

# Perceptual integration and differentiation of spectral cues for intervocalic stop consonants

BRUNO H. REPP

Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06510

The place of articulation of intervocalic stop consonants is conveyed by temporally distributed spectral information, viz, the formant transitions preceding and following the silent closure interval (VC and CV transitions). Experiment 1 shows that more than 200 msec of silent closure is needed to hear VC and CV formant transitions as separate phonemic events (geminate stops). As closure duration is reduced, these cues are integrated into a single phonemic percept, and the VC transitions become increasingly redundant (Experiments 2 and 3). VC and CV transitions conveying different places of articulation, on the other hand, are heard as separate phonemes at closure durations as short as 100 msec. If closure duration is further reduced, a single stop is heard whose place of articulation corresponds to the CV transitions (Experiment 3). Even in the absence of CV transitions, VC transitions carry little perceptual weight at very short closure durations (Experiment 4). Despite their apparent redundancy, however, the VC transitions exert a positive bias on the perception of CV transitions at very short closure durations. At closure durations beyond 100 msec, on the other hand, VC and CV transitions interact contrastively in perception and tend to be heard as different phonemes (Experiments 5 and 6). The results of these experiments suggest two different processes of temporal integration in phonetic perception, one taking place at a precategorical level, the other combining identical phoneme categories within a certain time span.

An intervocalic stop consonant—for example, the /d/ in /ade/—is produced by rapidly constricting the vocal tract at a point appropriate for the place of articulation of the consonant (e.g., by bringing the tip of the tongue against the alveolar ridge), thereby interrupting the airflow from the lungs for a very brief period, and then releasing the constriction and rapidly returning the vocal tract to a more open configuration. The acoustic consequences of such an articulatory act are schematically illustrated in Figure 1. This stylized spectrogram of /ade/ begins with a steady-state formant pattern specifying the vowel /a/ and ends with a different steady state specifying the vowel /e/. The silent interval in the center represents the *closure* of the vocal tract. (In real utterances, the closure interval may be filled with low-amplitude periodicity when the stop consonant is voiced, but silence may be substituted in synthetic utterances without any significant perceptual consequences.) On each side of the closure period, the

formants change rapidly towards the pattern characteristic of the point of occlusion. The transitions from the initial vocalic portion into the closure will be called *VC transitions*; the transitions from the closure into the final vocalic portion, *CV transitions*.

Thus, the information about manner and place of articulation of an intervocalic stop consonant is conveyed partially by temporal information (the closure period) and partially by spectral information (formant transitions) preceding and following the closure interval. Although spectral information is necessarily distributed over time and thus always contains some temporal information as well, in this special case it

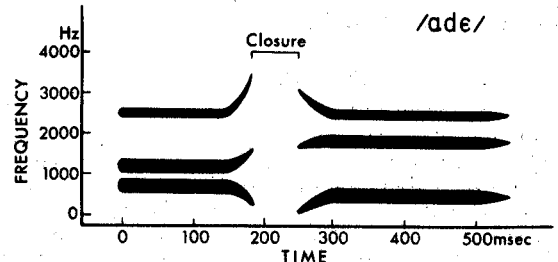


Figure 1. Schematic spectrogram of the synthesized VCV utterance /ade/.

This research was supported by NICHD Grant HD01994 to the Haskins Laboratories. Preliminary reports of these data appeared in Repp (Notes 5-7), where also more detailed discussions of some side issues may be found. I thank Alvin Liberman, Quentin Summerfield, and David Isenberg for helpful comments on an earlier draft.

consists of two distinct parts separated by silence, both conveying information about the same phoneme. Presumably, it is our implicit knowledge about articulatory events that makes us integrate these widely distributed acoustic cues into a single phonemic percept—that of a stop consonant embedded between two vowels.

It is important to note that, in general, both vocalic portions of a vowel-consonant-vowel (VCV) utterance are sufficient by themselves to identify the stop consonant. Thus, /adε/ may be considered as the VC syllable /ad/ followed by the CV syllable /dε/ after a brief silent interval. Obviously, there must be a temporal limit beyond which VC and CV transitions are no longer integrated into a single phonemic percept. When the closure duration exceeds this limit, a double or geminate consonant is heard (/ad-dε/), which is equivalent to hearing the two syllables, /ad/ and /dε/, separately and in sequence. The critical closure duration that separates single-consonant from double-consonant percepts—the *single-geminate boundary*—is an interesting parameter, since, in view of the exceptionally spread-out information, it may provide an estimate of the maximal interval over which cues for a single phoneme can be integrated.

Pickett and Decker (1960), in a tape-splicing experiment with natural utterances (TOPIC vs. TOP PICK) found the single-geminate boundary to be in the vicinity of 200 msec of closure duration at a normal rate of speech. (However, the boundary varied between 150 and 300 msec as a function of the rate of articulation of the sentence frame in which the test word was presented.) More recently, Fujisaki, Nakamura, and Imoto (1975) reported single-geminate boundaries of 164 and 169 msec for the synthesized Japanese utterances /ita/ vs. /itta/, in isolation and in a sentence frame, respectively. Both these studies examined only voiceless stop consonants. The present first experiment was conducted to determine the single-geminate boundary for voiced intervocalic stop consonants in isolated synthetic VCV utterances. Besides its general interest, this critical closure duration was needed as a guideline for later experiments reported here and elsewhere (Repp, Note 1), that employed similar stimuli.

## EXPERIMENT 1

### Method

**Subjects.** Nine volunteers were recruited from the Yale academic community and paid for their participation. They had little experience in listening to synthetic speech, although a few had participated in earlier experiments. All were native speakers of English. The author routinely served as an additional subject. Although being a highly experienced listener and a native speaker of German, his responses did not seem fundamentally different from those of the other subjects, except for being more accurate. Therefore, his results were included in the data reported below.

Table 1  
Onset and Offset Frequencies (Hertz) and Durations (Milliseconds) of Formant Transitions of the Stimuli Used in Experiments 1-4

Stim	F <sub>1</sub>			F <sub>2</sub>			F <sub>3</sub>		
	On	Off	Dur	On	Off	Dur	On	Off	Dur
/ab/	769	562	20	1232	1075	25	2525	2525	0
/ad/	769	386	30	1232	1465	30	2525	3363	30
/be/	208	537	50	1075	1845	45	1690	2525	40
/dε/	181	537	55	1845	1845	0	3026	2525	50
/bi/	208	286	40	1465	2307	45	2180	3026	35
/di/	208	286	40	2078	2307	35	3698	3026	40

**Stimuli.** Four basic VCV utterances were generated on the Haskins Laboratories parallel resonance synthesizer: /abε/, /abi/, /adε/, and /adi/. These three-formant stimuli consisted of two vocalic segments, 185 and 300 msec long, respectively, that were separated by a completely silent interval of variable duration (cf. Figure 1). The first (VC) segment (/ab/ or /ad/) consisted of an initial steady state followed by stepwise linear formant transitions. The second (CV) segment (/be/, /bi/, /dε/, or /di/) began with stepwise linear formant transitions and ended in a steady state. The onset and offset frequencies and durations of the formant transitions are listed in Table 1. All stimuli had the same constant fundamental frequency (90 Hz). Closure duration was varied between 65 and 515 msec in 10 steps of 50 msec. The resulting 40 stimuli were recorded in three different randomizations, with interstimulus intervals (ISIs) of 3 sec.

**Procedure.** To illustrate the single-geminate distinction, the four basic VCVs were first presented once with 65-msec closure periods and once with 515-msec closure periods. All subjects agreed that the latter contained two identical consonants, while the former contained only one. The subjects were then told that they would hear a series of intermediate cases and that they should write down "1" when they heard one consonant and "2" when they heard two. They listened over Telephonics TDH-39 earphones in a quiet room. The stimulus tapes were played back on an Ampex AG-500 tape recorder at a comfortable intensity.

## Results and Discussion

The single-geminate boundary was determined for each subject by linear interpolation. Although the single-geminate distinction does not occur within English words, all but two subjects gave reasonably consistent responses, and a boundary could be determined for all subjects. The mean boundary was at 213-msec closure duration, with a standard deviation of 28 msec across subjects. There was no significant variation as a function of either consonant (/b/ vs. /d/) or final vowel (/ε/ vs. /i/).

Since Pickett and Decker (1960) have shown that the single-geminate boundary is very sensitive to contextual influences such as speaking rate, the boundary estimate obtained here may well be valid only for the specific stimuli used, i.e., for voiced labial and alveolar stops in isolated synthetic syllables of a given duration and presented at a certain rate. Despite these limitations, the estimate is useful as a reference value within the present series of experiments.

## EXPERIMENT 2

Experiment 1 dealt with the temporal limit of cue integration. When integration occurs—at closure durations shorter than the single-geminate boundary—a second question emerges concerning the relative perceptual weights given to the spectral information preceding and following the closure period. There is evidence that CV transitions are perceptually more salient than VC transitions. Lisker (1957, Footnote 12) first observed that when a VC portion with transitions specifying one place of articulation is spliced together with a CV portion containing transitions appropriate for a different place of articulation, the resulting conflict of cues is resolved in favor of the CV transitions, given that the closure period is rather short. (Otherwise, a cluster of two different stop consonants is heard.) This perceptual primacy of the CV transitions has more recently been confirmed by Abbs (1971), Dorman, Raphael, and Liberman (Note 2), and Fujimura (Note 3). Massaro (1975) has ascribed it to backward masking (interruption of processing) of the VC transitions by the CV transitions. Alternatively, the effect may be considered as evidence of perceptual integration over time, with higher weighting of more recent information. Later in this paper, some findings will be reported that favor the second interpretation. At this time, we take note of the fact that the closure duration needed to hear two different stop consonants when phonetically conflicting VC and CV transitions are juxtaposed is only 60-80 msec (Dorman et al., Note 2), and thus much shorter than the single-geminate boundary. Thus, the perceptual limitation that prevents listeners from hearing VC transitions followed by conflicting CV transitions as a separate phoneme is different from the limitation that prevents them from hearing VC transitions as a separate phoneme when followed by compatible CV transitions. While the former limitation may conceivably be due to backward masking, the latter most likely is not. There is no reason why the temporal range of backward masking (presumably a precategorical form of interference) should depend on the phonetic compatibility of the stimuli involved.

Given that compatible VC and CV transitions are heard as a single phoneme at closure durations shorter than 200 msec, and given that VC transitions are perceptually less salient than CV transitions in a conflicting situation, we may ask whether VC transitions followed by compatible CV transitions have any perceptual significance at all. One way of finding out is to ask listeners to discriminate VCV utterances from similar utterances in which the VC transitions have been removed. If the only factor limiting the detection of the presence (or absence) of VC transitions were a backward masking effect of CV transitions on VC transitions at very short closure durations,

then we should expect listeners to be able to discriminate utterances with and without VC transitions as soon as masking ceases, i.e., at closure durations beyond 60-80 msec. However, utterances with and without VC transitions presumably have identical phonemic structure (VCV) until closure duration reaches the single-geminate boundary. Knowing the difficulty of subphonemic discrimination from studies of categorical perception (e.g., Pisoni, 1971), we might predict that listeners will find such utterances difficult to tell apart up to closure durations of 200 msec. This prediction was tested in Experiment 2.

## Method

**Subjects.** The same subjects as in Experiment 1 participated.

**Stimuli.** The same basic stimuli as in Experiment 1 were used, except that closure duration was varied from 65 to 265 msec in steps of 25 msec (omitting the 240-msec interval). For each of the resulting 32 VCV stimuli, a corresponding stimulus without VC transitions (V-CV) was synthesized by replacing the transitions with the steady-state formant frequencies of the initial vowel /a/, leaving all amplitude settings unchanged. VCV stimuli were paired with their V-CV counterparts in AXB discrimination triads. Each pairing occurred in all four AXB configurations (AAB, ABB, BAA, BBA), resulting in 128 stimulus triads that were recorded with ISIs of 1 sec within triads and 3 sec between triads.

**Procedure.** This experiment was the last of three conducted in a single experimental session of about 90 min duration. Both earlier experiments (the present Experiment 1 and Experiment V of Repp, Note 1), involved similar stimuli. A practice series of 16 AXB triads was presented first, containing only stimuli with closure durations of 265 msec. The subjects were instructed to write down "A" whenever the second utterance in a triad was equal to the first, and "B" when it was equal to the third utterance. The nature of the difference to listen for was pointed out, and all subjects agreed after listening to the practice trials that they knew where and what the relevant difference was. They were warned that the experimental trials might be considerably more difficult.

## Results

One subject performed at chance level at all closure durations; her data were excluded. The average results of the remaining nine subjects are shown in Figure 2. Error percentages are shown as a function of closure duration, separately for stimuli containing /b/ (B) and stimuli containing /d/ (D). Fifty percent errors represents chance performance. It can be seen that errors decreased as closure duration increased [ $F(7,56) = 8.0, p < .001$ ] and that more errors were made with B than with D stimuli [ $F(1,8) = 24.0, p < .001$ ]. Performance was better than chance at all closure durations, except perhaps for B stimuli with the shortest closure (65 msec). However, the task was clearly a difficult one, and there was no evidence to support the hypothesis that errors would asymptote around 60-80 msec of closure duration. Instead, a minimum in the error function was reached only at 165 msec. Surprisingly, errors increased subsequently, so that, instead of reaching its minimum around 200 msec (the single-geminate boundary), as expected, the error function showed a local peak in this region.

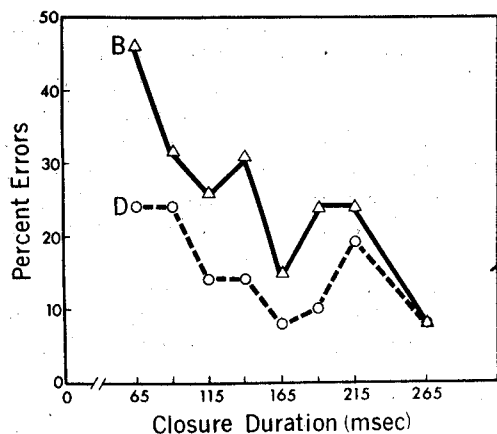


Figure 2. AXB discrimination of VCV and V-CV stimuli: Percentage of errors as a function of closure duration, shown separately for stimuli containing a /b/ (B) and stimuli containing a /d/ (D).

A substantial percentage of errors was still obtained at the longest closure interval (265 msec).

### Discussion

The relative difficulty of the task (which was experienced subjectively to an even greater degree than the results show) proved that the VC transitions carry little perceptual weight when followed by phonetically compatible CV transitions. With one possible exception (/adi/, according to informal observations), the intelligibility of the utterances was unaffected by removal of the VC transitions. (This may be contrasted with the dramatic effect of removing the CV transitions; see Experiment 4 below.) However, Repp (Note 1, Experiment V) has demonstrated in a reaction-time task that the phonetic information conveyed by VC transitions that are followed by compatible CV transitions can nevertheless be utilized by listeners to make fast decisions about the identity of the medial stop consonant, at least at closure durations beyond 65 msec. Thus, while it may be the case that CV transitions interrupt the processing of VC transitions at very short closure durations, such masking does not occur at longer closure durations. Repp's (Note 1) data suggest that VC and CV transitions are both phonetically interpreted and subsequently integrated into a single phonemic percept. Apparently, it is this integrative process, together with the higher perceptual weight of CV transitions, that makes VCV and V-CV utterances difficult to distinguish, particularly at closure durations beyond 60-80 msec, where a backward masking effect does not operate.

The difference between B and D stimuli is best explained by the acoustic structure of the stimuli: the former happened to have shallower VC transitions

of both the first and the third formant than the latter (cf. Table 1).<sup>1</sup> Therefore, labial VC transitions were probably more difficult to detect as an auditory event. Of course, the discrimination required here was basically auditory in nature; only beyond the single-geminate boundary was a phonetic distinction available (e.g., /a-bi/ vs. /ab-bi/). At closure durations shorter than 200 msec, the task amounted to detecting a slight discontinuity within the V-CV utterances in an AXB triad; this was facilitated by pairing these stimuli with VCVs containing relatively extreme VC transitions.

The unexpected increase in errors in the region of the single-geminate boundary might represent a change in the subjects' strategy: Instead of relying on the auditory distinction just described, they may, instead, have paid attention to the phonetic contrast that became available but was still unreliable around the single-geminate boundary. Such a change in strategy may have been precipitated by the occurrence of a geminate percept early in an AXB triad. In particular, the increase in errors may reflect a tendency (suggested by the author's experience as a listener) to misperceive V-CV stimuli as containing geminates at closure durations around 200 msec.

Due to this local peak in the error function and the relatively high error rates at the longest closure durations used here, the hypothesis that the error function would reach its asymptote at approximately the single-geminate boundary did not receive clear support, but the data were certainly closer to this hypothesis than to the alternative prediction of an asymptote around 60-80 msec. Thus, the results demonstrate the difficulty of making auditory discriminations in the presence of phonemic identity. On the other hand, there was a clear increase in performance as closure duration increased, indicating that the perception of these stimuli was not categorical in a strict sense. Either it was the case that the auditory differences were unusually salient (relative to other types of subphonemic differences), or perhaps the differences being discriminated (VCV vs. V-CV) should be considered phonetic in nature (and hence more easily accessible to the listeners than purely auditory differences) despite the phonemic identity of the utterances—involving, as they did, different "offset qualities" of the initial vowel.

### EXPERIMENT 3

The purpose of the next experiment was to examine performance in the discrimination task of Experiment 2 at shorter closure durations (0-100 msec), and to compare the results with those obtained from stimuli differing with regard to the presence or absence of VC transitions incompatible with the

following CV transitions. If the backward masking (interruption) hypothesis is correct, then it should not matter whether the VC transitions whose presence (or absence) is to be detected are phonetically compatible or incompatible with the CV transitions. Thus, similarly poor performance would be predicted in the two conditions at very short closure durations. A third condition was included in which utterances with compatible VC and CV transitions were to be discriminated from utterances with incompatible VC and CV transitions, the difference always being located in the VC portion. This condition provided a test of the robustness of the interference effect described in the introduction to Experiment 2: Given that only a single consonant is heard when conflicting VC and CV transitions are separated by a very short closure interval, does the utterance sound natural, or is there some auditory residue of the VC transitions that would enable listeners to discriminate these stimuli from utterances with compatible VC and CV transitions? (Cf. Rudnicky and Cole, Note 4, for a description of such an "auditory residue" effect.)

## Method

**Subjects.** Ten new volunteers and the author participated as subjects.

**Stimuli.** The four basic VCV stimuli (denoted here by  $VC_i-C_jV$ , to indicate the phonetic compatibility of VC and CV transitions) were again used, with closure durations of 0, 25, 50, 75, 100, and 250 msec, the longest value being used for practice only. Corresponding  $V-C_jV$  utterances were generated as in Experiment 2. A third set (denoted  $VC_i-C_jV; i \neq j$ ) was created by interchanging the VC portions of  $VC_i-C_jV$  stimuli with different consonants: /ab-di/, /ab-de/, and /ad-bi/, /ad-be/. Each utterance type was paired with each of the others, resulting in three discrimination conditions:  $VC_i-C_jV$  vs.  $V-C_jV$  (compatible vs. absent VC transitions, as in Experiment 2),  $VC_i-C_jV$  vs.  $VC_i-C_jV$  (compatible vs. conflicting), and  $V-C_jV$  vs.  $VC_i-C_jV$  (absent vs. conflicting). In each condition, there were 80 AXB triads, resulting from four stimuli with five closure durations in four AXB configurations. The stimuli in a triad always had the same closure duration and the same CV portion; they differed only in the acoustic information immediately preceding the closure interval. Each discrimination series was randomized and recorded as a separate block, preceded by 16 practice trials (stimuli with 250-msec closure intervals). The within-triad ISI was 1 sec, the between-triad ISI, 3 sec.

**Procedure.** This experiment was preceded by the brief test described as Experiment 4 below. The sequence of the three discrimination conditions was approximately counterbalanced across subjects. The structure of the stimuli was explained before each condition, so that the subjects knew quite well what they were listening to and what they were trying to discriminate. Other procedural details were the same as in the earlier experiments.

## Results

Not unexpectedly, the  $VC_i-C_jV$  vs.  $V-C_jV$  condition was the most difficult. This was already evident in the practice trials, where the average error rate was 16.5%, as compared to 6.8% and 6.3%, respectively, in the other two conditions. The results are shown in Figure 3. The  $VC_i-C_jV$  vs.  $V-C_jV$  condition was

more difficult than the  $VC_i-C_jV$  vs.  $V-C_jV$  condition, which, in turn, was more difficult than the  $VC_j-C_jV$  vs.  $VC_i-C_jV$  condition. Error rates declined steadily as closure duration increased, but were still considerably above the practice trial error rate at the longest closure duration, suggesting that errors would have further decreased if closure durations had been extended beyond 100 msec.

A four-way analysis of variance showed not only highly significant effects of discrimination condition [ $F(2,20) = 17.3, p < .01$ ] and closure duration [ $F(4,40) = 38.1, p < .001$ ], but also a significant effect of the final vowel [ $F(1,10) = 7.9, p < .05$ ] and a significant Condition by Consonant (CV transitions) interaction [ $F(2,20) = 13.6, p < .01$ ]. Since none of these effects interacted with closure duration, the data were collapsed over this factor, and each condition was analyzed separately in a two-way analysis of variance, with consonant and final vowel as factors. In the  $VC_j-C_jV$  vs.  $V-C_jV$  condition, there was a significant effect of consonant [ $F(1,10) = 13.1, p < .01$ ], which replicated the effect found in Experiment 2 (cf. Figure 2): For B stimuli, performance remained at chance level up to a closure duration of 50 msec, while for D stimuli, the error percentage decreased almost linearly with increasing closure duration. A similar consonant effect was found in the  $VC_i-C_jV$  vs.  $V-C_jV$  condition [ $F(1,10) = 8.9, p < .02$ ]: Here, too, the presence (absence) of labial VC transitions was more difficult to detect than that of alveolar VC transitions. In theory, the consonant effects in these two conditions could have been due to opposed effects of the CV transitions (cf. the statistical interaction between conditions and

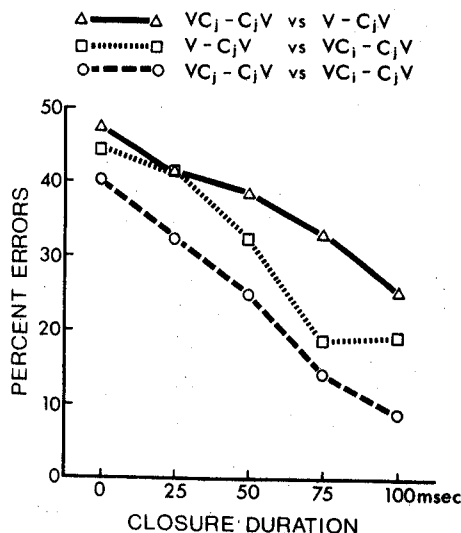


Figure 3. Three AXB discrimination tasks: Percentage of errors as a function of closure duration.

consonants referred to above). However, in the VC<sub>i</sub>-C<sub>j</sub>V vs. VC<sub>i</sub>-C<sub>j</sub>V condition, where labial and alveolar VC transitions were to be discriminated from each other on each trial and only the CV transitions varied from trial to trial, there was no significant consonant effect; this indirectly confirmed that both consonant effects in the other two conditions were due to a difference in detectability between labial and alveolar VC transitions. Effects of the final vowel were only marginally significant, irregular, and difficult to interpret.

### Discussion

The results of the VC<sub>i</sub>-C<sub>j</sub>V vs. VC<sub>i</sub>-C<sub>j</sub>V discrimination task confirm the interference effect found earlier by Dorman et al. (Note 2) and others in identification tasks. As closure duration was reduced, VC<sub>i</sub>-C<sub>j</sub>V stimuli sounded more and more like VC<sub>j</sub>-C<sub>j</sub>V stimuli, i.e., only one consonant was heard whose perceived place of articulation was determined by the CV transitions. The high error rates obtained here show that the effect was robust enough to hold up in a discrimination paradigm. However, discrimination performance was somewhat better than chance even at zero closure duration. This may indicate that some residual auditory cues remained of the VC transitions in VC<sub>i</sub>-C<sub>j</sub>V stimuli. Alternatively, the VC transitions may have escaped interference on a small number of trials which increased with closure duration. In the absence of parallel identification data, these two possibilities cannot be distinguished.

The poor performance at very short closure durations is in agreement with the recognition backward masking hypothesis (Massaro, 1975), and so is the fact that, at the two shortest closure durations, the detectability of the VC transitions did not seem to depend on whether they were phonetically compatible with the CV transitions (triangles vs. squares in Figure 3). As closure duration increased, the difference in phonemic structure between VC<sub>i</sub>-C<sub>j</sub>V and VC<sub>j</sub>-C<sub>j</sub>V (and V-C<sub>j</sub>V) stimuli began to emerge as the interference effect subsided, leading to a rapid reduction in discrimination errors. In distinguishing VC<sub>j</sub>-C<sub>j</sub>V from V-C<sub>j</sub>V utterances, on the other hand, the listeners presumably had to rely on subphonemic cues throughout; hence the poorer performance in this condition. The result that VC<sub>i</sub>-C<sub>j</sub>V stimuli were more easily discriminated from VC<sub>j</sub>-C<sub>j</sub>V stimuli than from V-C<sub>j</sub>V stimuli was probably due to the greater auditory difference in the first case. Thus, the listeners apparently made use of both auditory and phonetic differences in discriminating the stimuli.

The present results suggest a somewhat longer time range of the interference effect than the data of Dorman et al. (Note 2). While the earlier results indicated a range of 60-80 msec, the asymptotes of the present error functions seemed to lie beyond 100 msec.

## EXPERIMENT 4

Experiments 2 and 3 demonstrated the low perceptual salience of the VC transitions in VCV (VC<sub>i</sub>-C<sub>j</sub>V) utterances. When closure duration is short, the VC transitions can be removed without affecting the intelligibility of the stop consonant; in fact, their absence is hardly perceived. Presumably, removal of the CV transitions from a VCV utterance would have a more detrimental effect on perception. This was to be demonstrated in Experiment 4.

Experiment 4 also bears on the question of whether the interference of CV transitions with the perception of phonetically conflicting VC transitions is a direct consequence of phonetic incompatibility on the place of articulation dimension, or whether it reflects a more general auditory interference. To the extent that a steady-state vowel following VC transitions after a short interval of silence interferes with their perception, it may be argued that interference takes place at a prephonetic level. To investigate this issue, Experiment 4 presented stimuli without CV transitions (VC-V stimuli) in a forced-choice identification task.

### Method

**Subjects.** The same subjects as in Experiment 3 participated. (Experiment 4 preceded Experiment 3 in the same single session.)

**Stimuli.** The four basic VCV stimuli were again used. However, their CV transitions were replaced with steady-state formants appropriate for the final vowel, leaving all amplitude settings unchanged. Thus, /ab/ and /ad/ were followed by either /ε/ or /i/ after a variable silent "closure" interval. Closure durations ranged from 0 to 125 msec in 25-msec steps. The resulting 24 VC-V stimuli were recorded in five different randomizations with ISIs of 3 sec. A random series of 10 /ab/ and 10 /ad/ syllables preceded the test series.

**Procedure.** After identifying the VC syllables in the short practice list, the subjects listened twice to the VC-V identification series. In Run 1, they were instructed to write down "B" when they heard /ab-ε/ or /ab-i/, "D" when they heard /ad-ε/ or /ad-i/, and "O" when they heard no consonant at all, i.e., /a-ε/ or /a-i/. In Run 2, null responses were no longer permitted, and a forced choice between B and D had to be made for each stimulus.

### Results

All subjects identified the 20 practice VC syllables without difficulty. (Only a single error was committed.) The results of the first run through the VC-V identification test are shown in Figure 4a. It can be seen that two-thirds of the stimuli with zero closure interval received null responses. The percentage of null responses was much reduced at a closure duration of 25 msec and then declined further as closure duration increased. The difference between the total error rate (solid line) and the percentage of null responses (dotted line) represents the percentage of confusion errors on the place of articulation dimension; surpris-

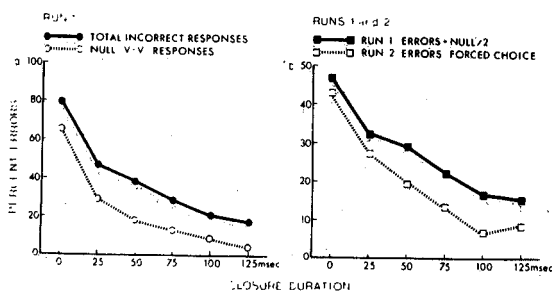


Figure 4. Results of the VC-V identification task: Error percentages as a function of closure duration.

ingly, this percentage did not change at all with closure duration.

Estimates of /b/-/d/ discrimination performance in Run 1 may be obtained by assuming that, if forced to guess instead of responding 0 (null), the subjects would have been correct on 50% of the trials. These estimates are shown in Figure 4b together with the results of Run 2 (where null responses were not permitted). It can be seen that performance was close to chance when there was no closure interval at all; however, performance improved rapidly as closure duration increased. An asymptote seemed to be reached at a closure duration of 100 msec; note, however, that the asymptotic error rate was much higher than for VC syllables in isolation, which were perfectly discriminated! Performance in Run 2 was better than in Run 1. This may reflect not only a genuine improvement due to practice, but also the possible incorrectness of the assumption that all null responses in Run 1 were equivalent to random guesses.

A four-way analysis of variance was performed on the data in Figure 4b, with consonant and final vowel as additional factors. The effects of runs [ $F(1,10) = 26.6, p < .01$ ] and of closure duration [ $F(5,10) = 33.5, p < .001$ ] were highly significant. In addition, there was a significant effect of consonant [ $F(1,10) = 5.5, p < .05$ ] and a highly significant Consonant by Vowel interaction [ $F(1,10) = 17.2, p < .01$ ]. When followed by a steady-state vowel, /ab/ was easier to identify than /ad/, especially at longer closure durations, where /ad/ stimuli were solely responsible for the high error rates. In addition, /ad/ was much more difficult to identify when followed by /i/ than when followed by /ε/, but /ab/ was more difficult to identify when followed by /ε/ than when followed by /i/.

## Discussion

The results of Experiment 4 confirm the perceptual importance of the CV transitions: While removal of the VC transitions from a VCV utterance has a negligible effect, so that their absence can hardly be detected (Experiments 2 and 3), removal of the CV

transitions dramatically reduced the identifiability of the medial consonant. The poor identification performance at short closure durations shows that it is not absolutely necessary to follow VC transitions with conflicting CV transitions to obtain interference—a steady-state vowel is sufficient to interfere with their perception.<sup>2</sup> The “Run 2” function in Figure 4b is similar to the VC<sub>i</sub>-C<sub>j</sub>V vs. VC<sub>i</sub>-C<sub>j</sub>V discrimination function in Figure 3, indicating a comparable time course of the VC<sub>i</sub>-C<sub>j</sub>V and VC-V interference effects. The similarity of the two effects suggests that the interference does not take place in a mechanism specialized for the perception of stop consonants or place of articulation, since a steady-state vowel presumably does not engage this processing mechanism.<sup>3</sup> Rather, the interference effect seems to represent a more general form of backward masking, due to either interruption or integration. The results are compatible with either of these two familiar versions of the backward masking hypothesis. Thus, the stimulus portion following the silent closure interval either interrupts the phonetic processing of the VC transitions, so that information is lost, or the phonetic information preceding and following the closure is integrated prior to phonetic categorization, with heavy dominance of post-closure information. In the latter case, the nonconsonantal nature of the final steady-state vowel in VC-V utterances would naturally make the medial stop difficult to identify.

The consonant effect in the present data was in the opposite direction from that observed earlier: In Experiments 2 and 3, when CV transitions followed the closure period, labial VC transitions were more difficult to detect than alveolar VC transitions; however, they were easier to identify than alveolar transitions when followed by a steady-state vowel (Experiment 4). Inspection of the data from Run 1 showed that the difference between consonants was primarily due to misidentification of /d/ as /b/, i.e., to place of articulation confusions, and not so much to null responses (which reflected failures to detect the cues for stop consonant manner). This is reasonable, since stop manner was primarily signaled by the presence of a closure interval (which did not depend on the nature of the preceding transitions), while place of articulation was conveyed by the VC transitions. It was surprising that there was virtually no decline in place confusions as closure duration increased, given that the VC portions could be perfectly identified in isolation. Repp (Note 1, Experiment VI) found high confusion rates in analogous stimuli at even longer closure durations. At present, there is no explanation for this finding. The significant consonant-vowel interaction requires further research; it suggests that either the final vowels contributed misleading place-of-articulation information, or that psychoacoustic factors, such as



the relative continuity of formants across the silent interval, played a role.

## EXPERIMENTS 5 AND 6

The preceding experiments demonstrated that perceptual interference occurs when the spectral cues preceding and following a short closure interval are brought into conflict. Experiments 5 and 6 were concerned with perceptual interactions between VC and CV transitions beyond the presence or absence of interference. For example, we know that, in the perceptual competition between conflicting VC and CV transitions at short closure durations, the latter generally win. Is the phonetic information conveyed by the VC transitions completely lost, as the interruption hypothesis would predict? Or do the VC transitions affect the perception of the CV transitions, even if the former are not heard as a separate phonemic event, as might be expected if integration occurs?

In order to answer this question, the following procedure was devised: An acoustic continuum of CV stimuli ranging from /bɛ/ to /dɛ/ was constructed by varying the CV formant transitions in a number of steps between endpoint patterns characteristic of labial and alveolar places of articulation, respectively. The phoneme boundary on this continuum was determined for the stimuli in isolation and when immediately preceded by a VC portion with labial or alveolar formant transitions. To further increase the sensitivity of the design, the subjects used a 6-point rating scale to judge the "category goodness" of the stop consonant heard. If the processing of the VC transitions were truly interrupted by the CV transitions, the perception of the CV transitions should remain unaffected by the nature of the VC transitions. So, for example, CV stimuli chosen to be ambiguous between /bɛ/ and /dɛ/ when presented in isolation should be just as ambiguous when preceded by /ab/ or /ad/ at a very short closure period; i.e., the resulting VCV stimulus should be ambiguous between /abɛ/ and /adɛ/. If, on the other hand, VC cues are integrated with CV cues, the preceding VC stimulus should bias the perception of place of articulation in a positive direction; i.e., listeners should be more likely to hear /abɛ/ when the precursor is /ab/, and /adɛ/ when the precursor is /ad/. According to this latter view, phonetic categorization is based on integration of auditory information over time, with more weight given to the more recent cues.

When the closure period is extended beyond 100 msec, say (but not beyond 200 msec), conflicting VC and CV transitions are perceived as separate phonemes, while compatible transitions are still perceived as a single phoneme. The present paradigm permits us to investigate whether there are any

perceptual interactions between VC and CV transitions at these longer temporal separations. If a CV portion ambiguous between /bɛ/ and /dɛ/ is preceded by /ab/, will the stimulus be perceived more often as /abɛ/, or as /ab-dɛ/, or will it be ambiguous between the two? It was considered possible that a contrastive effect would be obtained in this situation, i.e., that the VC transitions might bias the perception of the CV transitions toward a different phoneme. Such an effect would be in accord with various context, anchoring, and adaptation effects observed in other paradigms involving speech sounds (e.g., Ades, 1976; Cooper, 1975; Simon, 1977). However, Ades (1974) has reported that no cross-adaptation effects exist between VC and CV syllables. Moreover, the fact that, at intermediate closure durations, phonetically compatible VC and CV transitions are still integrated into a single phonemic percept suggests an integrative tendency that may be reflected in an assimilative (positive) interaction between the spectral cues. Thus, the answer to the question posed at the beginning of this paragraph was quite uncertain before Experiments 5 and 6 were conducted. In any case, the single-geminate boundary (~200 msec of closure duration) was expected to constitute the limit to the range of whatever perceptual interactions might occur across the closure interval.

Perceptual bias may operate not only in a forward direction (from VC transitions to CV transitions), but also backward in time; that is, the perception of the VC transitions may be influenced by the CV transitions, even when the VC transitions are perceived as a separate phoneme. This issue was investigated by following VC stimuli from an /ab/-/ad/ continuum with either /bɛ/ or /dɛ/, after appropriate periods of silence, and by determining whether the perception of ambiguous VC transitions was influenced by the CV transitions. Backward effects are especially interesting because they indicate suspension or revision of decisions about earlier perceptual cues in the light of later information, and therefore are a direct indicator of temporal integration in phonetic processing. Forward effects, on the other hand, may contain a perseverative (adaptation or memory) component that distorts the temporal range of the perceptual interaction.

Experiments 5 and 6 were designed as a first exploration of perceptual interactions between VC and CV stimuli in the experimental space defined by the factors of closure duration and direction in time (forward vs. backward). The two studies are reported together, since they used the same stimuli and generally complemented each other, although different groups of subjects were used.

## Method

**Subjects.** Ten new volunteer subjects participated in Experiment 5, and nine different subjects in Experiment 6. (The author



also served as a subject, but—in contrast to the earlier experiments and for reasons not entirely clear—his data deviated considerably from those of the less experienced subjects and therefore were not included in the group data.)

**Stimuli.** Two continua of seven stimuli were constructed by varying the offset (onset) frequencies of the second and third formants in equally spaced steps: a VC continuum from /ab/ to /ad/, and a CV continuum from /be/ to /de/. The offset (onset) frequencies of the stepwise-linear formant transitions are listed in Table 2. They were selected so that the phoneme boundaries would fall approximately in the center of each continuum. All other spectral and temporal characteristics of the stimuli were constant across each continuum and identical with the specifications of the stimuli in the earlier experiments. (The transitions of the first formant in the VC series were the same as for /ab/ in the earlier experiments; see Table 1.)

Experimental conditions will be abbreviated as F (forward) or B (backward) followed by a number indicating the closure duration in milliseconds.

Experiment 5 included the following conditions:

(1) *F control*. This was a random series of 75 CV syllables, consisting of five repetitions of a (1,2,3,3,3,2,1) frequency distribution of the seven stimuli from the CV continuum. This distribution was used to obtain more reliable information in the region of the phoneme boundary; it was maintained in all other conditions. The F control condition provided baseline data for the identification of CV syllables in isolation.

(2) *F25*. This series of 150 stimuli consisted of the same CVs preceded by either /ab/ or /ad/, the two endpoint stimuli of the VC continuum, with a closure interval of 25 msec intervening. The sequence was arranged in successive blocks of 30 stimuli, each containing one cycle of all VC-CV combinations, with the basic frequency distribution described above. The F25 condition served to investigate whether the VC transitions affect the perception of the CV transitions despite being perceptually dominated by them.

(3) *F265*. Analogous to F25, but with a closure duration of 265 msec. This condition was intended to establish that there are no forward effects at a closure duration beyond the single-geminate boundary.

(4) *B control*. Analogous to the F control, with VC stimuli replacing CV stimuli. The corresponding baseline condition.

(5) *B115*. Analogous to the F25 and F265 conditions, with reversed roles of VC and CV stimuli and a closure duration of 115 msec. Thus, each stimulus from the VC continuum was followed by either /be/ or /de/, the two endpoint stimuli of the CV continuum. This condition investigated whether the CV por-

tion affected the perception of the VC portion at a closure interval near the end of the temporal range of the interference effect.

(6) *B265*. Analogous to B115, but with a closure duration of 265 msec. This condition was intended to establish that no backward effects exist at a closure duration beyond the single-geminate boundary.

Experiment 6 included the following conditions:

(1) *F control*. As in Experiment 5.

(2) *F15*. A replication of the F25 condition with an even shorter closure duration. In contrast to the F25 condition, this sequence included four different precursors: stimuli 1, 3, 6, and 8 from the VC continuum.<sup>4</sup> The stimuli were arranged in three blocks of 60, each block containing one cycle of all VC-CV combinations, with the basic frequency distribution described earlier. (The rationale for using four precursors will be explained below.)

(3) *B control*. As in Experiment 5.

(4) *B140*. A replication of B115, with a closure duration that was presumably beyond the range of the interference effect.

(5) *F140*. Analogous to F25 and F265, but with an intermediate closure duration. This condition examined forward effects beyond the temporal range of interference.

**Procedure.** All subjects received the conditions in the order described. (The F control and F15 conditions in Experiment 6 were repeated once.) In the control conditions, subjects were instructed to rate each consonant on a "D-ness" scale ranging from 1 to 6, where 1 represented "a very clear B," 3 "ambiguous, but more like a B," 4 "ambiguous, but more like a D," and 6 a "very clear D." The subjects were urged to use the extreme ratings at least occasionally, i.e., to make their judgments according to the relative goodness of the stimuli and not according to how they compared with real speech. Each test series was preceded by five alternating tokens of the two endpoint stimuli, and the subjects were told that these were the "best" stimuli they would hear.

In the other conditions, the subjects were asked to maintain the criteria established during the control series, i.e., to give generally poorer ratings if all stimuli sounded poorer and generally better ratings if all stimuli sounded better than in the control series. In the F15 and F25 conditions, the subjects were merely told that each CV syllable would be preceded by the vowel /a/; nothing was mentioned about a second consonant. In the F265 condition, the subjects were told that each CV syllable would be preceded by either /ab/ or /ad/; they were asked to ignore the first consonant heard and evaluate the second consonant only. Conversely, in the B265 condition, subjects rated the first consonant heard, ignoring the second one.

In the conditions with intermediate closure durations (F115, F140, B140), the subjects did not give ratings but merely indicated whether they heard a single consonant (/abε/ or /adε/) or two different consonants (/ab-de/ or /ad-be/) by writing down "1" or "2."

## Results and Discussion

**Forward conditions.** The results of the F25 condition shall be discussed first; they are shown in Figure 5b. The dotted line connecting triangles represents the average "D-ness" ratings of the CV syllables in isolation (F control). The two solid functions represent the average ratings of the same syllables when preceded by /ab/ or /ad/, respectively. It is obvious that the VC precursors had an effect on the subjects' responses; the difference between the two solid functions was highly significant [ $F(1,9) = 45.9, p < .001$ ]. The effect was assimilative, as expected. Thus, subjects were more likely to hear /abε/ when the precursor was /ab/, and /adε/ when the precursor was /ad/, even though the precursors themselves did not give

Table 2  
Transition Offset (Onset) Frequencies in the  
VC and CV Continua

Stimulus Number	VC Series Transition Offset (Hz)		CV Series Transition Onset (Hz)	
	F <sub>2</sub>	F <sub>3</sub>	F <sub>2</sub>	F <sub>3</sub>
1	996	2180	1312	2018
2	1075	2348	1386	2180
3	1155	2525	1465	2348
4	1232	2694	1541	2525
5	1312	2862	1620	2694
6	1386	3026	1695	2862
7	1465	3195	1772	3026
8*	1541	3363		
Steady States	1232	2525	1845	2525

\*Used only as a precursor (F15 condition).

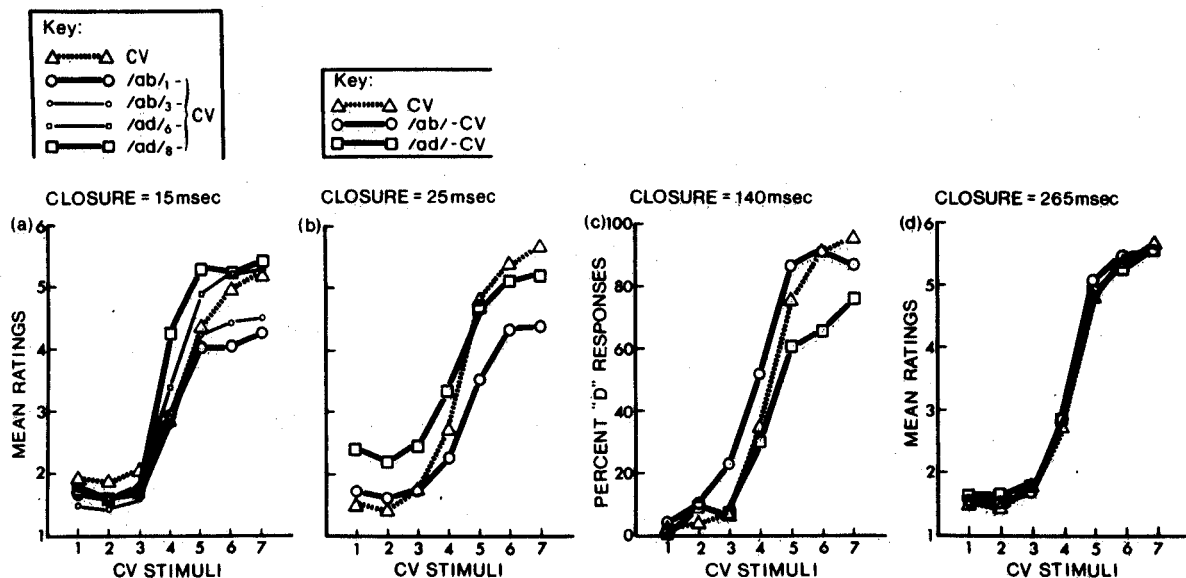


Figure 5. Results of forward conditions: F15, F25, F140, and F265. Each panel shows the results for CV syllables in isolation (dotted line) and when preceded by different precursors (solid lines). Different groups of subjects participated in (F15, F140) and (F25, F265).

rise to a separate phonemic percept. It can be seen that the rating function for isolated CV syllables was steeper than the two precursor functions. This was a consequence of the fact that only VC transitions conflicting with the CV transitions had a substantial influence on the ratings; phonetically compatible VC transitions did not improve the category goodness of the consonants heard relative to the control condition (although there was some room for improvement), but instead lowered it slightly. This outcome was probably due to the unnaturally short closure periods of the stimuli and the absence of any modeling of coarticulation effects that may occur in natural speech across the closure period (Öhman, 1966).

When the results shown in Figure 5b were plotted in terms of the percentage of D responses (i.e., ratings between 4 and 6), exactly the same pattern of results was obtained. This raised the question of whether the VC transitions exerted a true biasing effect on the CV transitions, or whether they simply dominated the CV transitions on a certain number of trials.<sup>5</sup> For this purpose, separate analyses of variance were conducted on B and D responses including (the first or last) four positions on the CV continuum. Both B and D responses showed significant effects of position [ $F(3,27) = 9.3$ ,  $p < .01$ , and  $F(3,27) = 6.7$ ,  $p < .01$ , respectively] indicating that the subjects were sensitive to acoustic within-category variations. The precursor effect was significant for D responses [ $F(1,9) = 9.5$ ,  $p < .025$ ], but it fell short of significance for B responses. Thus, the /d/ in utterances perceived as /ade/ received higher ratings when the

precursor was /ad/ than when it was /ab/. This result gives moderate support to the hypothesis that the VC transitions indeed exerted a continuous bias on the perception of the CV transitions, and that the phonemic decisions were based on temporal integration of spectral cues. Backward masking of the VC transitions through interruption of processing seems less likely now.

Although the assimilative effect obtained in the F25 condition was most likely due to temporal integration at a precategorical level, it was considered possible that it represented some form of phonetic bias resulting from covert categorization of the VC transitions. In order to rule out this hypothesis, the F15 condition used four acoustically different VC precursors, two from each category. If the assimilative effect is mediated by a covert categorization of the VC transitions, then acoustic within-category differences between precursors should have little effect. On the other hand, if the assimilation effect is due to precategorical integration of auditory cues, the extent of the bias should vary with the acoustic characteristics of the precursors; thus, a precursor with more extreme transitions should have a stronger effect on perception than a precursor with less extreme transitions.

The results of the F15 condition are shown in Figure 5a. For reasons that are not clear, the VC precursors had no differential effect at the /b/-end of the CV continuum, except for shifting the ratings slightly towards /b/. Towards the /d/-end of the continuum, however, a strong assimilative precursor

effect emerged, in agreement with the earlier F25 results. In addition, those precursors that were closer to the /ab/-/ad/ boundary tended to produce a weaker effect than the precursors from the ends of the VC continuum, as predicted by the perceptual integration hypothesis.

The statistical analysis of the F15 results showed a significant between-category effect of VC precursors (precursors 1 and 3 vs. 6 and 8) [ $F(1,8) = 8.9$ ,  $p < .025$ ]; an interaction of this effect with position on the CV continuum [ $F(4,32) = 4.0$ ,  $p < .025$ ] which reflected the absence of a precursor effect at the /b/-end of the continuum; a significant interaction of the within-category precursor effect (precursors 1 and 6 vs. 3 and 8) with position [ $F(4,32) = 4.6$ ,  $p < .01$ ] caused by the same pattern of results as the preceding interaction; and a significant triple interaction [ $F(4,32) = 5.5$ ,  $p < .01$ ]. In order to clarify these results further, weighted average ratings were computed separately for the two response categories (as in the F25 condition), and separate analyses of variance were conducted on B and D responses. B responses showed only a highly significant effect of position [ $F(3,24) = 43.3$ ,  $p < .001$ ] due to an increase in ratings between positions 3 and 4 on the CV continuum. D responses showed significant effects of position [ $F(3,24) = 3.9$ ,  $p < .025$ ], between precursor categories [ $F(1,8) = 15.6$ ,  $p < .01$ ], and within precursor categories [ $F(1,8) = 6.0$ ,  $p < .05$ ]. Thus, the results for D responses support the prediction that within-category acoustic differences between VC precursors would affect the degree of the assimilative bias. Such a gradual effect most likely reflects precategorical integration of the phonetic information conveyed by VC and CV transitions; implicit categorization of the VC transitions can be ruled out as a factor. The between-category precursor effect was somewhat larger than the within-category effect, but this may have reflected the fact that, in the particular precursors chosen, the acoustic difference between categories was larger than within categories.

The results of the F140 condition are shown in Figure 5c. The ordinate is labeled here "percent 'D' responses," which means: the percentage of ratings between 4 and 6 for CV syllables in isolation (F control), the percentage of "2" responses for stimuli preceded by /ab/ (indicative of /ab-dε/ percepts), and the percentage of "1" responses for stimuli preceded by /ad/ (indicative of /adε/ percepts). In contrast to the conditions with very short closure durations, F15 and F25, this condition showed a pronounced negative (contrastive) effect of the VC precursors on the perception of the CV portions [ $F(1,8) = 11.5$ ,  $p < .01$ ]. Thus, listeners were more likely to perceive the CV portion as /dε/ when the precursor was /ab/ than when it was /ad/.

Finally, Figure 5d shows the results of the F265 condition. At this closure duration, which exceeded the single-geminate boundary, there was no significant precursor effect, as expected. Both precursor functions fell on top of the function for CV syllables in isolation. (However, there was considerable variation between subjects.)

**Backward conditions.** At very short closure durations, backward bias was highly assimilative in a trivial sense, because of the strong perceptual dominance of CV transitions over VC transitions. It is the region beyond the range of this interference effect that is of interest here. Figure 6a shows the results of the B115 condition. It is evident that the two CV postcursors differed little in their effects, although individual results varied considerably. Judgments in this condition were difficult for the subjects, and the absence of a postcursor effect suggests that the closure duration chosen just fell in the region where, on the average, backward bias ceased to be assimilative. Surprisingly, there was a consistent difference between the postcursor functions and the function for VC syllables in isolation (B control) [ $F(1,9) = 5.9$ ,  $p < .05$ , for the main effect;  $F(4,36) = 13.3$ ,  $p < .01$  for the interaction with position on the continuum].<sup>6</sup> In part, this difference was due to reduced consistency of listeners' responses in the B115 condition. In addition, however, VC portions heard as /ad/ in isolation were more likely to be perceived as /ab/ when followed by *any* CV postcursor. This effect was reminiscent of the high alveolar-to-labial confusion rate in Experiment 4. Thus, the perception of the VC transitions was not completely independent of the CV portion.

The results of the B140 condition are shown in Figure 6b. Here, there was a clear difference between the two CV postcursor functions in the direction of a contrast effect, as in the corresponding F140 condition [ $F(1,8) = 13.6$ ,  $p < .01$ ]. At the same time, the tendency to classify the VC transitions as labial stops when followed by either of the two postcursors (B115 condition) was replicated with a different group of subjects.

Finally, Figure 6c shows the results of the B265 condition. (The data of one subject had to be excluded here because he apparently responded to the CV portions of these stimuli.) The effect of the CV postcursors on perception of the VC transitions was contrastive and rather small but significant [ $F(1,8) = 8.0$ ,  $p < .05$ ], suggesting that, at least occasionally, decisions about the VC transitions were delayed or remained modifiable for almost 300 msec.

**Summary.** A schematic summary of the results of Experiments 5 and 6 is provided in Figure 7. The average magnitude of perceptual bias (on an arbitrary scale) is plotted as a function of closure duration. The data points resulting from the various conditions

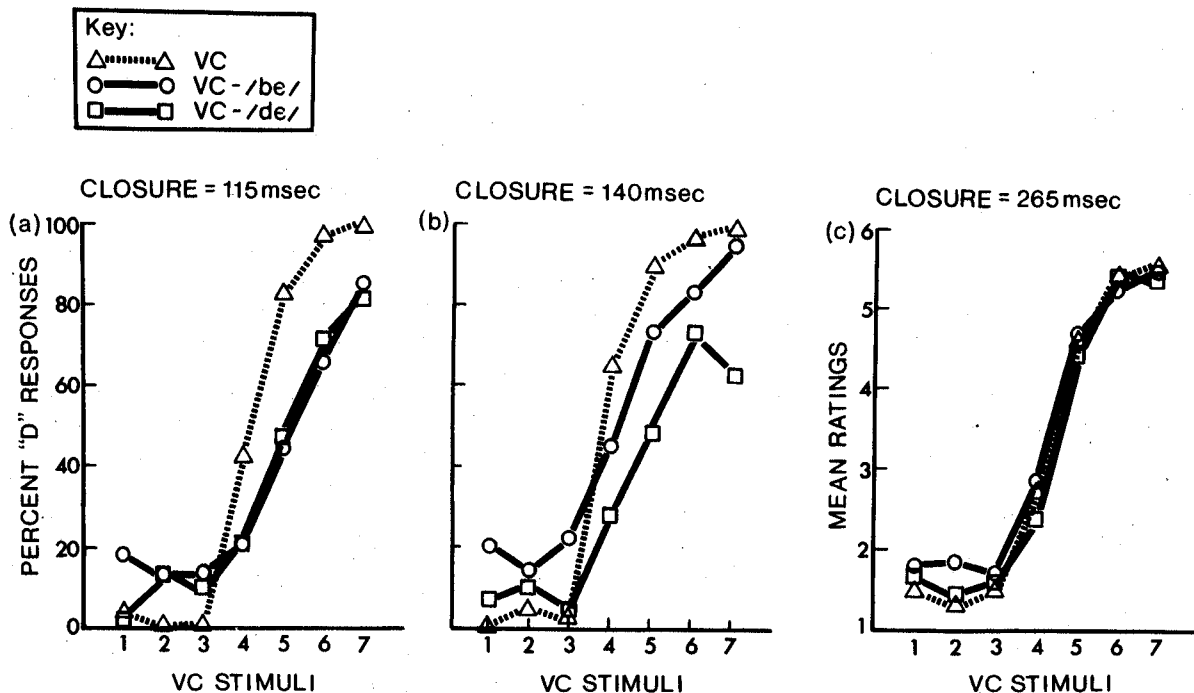


Figure 6. Results of backward conditions: B115, B140, and B265. Each panel shows the results for VC syllables in isolation (dotted line) and when followed by different postcursors (solid lines). Different groups of subjects participated in (B115, B265) and B140.

have been connected by hypothetical functions representing the time course of forward (F) and backward (B) bias. The F function begins at a moderate level of assimilation, changes to contrastive at perhaps 90 msec of closure duration, and reaches the zero asymptote somewhat beyond the single-geminate boundary (perhaps at 250 msec). The B function starts out as strongly assimilative (due to the interference effect at short closure durations) and consequently changes to contrastive at a somewhat longer closure duration (about 115 msec) than the F function. The significant contrast effect in the B265 condition suggests that the zero asymptote of the B function also occurs later than that of the F function (perhaps around 300 msec), so that backward contrast exceeds forward contrast at closure durations beyond about 150 msec.

The schematic diagram in Figure 7 constitutes the hypotheses that now require more detailed confirmation. Of special interest are the exact crossover points of the F and B functions as they change from assimilative to contrastive, as well as their final asymptotes. In addition, two kinds of asymmetries need to be clarified: why alveolar CV transitions were more affected by forward assimilation than labial CV transitions (F15 and F25 conditions), and why the perception of VC transitions was biased towards labial place of articulation when followed by any CV por-

tion (B115 and B140 conditions). It also remains to be seen whether the contrast effects at intermediate closure durations contain a simple response bias component; after all, a contrast effect could result simply from a tendency to give many "2" (i.e., "two different consonants") responses. Although response biases usually play a minor role in speech perception experiments, it may be useful to randomize closure durations in future studies in order to demonstrate changes in contrast effects with increasing closure duration independently of any response biases that may exist. Such experiments are now in progress.

## GENERAL DISCUSSION

The formant transitions surrounding the closure period of an intervocalic stop consonant provide a good example of temporally distributed information for a single phoneme. In order to perceive this information as a single event, an integrative perceptual process is needed. The present experiments suggest that there are two levels at which phonetic perception integrates over time. One level is precategorical, the other postcategorical. This distinction corresponds to that between auditory and phonetic levels of processing (see Studdert-Kennedy, 1976), but it must be realized that even precategorical cue integration is probably motivated by phonetic principles.

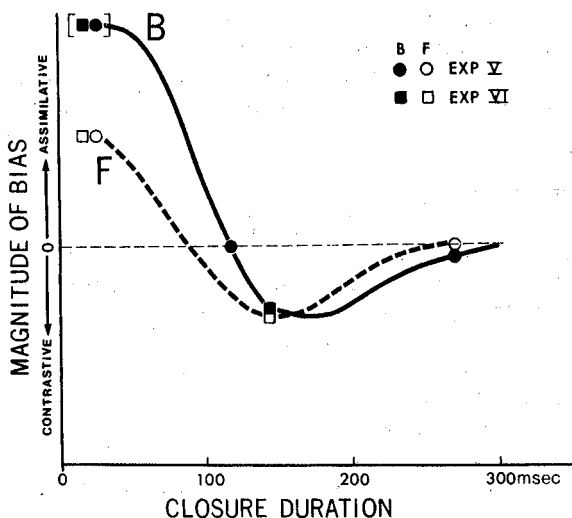


Figure 7. Schematic summary of the data in Figures 5 and 6. The ordinate is in arbitrary units. The data points are pure approximations; they are not based on numerical estimates.

While the internal representation of cues not yet categorized is presumably in an auditory code, the fact that quite diverse auditory cues are integrated into a single phonemic percept seems to require more than a psychoacoustic rationale (cf. Repp, Liberman, Eccardt, and Pesetsky, in press).

### Precategorical Integration

Precategorical integration of temporally distributed cues is apparent in the situation where conflicting VC and CV transitions separated by a very brief closure period lead to the perception of a single stop consonant (Dorman et al., Note 2). This interference effect was replicated in Experiment 3 using a discrimination task, and in Experiments 5 and 6 using an identification-judgment task. The assimilative bias of VC transitions on CV transitions in Experiments 5 and 6, and especially the within-category precursor effect in the F15 condition, specifically supported the notion of precategorical temporal integration with higher weighting of the more recent cues. The backward interference effect and the forward assimilation effect are both consequences of this precategorical integration.

In addition, Experiment 4 demonstrated that a steady-state vowel interferes with the perception of preceding VC transitions. Since steady-state resonances generally do not convey consonantal information, this interference effect may also be ascribed to weighted integration of information preceding and following the closure interval. The extreme difficulty subjects experienced in discriminating VCV stimuli with and without VC transitions at short closure durations (Experiments 2 and 3) was further evidence

for a temporal integration process that assigns little weight to early events. At very brief temporal separations, phonetically incompatible VC and CV transitions presumably are integrated just as strongly as compatible transitions, since phonetic interpretation does not play any role at this level of integration. On the other hand, the temporal range of precategorical integration is probably sensitive to acoustic stimulus properties, particularly those of the VC transitions, which may influence the time required to categorize these cues. Evidence supporting this prediction will be reported in a forthcoming paper.

The temporal limits of precategorical integration may have just been reached in the B115 condition of Experiment 5. An estimate of the total integration period may be obtained by adding the duration of the spectral information being integrated to the maximal temporal separation. VC transitions (30 msec), closure (115 msec), and CV transitions (45 msec) yield a total duration of approximately 200 msec, which is of the same magnitude as Huggins' (1975) estimate of the range of temporal integration in the perception of temporally segmented speech. Huggins hypothesized that "the ear tries to integrate into a single percept any two relatively similar events that coexist in echoic storage, and only becomes able to treat them as separate events if they do *not* coexist in echoic storage" (p. 156). If it is accepted that *any* VC and CV transitions are "relatively similar" (due to their periodic nature, equal fundamental frequencies, etc.), the 200-msec integration period obtained here may well reflect the duration of an echoic store.<sup>7</sup> There are a number of related findings in the auditory literature suggesting a temporal integration period of about 200 msec (see Huggins, 1974, 1975, for a brief discussion). This may or may not be a coincidence. Certainly, stimulus and task factors play a role in determining the actual period of integration, and it is unlikely that a single fixed time constant underlies all these diverse phenomena.

The duration of an echoic store may limit the temporal range of precategorical cue integration, but it does not provide a sufficient explanation for why certain cues are integrated into a single phonemic percept, once they coexist in echoic storage. While a vague appeal to auditory similarity may suffice in the present case, there are other cases of precategorical integration in speech perception that are difficult to explain in these terms. Repp et al. (in press) have described such cases and found it necessary to refer to the listener's intrinsic knowledge of articulatory dynamics in order to explain their findings. In their view, different cues are integrated in perception because they are perceived as specifying a single underlying articulatory act. A similar argument with regard to the VC-CV interference effect has been set forth by Dorman et al. (Note 2). For a detailed dis-

cussion of the question as to why perceptual integration occurs, the reader is referred to these two recent articles.

### Postcategorical Integration

A different process of temporal integration must be postulated to account for the perceptual distinction between single and geminate stop consonants. When VC and CV transitions signaling the same place of articulation are separated by less than about 200 msec of silence, they are perceived as a single phoneme (Experiment 1). This integrative process, whose "time window" clearly exceeds that of precategorical temporal integration, most likely takes place at a higher level. Presumably, a phonemic category is established on the basis of the VC transitions when followed by a sufficient amount of silence, but if an identical phonemic percept is arrived at within a certain time span (on the basis of CV transitions occurring 100-200 msec later), the two are combined into a single percept. This perceptual rule probably represents the listener's knowledge of articulatory timing in geminate production, and therefore it may have no parallel outside speech perception. Moreover, Pickett and Decker (1960) have shown the extreme sensitivity of the single-geminate distinction to rate of speech, both in perception and in production. This befits a higher-level phonetic process; similarly large variability of temporal parameters would not be expected at a lower level of processing, such as that of precategorical cue integration.

The process of postcategorical integration is paralleled by the VC-CV contrast effects observed in Experiment 6, which were hypothesized to extend from roughly 120-250 msec of closure duration. These contrast effects are logically prior to postcategorical integration: Integration is possible only if VC and CV transitions are perceived as signaling the same place of articulation. The nonindependence of the perception of these spectral cues, both forward and backward in time, is evidence that phonetic decisions rely on fairly long temporal stretches of the speech signal. Perceptual contrast effects probably reflect a natural tendency of the perceptual system to differentiate; integration follows if differentiation does not occur. The forward contrast effects are evidence that VC transitions followed by more than about 100 msec of silence are translated into a preliminary phonemic code, whether or not they are phonetically compatible with the following CV transitions. The backward contrast effects show that these preliminary codes are subject to revision in the light of phonetic information received later in time. It is not clear from the present results whether the temporal range of the contrast effects is limited by the single-geminate boundary, as originally hypothesized. Nor is it really necessary for this to be the

case. Perceptual contrast and integration may involve different processes that apply in sequence; they may have different time characteristics, unless they are limited by some—as yet unknown—common mechanism.

### Conclusion

Since speech signals specify the behavior of an articulatory system, it is likely that both types of integration discussed here ultimately reflect the listener's apprehension of the underlying articulatory events and relationships. The temporal parameters of precategorical integration in perception reflect the characteristic timing of certain articulatory movements, such as closing and opening of the vocal tract (Dorman et al., Note 2). In unnatural stimuli, such as VC<sub>i</sub>-C<sub>j</sub>V utterances with very short closure periods, precategorical integration of otherwise incompatible cues may be forced precisely because they could not have resulted from two distinct articulatory acts. Thus, the temporal range of precategorical cue integration may be not only speech specific but also specific to the particular articulatory event under study. However, the temporal limits of precategorical integration may also reflect basic limitations of the auditory system. Postcategorical integration, on the other hand, is entirely speech specific and presumably reflects the listener's knowledge about certain "normal" temporal characteristics of speech that are beyond any temporal limitations of either articulation or perception. Undoubtedly, higher levels of processing (syntactic, semantic, etc.) integrate over even longer time spans than the two levels considered here. Thus, speech perception may be considered to proceed on a number of different levels with increasingly larger integrative capacity.

### REFERENCE NOTES

1. Repp, B. H. *Temporal integration and "coperception" in speech perception: Experiments with intervocalic stop consonants*. Unpublished manuscript, 1977 (available from the author).
2. Dorman, M. F., Raphael, L. J., & Liberman, A. M. *Some experiments on the sound of silence in phonetic perception*. Manuscript submitted for publication, 1978.
3. Fujimura, O. *A look into the effects of context—some articulatory and perceptual findings*. Paper presented at the 8th International Congress of Phonetic Science, Leeds, England, 1975.
4. Rudnicki, A. I., & Cole, R. A. *Vowel identification and subsequent context*. Paper presented at the 93rd Meeting of the Acoustical Society of America, University Park, Pennsylvania, June 1977. [Abstracted in *Journal of the Acoustical Society of America*, 1977, 61 (Supplement No. 1), S39.]
5. Repp, B. H. Perception of implosive transitions in VCV utterances. *Haskins Laboratories Status Report on Speech Research*, 1976, **SR-48**, 209-233.
6. Repp, B. H. Perceptual integration and selective attention in speech perception: Further experiments on intervocalic stop consonants. *Haskins Laboratories Status Report on Speech Research*, 1977, **SR-49**, 37-69.
7. Repp, B. H. Perceptual integration and differentiation of

spectral information across intervocalic stop closure intervals. *Haskins Laboratories Status Report on Speech Research*, 1977, SR-51/52, 131-138.

## REFERENCES

- ABBS, M. H. *A study of cues for the identification of voiced stop consonants in intervocalic contexts*. Unpublished PhD dissertation, University of Wisconsin, 1971.
- ADES, A. E. How phonetic is selective adaptation? Experiments on syllable position and vowel environment. *Perception & Psychophysics*, 1974, 16, 61-66.
- ADES, A. E. Adapting the feature detectors for speech perception. In E. C. T. Walker & R. J. Wales (Eds.), *New approaches to language mechanisms*. The Hague: North Holland, 1976.
- COOPER, W. E. Selective adaptation to speech. In F. Restle et al. (Eds.), *Cognitive theory* (Vol 1). Hillsdale, N.J.: Erlbaum, 1975. Pp. 23-54.
- FUJISAKI, H., NAKAMURA, K., & IMOTO, T. Auditory perception of duration of speech and non-speech stimuli. In G. Fant & M. A. A. Tatham (Eds.), *Auditory analysis and perception of speech*. London: Academic Press, 1975. Pp. 197-219.
- HUGGINS, A. W. F. On perceptual integration of dichotically alternated pulse trains. *Journal of the Acoustical Society of America*, 1974, 56, 939-943.
- HUGGINS, A. W. F. Temporally segmented speech. *Perception & Psychophysics*, 1975, 18, 149-157.
- LISKER, L. Closure duration and the intervocalic voiced-voiceless distinction in English. *Language*, 1957, 33, 42-49.
- MALMBERG, B. The phonetic basis for syllable division. *Studia Linguistica*, 1955, 9, 80-87.
- MASSARO, D. W. Preperceptual images, processing time, and perceptual units in speech perception. In D. W. Massaro (Ed.), *Understanding language. An information-processing analysis of speech perception, reading, and psycholinguistics*. New York: Academic Press, 1975. Pp. 125-150.
- ÖHMAN, S. E. G. Coarticulation in VCV utterances: Spectrographic measurements. *Journal of the Acoustical Society of America*, 1966, 39, 151-168.
- PICKETT, J. M., & DECKER, L. R. Time factors in perception of a double consonant. *Language and Speech*, 1960, 3, 11-17.
- PISONI, D. B. *On the nature of the categorical perception of speech sounds*. Unpublished doctoral dissertation, University of Michigan, 1971.
- REPP, B. H., LIBERMAN, A. M., ECCARDT, T., & PESETSKY, D. Perceptual integration of temporal cues for stop, fricative, and affricate manner. *Journal of Experimental Psychology: Human Perception and Performance*, in press.
- SIMON, H. J. *Anchoring and selective adaptation of phonetic and*

*nonphonetic categories in speech perception*. Unpublished doctoral dissertation, CUNY, 1977.

## NOTES

1. These differences were accidental—the stimuli were drawn from a computer file created by an earlier investigator—and were discovered only after Experiments 1-4 had been conducted.
2. Malmberg (1955) reported long ago that listeners grouped the consonant of VC-V stimuli with the final vowel when the closure duration was made very short; however, he did not mention any interference effect. Perhaps the absence of interference was due to his particular synthetic stimuli which had identical initial and final vowels, or to his use of sophisticated listeners. The VC-V interference effect may well be less stable than the VC<sub>i</sub>-C<sub>j</sub>V interference effect: Informal observations suggest that the former may be reduced after practice (cf. also Figure 4b), while the latter generally persists even after many trials.
3. One reviewer (Bill Huggins) pointed out that this statement may not be valid because a steady-state vowel with a relatively abrupt onset may be perceived as beginning with a glottal stop. However, the present vowels had an amplitude ramp, so that their onsets were somewhat gradual. Also, glottal stops are not a phonemic category in English, and since their acoustic correlate is the simple auditory attribute of abruptness of onset, it seems permissible to assume that English-speaking listeners do not process them at the phonetic level.
4. Stimulus 8 was an additional syllable especially constructed for this purpose and having even more extreme formant transition offset frequencies than Stimulus 7 (cf. Table 2). The VC precursors were very consistently identified in isolation (cf. Figure 6b): Stimuli 1 and 3 received 100 and 98% B responses, respectively, and Stimuli 6 and 7 received 98 and 100% D responses, respectively, so that it was safe to assume that Stimulus 8 would have received 100% D responses, too.
5. One subject indeed heard /ade/ whenever the precursor was /ad/. Another subject reported hearing /ab-de/ on a number of trials.
6. In this statistical analysis, the two shortest and the two longest closure periods were combined, leading to five levels with equal numbers of observations in each.
7. Precategorical integration can be prevented by making the events before and after the closure period more dissimilar, e.g., by having them pronounced by two different speakers (Dorman et al., Note 2).

(Received for publication November 23, 1977;  
accepted August 9, 1978.)