant ($F1$) following the silence. This trading relation, which is similar to that for 'slit-split' studied by Fitch et al. [5], has been previously investigated by Best et al. [1]: Less silence is needed to change 'say' to 'stay' when $F1$ starts at a lower frequency. Best et al. also obtained this trading relation in two different discrimination tests (oddity and variable-standard AX). These tests actually included some within-category trials along with between-category trials, and the trading relation seemed to disappear within the 'stay' category. However, this result is not conclusive, since it may reflect a floor effect and is based on rather few responses. It is interesting to note, however, that the similar data of Fitch et al. [5] for the 'slit-split' contrast, although they are open to the same objections, actually suggest a reversal of the trading relation in the within-category regions: Whereas the ordering of performance on the three types of trials was cooperating cues > one cue > conflicting cues in the phonetic boundary region, it changed to cooperating cues = conflicting cues > one cue (at chance) within categories. This is exactly the pattern one should expect from a trading relation that is specific to phonetic categorization.

This expectation was further confirmed by Best et al. [1] in an elegant study with 'sine-wave analogs' of 'say-stay' stimuli. Subjects who reported that they heard the sine-wave stimuli as (highly unnatural) tokens of 'say' or 'stay' exhibited the same trading relation between silence duration and $F1$(–analog) onset frequency as was observed in speech stimuli, whereas those subjects who heard the sine-wave stimuli as nonspeech showed a radically different pattern of responses which suggested that they paid selective attention to variations in one or the other cue but did not integrate them into a unitary percept, as phonetic listeners did.

Given these rather convincing results, the present re-investigation of the 'say-stay' contrast aimed at replicating the findings of Best et al. and thereby validating the present procedure. The prediction was, then, that the trading relation between silence duration and $F1$ onset frequency would be observed only in the Between condition but not in the Within condition.

3.1. Method

3.1.1. Subjects

Eleven volunteers were recruited by announcements on Yale University campus and were paid for their participation. Most of them had served in earlier speech perception experiments. A different group of 9 subjects took the brief labeling test.

3.1.2. Stimuli

The stimuli were hybrids composed of a natu-

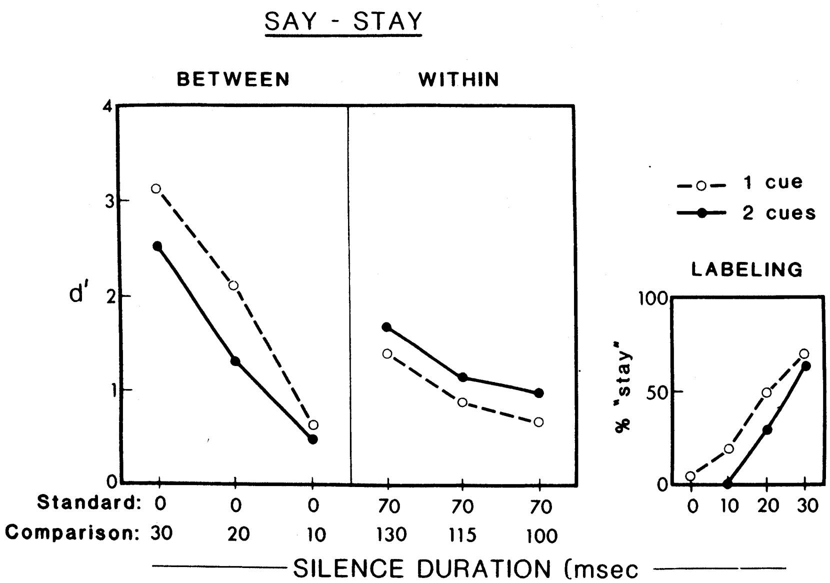

Fig. 2. Results of Experiment 1. For the labeling data, the dashed and solid lines represent stimuli with low and high $F1$ onset, respectively.

ral-speech [s] noise followed by a synthetic periodic portion. The [s] noise, derived from a male speaker's utterance of [sa], was 212 msec long, with a gradually rising amplitude over the first 170 msec and a rapid fall thereafter. The periodic portion was produced on the OVE IIIc serial resonance synthesizer at Haskins Laboratories, following formant specifications provided by Best et al. [1] in their Fig. 1 (speaker SSB). Its duration was 300 msec, its amplitude rose with $F1$, and its fundamental frequency fell linearly from 110 to 80 Hz.

The two stimulus portions were concatenated after both had been digitized at 10 kHz using the Haskins Laboratories PCM system. The *primary cue* was the amount of silence between them. In the Between condition, the standard stimulus had no silence at all ('say'), and the comparison stimuli on 'different' trials had 30, 20, and 10 msec, respectively, in the three test blocks. In the Within condition, the standard had 70 msec of silence ('stay'), and the comparison values were 130, 115, and 100 msec. The 'say-stay' boundary was expected to be in the vicinity of 20 msec of silence. The *secondary cue* was the onset frequency of $F1$ in the periodic portion. On 1-cue trials, it was 200 Hz, whereas, on 2-cue trials, it was raised to 299 Hz—a cue favoring 'say' and thus 'conflicting' with the longer silence cue in the comparison stimuli. The duration of the $F1$ transition was 40 msec.

3.2. Results

The results are shown in Fig. 2. The first panel shows average d′ scores for 1-cue and 2-cue trials as a function of test blocks in the Between condition. The second panel shows the corresponding data for the Within condition. As predicted, performance was higher on 1-cue than on 2-cue trials in the Between condition, but it was higher on 2-cue than on 1-cue trials in the Within condition. This reversal was confirmed by a significant Conditions by Cues interaction, $F(1, 10) = 6.6$, $p < 0.05$. Separate tests of the Cues main effects in each condition revealed that the trading relation in the Between condition was marginal, $F(1, 10) = 3.7$, $p < 0.10$, while the advantage for 2-cue trials in the Within condition was reliable, $F(1, 10) = 12.1$, $p < 0.01$.

The third panel of Fig. 2 shows the labeling data for the stimuli used in the Between condition.

One listener perceived all stimuli as 'say' and was excluded. The data of the remaining eight listeners confirm that the standard stimulus (no silence) was head as 'say' and that the 'say-stay' boundary fell between 20–25 msec, as expected. The labeling data also exhibit the trading relation between the two cues, with fewer 'stay' responses to the high $F1$ onset stimuli, although the difference was rather small and fell short of significance, $F(1, 7) = 4.0$, $p < 0.10$.

3.3. Discussion

Basically, the results confirmed the predictions: A trading relation between the two cues appeared, though not very reliably, in the region of the 'say-stay' boundary, whereas it was reversed within the 'stay' category. This suggests, in accordance with the findings of Best et al. [1], that the trading relation between silence duration and $F1$ onset frequency is phonetic, rather than auditory, in origin.

The weakness of the trading relation in the Between condition and in the labeling task may be attributed to three factors: (1) The difference in the secondary cue was rather small. (2) The discrimination paradigm may have facilitated the detection of auditory stimulus differences in the Between condition, particularly since this condition was preceded by the Within condition which required auditory discrimination of similar differences. Any phonetic trading relation between the relevant cues (or, rather, its manifestation in the data) would be weakened by auditory discrimination beyond the detection of phonetic differences. (3) There was considerable between-subject variability, in part due to the openness of the instructions which permitted a variety of different strategies, and in part due to variations in subjects' overall performance level which introduced floor effects. In any case, the critical result—the change across conditions in the relation between 1-cue and 2-cue discrimination (the Conditions by Cues interaction)—was significant.

It is conceivable, of course, that an *auditory* trading relation between silence duration and $F1$ onset frequency exists when the silence is short but not when it is long. (Recall that this 'local psychoacoustic interaction' hypothesis cannot be distinguished from the 'phonetic' hypothesis by the present experiments.) For example, the presence of a silent interval might be more difficult to detect when $F1$ has a higher onset, but the perceived duration of longer silent intervals may not be affected by $F1$ onset frequency. This explanation is consistent with the present data, but it seems less convincing in view of the Best et al. [1] findings. Specifically, these authors found that subjects who perceived sine-wave analogs of 'say-stay' stimuli nonphonetically *and* focused on the silence cue were not at all affected by $F1$(–analog) onset frequency, even when the silence durations were in the short range.

Best et al. [1] found that listeners who followed an auditory strategy focused on one cue and ignored the other. In the present Within condition, selective attention to the silence cue would have resulted in equal scores on 1-cue and 2-cue trials, both declining over test blocks, whereas selective attention to the spectral cue would have resulted in much better performance on 2-cue than on 1-cue trials, with no decline in performance over blocks. However, no subject exhibited this second pattern, and few exhibited the first. Thus, the average data (Fig. 2) are fairly typical of the individual subject; they are not an artifact of averaging over subjects with radically different strategies. It seems likely, then, that the present subjects took both cues into account, even though the practice trials encouraged selective attention to the primary cue. In that case, the higher scores on 2-cue than on 1-cue trials simply show that stimuli differing on two dimensions are easier to discriminate than stimuli differing on one dimension only, which is perfectly plausible and consistent with the relative auditory independence of the two cues shown by Best et al. [1]. Their finding that subjects paid selective attention to one or the other cue was probably due to their paradigm, an AXB classification task in which the two cues were perfectly correlated in the reference stimuli (A, B). Thus, their subjects were encouraged to select one cue and ignore the other, redundant one; in fact, this strategy simplified the subjects' task. The present AX discrimination task, on the other hand, while it emphasized the silence cue, encouraged listeners to pay attention to all possible stimulus differences. The ability of subjects to make use of both cue dimensions in one task is not inconsistent with their ability to select only one of them in a different task, since either strategy may be followed with independent auditory dimensions.

It should be noted that the advantage of 2-cue over 1-cue trials in the Within condition did not increase over blocks (as might be expected if subjects began to direct their attention to the secondary cue as the difference in the primary cue got smaller) but remained constant at about 0.3 d', which provides an estimate of the (rather poor) discriminability of the secondary-cue difference, assuming that the discriminabilities of the two cues were additive. Another feature of the data worth mentioning is the apparent convergence of the 1-cue and 2-cue scores in the last block of the Between condition. Although this effect was not significant, it was quite clearly exhibited by several individual subjects (without constituting a floor effect). Note that the trading relation between the cues is expected to disappear not only within the 'stay' category but also within the 'say' category—a situation approximated by the third block of the Between condition. This issue will be further addressed in Experiment 4.

4. Experiment 2: 'say shop'-'say chop'

The trading relation investigated in this experiment involved the same primary cue as in Experiment 1, viz., duration of silence, but a different secondary cue—the duration of the fricative noise following the silence. The trading relation between these two cues was demonstrated by Repp et al. [26]: More silence was needed to turn 'say shop' into 'say chop' when the fricative noise was long than when it was short.

This trading relation has much in common with that of Experiment 1; however, it does involve two cues varying along the same physical dimension (duration), which makes a nonphonetic perceptual interaction perhaps more likely than between a temporal and a spectral dimension. For example, there may be a contrastive effect, such that a long fricative noise makes the preceding silence sound relatively short (or vice versa), which would lead to the observed trading relation. If there is such an interaction between the two temporal cues, then it should surface regardless of whether or not subjects perceive phonetic contrasts. Noting that the silence durations in the Within condition of Experiment 1 had been rather far from the category boundary, somewhat shorter durations were chosen for the present Within condition.

4.1. Method

4.1.1. Subjects

Ten volunteers participated, two of whom had also been subjects in Experiment 1, and six of whom had previously been subjects in Experiment 3b.

4.1.2. Stimuli

The stimuli were created on the OVE IIIc synthesizer. Formant parameters were copied from a spectrogram of 'say shop' produced by a male speaker (as used in [26]). The initial 240-msec 'say' portion was followed by a variable silent interval, a fricative noise of variable duration, and a 125-msec final periodic portion ('op') whose first 10 msec overlapped the last 10 msec of the fricative noise. The fricative noise reached maximum amplitude after 50 msec. Fundamental frequency rose from 85 to 100 Hz during the 'ay' portion and fell from 100 to 90 Hz during the 'op' portion.

The *primary cue* was the amount of silence preceding the fricative noise. In the Between condition, the standard stimulus had no silence at all ('say shop'), and the comparison stimuli on 'different' trials had 30, 20, and 10 msec, respectively, in the three test blocks, just as in Experiment 1. In the Within condition, the standard had 40 msec of silence ('say chop'), and the comparison values were 100, 80, and 60 msec. The 'say shop-say chop' boundary was expected to be in the vicinity of 20 msec of silence. The *secondary cue* was the duration of the fricative noise in the second syllable. On 1-cue trials, its duration was 110 msec, whereas, on 2-cue trials, it was 130 msec, thus biasing perception more toward 'say shop'. The duration of the noise was changed at the synthesis stage by extending its central steady-state portion. The stimulus tapes were recorded directly from the synthesizer, without digitization of stimuli, so the fricative noise waveforms exhibited natural random variability across tokens.

4.2. Results

The results are shown in Fig. 3. There was no difference between 1-cue and 2-cue trials in the Between condition but a considerable advantage of 2-cue over 1-cue trials in the Within condition. The Conditions by Cues interaction was significant, $F(1, 9) = 22.4$, $p < 0.002$, confirming the different effects that addition of a secondary cue had in the two conditions. The 2-cue advantage in the Within condition was highly reliable, $F(1, 9) = 32.3$, $p < 0.001$.

The labeling results (third panel of Fig. 3) revealed that the standard was always heard as 'say shop' and that the phonetic category boundary fell between 20-25 msec, as expected. However, there was also the expected trading relation, with more 'say chop' responses to stimuli containing the shorter noise, $F(1, 9) = 16.9$, $p < 0.01$. Thus, the trading relation was exhibited in labeling but not in Between discrimination.

4.3. Discussion

Except for the complete absence of a trading relation in the Between discrimination results, the present data are quite similar to those of Experiment 1, suggesting that the trading relation between silence and fricative noise duration is similar to that between silence duration and $F1$ onset frequency, and that both are phonetic in origin. Both, of course, concern the perception of similar phonetic contrasts—stop and affricate manner, respectively, both of which involve the presence vs. absence of oral closure.

As in Experiment 1, the critical finding is the Conditions by Cues interaction, which reflects the change in the difference between 1-cue and 2-cue trials across conditions. The absence of a trading relation in the Between condition is probably due to listeners' detection of auditory differences in addition to the phonetic contrast. Since the difference in the secondary cue was more noticeable here than in Experiment 1 (as suggested by the larger difference between 1-cue and 2-cue trials in the Within condition), the resulting auditory advantage for 2-cue trials may have cancelled the phonetic trading relation in the Between condition.

The difference between the 1-cue and 2-cue d' functions in the Within condition suggests that the discriminability of the secondary cue difference was about 0.4 d' at the outset and increased to 0.9 d' in the last block, where discrimination on 1-cue trials was at chance. Although this increase did not reach significance, it does suggest that some subjects directed their attention towards the noise duration difference as the silence duration dif-

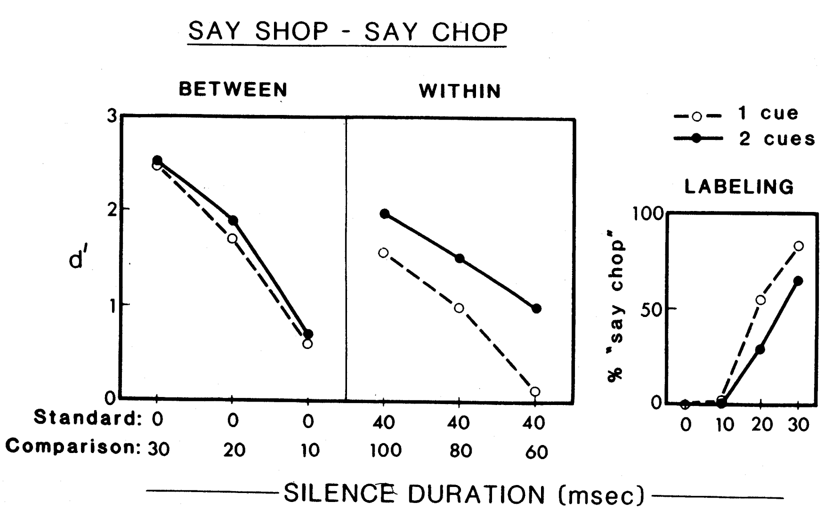

Fig. 3. Results of Experiment 2. For the labeling data, the dashed and solid lines represent stimuli with short and long fricative noise, respectively.

ference decreased. The data also suggest, surprisingly, that the difference between a 110-msec and a 130-msec noise was much easier to detect than the difference between a 40-msec and a 60-msec silence (Within condition, last block). Since this finding contradicts Weber's Law, it indicates that silence and noise durations were not equally accessible in the auditory signal transforms on which the subjects based their judgments.

5. Experiment 3: 'goat'-'coat'

This study was concerned with a trading relation reported by Repp [22]: When acoustic voice onset time (VOT) is used as the primary cue to the voicing of an utterance-initial stop consonant, less of an increase in VOT is needed to turn a voiced stop into a voiceless one when the amplitude of the aspiration noise (whose duration is the VOT) is reduced. This trading relation is different in two important respects from those investigated in Experiments 1 and 2. First, the two interacting cues are both properties of the same signal portion, viz., of the aspiration noise that precedes voicing onset. Second, it appears that, in contrast to virtually all other trading relations in phonetic perception, there is not good articulatory rationale for this trading relation. Although the relevant measurements have not been done, it seems likely that the amplitude of aspiration, measured at a fixed distance from the release, would be about the same in voiced and voiceless stops. It is true, of course, that voiced stops have a much shorter period of aspiration, and this necessary covariation of aspiration duration and time-integrated amplitude may be sufficient to account for the perceptual trading relation. Still, the articulatory explanation seems less compelling than that for other effects, where different cues can be shown to be acoustically diverse consequences of the same articulatory act (cf [26]). Moreover, there are well-known instances of trade-offs between duration and amplitude at the auditory threshold and in judgments of loudness (e.g., [9,29]). For these reasons, the present trading relation was hypothesized to be auditory in origin. Thus it was predicted to occur in both conditions of Experiment 3; that is, performance was expected to be higher on 1-cue than on 2-cue trials in both the Between and Within conditions.

Experiment 3 was run twice. The first run (Exp. 3a) was only partially successful because the stimuli in the Between condition turned out to have missed the boundary (their VOTs were too long), so that the Between condition was effectively another Within condition. Also, the VOT differences were rather small, so that the task was very difficult. Therefore, a replication (Exp. 3b) was conducted with shorter VOT values in the Between condition and larger VOT differences. Results from both runs will be reported.

5.1. Method

5.1.1. Subjects

Eight volunteers participated in Experiment 3a. All of them had previously been subjects in Experiment 1. There were nine subjects in Experiment 3b, two of whom had also been in Experiment 3a. Six of them, plus a research assistant, took the labeling test.

5.1.2. Stimuli

The stimuli were derived from natural speech. A female speaker recorded the words 'goat' and 'coat'. They were digitized at 10 kHz, and a VOT continuum was constructed by first replacing the burst and aspiration portions of 'goat' (22 msec) with the first 22 msec of 'coat' and by then substituting additional equivalent amounts of aspiration noise from 'coat' (VOT = 66 msec) for each successive pitch period of 'goat'. For a detailed description of this procedure, see the appendix in [8].

Stimuli from this continuum were used in the Between condition only. For the Within condition, where VOTs longer than that of the natural 'coat' were required, the stimuli were generated by a different procedure. Note that, in the method described above, total stimulus duration remains constant as VOT is increased while the periodic stimulus portion is progressively shortened. This is a standard feature of VOT continua and probably does not matter when relatively short VOTs are to be discriminated. However, when VOTs are made rather long, little is left of the periodic portion, and removal of even a single pitch period may become perceptually quite salient. That is, subjects may discriminate such stimuli not on the basis of VOT but on the basis of changes in the duration and intonation of the 'vowel' (cf. [33]). To prevent this from happening in the present Within condition, the periodic stimulus portion was held constant, and VOT was further increased by using a waveform editor to duplicate randomly selected segments of the final portion of the aspiration noise, where the formant transitions presumably were close to asymptote.

Thus, the stimuli in the Between condition had a total duration of 228 msec (VOT plus periodic portion), with the periodic portion diminishing as VOT increased, whereas the stimuli in the Within condition had a constant periodic portion of 155 msec and a total duration that increased with VOT. All stimuli included, in addition, a rather powerful final [t] release burst of approximately 112 msec duration, which was separated from the end of the periodic portion by a 133-msec silent closure interval.

The *primary cue* in this study was, of course, VOT (i.e., the duration of the aperiodic portion at stimulus onset). In the Between condition of Experiment 3b (that of Experiment 3a will not concern us here, since performance was at chance), the standard had a VOT of 38 msec (which seems rather long for a voiced stop but was still heard as 'goat'), and the comparison stimuli had VOTs of 55, 49, and 44 msec, respectively. In the Within condition of Experiment 3a, the standard had a VOT of 73 msec ('coat'), and the comparison stimuli had values of 98, 91, and 85 msec, respectively. In the Within condition of Experiment 3b, the same standard was used, but the comparison stimuli had values of 108, 98, and 85 msec. The *secondary cue* was the amplitude of the aperiodic stimulus portion. On 2-cue trials, it was reduced by 6 dB SPL in the comparison stimulus, counteracting the longer VOT of that stimulus. This manipulation was performed on the digitized waveform, using computer instructions.

5.2. Results

In the Between condition of Experiment 3a, because of inappropriately long VOTs, all but one subject heard only 'coat' (as determined by postexperimental interviews) and performed at chance level. The single subject who appeared to be able to make use of phonetic contrasts in this condition performed quite well and had higher scores on 1-cue than on 2-cue trials, in accord with the expected trading relation. The choice of VOT values for the Between condition was more successful

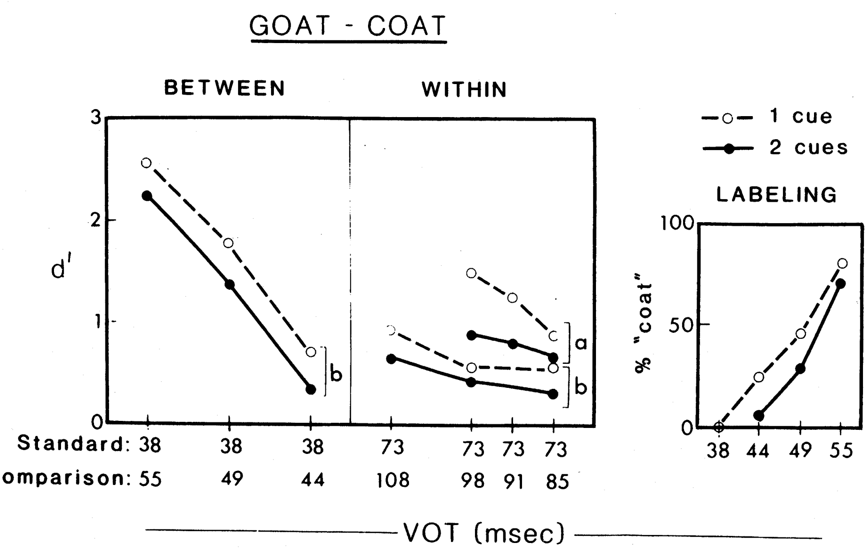

Fig. 4. Results of Experiment 3. For the labelling data, the dashed and solid lines represent stimuli with intense and less intense aspiration noise, respectively.

in Experiment 3b; these results are shown in the first panel of Fig. 4. Scores were higher on 1-cue than on 2-cue trials, $F(1, 8) = 5.5$, $p < 0.01$, which reflects the expected trading relation.

Within-category discrimination of the 'goat-coat' stimuli proved to be a difficult task for naive subjects. Their problem seemed to be to discover the dimension on which the stimuli differed. (Recall that the nature of the difference was not revealed in the instructions but had to be detected during the practice block.) In both parts of the experiment, performance in the first presentation of the Within condition was close to chance, and there was no difference between 1-cue and 2-cue trials. Therefore, these data were discarded. Prompted by subjects' complaints over the difficulty of the task, the experimenter told them before the repetition of the Within condition what kind of difference to listen for, and he produced exaggerated examples of stops with different amounts of aspiration to illustrate the point. This improved subjects' performance, especially in Experiment 3a. The results from this final condition of Experiment 3a are presented in the second panel of Fig. 4 (the functions labeled 'a'). It can be seen that performance was better on 1-cue than on 2-cue trials, $F(1, 7) = 5.7$, $p < 0.05$. This pattern contrasts with that obtained in the Within conditions of Experiments 1 and 2, where the opposite difference was observed.

The corresponding results from Experiment 3b are shown as the function labeled 'b'. Although better than chance, on the average, scores were low and highly variable. However, there was a (nonsignificant) tendency for 1-cue discrimination to be better than 2-cue discrimination, which supports the results of Experiment 3a. The Cues by Conditions interaction in the overall analysis of variance for Experiment 3b was nonsignificant.

The third panel of Fig. 4 shows the labeling data. These data confirm that the standard stimulus (VOT = 38 msec) was perceived as 'goat', and they also show the expected trading relation, although it fell short of significance, $F(1, 6) = 5.5$ $p < 0.10$.

5.3. Discussion

The results of this experiment are strongest in terms of what they do not show. The most significant finding is the *absence* of an advantage for 2-cue trials in the Within condition. The data suggest that, on the contrary, there was an advantage for 1-cue trials in both the Within and Between conditions. This pattern of results is the one expected for a trading relation of psychoacoustic origin. Thus, the interaction between the duration of aspiration noise and its amplitude may be similar to other kinds of auditory time-intensity trade-offs. In this sense, it is a curiosity among trading relations in speech perception, for virtually all others have articulatory correlates that make them likely to be specific to phonetic perception. Two additional such trading relations are investigated in the following experiments.

6. Experiment 4: 'slit-split'

Experiments 1–3 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, particularly in Experiments 1–2, were located at relatively short durations of the temporal cue and did not leave sufficient 'room' for above-chance discrimination at the short end of the continuum. Also, to the extent that the boundary coincided with a psychoacoustic threshold of some sort (cf. [16,19,21]), 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 desirable. 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) thought to underly a psychoacoustic 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 that equates category boundaries with auditory thresholds.

To pursue this possibility, it was necessary to find a stimulus continuum on which the boundary is at somewhat longer durations of a temporal cue. The 'slit-split' distinction fit the bill. In the study by Fitch et al. [5], 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 gave rise to the hope of obtaining above-chance discrimination scores strictly within the 'slit' category.

Experiment 4 was conducted in two parts. Experiment 4a included the Between condition and the Within ('slit') condition just described. Experiment 4b 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. In contrast to the earlier experiments in this series, formal identification tests were collected prior to the experiment to facilitate stimulus selection.

6.1. Method

6.1.1. Subjects

Nine paid volunteers participated in Experiment 4a, and seven different ones in Experiment 4b. The identification test was taken by nine subjects, four of whom also participated in Experiment 4a.

6.1.2. Stimuli

A female speaker recorded the utterance 'split', which was digitized at 20 kHz. The pre-closure 's' 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' which retained the final 12 msec of the release burst, and a weak 'blit' which had no release burst. Thus, the *primary cue* was the duration of the silence preceding the 'blit' while the *secondary cue* was the presence or absence of the labial release burst. (Note that Fitch et al. [5] used a different secondary cue—presence versus absence of formant transitions.)

In the Between condition of Experiment 4a, the standard had a closure silence of 40 msec, and the comparison stimuli in successive test blocks had silences of 80, 70, and 60 msec. In the Within ('slit') condition of Experiment 4a, the standard had no silence, while the comparisons had silences of 40, 30, and 20 msec. In the Within ('split') condition of Experiment 4b, the standard had 140 msec of silence, while the comparisons had silences of 200, 180, and 160 msec. In all these conditions, the standard always contained the strong 'blit' portion. The Between and Within ('slit') conditions of Experiment 4b reversed the roles of the standard and comparison stimuli of the corresponding conditions in Experiment 4a: In the Between condition, the standard initially had 80 msec of silence, 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 20 msec, while that of the comparison remained fixed at 0 msec. Here the standard always contained the weak 'blit' portion. The reason for these changes will become apparent below. The subjects in Experiment 4b listened to the Within ('split') tape at the end of the session.

The identification test was more extensive than in Experiments 1–3 and 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'.

6.2. Results and discussion

The results of the identification test are shown in Fig. 5. They proved to be very orderly. The category boundaries (50-percent cross-over points) were at 49 and 70 msec for the strong and weak 'blit', respectively. Note that the standards used in the Within 'slit' (Experiment 4a) and 'split' (Experiment 4b) conditions, with silences of 0 and 140 msec, were unambiguous instances of 'slit' and 'split', respectively, as intended.

The average results of the discrimination tests of Experiment 4a are shown in Fig. 6. 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 [$F(2, 16) = 4.6$, $p < 0.05$, for the Cues by Blocks interaction; $F(1, 8) = 3.3$, $p > 0.05$, 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, $F(1, 8) = 8.1$, $p < 0.05$. The Conditions by Cues interaction was not significant, $F(1, 8) = 1.7$.

At first blush, these results look exactly like those expected if the trading relation had a purely psychoacoustic basis. However, the abnormally high performance level in the Within condition gives rise to suspicion. Indeed, the author's observations as a pilot subject suggested 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 the perception of phonetic contrasts, rather than from a psychoacoustic interaction. It was for this reason that Experiment 4b

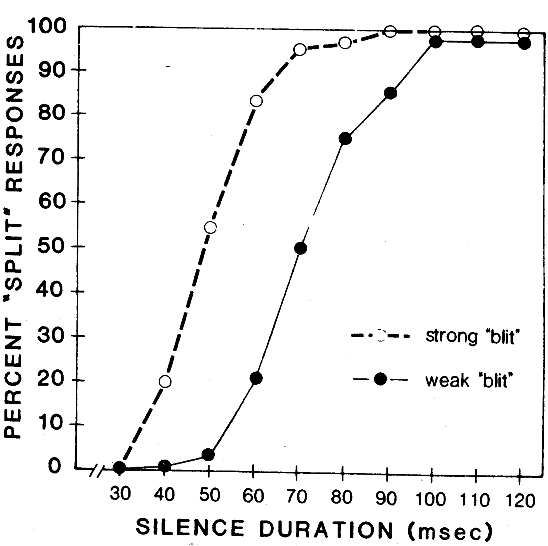

Fig. 5. Identification results of Experiment 4.

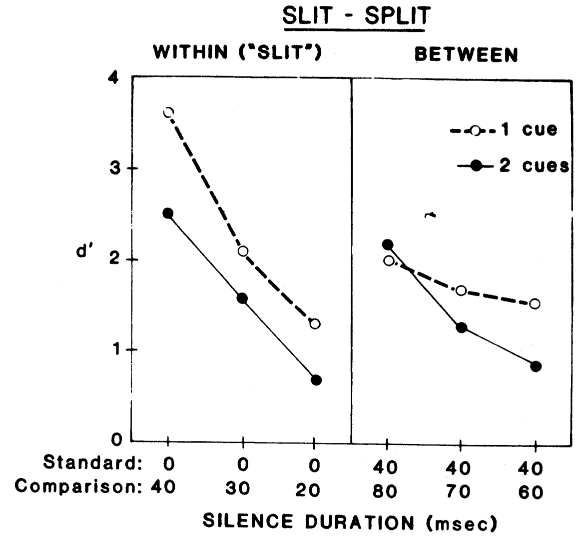

Fig. 6. Discrimination results of Experiment 4a.

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 Experiments 1–3.

The results of Experiment 4b are shown in Fig. 7. The conditions in the two panels on the left correspond to those in Fig. 6. The change in standards had a quite dramatic effect. In the Between condition, performance was better than previously and exhibited a clear trading relation, $F(1, 6) = 8.0$, $p < 0.05$. 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

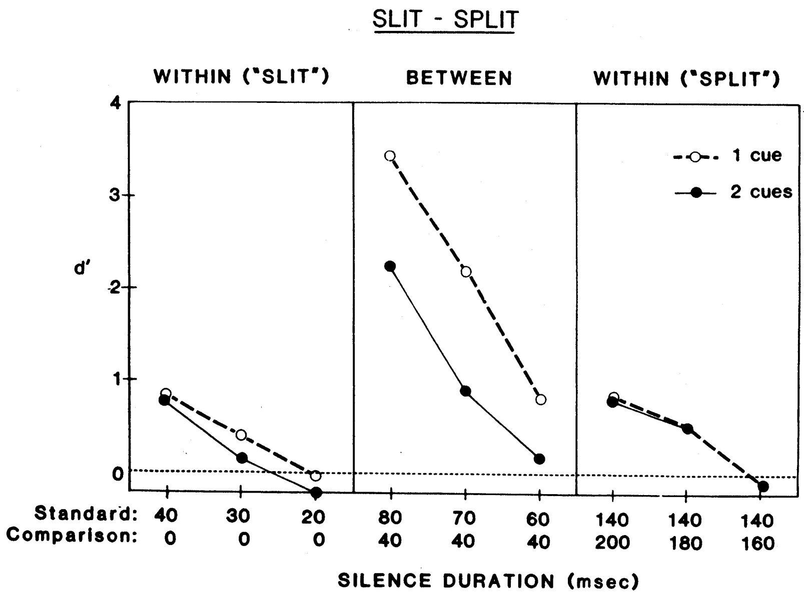

Fig. 7. Discrimination results of Experiment 4b.

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 Fig. 7), however, which strongly resemble those of the Within ('slit') 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 in the Between condition where that cue made a large difference (cf. also labeling data in Fig. 5). Therefore, it seems possible that the lack of any secondary-cue effect in the Within ('slit') condition was likewise due to lack of attention, although the possibility of masking remains.

7. Experiment 7: 'ga'-'ka'

In the final experiment of this series, a second attempt was made to assess within-category discrimination at the short end of a temporal continuum, using cues that seem less likely to engage in a masking interaction than those employed in Experiment 4. The stimuli were taken from 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 [14]. Since the secondary cue was to be the onset frequency of the $F1$ transition (cf. [15,32]), synthetic stimuli had to be used.

7.1. Method

7.1.1. Subjects

Ten paid volunteers participated, four of whom had also taken part in Experiment 4b. Five subjects took version A, and five took version B of the test (see below).

7.1.2. Stimuli

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: 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 frequencies. The two alternative $F1$ trajectories were chosen so as to have the same slope, making the magnitude of the difference in the secondary cue ($F1$ onset frequency) constant for different values of the primary cue (VOT) in the range of 0–50 msec.

Because Experiment 4 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 4b in that the standards were held constant through all test blocks while the comparisons 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.

7.2. Results

Figure 8 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 Fig. 9. They are plotted separately for versions A (top panels) and B (bottom panels) of the tests, not only because the VOT comparisons were

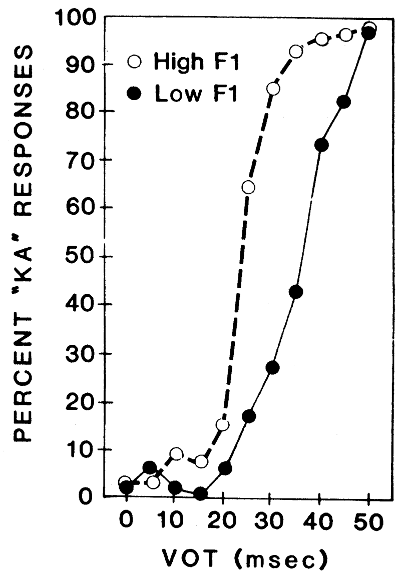

Fig. 8. Identification results of Experiment 5.

slightly different but also because one of the strongest effects in the overall analysis of variance was the Cues by Versions interaction, $F(1, 8) = 26.9$, $p < 0.001$, 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, $F(1, 8) = 33.6$, $p < 0.001$, which indicates that the pattern of results was different for the Within and Between conditions.

Both of these effects are clearly evident in Fig. 9. 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 also 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, however, it was small and nonsignificant in version A but large and significant, $F(1, 4) = 12.4$, $p < 0.05$, in version B. In the Within condition, on the other hand, there was a large 2-cue superiority in version A, $F(1, 4) = 52.9$, $p < 0.01$, but no difference whatsoever in version B. Note also the unexpectedly high level of performance in the Within condition in both versions.

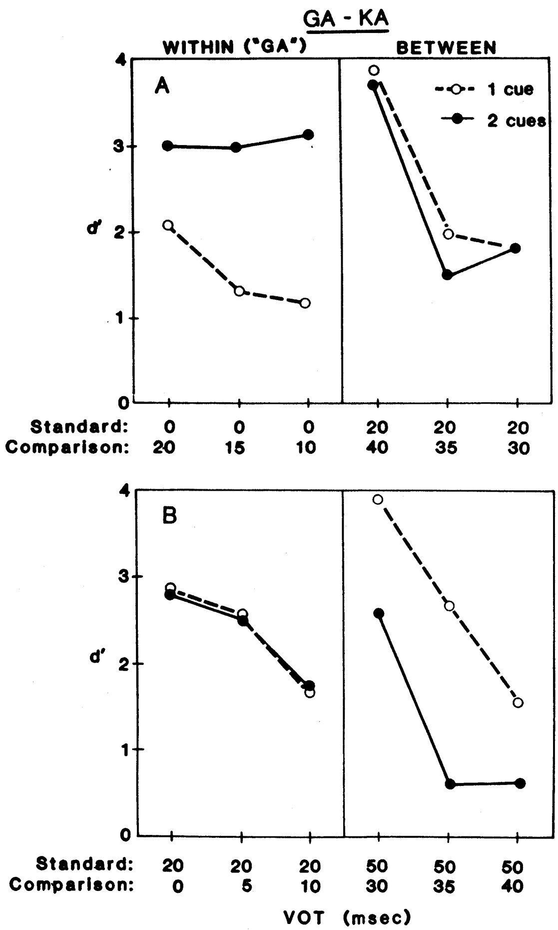

Fig. 9. Discrimination results of Experiment 5.

7.3. Discussion

These data present some problems for detailed 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, confirming Summerfield [31]. (However, see [10].)

One possible explanation for the strong Cues by Versions interaction is that the standard with the 20-msec VOT was in fact more ambiguous than the identification data (Fig. 7) suggest and oscillated between 'ga' and 'ka'. If so, the results would be a mixture of auditory and phonetic discrimination, and the phonetic trading relation would be canceled by the auditory advantage for 2-cue trials (cf. the Between condition of Experiment 2). This may explain the lack of a Cue effect in the lower left and upper right panels of Fig. 9, and it may then be legitimate to discount these conditions and focus on the upper left and lower right panels. In that case, of course, the results are crystal-clear.

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 thest 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 [32]. 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, remains a mystery.

8. General Discussion

The cumulative evidence from the present studies favors 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 which 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 (Exp. 3), which also was the only trading relation that was expected to be psychoacoustic in nature. Trading relations of this kind do not seem to play an important role in speech perception.

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 4 and 5, 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 [19,21,28]. 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 [34]. That is, the location of the boundary is itself phonetically conditioned and therefore cannot be part of a purely psychoacoustic hypothesis.

Together with the psychoacoustic hypothesis we can also dismiss its companion, the 'generalized phonetic processing' hypothesis. According to this hypothesis, trading relations characterize a speech mode of perception that is elicited by speech stimuli regardless of whether or not they are phonetically distinctive. In other words, this hypothesis postulates that distinctions within phonetic categories can be accomplished in a phonetic mode that is characterized by cue trading relations. The present results suggest, on the contrary, that trading relations are a direct consequence of phonetic categorization, and that listeners must resort to a different perceptual mode (an auditory mode) to accomplish within-category discrimination.

In conclusion, then, the present data lend support to the classic dual-process view of speech perception, as proposed by Fujisaki and Kawashima [6,7] and Pisoni [20] and reaffirmed by such authors as Samuel [27], Soli [30], and Repp [25]. Within the confines of the auditory perceptual system, the dual processes represent 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, however, 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 their internal representation.

Acknowledgments

This research was supported by NICHD Grant HD01994 and BRS Grant RR05596 to Haskins Laboratories. In am grateful to Janette Henderson and Sharon Manuel for assistance in running subjects, to Robert Crowder, Virginia Mann, Richard Pastore, and Sigfrid Soli for helpful comments on earlier versions of the paper, and to Michèle Gentil for help with the French abstract.

References

[1] C.T. Best, B. Morrongiello and R. Robson, "The perceptual equivalence of two acoustic cues for a speech contrast is specific to phonetic perception", *Perception and Psychophysics,* Vol. 29, 1981, pp. 191–211.

[2] B. Delgutte, "Representation of speech-like sounds in the discharge patterns of auditory-nerve fibers", *J. Acoust. Soc. Am.,* Vol. 68, 1980, pp. 843–857.

[3] P.D. Eimas and J.L. Miller, "Contextual effects in infant speech perception", *Science,* Vol. 209, 1980, pp. 1140–1141.

[4] B. Espinoza-Varas, "Optional processing strategies in multidimensional discrimination: Integration versus selective attention", presented at the 97th Meeting of the Acoustical Society of America, Cambridge, MA, June 1979.

[5] H.L. Fitch, T. Halwes, D.M. Erickson and A.M. Liberman, "Perceptual equivalence of two acoustic cues for stop-consonant manner", *Perception and Psychophysics,* Vol. 27, 1980, pp. 343–350.

[6] H. Fujisaki and T. Kawashima, "On the modes and mechanisms of speech perception", *Annual Report of the Engineering Research Insititute* (Faculty of Engineering, University of Tokyo), Vol. 28, 1969, pp. 67–73.

[7] H. Fujisaki and T. Kawashima, "Some experiments on speech perception and a model for the perceptual mechanism", *Annual Report of the Engineering Research Institute* (Faculty of Engineering, University of Tokyo), Vol. 29, 1970, pp. 207–214.

[8] W.F. Ganong III, "Selective adaptation and speech perception", unpublished Ph.D. dissertation, M.I.T., 1977.

[9] W.R. Garner and G.A. Miller, "The masked threshold of pure tones as a function of duration", *J. Exp. Psychol.,* Vol. 37, 1947, pp. 293–303.

[10] J. Hillenbrand, "Auditory perception experiments with sine-wave analogs to the voice onset time dimension", J. Acoust. Soc. Am., Vol. 74, 1983 (in press).

[11] P.K. Kuhl, "Discrimination of speech by nonhuman animals: Basic auditory sensitivities conducive to the perception of speech-sound categories", *J. Acoust. Soc. Am.,* Vol. 70, 1981, pp. 340–349.

[12] P.K. Kuhl and J.D. Miller, "Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli", *J. Acoust. Soc. Am.,* Vol. 63, 1978, pp. 905–917.

[13] A.M. Liberman and M. Studdert-Kennedy, "Phonetic perception", in: R. Held, H. Leibowitz and H.-L. Teuber, eds., *Handbook of sensory physiology,* Vol. VIII. Springer, Heidelberg, 1978.

[14] L. Lisker and A.S. Abramson, "The voicing dimension: Some experiments in comparative phonetics", *Proc. 6th Int. Cong. Phon. Sci.,* Academia, Prague, 1970, pp. 563–567.

[15] L. Lisker, A.M. Liberman, D.M. Erickson, D. Dechovitz and R. Mandler, "On pushing the voice-onset-time (VOT) boundary about", *Language and Speech,* Vol. 20, 1977, pp. 209–216.

[16] J.D. Miller, C.C. Wier, R.E. Pastore, W.J. Kelly and R.J. Dooling, "Discrimination and labeling of noise-buzz sequences with varying noise-lead times: An example of categorical perception", *J. Acoust. Soc. Am.,* Vol. 60, 1976, pp. 410–417.

[17] J.L. Miller and P.D. Eimas, "Studies on the categorization of speech by infants", *Cognition*, Vol. 13, 1983, pp. 135–165.

[18] R.E. Pastore, "Possible psychoacoustic factors in speech perception", in: P.D. Eimas and J.L. Miller, eds., *Perspectives on the study of speech,* Erlbaum, Hillsdale, NJ, 1981.

[19] R.E. Pastore, W.A. Ahroon, K.J. Baffuto, C. Friedman, J.S. Puleo and E.A. Fink, "Common-factor model of categorical perception", *J. Exp. Psychol. (HPP),* Vol. 3, 1977, pp. 686–696.

[20] D.B. Pisoni, "Auditory and phonetic memory codes in the discrimination of consonants and vowels", *Perception and Psychophysics,* Vol. 13, 1973, pp. 253–260.

[21] D.B. Pisoni, "Identification and discrimination of the relative onset of two component tones: Implications for the perception of voicing in stops", *J. Acoust. Soc. Am.,* Vol. 61, 1977, pp. 1352–1361.

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

[23] B.H. Repp, "Two strategies in fricative discrimination", *Perception and Psychophysics,* Vol. 30, 1981, pp. 217–227.

[24] B.H. Repp, "Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception", *Psychol. Bull.,* Vol. 92, 1982, pp. 81–110.

[25] B.H. Repp, "Categorical perception: issues, methods, findings", in: N.J. Lass, ed., *Speech and language: Advances in theory and practice,* Vol. 10. Academic Press, New York, 1983.

[26] B.H. Repp, A.M. Liberman, T. Eccardt and D. Pesetsky, "Perceptual integration of acoustic cues for stop, fricative and affricate manner", *J. Exp. Psychol. (HPP),* Vol. 4, 1978, pp. 621–637.

[27] A.G. Samuel, "The effect of discrimination training on speech perception: noncategorical perception", *Perception and Psychophysics,* Vol. 22, 1977, pp. 321–330.

[28] M.E.H. Schouten, "The case against a speech mode of perception", *Acta Psychologica,* Vol. 44, 1980, pp. 71–98.

[29] A.M. Small, Jr., J.F. Brandt and P.G. Cox, "Loudness as a function of signal duration", *J. Acoust. Soc. Am.,* Vol. 34, 1962, pp. 513–514.

[30] S.D. Soli, "The role of spectral cues in discrimination of voice onset time differences", *J. Acoust. Soc. Am.,* Vol. 73, 1983, pp. 2150–2165.

[31] Q. Summerfield, "Differences between spectral dependencies in auditory and phonetic temporal processing: relevance to the perception of voicing in initial stops", *J. Acoust. Soc. Am.,* Vol. 72, 1982, pp. 51–61.

[32] Q. Summerfield and M. Haggard, "On the dissociation of spectral and temporal cues to the voicing distinction in initial stop consonants", *J. Acoust. Soc. Am.*, Vol. 62, 1977, pp. 436–448.

[33] C.C. Wood, "Discriminability, response bias, and phoneme categories in discrimination of voice onset time", *J. Acoust. Soc. Am.*, Vol. 60, 1976, pp. 1381–1389.

[34] B.H. Repp and A.M. Liberman, "Phonetic category boundaries are flexible", in: S.N. Harnad, ed., *Categorical Perception* (working title), Cambridge University Press, in preparation.