422

Coarticulation in sequences of two nonhomorganic stop consonants: Perceptual and acoustic evidence

Bruno H. Repp
Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06511

(Received 21 June 1982; accepted for publication 18 April 1983)

This study investigated whether any perceptually useful coarticulatory information is carried by the release burst of the first of two successive, nonhomorganic stop consonants. The CV portions of natural VCCV utterances were replaced with matched synthetic stimuli from a continuum spanning the three places of stop articulation. There was a sizable effect of coarticulatory cues in the natural-speech portion on the perception of the second stop consonant. Moreover, when the natural VC portions including the final release burst were presented in isolation, listeners were significantly better than chance at guessing the identity of the following, "missing" syllable-initial stop. The hypothesis that the release burst of a syllable-final stop contains significant coarticulatory information about the place of articulation of a following, nonhomorganic stop was further confirmed in acoustic analyses which revealed significant effects of CV context on the spectral properties of the release bursts. The relationship between acoustic stimulus properties and listeners' perceptual responses was not straightforward, however.

PACS numbers: 43.70.Dn, 43.70.Gr

INTRODUCTION

Because of anticipatory coarticulation, the speech signal frequently carries advance information about phonetic segments whose primary acoustic correlates occur later in time. This advance information may be helpful to listeners, especially when the principal cues for a segment are ambiguous or when a fast decision is required (e.g., Mann, 1980; Martin and Bunnell, 1981; Repp, 1980a; Streeter and Nigro, 1979). In some cases, anticipatory coarticulation is highly predictable: For example, when oral closure is established during an utterance, the spectrum of the preceding signal portion will show characteristic changes due to the rapid narrowing of the vocal tract. Perhaps such immutable effects should not be considered "coarticulation" at all, for they are necessary consequences of the temporal unfolding and overlap of different articulatory gestures. In other cases, however, coarticulatory effects may be variable in degree and extent or entirely optional, as well as more subtle at the acoustic level. The presence of these effects must be demonstrated empirically.

From a perceptual viewpoint, too, the second type of effect is the more interesting one. Whereas the acoustic manifestation of obligatory anticipatory "coarticulation" generally provides a rather conspicuous cue to the target segment (e.g., the formant transitions into a stop closure are an important cue for stop consonant perception, while they contribute little to the perception of the preceding segment), in cases of more variable or optional coarticulation the advance cues for the target segment are often contained in acoustic properties that primarily function as a cue for a preceding segment. Thus, for example, the spectrum of a fricative noise is a primary cue for fricative identity but, at the same time, may provide information about the following vowel (Yeni-Komshian and Soli, 1981). Thus this signal portion does double duty, as a primary cue for one segment and as a secondary cue for another. Whether (and how) listeners are able to sort out the two kinds of information is an intriguing question for the student of speech perception.

Obviously, there are limits to coarticulation. To cite a relevant example: It appears from perceptual evidence (Repp, 1983) that the formant transitions associated with two successive, nonhomorganic stop consonants carry little coarticulatory information; that is, the opening or closing phase of a stop consonant does not seem to be systematically influenced by that of a preceding or following stop (although it may vary in the context of liquids or fricatives—Mann, 1980; Repp and Mann, 1982). Acoustic measurements support this conclusion, especially with regard to the absence of anticipatory coarticulation (see below). Thus although speakers have, in principle, the opportunity to make assimilatory adjustments in their articulation of two successive stops, such adjustments do not seem to occur, at least not in carefully produced utterances. One reason for this is that the closure interval separating the closing and opening phases of the two-stop sequence is sufficiently long (Repp, 1980b) to accommodate the articulatory transition from one place of stop occlusion to the next; therefore, further coarticulatory adjustments are entirely optional.

The closure period separating the two vocalic segments in a VCCV utterance (where CC is a sequence of two nonhomorganic stop consonants, C1 and C2) is largely silent or filled with low-amplitude voicing pulses that convey no information about ongoing articulatory maneuvers inside the closed cavity. However, the interval also frequently contains a noisy burst generated by the articulatory release of the first stop (henceforth, *C1 release burst*). This acoustic event, which occurs roughly in the middle of the closure interval, tends to be shorter and of lower amplitude than the typical release burst of utterance-final stops (Abercrombie, 1967; Henderson and Repp, 1982; Repp, 1980b). Nevertheless, it seems likely that it carries some perceptual information, both in its spectral properties and in its timing within the two-stop closure. Since the burst derives from the release of the first stop, it obviously contains some information specific to that stop's place of articulation—the question could only be how important that information is to a listener. The more interesting questions, investigated in the present study, were whether the burst also contains information about the place of articulation of the following stop consonant, and whether listeners are sensitive to that information. Since the articulatory closure for the second stop is established very soon after the time at which the closure of the first stop is released (or earlier, depending on the precise sequence of places of articulation), it was considered possible that acoustic properties of the C1 release burst might reflect the anticipated vocal tract configuration for the second stop as well as the place of articulation of the first stop.

Following the premise that coarticulatory effects are of little interest unless they are of some use to listeners, the emphasis of the present study was on listeners' sensitivity to (potential) coarticulatory information in C1 release bursts. To investigate this issue, the VC and closure portions (containing the C1 release bursts) of natural VCCV utterances were presented to listeners for identification of the "missing," following stop. In a second condition, the same natural stimulus portions were followed by (more or less ambiguous) synthetic CV portions, to determine whether the perception of the second, syllable-initial stop (cued primarily by the synthetic formant transitions) would be influenced by coarticulatory cues in the C1 release burst. Following this perceptual study, an acoustic analysis of the stimuli was conducted to reveal the nature of the coarticulatory variation.

I. PERCEPTUAL EXPERIMENT

A. Method

1. 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. Thus there were data from eight subjects for each stimulus set. All volunteers were native speakers of American English. The author and the research assistant are native speakers of Austrian German and Scots English, respectively. All subjects had participated in an earlier experiment using the same stimulus components (Repp, 1983; Exp. 3).

The author (BR) and a linguist colleague (GC), a native speaker of American English, produced the original sets of utterances. It was considered unlikely that the author's native German would render either his production 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 findings, two parallel sets of stimuli were used.

2. Stimuli

a. Natural utterances. Speakers GC and BR each recorded a set of nonsense utterances which included five tokens each of /abda/, /abga/, /adba/, /adga/, /agba/, and /agda/. The utterances were produced with intended stress on the first syllable, so as to prevent reduction of the first vowel. The speakers read at a steady pace from a randomized list into a Sennheiser MKH 415T microphone whose response was recorded by a Crown 822 tape recorder.

A representative token of a VCCV utterance (/adga/ produced by GC) is shown in Fig. 1 in the form of an oscillographic trace. It contains three major acoustic segments: The segment from onset to the beginning of the closure (the *VC portion*); the *closure period*; and the segment from the release of the closure to the end (the *CV portion*). Roughly in the middle of the closure period, there is a brief noise burst deriving from the articulatory release of the first stop consonant (the *C1 release burst*). Although this burst is sometimes absent in fluent speech (Henderson and Repp, 1982), it is generally found in carefully articulated utterances (see also Repp, 1980b). All but two of BR's and all but one of GC's VCCV tokens contained C1 release bursts. The average durations of the three major segments (VC, closure, CV) were 122, 132, and 299 ms for GC and 165, 150, and 240 ms for BR. The average durations of the C1 release bursts were 22 and 21 ms, respectively. For both speakers, closure voicing usually ceased prior to the C1 release, resumed briefly during the release burst, and was absent during the closure following the burst. (The token in Fig. 1 is representative in that regard.)

All utterance tokens were digitized at 10 kHz using the Haskins Laboratories pulse code modulation system. The onset of the CV portion was determined in each token, and the signal portion preceding that point (i.e., the VC portion plus closure) was stored separately for use in the perceptual experiment.

b. Synthetic stimuli. Four continua of synthetic syllables were generated on the OVE IIIc synthesizer at Haskins Laboratories. The formant trajectories of the endpoint stimuli /ba/, /da/, and /ga/ were derived from spectrograms of good-sounding natural tokens produced by each speaker. The stimuli were fully periodic and had no release bursts.

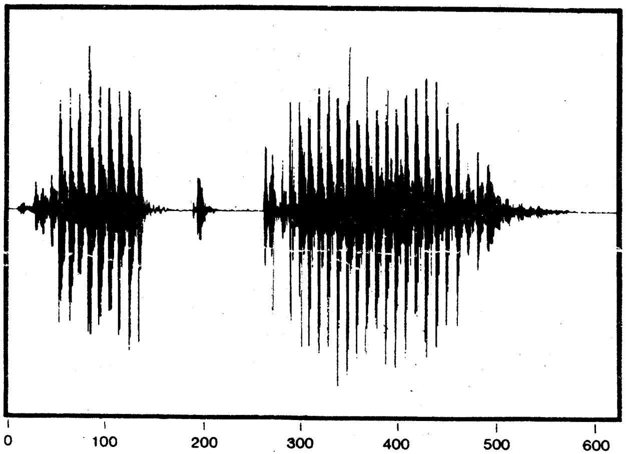

FIG. 1. Oscillogram of a VCCV utterance (/adga/ produced by speaker GC).

The stimuli for each speaker all were given 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 within each speaker set differed only in the transitions of the second and/or third formant. Seven-member synthetic continua from /ba/ to /da/ and from /da/ to /ga/ were produced for each speaker by linear interpolation between the formant tracks of the two respective endpoint stimuli, in roughly equal steps. All stimuli were digitized at 10 kHz. (For additional details, see Repp, 1982.)

c. Experimental conditions. Two parallel sets of tapes were recorded, one using the GC stimuli and one using the BR stimuli. Within each set, there were three tapes containing the following:

(1) Each of the 30 natural VC portions followed by the original closure period (which contained the C1 release burst) and by each of the seven synthetic /ba/–/da/ stimuli in all possible combinations, a total of 210 "hybrid VC–CV" stimuli.

(2) As (1), with the synthetic /da/–/ga/ stimuli.

(3) The 30 natural VC portions followed only by the original closure period, repeated five times.

The interstimulus interval was 3 s on the first two tapes and 4 s on the third. The order of tapes 1 and 2 was varied across subjects; tape 3 was always last.

3. Procedure

The stimulus tapes were played back on an Ampex AG500 tape recorder, and the subjects listened over TDH-39 earphones in a quiet room. When listening to tapes 1 and 2, the subjects chose from nine alternatives: "bb," "bd," "bg," "db," "dd," "dg," "gb,""gd," and "gg." The subjects were told that the stimuli consisted of concatenated VC and CV syllables, that the stop consonants in both syllables were to be identified, and that these consonants could be either the same or different. If only a single consonant was heard, the response was to be "bb," "dd," or "gg." For tape 3, the subjects were instructed to identify the syllable-final stop *and* the stop that might have followed it in the original VCCV utterance, guessing if necessary, with the restriction that the two stops always be different from each other. Thus subjects chose from six response alternatives here ("bd," "bg," "db," "dg," "gb," and "gd").

B. Results and discussion

1. Identification of syllable-final stops

Table I shows the percentages of correct identification of syllable-final stops in three conditions: without C1 release burst (data for the same subjects from Repp, 1983), with it (tape 3), and with added synthetic CV context (tapes 1 and 2). As can be seen, the results differed for the two stimulus sets: Inclusion of the release burst improved performance in the GC set but not in the BR set, and addition of the synthetic CV portion likewise had no consistent effect. The main point of the table is that identification of these natural-speech stimuli was accurate, though not perfect. Errors were presumably caused by some poorly articulated tokens and by the absence of the original CV context.

2. Identification of following stop from VC portion plus C1 release burst

The subjects' success in identifying the missing syllable-initial stop (C2) on tape 3 was assessed by considering only those trials on which the first stop (C1) was identified correctly, for the subjects had been told that C2 was always different from C1. Table I shows that about 92% of the trials met that requirement.

Table II shows confusion matrices as well as percent correct scores (50% correct is chance level). It is evident that performance was much better than chance for both sets of stimuli as a whole, and also for each individual place of articulation of C1, although scores varied significantly with place of articulation [$F(2,14) = 6.3, p < 0.05$, in the GC set; $F(2,14) = 5.7, p < 0.05$, in the BR set], mainly due to lower scores for labial than for alveolar or velar stops. Clearly, the stimuli did contain information about the place of articulation of C2. This information was almost certainly conveyed by the C1 release burst, despite its short duration. Although the present study did not include a condition in which subjects were asked to identify C2 from VC portions without release bursts, performance in such a task would hardly have exceeded chance.[1] Acoustic analysis showed that, in general, the VC formant transitions in these stimuli did not vary systematically with C2 (see Table V below).

3. Identification of second stop in hybrid VC–CV stimuli

The purpose of this condition was to examine whether the coarticulatory information conveyed by C1 release bursts would continue to be perceptually salient in the presence of other, more powerful (albeit sometimes ambiguous) cues to C2 place of articulation—the synthetic, vocalic formant transitions. Response percentages for tapes 1 and 2, averaged over all seven members of the synthetic CV continua, are shown in Table III. It can be seen that, for each particular VC type, the percentage of responses in the category corresponding to the original CV context (underlined in the table) was increased relative to responses in the same category when it did not correspond to the original CV context, even when that category was extraneous to the synthetic CV continuum. This coarticulatory effect was significant ($p < 0.01$) in separate analyses of variance for each of the four stimulus series. Thus the natural signal portion provided significant information about the original C2, and listeners apparently integrated this coarticulatory information with the

TABLE I. Identification of syllable-final stops (percent correct).

Stimulus set	VC	VC + burst	VC + burst + CV
GC	81.2	94.2	91.2
BR	96.3	90.5	95.8

TABLE II. Identification of second stop, given correct identification of first stop, from isolated VC portions including release burst.

Stimulus	GC "b"	GC "d"	GC "g"	GC Correct	BR "b"	BR "d"	BR "g"	BR Correct
	Response (percent)							
ab(da)	...	81	19	69	...	80	20	63
ab(ga)	...	43	57		...	54	46	
ad(ba)	69	...	31	79	65	...	35	73
ad(ga)	11	...	89		19	...	81	
ag(ba)	80	20	...	78	70	30	...	81
ag(da)	24	76	...		9	91	...	
Average				75				72

formant transition cues contained in the synthetic CV portion into a unitary C2 percept.

The "percent correct" scores[2] in Table III ("correct" with regard to the original CV context) reveal that, in all four stimulus series, coarticulatory effects were strongest when C1 was labial and, in three out of four, weakest when it was velar. [These place-of-articulation effects reached significance only for the two GC continua: $F(2,14) = 5.5$ and 7.2, $p < 0.05$ and $p < 0.01$, respectively.] This finding is unexpected because, in the condition where C2 was to be identified from the VC portion plus release burst alone (tape 3), subjects were most accurate with velar bursts and least accurate with labial bursts (see Table II). This ordering would also be predicted on the basis of the relative acoustic prominence of the burst (see below). The reversal in the presence of a synthetic CV portion is curious and remains unexplained.[3]

That the coarticulatory cues were in fact contained in the C1 release burst, and not in the VC formant transitions, is evident from a comparison of the present results with those of a condition in which the natural VC portions (without release bursts) were separated from the synthetic CV portions by a fixed silent interval (Repp, 1983). In that condition, there was no effect whatsoever of the original CV context on the perception of C2, suggesting that the VC formant transitions contained no coarticulatory information. Further support for this conclusion comes from an acoustic analysis of the stimuli (see below).

The perception of C2 was influenced not only by the coarticulatory cues in the C1 release burst but also by the place of articulation of C1. For example, consider the /ba/–/da/ continua in Table III and compare the C2 response frequencies for /ab(ga)/ and /ad(ga)/. Both of these VC precursors contained coarticulatory cues for /g/; therefore whatever different effects they had on C2 perception must have been due primarily to C1. It is evident from Table III that, in both simulus sets, there were more "d" responses in the context of /ab(ga)/ and more "b" responses in the context of /ad(ga)/. That is, there was a strong tendency to assign C2 to a category different from that of C1, which may be interpreted as a proactive contrast effect of C1 on the perception of C2. Similar comparisons among the other stimulus combinations reveal that, with one exception, contrast effects were present throughout; they were significant ($p < 0.01$) for three of the four stimulus continua in separate analyses of variance. Repp (1983) observed very similar effects when the same VC and CV portions were separated by a fixed silent interval. Thus the presence of a C1 release burst between the two signal portions by no means reduced the contrast. In fact, Repp (1983) has argued that what appears to be contrast is actually a perceptual differentiation induced by the closure duration cue: Given some ambiguity in the signal and given that the closure duration is in the range

TABLE III. Identification of synthetic CV syllables in the context of natural VC portions that include the original closure and release burst.

VC context	/ba/–/da/ continuum "b"	"d"	"g"	"correct"	/da/–/ga/ continuum "b"	"d"	"g"	"correct"
	Response to CV portion (percent)							
GC								
ab(da)	37	48	15	70	0	57	43	65
ab(ga)	26	27	47		0	27	73	
ad(ba)	62	24	14	68	6	33	61	54
ad(ga)	43	6	51		1	10	89	
ag(ba)	52	46	2	56	12	59	29	58
ag(da)	41	57	2		1	72	27	
BR								
ab(da)	49	50	1	71	5	53	42	64
ab(ga)	40	34	26		4	26	70	
ad(ba)	75	22	3	62	24	22	54	59
ad(ga)	55	23	22		10	18	72	
ag(ba)	47	52	1	59	8	65	27	55
ag(da)	29	70	1		1	81	18	

most appropriate for sequences of two nonhomorganic stop consonants, listeners will tend to perceive such sequences rather than either single or geminate stops. In addition, the infrequent occurrence of geminate stops in English may introduce a bias against geminate responses, even when listeners are aware of the concatenated nature of the stimuli.

The results shown in Table III can be summarized in three numbers: the percentages of trials on which subjects assigned C2 in a hybrid VC–CV stimulus to (a) the same category as C1, (b) the same category as C2 in the original utterance, and (c) the third, "neutral" category. Averaged over all stimuli and subjects (a total of 6720 responses), these three percentages are (a) 17.7, (b) 51.0, and (c) 31.3. The difference between (b) and (c) represents the effect on C2 perception of coarticulatory information in the C1 release burst, and the difference between (a) and (c) represents the "contrast" effect due to the identity of C1 in conjunction with closure duration as a cue for a change in stop place of articulation from C1 to C2 (plus a possible bias against geminate responses). Both effects, coarticulatory and contrast, can be seen to be substantial.

II. ACOUSTIC ANALYSIS

Detailed analyses of the natural speech stimuli were conducted to determine the acoustic correlates of the coarticulatory cues revealed in the perceptual results. Although C1 release burst spectrum and amplitude were expected to be most relevant, it was considered possible that the timing of the burst in the closure period may have provided additional information. This possibility is examined first.

A. Temporal measurements

1. Method

The duration of (a) the closure interval preceding the C1 release burst (C1 closure), (b) the C1 release burst itself, and (c) the closure interval following the C1 release burst (C2 closure) were measured on a large-scale oscillographic display to the nearest millisecond. There was generally little uncertainty about the beginning of (b) and about the end of (c). The precise beginnings of (a) and (c) were somewhat more difficult to define (cf. Fig. 1), but an attempt was made to follow consistent criteria: a significant reduction in voicing amplitude for the onset of (a), and a return to near-baseline energy for the offset of (b), i.e., the onset of (c). The sum of the three measures yielded (d) the total closure duration.

One-way analyses of variance were conducted on each of the four measures separately for each speaker, using the between-token variability as error estimate. Since the places of articulation of C1 and C2 were not orthogonal factors, their effects on the segment durations of interest were evaluated by means of Newman–Keuls tests.[4] Effects of C1 were assessed by comparing pairs of utterances in which C2 did not vary (/adba/–/agba/, /abda/–/agda/, /abga/–/adga/), and effects of C2 were assessed by comparing utterances in which C1 was constant (/abda/–/abga/, /adba/–/adga/, /agba/–/agda/).

2. Results and discussion

Mean durations, standard deviations calculated from the five (occasionally four) tokens of each utterance, and the results of the significance tests are displayed in Table IV. The results are in close agreement with earlier measurements of similar utterances reported by Repp (1980b).

The duration of the C1 closure was affected by the place of articulation of C1, being longest for /b/, but not by that of C2. Thus this portion of the closure did not convey any significant coarticulatory information.

The duration of the C1 release burst also depended primarily on C1, being shortest for /b/. It seems that this variable, too, contained little specific information about C2. The short duration of labial bursts may account in part for the lower C2 recognition scores from coarticulatory cues when

TABLE IV. Average durations of C1 release bursts and closure intervals in milliseconds (standard deviations in parentheses).

Utterance	C1 closure	C1 release burst	C2 closure	Total closure
GC				
abda	77 (20) [a]	16 (4) [a]	45 (5)	138 (21)
abga	71 (11) [a]	20 (6)	44 (10)	135 (18)
adga	52 (17) [a]	26 (8)	38 (12) [a]	116 (11)
adba	53 (13)	20 (10)	56 (19) [a]	126 (12)
agba	60 (13)	19 (9) [a]	59 (12)	138 (17)
agda	56 (5) [a]	29 (9) [a] [a]	53 (9)	138 (6)
Average	62 (10)	22 (5)	49 (8)	132 (9)
BR				
abda	66 (6) [a]	14 (10)	67 (5)	147 (5)
abga	68 (15)	9 (2) [b]	71 (6) [a]	149 (9)
adga	57 (9) [a]	27 (7) [b]	57 (11) [a] [c]	141 (10) [c]
adba	57 (4) [a]	19 (2) [a]	93 (5) [c]	169 (7) [c]
agba	41 (12)	29 (13) [a]	89 (10) [b]	159 (8)
agda	53 (7) [a]	21 (4)	75 (6) [b]	150 (9)
Average	57 (7)	21 (4)	75 (6)	150 (9)

[a] $p < 0.05$.
[b] $p < 0.01$.
[c] $p < 0.001$.

C1 was labial (Table II), but it makes the large effect of labial bursts in hybrid VC–CV utterances (Table III) seem even more surprising.

The duration of the C2 closure, on the other hand, was strongly influenced by the place of articulation of C2, being longest for /b/, especially in BR's utterances. Thus this portion of the closure may have provided a cue to the place of articulation of C2 in the VC–CV hybrid stimuli. (In the condition without synthetic CV portions, C2 closure was of course undefined.) However, note that the strongest coarticulatory effects were obtained when C1 was labial, whereas Table IV shows that precisely in this case (/abda/ versus /abga/) the C2 closure provided little information about C2. This observation suggests that spectral properties of the C1 release burst were the principal source of coarticulatory information.

It might be hypothesized that whatever spectral cues a burst contains will be more effectively perceived the longer the burst lasts. To test this hypothesis, the burst duration measurements for the five (sometimes four) individual tokens of each utterance were correlated with the average response percentages in the relevant category to the same tokens in the VC-only perceptual test (tape 3). There was some relationship in the GC set (average $r = 0.45$) but not in the BR set (average $r = -0.05$), suggesting that long bursts conveyed only little more information than short bursts.

B. Amplitude measurements

A measure of C1 burst rms amplitude (the RO parameter of the ILS analysis program)[5] was obtained from the first 15 ms of each burst. The burst amplitudes showed surprisingly little relation to the burst durations ($r = -0.17$ in the GC set; $r = 0.38, p < 0.05$, in the BR set). For both speakers, labial bursts were significantly weaker than alveolar and velar burst, and while GC produced stronger alveolar than velar bursts, BR did the opposite. These differences are obviously related to the average percent-correct scores for C2 shown in Table II. Correlations computed over tokens within each utterance revealed moderate relationships between individual burst amplitudes and C2 recognition on tape 3(average $r = 0.50$ in the GC set; average $r = 0.30$ in the BR set). This suggests that, in the absence of a following synthetic CV portion, listeners were able to extract more coarticulatory information from strong bursts than from weak ones. That the relationship was not very strong, however, is suggested by the fact that BR's bursts were generally much weaker than GC's; nevertheless, both sets of stimuli led to nearly equal perceptual effects (Tables II and III). Also, it will be recalled that in hybrid VC–CV stimuli the weakest bursts had the strongest effects.

C. Spectral measurements

1. Method

The spectrum of the initial 15 ms of each C1 release burst was obtained using an FFT program (FDI of the ILS package)[5] with a 20-ms Hamming window whose left edge was placed 5 ms before burst onset. No pre-emphasis was applied. For purposes of graphic display, the spectra were smoothed by linearly averaging over approximately 400 Hz, moving across the frequency scale in steps of roughly 20 Hz. The spectra were also amplitude normalized, and average spectra were computed from all tokens of a given utterance. Estimates of the formant frequencies in the vocalic portions immediately preceding and following the closure had been obtained previously using a UA-A6 Federal Scientific spectrum analyzer (see Repp and Mann, 1982, for details of this method).

2. Results and discussion

Figure 2 compares the smoothed average spectra of release bursts for the same C1 in different C2 contexts. All burst spectra contained significant amounts of energy in the region of the first formant ($F1$), which may indicate the presence of voicing during the burst. These $F1$ peaks were not sensitive to C2 context, however. In contrast, it can be seen that coarticulatory information about C2 resided in the second-formant ($F2$) region, between 1000 and 2000 Hz. The most striking difference occurred for velar bursts: /g(b)/ bursts had $F2$ peaks at considerably lower frequencies than did /g(d)/ bursts. Similarly, /b(g)/ bursts had $F2$ peaks at lower frequencies than /b(d)/ bursts. No such difference is evident for /d/ bursts, but /d(g)/ bursts had a pronounced energy minimum around 1000–1200 Hz, whereas /d(b)/ bursts did not.

Table V lists, in its center column, the average $F2$ peak frequencies of the various bursts. Despite the small number of tokens and considerable variability, the effects of C2 context on labial and velar bursts were highly significant in t tests. At the same time, of course, the $F2$ frequencies reflected the place of articulation of C1, being lowest for /b/ and highest for /d/. (The statistical results for these C1 effects, most of which were highly significant, are omitted from the table for the sake of clarity.) Note, however, that the effects of C2 on the $F2$ peak frequency were at least as large as those of C1.

Table V also lists, for comparison, the frequencies of $F2$ in the voiced signal portions immediately preceding and fol-

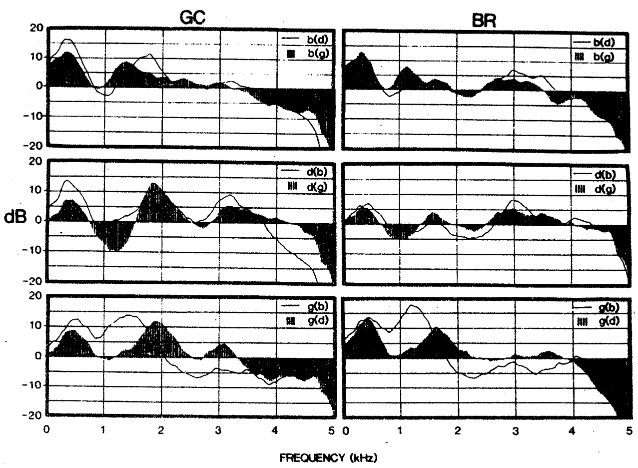

FIG. 2. Effects of CV context on C1 release burst spectra.

TABLE V. Average $F2$ frequencies in C1 release bursts and in vocalic portions immediately preceding and following the closure, in Hz (standard deviations in parentheses).

Utterance	VC offset	C1 burst	CV onset
GC			
abda	1444 (65)	[b] 1734 (77)	1868 (46)
abga	1412 (46)	1383 (106)	[a] 1772 (59)
adga	1732 (33)	1920 (138)	1860 (35)
adba	1728 (36)	2070 (241)	[a] 1516 (26)
agba	1652 (27)	[a] 1511 (138)	1424 (71)
agda	1652 (33)	1886 (99)	1840 (28)
BR			
abda	[b] 1012 (23)	[b] 1586 (167)	1416 (61)
abga	1084 (30)	1183 (107)	[b] 1400 (42)
adga	1296 (43)	1570 (99)	1532 (27)
adba	1292 (30)	1629 (58)	1100 (47)
agba	1276 (61)	[b] 1274 (97)	1084 (56)
agda	1316 (36)	1658 (41)	1415 (38)

[a] $p < 0.01$.
[b] $p < 0.001$.

lowing the closure interval. It is evident that, in general, the $F2$ frequency of the burst did not lie on a trajectory between the VC and CV frequencies. It can also be seen that, while VC and CV frequencies primarily reflected the places of articulation of C1 and C2, respectively, there were some coarticulatory effects that were significant in t tests. One of them, a lower onset frequency of $F2$ in /(ab)ga/ than in /(ad)ga/, was obtained for both speakers. However, these coarticulatory variations were apparently not effective as perceptual cues (Repp, 1983). The only anticipatory effect, a lower $F2$ offset in /ab(da)/ than in /ab(ga)/, was observed only for speaker GC and therefore does little to explain the present perceptual findings.

Given the systematic spectral differences among C1 bursts, some relation might be expected between $F2$ peak frequency and listeners' responses in the perceptual experiments. For example, /g(b)/ bursts with very low $F2$ frequencies should have led to especially high proportion of "gb" responses, and /g(d)/ bursts with very high $F2$ frequencies should have led to the highest proportions of "gd" responses. Unfortunately, this hypothesis found no support in a correlational analysis. This leaves open the question of what aspect of the C1 release bursts actually conveyed the coarticulatory information. It may have been some more complex spectral property than the $F2$ peaks considered here. This is also suggested by the fact that alveolar bursts, which did not vary significantly in $F2$ frequency, did transmit coarticulatory information.

The possibility was considered that C1 release bursts conveyed dynamic information in the form of significant spectral changes over their short duration. Spectra of the second 15-ms portion of each burst were obtained and compared to the spectra of the first 15 ms shown in Fig. 2. In general, $F2$ peaks either persisted at the same frequency or broadened or disappeared altogether, suggesting complete oral closure. $F1$ exhibited a clear downward shift, which also suggests oral closure. Speaker GC's /d(g)/ and /g(d)/ bursts seemed to show a dramatic upward shift of a spectral peak in the $F3$ region, but BR's tokens showed no such effect. On the whole, these comparisons revealed no additional clues to the nature of the coarticulatory information. Essentially, the C1 release burst seems to be a static cue that provides an acoustic snapshot of the momentary state of the vocal tract.

III. CONCLUSION

There is coarticulatory information about a following, nonhomorganic stop in the release burst of a syllable-final stop, and this information can be used by listeners despite the relative weakness of the burst as an acoustic event, and even when additional cues to the second stop are available. Although the burst derives from the release of the occlusion for the first stop, its spectrum is apparently influenced by the configuration of the articulators as they move toward (or have already attained) the occlusion for the second stop. The perceptual salience of the acoustic changes wrought by this form of anticipatory coarticulation illustrates once again the multiplicity of cues to stop place of articulation (cf. Dorman and Raphael, 1980) and listeners' sensitivity to the detailed spectral properties of the speech signal.

ACKNOWLEDGMENTS

This research was supported by NICHD Grant HD01994 and BRS Grant RR05596 to Haskins Laboratories. I am grateful to Janette Henderson for her extensive assistance, and to Leigh Lisker, Sigfrid Soli, and an anonymous reviewer for helpful comments.

---

[1] This conjecture is based on the author's performance when trying to guess C2 while listening to the VC portions only. His scores were 51% and 54% correct for the GC and BR sets, respectively, which is not significantly better than chance; his scores had been 84% correct for both stimulus sets when C1 release bursts were included (tape 3).

[2] These scores represent the average of the two underlined percentages for each VC type, each rescaled after exclusion of the irrelevant C2 = C1 response category. A score of 50% represents chance—i.e., no coarticulatory effect. Because of the constraints on responses introduced by the synthetic formant transition cues, these values are not directly comparable to the percent correct scores in Table II.

[3] One might expect labial bursts to be most effective with synthetic CV portions from the /da/–/ga/ continuum (as they were), for in this case both of the original CV contexts (/da/,/ga/) are intrinsic to the continuum, so that influences of coarticulatory cues on the interpretation of ambiguous synthetic formant transition cues may be maximized. For the same reason, however, velar bursts should have been most effective with synthetic CV portions from the /ba/–/da/ continuum, and clearly they were not. The fact that listeners were more willing to give "g" responses to stimuli from the /ba/–/da/ continuum than they were to give "b" responses to stimuli from the /da/–/ga/ continuum (Table III) likewise cannot explain the finding that the rank order of effectiveness of coarticulatory cues was labial > alveolar > velar (place of C1) for both types of synthetic continua. The possibility that velar C1 release bursts were followed by longer closure intervals than alveolar and labial bursts and therefore had less influence on C2 perception is not convincingly supported by the acoustic analysis (Table IV).

[4] These tests were conservative because only six of the possible 15 comparisons between individual means were of interest. Because of the planned nature of the comparisons, simple F or t tests would have been legitimate; this was in fact the method used with the amplitude and spectrum data.

[5] Interactive Laboratory System V4.0, developed by Signal Technology Inc., 15 W. De La Guerra, Santa Barbara, CA 93101.

Abercrombie, D. (**1967**). *Elements of General Phonetics* (Aldine, Chicago).

Dorman, M. F., and Raphael, L. J. (**1980**). "Distribution of acoustic cues for stop consonant place of articulation in VCV syllables," J. Acoust. Soc. Am. **67**, 1333–1335.

Henderson, J. B., and Repp, B. H. (**1982**). "Is a stop consonant released when followed by another stop consonant?," Phonetica **39**, 71–82.

Mann, V. A. (**1980**), "Influence of preceding liquid on stop consonant perception," Percept. Psychophys. **28**, 407–412.

Martin, J. G., and Bunnell, H. T. (**1981**). "Perception of anticipatory coarticulation effects," J. Acoust. Soc. Am. **69**, 559–567.

Repp, B.H. (**1980a**). "Accessing phonetic information during perceptual integration of temporally distributed cues," J. Phon. **8** 185–194.

Repp, B. H. (**1980b**). "Perception and production of two-stop-consonant sequences," Haskins Lab. Status Rep. Speech Res. **SR-63/64**, 177–194.

Repp, B. H. (**1982**). "Perceptual assessment of coarticulation in sequences of two stop consonants," Haskins Lab. Status Rep. Speech Res. **SR-71/72**, 131–166.

Repp, B. H. (**1983**). "Bidirectional contrast effects in the perception of VC–CV sequences," Percept. Psychophys. **33**, 147–155.

Repp, B. H., and Mann, V. A. (**1982**). "Fricative-stop coarticulation: Acoustic and perceptual evidence," J. Acoust. Soc. Am. **71**, 1562–1567.

Streeter, L. A., and Nigro, G. N. (**1979**). "The role of medial consonant transitions in word perception," J. Acoust. Soc. Am. **65**, 1533–1541.

Yeni-Komshian, G. H., and Soli, S. D. (**1981**). "Recognition of vowels from information in fricatives: Perceptual evidence of fricative-vowel coarticulation," J. Acoust. Soc. Am. **70**, 966–975.