Perception & Psychophysics
1983, 33 (2), 147-155

Bidirectional contrast effects in the perception of VC-CV sequences

BRUNO H. REPP
Haskins Laboratories, New Haven, Connecticut

The perceived places of articulation of two successive stop consonants are not independent: Given some ambiguity in the formant transition cues and a closure duration between 100 and 200 msec, contrastive perceptual interactions in both directions have been observed in identification tasks. Retroactive contrast declines as the closure interval is lengthened and is strongly influenced by the range of closure durations employed, whereas proactive contrast appears to be less sensitive to these factors (Experiment 1). Reduced contrast and no effects of closure duration are obtained in a discrimination task with selective attention to one stimulus portion; this suggests that the effects in identification arise largely at a higher level of (phonetic) perception (Experiment 2). The contrast effects do not seem to represent a perceptual compensation for coarticulatory dependencies between stops produced in sequence, for there appears to be little coarticulation as far as place of articulation is concerned (Experiment 3). The most plausible hypothesis is that the presumed contrast effects do not result from any direct interaction of spectral cues across the closure interval but are due to perceptual information conveyed by the closure itself: Closure durations of 100-200 msec happen to be most appropriate for sequences of two nonhomorganic stops. Here, it seems, is another case in which listeners' tacit knowledge of canonical speech patterns determines perception.

Recent studies have revealed that stop consonants are sensitive to other consonantal segments in their vicinity, not only with regard to voicing (e.g., Klatt, 1975), but also with regard to place of articulation. This evidence has come primarily from perceptual studies. Thus, Repp (1978) has shown that the perception of a syllable-initial stop may be influenced by a preceding, syllable-final stop (and vice versa), Mann (1980) found an influence of a preceding, syllable-final liquid, and Mann and Repp (1981) found an influence of a preceding fricative: Listeners are more likely to perceive a syllable ambiguous between /da/ and /ga/ as "ga" when it is preceded by /d/, /s/, or /l/ than when it is preceded by /g/, /ʃ/, or /r/. The general principle seems to be that an ambiguous stop is more likely to be perceived as having a posterior place of articulation when it is preceded by a consonantal segment that has an anterior place of articulation (relative to some other possible context). There are two types of explanation for these findings:

(1) The perceptual interaction between the context and the target segment may take place at a purely auditory level of processing: For example, the spectral properties of the acoustic segment preceding the stop closure interval may prime the auditory system in a way that modifies the internal spectral representation of the signal onset following the closure, which contains the important cues for the perception of stop place of articulation. If such an auditory interaction takes place, it is likely to be contrastive; prominent spectral components of the preceding segment would adapt the neurons sensitive to these frequencies, so that they would respond more weakly to the following segment. Indeed, there is evidence, from physiological studies in animals, that adaptation does take place in the auditory nerve (Delgutte, 1980; Harris & Dallos, 1979). Considering the spectral complexity of the speech stimuli used in the various perceptual studies, it is not clear whether auditory adaptation of this sort really could account for the contrast effects obtained, but the possibility certainly deserves continued attention.

(2) The other possibility is that perceptual contrast originates in phonetic, as distinct from general auditory, properties of the stimuli. Assuming that simple response biases reflecting the frequencies of occurrence of particular consonant sequences or the subjects' interpretation of the experimental task can be ruled out, the major hypothesis to be considered relates perception to coarticulation of stop consonants with neighboring segments. Since coarticulation is invariably assimilatory in nature, listeners' perceptual compensation for such effects would naturally be contrastive. Some support for the coarticulation-compensation hypothesis has been obtained in

This research was supported by NICHD Grant HD01994 and by BRS Grant RR05596 to Haskins Laboratories. I am grateful to Janette Henderson, Sarah Peck, and Patti Jo Price for assistance, and to Virginia Mann for helpful comments. My mailing address is: Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06510.

studies of fricative-stop and liquid-stop sequences (Mann, 1980; Repp & Mann, 1981, 1982).

The present series of experiments was concerned with the contrastive influence of one stop consonant on the perception of another (preceding or following) stop consonant. The phenomenon of interest was first reported by Repp (1978, Experiments 5 and 6). He preceded synthetic syllables ambiguous between /bɛ/ and /dɛ/ with either an unambiguous /ab/ or an unambiguous /ad/ and found that, when the silent closure interval separating the two syllables was between roughly 100 and 200 msec, listeners tended to report two different stops (/abdɛ/, /adbɛ/) more often than a single stop (/abɛ/, /adɛ/). A similar contrastive effect was found when syllables ambiguous between /ab/ and /ad/ were followed by either /bɛ/ or /dɛ/. Thus, there were both proactive and retroactive contrast effects.

Applying the hypotheses outlined above to this specific case, the contrast effects between the two stop consonants, C_1 and C_2, could be due to (1) psychoacoustic interactions between the two sets of formant transitions that provide the place of articulation information, or to (2) perceptual compensation for a coarticulatory dependency between successive stop consonants. The presence of strong retroactive effects over rather long temporal intervals (by psychoacoustic standards) suggests that the first hypothesis may not be sufficient, although it may account for part of the effects. This general conclusion receives support from the results of Experiments 1 and 2. Experiment 3 examines the coarticulation-compensation hypothesis, with negative results. In the final discussion, a new hypothesis is proposed that seems to provide the most convincing account of the experimental findings.

EXPERIMENT 1

Repp's (1978) experiments provided only a very rough sampling of different closure durations, and thus only very preliminary information about the possible time course of contrastive effects. It was the purpose of Experiment 1 to map out time functions in considerably more detail, in the hope of thereby constraining the possible explanations of the effects. For practical reasons, Experiment 1 was divided into three parts (1a, 1b, 1c), each covering one-third of the total range of closure durations (10-310 msec). Experiment 1b was conducted in advance of Experiments 1a and 1c.

Method

Subjects

Experiment 1b employed 12 subjects; they included 9 paid student volunteers with varying experience in listening to synthetic speech, 2 research assistants, and the author. Experiments 1a and 1c employed 9 subjects each, 7 of whom participated in both experiments. Only 2 subjects (the author and 1 research assistant) participated in all three experiments.

Stimuli

The stimuli consisted of two synthetic stimulus continua, generated on the OVE IIIc synthesizer at Haskins Laboratories. The *VC continuum* consisted of seven stimuli ranging from /ab/ to /ad/ and differing only in the final formant transitions. The F_1 transition had a constant offset frequency of 541 Hz, but changed in duration from 90 msec in Stimulus 1 to 30 msec in Stimulus 7. The F_2 and F_3 transitions had a constant duration of 50 msec but varied in offset frequency: F_2 offset changed from 1,060 Hz in Stimulus 1 to 1,297 Hz in Stimulus 7, and F_3 offset changed from 2,181 Hz in Stimulus 1 to 2,539 Hz in Stimulus 7, both in roughly equal steps. All transitions were stepwise-linear in 10-msec time segments. The formant frequencies of the initial steady-state portion were 777 Hz (F_1), 1,147 Hz (F_2), and 2,466 Hz (F_3). All VC stimuli had a duration of 180 msec, a constant fundamental frequency of 120 Hz, and an amplitude contour that increased over roughly two-thirds of the stimulus and then declined.

The *CV continuum* consisted of seven stimuli ranging from /ba/ to /da/ and differing only in the initial transitions of F_2 and F_3. The F_1 transition was constant, with an onset frequency of 459 Hz. F_2 onsets ranged from 1,099 Hz in Stimulus 1 to 1,635 Hz in Stimulus 7, and F_3 onset ranged from 2,262 Hz in Stimulus 1 to 2,500 Hz in Stimulus 7, both in roughly equal steps. All transitions were 50 msec long. The formant frequencies of the final steady-state portion were 728 Hz (F_1), 1,156 Hz (F_2), and 2,466 Hz (F_3). All CV stimuli had a duration of 290 msec, a fundamental frequency that was constant at 120 Hz over the first 90 msec and then fell linearly to 100 Hz, and an amplitude contour that rose slightly over the first 50 msec and then fell gradually until stimulus offset.

All stimuli were digitized at 10 kHz using the Haskins Laboratories pulse code modulation (PCM) system. Experimental sequences were recorded on magnetic tape using a special sequencing program. In each experiment, there were two conditions, called *proactive* and *retroactive*. In the proactive condition, each of the seven stimuli from the CV continuum was preceded by one of the two endpoint stimuli of the VC continuum, at various interstimulus intervals which are referred to here as closure durations. In the retroactive condition, each of the seven stimuli from the VC continuum was followed by one of the two endpoint stimuli of the CV continuum, with various closure durations in between. Thus, there were 14 basic stimulus combinations in each condition. To obtain more observations for ambiguous stimuli, a 1-2-3-3-3-2-1 frequency distribution was imposed on the seven-member continua, so that the basic test unit contained $2\times(1+2+3+3+3+2+1)=30$ stimuli. In each experiment, each VC-CV stimulus occurred with five different closure durations, in a random sequence containing $5\times30=150$ stimuli. Three such sequences of 150 stimuli were recorded on each experimental tape. The interval between successive VC-CV combinations was 3 sec.

The three parts of the experiment differed only in the range of closure durations. Within each part, closure durations varied in 25-msec steps over a 100-msec range. Experiment 1a covered the range of 10-110 msec, Experiment 1b that of 110-210 msec, and Experiment 1c that of 210-310 msec.

In addition, randomized sequences of isolated VC and CV syllables were recorded. Each of these two sequences contained 75 stimuli, resulting from five replications of the basic 15-stimulus unit due to the 1-2-3-3-3-2-1 frequency distribution of the 7 stimuli on each continuum. The interstimulus interval was 2 sec. These tapes were used in all three parts of the experiment.

Procedure

Each experiment required two sessions per subject of approximately 90 min duration. At the beginning of each session, the subject listened to the isolated CV and VC sequences, in that order. Then the VC-CV tapes were presented. The order of the proactive and retroactive conditions was counterbalanced between subjects and reversed between the first and second sessions. In each experiment, the most ambiguous stimuli (i.e., Stimuli 3-5 from a given continuum) received a total of 30 responses from each subject when presented as isolated monosyllables and 18 responses when presented in a specific VC-CV combination.

The response choices given to the subjects were the following: B and D for isolated syllables; B, D, BD, and DB for VC-CV combinations. In Experiment 1c, the choices B and D for VC-CV combinations were changed to BB and DD, respectively, since the closure durations were in the range in which listeners were expected to hear geminate instead of single stops (cf. Repp, 1978). The listeners were never required to distinguish between single (B, D) and geminate (BB, DD) stops; although such a distinction may have provided useful information, it was felt that it would have made the task too complicated. Although listeners were encouraged to note down any other consonants heard, there were hardly any occurrences of responses other than B and D and their combinations.

The tapes were played back at a comfortable intensity on an Ampex AG-500 tape recorder, and the subjects listened binaurally over TDH-39 earphones in a quiet room. The listeners were fully informed about the structure of the stimuli before each condition.

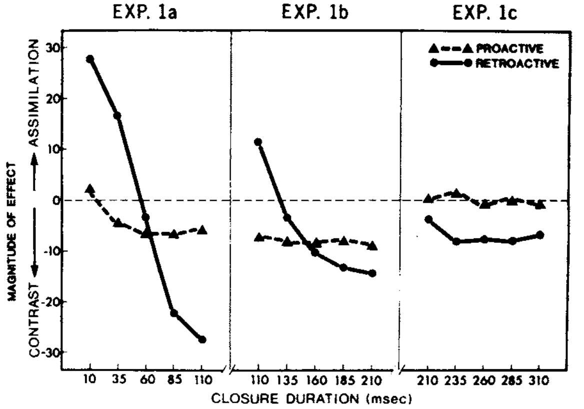

Figure 1. Proactive and retroactive interactions between C_1 and C_2 as a function of closure duration (Experiment 1).

Results and Discussion

A gross measure of the perceptual interaction between C_1 (VC) and C_2 (CV) is provided by

$$[(100/n)\Sigma_i(\text{responses of D, DD, or DB}) \text{ to } VC_i\text{-/ba/}]$$
$$- [(100/n)\Sigma_i(\text{responses of D, DD, or DB}) \text{ to } VC_i\text{-/da/}]$$

in the retroactive condition, and by

$$[(100/n)\Sigma_i(\text{responses of D, DD, or BD}) \text{ to /ab/-}CV_i]$$
$$- [(100/n)\Sigma_i(\text{responses of D, DD, or BD}) \text{ to /ad/-}CV_i]$$

in the proactive condition, where i indexes the seven stimuli on a given synthetic continuum and n is the total number of responses to the stimuli on a continuum. Thus, the index is a percentage difference and varies from −100 for maximal contrast to +100 for maximal assimilation. These indices of stimulus interaction are plotted as a function of closure duration in Figure 1, separately for the proactive and retroactive conditions. (For a detailed discussion of the actual identification functions, the reader is referred to an earlier report by Repp, Note 1.)

In Experiment 1a (Figure 1a), the retroactive effect was strongly assimilative at the shortest closure durations, as expected. It reflects a strong tendency to perceive only a single stop consonant, with a place of articulation corresponding to C_2.[1] As the closure duration increased, the retroactive effect changed rapidly from assimilative to contrastive, with the crossover occurring at about 55 msec of closure duration. Although such a crossover had been predicted, it occurred considerably earlier (i.e., at a shorter closure duration) than expected on the basis of earlier data (cf. Figure 7 of Repp, 1978). The crossover point marks the emergence of C_1 as a separate phonetic percept (if different from C_2), and the contrastive effect indicates that there was a strong tendency to perceive C_1 as different from C_2.

The proactive function in Experiment 1a, on the other hand, was considerably flatter than the retroactive function. In an analysis of variance, this was reflected in a highly significant interaction between the effects of condition (proactive vs. retroactive) and closure duration $[F(4,32)=21.1,\ p<.001]$, in addition to a highly significant main effect of closure duration $[F(4,32)=27.1,\ p<.001]$, which was primarily due to the retroactive function. There was a constant small proactive contrast effect at closure durations beyond 35 msec; it was absent only at the shortest closure duration (10 msec). The change in the proactive effect with closure duration was significant in a separate test $[F(4,32)=5.4,\ p<.01]$; however, the small assimilative effect at the shortest closure duration was not significantly different from zero. Repp (1978) found that the cues for C_1 influenced perception even when C_1 was not perceived as a separate phoneme. The present results provide only weak support for this earlier observation, as there was no absolute assimilative effect, only an absence of proactive contrast.

Experiment 1b (Figure 1b) examined the region of intermediate closure durations. The retroactive function followed largely the course expected on the basis of earlier data (Repp, 1978): An assimilative effect at the shortest closure duration (110 msec) shifted toward a pronounced contrastive effect at longer closure durations, with the crossover occurring at about 130 msec of closure duration. No return to the zero baseline was indicated at the longest closure duration, suggesting a temporal range of the retroactive effect substantially exceeding 210 msec—an unexpected finding. In contrast to the retroactive function, the proactive function was completely flat, showing a moderate contrast effect at all closure durations. The different shapes of the functions were reflected in a highly significant interaction of the effects of condition and closure duration [F(4,44) = 16.2, $p < .001$] as well as in a significant main effect of closure duration [F(4,44) = 20.6, $p < .001$], which was solely due to the retroactive function. There was no significant effect of closure duration on the magnitude of proactive contrast, as determined in a separate test [F(4,4) = .5].

The most unexpected result was the large discrepancy between the retroactive effects for the same closure duration (110 msec) in Experiments 1a and 1b: In Experiment 1b, there was an assimilative effect, whereas, in Experiment 1a, there was a contrast effect that actually exceeded the contrast effect at the longest interval (210 msec) in Experiment 1b. Instead of a single crossover from positive to negative retroactive effects (expected to be at approximately 115 msec, according to earlier findings), there were two: one at 55 msec in Experiment 1a and the other at 130 msec in Experiment 1b. These results are indicative of strong effects of the range of closure durations used in a given condition on the listeners' perception of the stimuli—more precisely, on their tendency to hear one vs. two (different) stop consonants (see also Repp, Note 2). Indeed, single-consonant responses to nominally conflicting sets of C_1 and C_2 cues did not occur at the 110-msec interval in Experiment 1a but appeared with some frequency at the same interval in Experiment 1b.

In Experiment 1c (Figure 1c), the retroactive effect was contrastive throughout, but there was a significant reduction in contrast at the shortest interval (210 msec) [F(4,32) = 4.7, $p < .01$], which was reminiscent of the more pronounced trends in the retroactive functions of Experiments 1a and 1b. The proactive condition, on the other hand, showed no contrast at all. The difference between proactive and retroactive effects was significant [F(1,8) = 8.3, $p < .05$]. The different magnitudes of the retroactive contrast effects at 210 msec in Experiments 1b and 1c again suggest a stimulus range effect. The cause of the difference in the amount of proactive contrast between the two experiments is less clear; perhaps the difference in response choices (B and D vs. BB and DD) played a role.

Despite the large stimulus range effects, the influence of closure duration is clearly evident in Figure 1. Retroactive contrast at the longest intervals in each range (110, 210, 310 msec) declined as closure duration increased, suggesting that the effect might disappear when closure durations reach 400-500 msec. Proactive contrast seemed to disappear earlier and was definitely less pronounced than retroactive contrast.

The high sensitivity of retroactive contrast to stimulus range casts doubt on purely sensory explanations of the effect. Evidently, some or all of it takes place at a level of processing at which the subjects' criteria are highly flexible; this suggests a phonetic origin. Retroactive contrast may arise largely in phonetic short-term memory, for subjects are required to label both stop consonants on each trial. The effect might be reduced when subjects are required to pay attention only to C_1 and to ignore C_2. There are large contrast effects in pairs of isolated vowels when both stimuli are to be labeled (Repp, Healy, & Crowder, 1979) but only small effects—proactive, in this case—when only the second stimulus receives attention (Crowder, 1982). Experiment 2 explored whether the contrast effects between successive stop consonants hold up under conditions of selective attention.

EXPERIMENT 2

To achieve an effective focusing of the subjects' attention on one or the other stimulus component, Experiment 2 employed a forced-choice discrimination task in which the irrelevant component was held constant. It was expected that unstable phonetic criteria would play a minimal role in this task, although, because of the strong tendency to perceive stops categorically, such influences could not be ruled out completely. In any case, the experiment provided a much more stringent test of contrast effects than did the labeling task of Experiment 1. Only the strictly obligatory components of contrast (be they psychoacoustic or phonetic in nature) should survive under these conditions.

Because of practical limitations, only two closure intervals were used (150 and 250 msec), both in the region in which strong contrast effects were found in Experiment 1. The task was set up so that listeners had to distinguish between members of either the VC or the CV continuum, either in isolation or in the presence of one or the other post- or precursor (the endpoints of the other continuum), which were held constant over a block of trials. It is well known that, on synthetic stop consonant continua, discrimination performance is high when the two stimuli to be compared fall on opposite sides of the category boundary, but very low when the two stimuli are from the same phonetic category. This is the familiar pattern of categorical perception. In the present study, the question was whether a pre- or postcursor would shift the discrimination peak and/or change within-category discrimination performance on a given continuum. A contrastive effect should shift the peak toward the category represented by the pre- or postcursor, and discrimination performance should be improved within that category.

Method

Subjects

Sixteen subjects participated, including 14 paid volunteers, 1 research assistant, and the author.

Stimuli

The stimuli were the same as in Experiment 1. There were 12 experimental conditions, resulting from the orthogonal combination of three factors: retroactive vs. proactive (i.e., VC vs. CV discrimination), closure duration (150 vs. 250 msec), and context (none vs. /b/ vs. /d/ pre- or postcursor). All conditions were blocked (in contrast to Experiment 1; however, closure duration was a blocked factor in Repp, 1978). As in Experiment 1, the pre- or postcursors were the endpoint stimuli of the VC and CV continua. Thus, in the proactive condition, the subjects' task was to discriminate stimuli from the CV continuum in isolation, preceded by /ab/, and preceded by /ad/ at two closure durations; in the retroactive condition, they had to discriminate stimuli from the VC continuum in isolation, followed by /ba/, and followed by /da/.

The stimuli were arranged in AXB triads, with interstimulus intervals of 500 msec in the pre- or postcursor conditions. Isolated VC or CV stimuli were separated by as much silence as equaled their temporal separation in the corresponding pre- or postcursor conditions (950 or 1,050 msec for VC stimuli and 840 or 940 msec for CV stimuli, depending on the closure duration condition). The interval between AXB triads was 3 sec in all cases.

The stimulus differences to be detected were two-step separations on the seven-member synthetic continua. Thus, there were five different contrasts (1-3, 2-4, 3-5, 4-6, 5-7), each of which appeared in four possible AXB arrangements (AAB, ABB, BAA, BBA), resulting in 20 triads, which were repeated five times in random order to give a total of 100. Each of the 12 experimental conditions contained such a set of 100 triads, preceded by four easy practice triads that served to illustrate the structure of the stimuli.

Procedure

Each subject participated in four 1-h sessions. The four major conditions resulting from the orthogonal combination of the proactive-retroactive and closure duration factors were presented on separate days in an order that was counterbalanced across subjects according to a Latin-square schedule. In each session, the isolated VC or CV condition was presented first; it served both as familiarization and as a baseline for comparison with the pre- or postcursor conditions that followed. The order of the following /b/ and /d/ pre- or postcursor conditions was counterbalanced across subjects.

The equipment was the same as in Experiment 1. The subjects indicated their choices by writing A or B, depending on whether the second stimulus sounded more similar to the first or to the third, guessing if necessary. The subjects were fully informed about the structure of the stimuli and knew whether the difference to be detected was located in the VC or the CV portion.

Results and Discussion

The results are shown in Figure 2, the proactive condition (CV discrimination) at the top and the retroactive condition (VC discrimination) at the bottom. The discrimination functions for isolated stimuli (dotted, triangles) had the familiar peaked shape. Performance in the pre- and postcursor conditions was only slightly lower than for isolated stimuli, indicating little interference due to the added stimulus component.

The main results are easy to summarize. In no case was there a shift of the discrimination peak as a function of pre- or postcursor condition. However, discrimination performance tended to be improved at the end of the continuum that corresponded to the category represented by the pre- or postcursors—a pattern indicative of a contrast effect. This effect, revealed as an interaction between the (highly significant) effect of position on the continuum (abscissa in Figure 2) and the effect of /b/ vs. /d/ pre- or postcursor, was significant both in the proactive condition [F(4,60) = 6.2, p < .001] and in the retroactive condition [F(4,60) = 2.6, p < .05]. Surprisingly, neither effect was influenced by closure duration.

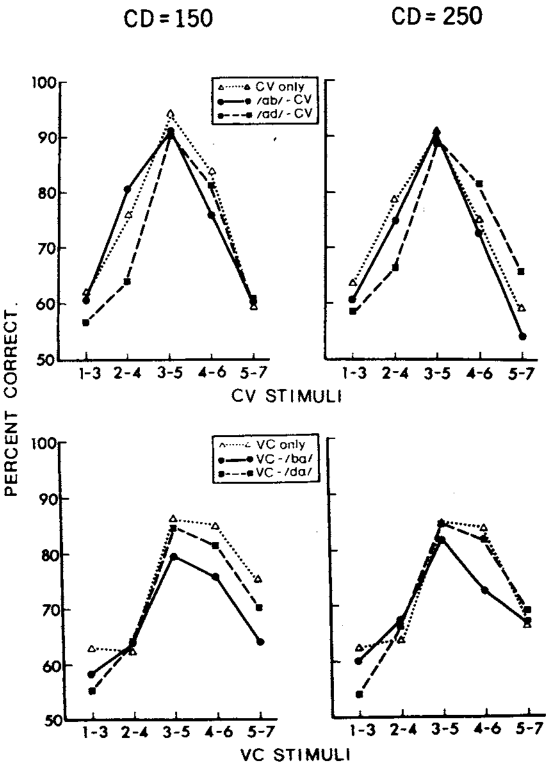

Figure 2. AXB discrimination performance for VC and CV stimuli in isolation and in context, as a function of context stimulus and closure duration (Experiment 2).

These results confirm the existence of seemingly obligatory perceptual contrast effects between C_1 and C_2, in both directions. However, retroactive effects, at least, were considerably smaller than those observed in the identification task (Experiment 1), since they were not sufficient to shift discrimination peaks.[2] Retroactive contrast was actually less reliable than proactive contrast, suggesting that the large retroactive effects in Experiment 1 derived primarily from phonetic decision rules in short-term memory. The finding that neither effect decreased as the closure duration was extended from 150 to 250 msec confirms that the large changes in retroactive contrast with closure duration (and with the range of this variable) in Experiment 1 represented higher order judgmental effects that did not come into play in the selective-attention discrimination task.

The residual contrast effects observed in Experiment 2 may arise at an early stage in processing. The possibility of a psychoacoustic interaction is perhaps favored by the fact that the spectral correlates of the same stop in initial and final position are roughly similar, though far from identical (especially not in different vocalic contexts, as in Repp, 1978). Although studies of selective adaptation, a phenomenon similar to contrast, have failed to find effects of VC adaptors on (mirror-image) CV test stimuli (Ades, 1974; Sawusch, 1977), adaptors and test stimuli in these studies were separated by several seconds of silence, which may have prevented the close-range interaction studied here. Since there is retroactive as well as proactive contrast, the interactive process would have to be bidirectional; it cannot be simple adaptation or masking. Crowder's (1981) lateral inhibition metaphor may provide an appropriate context in which to view these contrast effects, although the psychoacoustics of the case remain to be worked out in detail. The possibility of residual phonetic effects due to insufficient selective attention and covert categorization of the whole VC-CV stimulus cannot be excluded.

Whatever the precise explanation of the residual effects might be, it appears that contrast effects between successive stop consonants, retroactive effects in particular, have at least two sources: one at a relatively low and obligatory level and the other at a higher level, perhaps in phonetic short-term memory. What causes these higher level effects? One hypothesis that is encouraged by the findings on fricative-stop and liquid-stop sequences (Mann, 1980; Mann & Repp, 1981; Repp & Mann, 1981) was mentioned in the introduction. It is the possibility that the perceptual contrast effects derive from listeners' compensation for a coarticulatory dependency between two successive stop consonants. If the place of articulation of a stop shifted slightly toward that of a preceding or following stop, as it seems to do in the case of a preceding fricative or liquid, then a coarticulatory basis would exist for perceptual contrast. The difference between proactive and retroactive contrast may then correspond to a difference in the extent of proactive or retroactive coarticulation, and the decline of the perceptual effects over time may parallel a decline in the extent of coarticulatory shifts as the closure interval is lengthened. Experiment 3 addressed this general hypothesis.

EXPERIMENT 3

Quite apart from the question of whether coarticulation in two-stop sequences is the cause of perceptual contrast effects, which would be difficult to prove directly, we must ask whether such coarticulation exists at all. If evidence of coarticulation were found, the hypothesis that relates it to perception could be maintained; however, if no coarticulatory effects were found, the hypothesis would be eliminated (barring the possibility that coarticulatory variation was really present but not detected because, e.g., the methods of assessment were not sufficiently sensitive). The present study investigated coarticulation using an indirect, perceptual method that was used with some success by Mann (1980) and by Repp and Mann (1981). The basic technique is to replace a portion of a natural utterance with a matched synthetic segment that, however, is phonetically ambiguous, and to see whether listeners tend to interpret the ambiguous segment as matching the replaced segment. If so, it may be assumed that coarticulatory cues in the remaining natural signal portion provided clues to the segment that had been replaced. To supplement the perceptual results, an acoustic analysis was also conducted.

Besides probing for coarticulatory variation, the present study also investigated further the generality and nature of the perceptual interactions between two successive stop consonants. For this purpose, it used all three places of articulation; thus, for example, a stop ambiguous between /b/ and /d/ was preceded not only by an unambiguous /b/ or /d/ but also by a /g/. A change was also made in the response alternatives given to the subjects. In Experiment 1b, the subjects had had the choice of writing down two different stops or a single stop. This menu of alternatives may have been partially responsible for the contrast effects observed. In the present study, the subjects always wrote down two responses, one for the first and one for the second stop, and they were told that the two consonants could be either different or the same. Short of requiring identification of C_1 and C_2 in separate blocks of trials, these instructions nevertheless encouraged independent processing of the two sets of place-of-articulation cues.

Method

Subjects

A total of 12 subjects participated. Four of them—two paid student volunteers, the author, and a graduate research assistant—listened to both sets of tapes (described below). Each set was presented to four additional student volunteers who listened to one set only.

The author (B.R.), a native speaker of Austrian German, and a linguist colleague (G.C.), a native speaker of American English, produced the original sets of utterances. It was considered unlikely that the author's native language would render either his productions or his perception different from those of the other participants, since the study was concerned with phonetic distinctions that are similar in English and German. However, to forestall any possible objections to the author as a speaker, and to increase the generality of the results, two parallel sets of stimuli were used.

Stimuli

Natural utterances. Speakers G.C. and B.R. each recorded a set of nonsense utterances which included five tokens each of /abda/, /abga/, /adba/, /adga/, /agba/, /agda/ (as well as /aba/, /ada/, /aga/, which were not used in the perceptual experiment). The utterances were produced with stress on the first syllable, so as to prevent reduction of the first vowel. The speakers read at a steady speed from a randomized list into a Sennheiser MKH 415T microphone whose response was recorded by a Crown 822 tape recorder. The average durations of the three major segments (VC, closure, CV) were 122, 132, and 299 msec for G.C. and 165, 152, and 240 msec for B.R. All utterances were digitized at 10 kHz using the Haskins Laboratories pulse code modulation system. The VC and CV segments of each utterance were stored in separate computer files.

Synthetic stimuli. Eight continua of synthetic syllables were generated, four for each speaker. They ranged, respectively, from /ab/ to /ad/, from /ad/ to /ag/, from /ba/ to /da/, and from /da/ to /ga/. To match the endpoint stimuli as closely as possible to the corresponding segments of natural utterances, good-sounding natural tokens of the relevant segments were selected from the recorded VCV utterances and analyzed with the aid of a Federal Scientific UA-6A spectrum analyzer. The resulting computer spectrograms were displayed on an oscilloscope, and the three lowest formants were tracked by an automatic peak-picking procedure. The formant tracks were then traced with a light-pen whose output was automatically converted into frequency parameters for the OVE IIIc serial-resonance synthesizer. In this way, synthetic copies of /ab/, /ad/, /ag/, /ba/, /da/, and /ga/ were obtained for both G.C. and B.R.

Within each set of VC or CV utterances, all stimuli were assigned the same fundamental frequency contour, amplitude contour, and duration. The first-formant frequencies were also equalized at some compromise values, as were the steady-state vocalic portions. Thus, the stimuli differed only in the transitions of the second and/or third formant. Speaker differences in formant frequencies, fundamental frequency contour, and duration were preserved.

Seven-member synthetic continua from /ab/ to /ad/, /ad/ to /ag/, /ba/ to /da/, and /da/ to /ga/ for each speaker were produced by linear interpolation between the formant tracks of the two respective endpoint stimuli, in roughly equal steps. All stimuli were digitized at 10 kHz.

Experimental conditions. Two parallel sets of tapes were recorded, one using the G.C. stimuli and one using the B.R. stimuli. Within each set, there were two subsets of tapes corresponding to the retroactive and proactive conditions.

The *retroactive* condition investigated the influence of natural CV portions on the perception of synthetic VC portions. It included five tapes with random sequences of the following: (1) the 7 stimuli from the synthetic /ab/-/ad/ continuum, repeated 10 times; (2) the 7 stimuli from the synthetic /ad/-/ag/ continuum, repeated 10 times; (3) the 30 natural CV portions (3 syllables, each from 2 different VC contexts, 5 tokens of each), repeated 5 times; (4) the synthetic /ab/-/ad/ stimuli followed by the natural CV portions after a fixed silent interval, a total of 7×30=210 combinations; and (5) as (4), with the synthetic /ad/-/ag/ stimuli.

The *proactive* condition investigated the influence of natural VC portions on the perception of synthetic CV portions. It included five tapes analogous to those in the retroactive condition, with reversed roles of VC and CV portions.

The silent interval separating the VC and CV portions on Tapes 4 and 5 was 130 msec on the G.C. tapes and 150 msec on the B.R. tapes. These values matched the average VCCV closure durations of the two speakers. The interstimulus intervals were 2.5 sec on the tapes containing single VC or CV syllables and 3 sec on those containing VC-CV combinations, with longer intervals from time to time.

Procedure

The retroactive and proactive conditions were administered on different days in counterbalanced order. Within each condition, the tapes were presented in the order listed, except that the sequence of tapes differing only in the nature of the synthetic stimuli (/b-d/ vs. /d-g/) was varied across subjects.

When listening to tapes containing isolated VC or CV syllables, the subjects' task was to identify the stop consonants as "b," "d," or "g." All three alternatives were given, even when the stimuli were intended to cover only two categories. When listening to tapes containing VC-CV combinations, the subjects chose from nine alternatives: "bb," "bd," "bg," "db," "dd," "dg," "gb," "gd," and "gg." All nine responses were permitted even though only six were intended to be relevant to a given tape. The subjects were told that the stimuli consisted of the VC and CV components they had heard before, that the stop consonants in both components were to be identified, and that these consonants could be either the same or different. Single-consonant responses ("b," "d," "g") were not permitted and certainly not appropriate under these instructions.

Results and Discussion

Identification of Natural-Speech Stimuli

The natural CV portions were presented three times: once in isolation (5 repetitions) and twice preceded by synthetic VC portions (seven repetitions each time). Since there were five different tokens of each utterance, a total of 5×19=95 responses was obtained from each subject to each of the six basic syllables: /(ad)ba/, /(ag)ba/, /(ab)da/, /(ag)da/, /(ab)ga/, and /(ad)ga/. (The portion in parentheses indicates the original context.) The percentages of correct responses were 99.3 for the B.R. stimuli and 90.3 for the G.C. stimuli. Most of the G.C. errors derived from /ba/, and from /(ad)ba/ in particular, which was frequently misidentified as "d." Since these errors matched the original context (/ad/), they indicate a possible coarticulatory influence of a preceding stop on speaker G.C.'s production of /ba/.

A total of 95 responses was also obtained from each subject to each of the six basic VC syllables: /ab(da)/, /ab(ga)/, /ad(ba)/, /ad(ga)/, /ag(ba)/, and /ag(da)/. Overall, G.C.'s VC tokens were correctly identified on 86.4% of the trials, and B.R.'s tokens, on 93.8%. There were no strong suggestions of coarticulatory influences of a following stop in VC perception.

Identification of Synthetic Stimuli in Context

The results of the VC-CV conditions yielded a wealth of detail which the interested reader will find discussed in Repp (Note 3). Here, only the main findings are presented. They are summarized in Table 1, which collapses over speakers, tokens, and all utterance types. The first row, labeled Retroactive, shows the condition in which synthetic VC portions were followed by natural CV portions. The second row, labeled Proactive, shows the condition in which natural VC portions were followed by synthetic CV portions. In each condition, we have a synthetic stop, S_i, in the context of a natural stop, S_1, which was originally produced in the context of another stop, S_2, for which the synthetic stop was substituted. Since there were only three response choices (/b/, /d/, /g/), listeners could identify the synthetic stop, S_i, in one of three ways: They gave the response, S_1, that matched the other stop in the utterance, or they gave the response, S_2, that matched the excised segment, or they gave the response, S_3, that matched neither.

Now, if the natural signal portion contained any usable coarticulatory cues to the identity of the ex-

Table 1
Summary of VC-CV Data (Experiment 3)

Condition	Stimulus	S_i Perceived as (Percent): S_1	S_2	S_3
Retroactive	$S_i - (S_2)S_1$	29.3	35.5	35.2
Proactive	$S_1(S_2)-S_i$	28.2	35.8	36.0

cised segment, S_2, then S_2 responses should have been more frequent than the neutral S_3 responses. We see, however, that there was absolutely no difference between the percentages of S_2 and S_3 responses in both conditions. Contrast effects, on the other hand, should show up as a lower rate of S_1 responses, compared with S_2 and S_3 responses. It is evident that there were both proactive and retroactive contrast effects, with proactive effects being slightly larger. Statistical tests were conducted within each of the eight VC-CV conditions defined by the factors speaker, proactive/retroactive, and continuum (/b-d/ or /d-g/). Significant contrast effects were obtained in all but one (retroactive) condition. There were no significant coarticulatory effects anywhere.

Acoustic Analysis of Natural-Speech Stimuli

The frequencies of the second (F_2) and third (F_3) formants were determined at VC offset and at CV onset from spectral cross-sections generated with the help of a UA-A6 Federal Scientific spectrum analyzer. For each speaker's utterances, measurements for the 5 tokens of each syllable type were compared across the two different contexts by means of t tests.

The 12 individual comparisons for F_2 revealed four significant effects ($p < .01$). Only one difference occurred in both speakers' utterances: /(ad)ga/ had a higher F_2 onset (by 90-130 Hz) than /(ab)ga/. Speaker G.C. also showed a higher F_2 onset in /(ad)ba/ than in /(ag)ba/, which may be related to the high percentage of "d" responses to /(ad)ba/ mentioned earlier. These are all instances of perseverative coarticulation. The only anticipatory coarticulation effect was shown by speaker B.R., whose tokens of /ab(ga)/ had higher F_2 onsets than /ab(da)/.

The 12 comparisons for F_3 yielded only two significant effects, both for speaker B.R. and both perseverative. Clearly, therefore, the acoustic evidence for coarticulation was much less consistent than the occurrence of contrast effects in perception. Moreover, whatever coarticulatory effects were revealed in these acoustic analyses did not seem to have any perceptual salience.

GENERAL DISCUSSION

The results of Experiment 3 effectively rule out the hypothesis that contrast effects are due to perceptual compensation for coarticulatory interactions between successive stop consonants. How, then, are we to account for the contrast phenomenon? We have argued on the basis of Experiment 2 that there may be a low-level psychoacoustic component but that it can hardly account for the full extent of the effect, particularly of the retroactive one, in phonetic classification (Experiments 1 and 3). We propose now an explanation, not considered so far in this paper, that provides an elegant solution to the puzzle.

The explanation takes note of the fact that, in speech production, sequences of two different stop consonants have longer closure intervals than single intervocalic stops; in fact, the ratio of average durations is about two to one (Westbury, Note 4). Perceptual results suggest that the typical closure durations of double (geminate) stops, which occur only across word boundaries in English (e.g., "mad dog"), are likely to be even longer, at least in citation form (Dorman et al., 1979; Repp, 1978). It so happens that perceptual contrast effects are most pronounced precisely at those intervals that are characteristic of two-stop sequences produced in isolated disyllables. If these closure durations signaled to the listener that two stops have occurred rather than one, "contrast effects" would be a natural result: Listeners would automatically adjust their phonetic interpretation of an ambiguous stimulus portion so as to yield a place of articulation different from that conveyed by the less ambiguous portion. Effects of interval range on the magnitude of these contrastive effects may then be attributed to perceived changes in average speaking rate, and the bidirectionality and "time course" of the contrast effects are readily predicted.

Thus, according to this hypothesis, the silent closure interval, rather than merely separating the VC and CV portions, provides *information* about the number of stop consonants involved. Listeners are assumed to possess tacit knowledge about the temporal properties of speech and, specifically, of the fact that the closures of two-stop sequences are longer than those of single stops but shorter than those of double (geminate) stops. In this view, then, contrast effects do *not* derive from some perceptual interaction between the VC and CV portions, as a psychophysical view of speech perception would have it; rather, they are assumed to derive from the perceptual integration of information provided by the VC and CV formant transitions *and* by the closure interval itself. In other words, they derive from the fact that listeners interpret speech signals with reference to their knowledge of the normative properties of speech. This, after all, is the essence of phonetic perception (Repp, 1982).

The precise causes of different context effects may vary, of course. The effects of preceding fricatives and liquids on stop consonant perception (Mann, 1980; Repp & Mann, 1981) still suggest coarticulatory dependencies, since, in these cases, the duration of the stop closure interval seems to carry little information about changes in place of articulation, even though it may constitute a secondary cue to specific places of stop articulation (Bailey & Summerfield, 1980). In the case of two-stop sequences, however, the information conveyed by closure duration seems to be the major cause of (what has been mistakenly believed to be) perceptual contrast. If a closure duration of, say, 150 msec fits best the listeners' model of a sequence of two different stop consonants in an isolated VC-CV utterance, then that is what they will tend to perceive (unless other cues are powerful enough to override this tendency).

These considerations teach an important lesson: Speech perception cannot be understood without acknowledging that it occurs *with reference to* extensive internalized knowledge about all aspects of speech. Silence, as a purely acoustic event, is a meaningless vacuum separating the surrounding acoustic events. In speech, however, silence signifies something (viz, an articulatory closure), and it does so with reference to the listener's internal representation of the prototypical patterns of speech.

REFERENCE NOTES

1. Repp, B. H. *Bidirectional contrast effects in the perception of VC-CV sequences* (Status Report on Speech Research, SR-63/64). New Haven, Conn: Haskins Laboratories, 1980.

2. Repp, B. H. *A range-frequency effect on perception of silence in speech* (Status Report on Speech Research SR-61). New Haven, Conn: Haskins Laboratories, 1980.

3. Repp, B. H. *Perceptual assessment of coarticulation in sequences of two stop consonants* (Status Report on Speech Research SR-71/72). New Haven, Conn: Haskins Laboratories, 1982.

4. Westbury, J. R. *Temporal control of medial stop consonant clusters in English*. Paper presented at the 93rd Meeting of the Acoustical Society of America, State College, Pennsylvania, June 1977.

REFERENCES

Ades, A. E. How phonetic is selective adaptation? Experiments on syllable position and vowel environment. *Perception & Psychophysics*, 1974, **16**, 61-66.

Bailey, P. J., & Summerfield, Q. Information in speech: Observations on the perception of [s]-stop clusters. *Journal of Experimental Psychology: Human Perception and Performance*, 1980, **6**, 536-563.

Crowder, R. G. The role of auditory memory in speech perception and discrimination. In T. Myers, J. Laver, & J. Anderson (Eds.), *The cognitive representation of speech*. Amsterdam: North-Holland, 1981.

Crowder, R. G. Decay of auditory memory in vowel discrimination. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 1982, **8**, 153-162.

Delgutte, B. Representation of speech-like sounds in the discharge patterns of auditory-nerve fibers. *Journal of the Acoustical Society of America*, 1980, **68**, 843-857.

Dorman, M. F., Raphael, L. J., & Liberman, A. M. Some experiments on the sound of silence in phonetic perception. *Journal of the Acoustical Society of America*, 1979, **65**, 1518-1532.

Harris, D. M., & Dallos, P. Forward masking of auditory nerve fiber responses. *Journal of Neurophysiology*, 1979, **42**, 1083-1107.

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

Mann, V. A. Influence of preceding liquid on stop consonant perception. *Perception & Psychophysics*, 1980, **28**, 407-412.

Mann, V. A., & Repp, B. H. Influence of preceding fricative on stop consonant perception. *Journal of the Acoustical Society of America*, 1981, **69**, 548-558.

Repp, B. H. Perceptual integration and differentiation of spectral cues for intervocalic stop consonants. *Perception & Psychophysics*, 1978, **24**, 471-485.

Repp, B. H. Phonetic trading relations and context effects: New evidence for a phonetic mode of perception. *Psychological Bulletin*, 1982, **92**, 81-110.

Repp, B. H., Healy, A. F., & Crowder, R. G. Categories and context in the perception of isolated steady-state vowels. *Journal of Experimental Psychology: Human Perception and Performance*, 1979, **5**, 129-145.

Repp, B. H., & Mann, V. A. Perceptual assessment of fricative-stop coarticulation. *Journal of the Acoustical Society of America*, 1981, **69**, 1154-1163.

Repp, B. H., & Mann, V. A. Fricative-stop coarticulation: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America*, 1982, **71**, 1562-1567.

Sawusch, J. R. Peripheral and central processing in speech perception. *Journal of the Acoustical Society of America*, 1977, **62**, 738-750.

NOTES

1. At closure durations below approximately 70 msec, listeners generally do not perceive C_1; that is, they do not interpret the formant transitions leading into the closure as cues for a separate phonetic segment, even when those transitions specify a different place of articulation than the transitions out of the closure (Dorman, Raphael, & Liberman, 1979; Repp, 1978). This effect, which may be described as a strong assimilative (or disruptive) influence of C_2 on the perception of C_1, was not of prime interest in the present studies, although it was encountered in Experiment 1.

2. Because of the large range effects in Experiment 1, no attempt was made to calculate predicted discrimination functions from the labeling data. Even from just eyeballing the data, however, it seems clear that retroactive contrast, at least, was smaller in Experiment 2 than in Experiment 1.

(Manuscript received May 11, 1982;
revision accepted for publication November 10, 1982.)