Influence of following context on perception of the voiced–voiceless distinction in syllable-final stop consonants

Bruno H. Repp and David R. Williams[a)]
Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06511–6695

(Received 8 October 1984; accepted for publication 4 April 1985)

This paper reports acoustic measurements and results from a series of perceptual experiments on the voiced–voiceless distinction for syllable-final stop consonants in absolute final position and in the context of a following syllable beginning with a different stop consonant. The focus is on temporal cues to the distinction, with vowel duration and silent closure duration as the primary and secondary dimensions, respectively. The main results are that adding a second syllable to a monosyllable increases the number of voiced stop consonant responses, as does shortening of the closure duration in disyllables. Both of these effects are consistent with temporal regularities in speech production: Vowel durations are shorter in the first syllable of disyllables than in monosyllables, and closure durations are shorter for voiced than for voiceless stops in disyllabic utterances of this type. While the perceptual effects thus may derive from two separate sources of tacit phonetic knowledge available to listeners, the data are also consistent with an interpretation in terms of a single effect; one of temporal proximity of following context.

PACS numbers: 43.71.Es, 43.70.Fq

INTRODUCTION

Acoustic cues to the perception of the phonological voiced–voiceless distinction in American English syllable-final stop consonants have been investigated quite intensively in recent years. One important cue is "vowel duration" (i.e., the duration of the periodic stimulus portion taken to correspond to the vowel—see, e.g., Raphael, 1972; Raphael *et al.*, 1980) which is consistent with the commonly observed longer duration of vowels preceding voiced consonants in speech production (House and Fairbanks, 1953; Peterson and Lehiste, 1960). Other relevant perceptual cues include the offset characteristics (i.e., formant transitions and amplitude envelope) of the "vowel" (Wang, 1959; Wolf, 1978), its fundamental frequency contour (Lehiste, 1976; Gruenenfelder and Pisoni, 1980), and—if the stop consonant is released—the acoustic properties of the release (Malécot, 1958; Wolf, 1978), as well as the duration and voicing of the closure interval (Hogan and Rozsypal, 1980; Raphael, 1981). All these cues are also relevant for stop consonants in intervocalic position, where additional voicing information may be contained in the vocalic portion following the release burst (Lisker, 1978).

Although vowel duration is not always the most salient voicing cue (e.g., Wardrip-Fruin, 1982), it is nevertheless an acoustic dimension that has consistently been found to influence the perception of phonological stop consonant voicing in English. As a purely temporal cue, it has attracted researchers' attention because it offers an opportunity to study the sensitivity of phonetic perception to local and global changes in speaking rate—a topic of much theoretical interest (Miller, 1981; Port, 1981). A second temporal cue to the voicing distinction—the duration of the closure interval—is available in intervocalic and released final stops. Provided that closure voicing, an often overriding cue (e.g., Lisker, 1981), is eliminated, closure duration provides important voicing information for intervocalic stops (Lisker, 1957), though it is less salient in released utterance-final stops (Raphael, 1981). Port and Dalby (1982) have proposed that the joint perceptual influence of the two temporal variables, vowel duration and (silent) closure duration, is best expressed by a constant ratio rule (however, see Massaro and Cohen, 1983a). In production, too, the ratio of the two interval durations (the C/V ratio) seems to be fairly constant across changes in global speaking rate (Barry, 1979; Port, 1981). The ratio varies, however, across different utterance positions and as a function of other voicing cues (Barry, 1979).

The present study investigates the perception of phonological stop consonant voicing in a context that has not been studied previously, viz., when a syllable-final stop is followed by another syllable beginning with a stop consonant having a different place of articulation. Although the basic voicing cues are likely to be those already studied extensively in connection with stop consonants in intervocalic or absolute final position, sequences of two different stop consonants have several peculiar properties that warrant detailed investigation.

One consideration is that the total closure period for sequences of two nonhomorganic stop consonants is about twice as long as that for a single intervocalic stop (Westbury, 1977; Repp, 1982). When the first stop is unreleased, as is frequently the case (Henderson and Repp, 1982), there is no acoustic or perceptual basis for subdividing the total closure interval into portions pertaining to the two consecutive (perhaps overlapping) stop closures. This raises the question of whether closure duration is a salient cue for the perception of voicing in this context. Certainly, if it has any effect at all, one would expect to find different critical C/V ratios than for single intervocalic stops.

[a)] Also Department of Psychology, University of Connecticut, Storrs, CT 06268

 0001-4966/85/080445-13$00.80

When the first stop is released, the situation is similar to that for released stops in absolute final position, except that the release bursts of stops followed by another stop are generally much weaker (Henderson and Repp, 1982). This raises the question of whether such weak bursts can serve as a perceptual marker delimiting the closure interval pertaining to the syllable-final stop, thereby possibly changing the relative salience of the closure duration cue and with it the critical ratio to vowel duration. It has been shown that these weak release bursts carry considerable place of articulation information (Repp, 1983b), so it is not unreasonable to expect that they might have some influence on voicing perception as well.

Another prediction that could be made is that, in sequences of two nonhomorganic stop consonants, the onset characteristics of the second syllable will have less of an effect on the perceived voicing of the preceding syllable-final stop than they have in the case of single intervocalic stops. For one thing, the temporal separation of pre- and post-closure cues is greater because of the extended closure interval, which makes perceptual integration more difficult. In addition, the two stops have different places of articulation, which may result in a perceptual segregation of the respective voicing cues, syllable-initial cues pertaining only to the syllable-initial consonant. On the other hand, it has been observed in such VC_1C_2V stimuli that the perception of the place of articulation of one stop is influenced by that of the other (Repp, 1983a), so it may be asked whether there are similar (contrastive) interactions with regard to voicing perception. Repp (1983a) linked apparent perceptual contrast effects in place-of-articulation perception to listeners' intrinsic knowledge of systematic variations in closure duration in natural speech. However, apart from an unpublished study by Westbury (1977), little is known about the acoustic consequences of phonological voicing in nonhomorganic stop sequences. Therefore, the present study reports acoustic as well as perceptual data.

A final important goal of the present investigation was to demonstrate an effect of following context on voicing perception by comparing the absolute vowel durations required to change voiceless to voiced final stop percepts in monosyllables and in disyllables. Since it is known that, in production, the duration of a syllable decreases when a second syllable is added to form a disyllabic word (Lehiste, 1972; Klatt, 1973), it was predicted that addition of a second syllable would considerably reduce the absolute vowel duration at the voiced–voiceless boundary in the first syllable, while maintaining the perceptual relevance of the vowel duration cue. Indeed, an analogous effect on the perception of phonological vowel length was shown long ago by Nooteboom (1973). Given such a contextual effect, it might be asked further whether the effect is dependent on whether or not the listener considers the two syllables as parts of the same word, and by how much the temporal separation between the two syllables can be increased before the voicing boundary for the final stop of the first syllable approaches that for the final stop in an isolated monosyllable (cf. Nooteboom and Doodeman, 1980).

For the present experiments, two pairs of monosyllabic English words were sought, such that one ended and the other began with either a voiced or a voiceless stop consonant (in terms of spelling, at least), and that made sense in all four possible disyllabic combinations as well as in isolation. Such a set does not exist. Rather than using nonsense materials, an approximation was devised by using the words LAB/LAP and GOAT/COAT, which in combination yield the real word LABCOAT, as well as the novel but potentially meaningful compounds LAPCOAT, LABGOAT, and LAPGOAT. This set was considered more attractive to listeners than complete nonsense, although the possibility of semantic bias in phonetic perception (cf. Ganong, 1980) must be considered. As will be seen, however, it is unlikely that such a bias influenced the results in any significant way.

I. ACOUSTIC MEASUREMENTS

Acoustic measurements were obtained to get a general idea of the acoustic consequences of phonological stop consonant voicing in sequences of two nonhomorganic stop consonants, and particularly in the kinds of utterances used also in the perceptual experiments. The only relevant previous data were reported by Westbury (1977) who measured three speakers' productions of isolated CVC_1C_2VC nonsense utterances, in which C_1 and C_2 were stop consonants differing in both voicing and place of articulation. Westbury's acoustic measurements showed that, in voiced–voiceless stop sequences, the vowel in the first syllable was about 10 ms longer and the closure interval about 20 ms shorter than in voiceless–voiced sequences. Whether these differences were due to the voicing characteristics of the first or the second stop, or both, cannot be determined from Westbury's data. Also, C_1 in these utterances was either consistently unreleased, or Westbury ignored the C_1 release burst in his measurements.

In the present study, too, speakers produced isolated utterances. While these productions are not representative of fluent speech, and a certain amount of deliberate enhancement of phonetic differences may be expected, the data are appropriate for comparisons with perceptual responses to similarly isolated utterances, as collected in the subsequent experiments.

A. Method

1. Subjects

Four native speakers of American English, three females (CG, JM, AB) and one male (DW, the second author), served as talkers. CG and JM grew up in New York, AB in the Midwest, and DW in California.

2. Utterances

The utterances were LAB, LAP, GOAT, COAT, LABGOAT, LABCOAT, LAPGOAT, and LAPCOAT. Ten different random orders of these eight words were concatenated and printed on a sheet of paper in standard English spelling.

3. Recording procedure

Each talker read from the list after practicing for a few minutes. The instructions were to read at a steady rate, pausing after each word, and to speak clearly but naturally. The utterances were recorded in a sound-insulated booth using high-quality equipment.

4. Measurement procedures

Temporal properties of the utterances were measured from magnified CRT waveform displays. Durations to the nearest tenth of a millisecond were obtained between the following acoustic landmarks (described here with reference to a disyllabic utterance): (a) the onset of significant energy; (b) the point of change from [l] to [æ], as determined visually by a noticeable change in the waveform of the glottal cycle; (c) the beginning of the closure interval, as indicated by a significant damping or cessation of voicing pulses; (d) the onset of the C_1 release burst, if present; (e) the end of the C_1 burst; and (f) the onset of the C_2 release burst.

In this way, the durations of the following acoustic segments were obtained: (1) [l] resonance, (2) [æ] resonance ("vowel"),[1] and (3) total closure interval. If C_1 was released, (3) could be subdivided into (3a) C_1 closure, (3b) C_1 release burst, and (3c) C_2 closure. Monosyllabic LAB and LAP were always released by three talkers; talker CG did not release eight out of 20 utterances. In disyllabic context, labial release bursts were produced in all tokens by talkers JM and CG; there were 11/40 unreleased tokens for AB and 5/40 for DW.

For each talker, means and standard deviations of these acoustic segment durations were calculated from the ten repetitions of each utterance. True outliers were omitted; there were not more than a few for each talker. Analyses of variance were conducted separately for each dependent variable, using a repeated-measures design on the mean durations.

B. Results and discussion

One effect of interest is the shortening of the first syllable in a disyllabic word, as compared to its production in isolation. Although such shortening has been described previously (e.g., Lehiste, 1972; Klatt, 1973; Nooteboom, 1973), the added syllables in these studies were unstressed, while the present disyllables had a spondaic stress pattern. Nevertheless, shortening of the first syllable was exhibited consistently by all talkers. Table I compares the durations of [l], [æ], C_1 closure, and C_1 release burst in mono-and disyllables, averaging over the four talkers and over voicing distinctions (rows labeled "mean"). It is evident that most of the shortening took place during the [æ] portion—an average reduction of 73 ms (30 %), which was highly significant, $F(1,3) = 44.6$, $p < 0.007$. By contrast, the C_1 closure changed by only 10 ms (10 %), $F(1,3) = 16.5$, $p < 0.03$, and the [l] portion did not change significantly. In addition, a dramatic difference in C_1 release burst durations is evident, $F(1,3) = 71.5$, $p < 0.004$: While utterance-final labial release bursts contained significant amounts of aspiration or voicing, those in disyllabic utterances basically represented only the brief noise generated by the parting of the lips (cf. Henderson and Repp, 1982).

The amount of vowel shortening is comparable to that observed for trochaic words (Klatt, 1973). Klatt also observed about twice as much shortening when the first syllable ended in a voiced consonant than when it ended in a voiceless consonant. Such a trend was also found in the present data (32 % for LAB versus 26 % for LAP), though it was much smaller, primarily due to relatively less shortening of LAB than would be predicted from Klatt's data.

The second comparison of interest is that between phonologically voiced and voiceless syllable-final stop consonants (LAB vs LAP in Table I). It is evident that the [æ] portion was longer, $F(1,3) = 39.4$, $p < 0.009$, and the C_1 closure duration was shorter, $F(1,3) = 5.0$, $p < 0.12$, in LAB than in LAP. (The difference in C_1 closure duration was shown by all four subjects but varied considerably in magnitude; hence the low level of significance.) C_1 release bursts tended to be longer when the stop was voiceless; however, this difference was shown by only two talkers. The C_2 closure was also affected by C_1 voicing, being shorter for LAB than for LAP, $F(1,3) = 22.5$, $p < 0.02$. The duration of the initial [l], on the other hand, was completely unaffected by stop consonant voicing. Another difference, not shown in Table I, was that the C_1 closure of LAB usually contained low-amplitude voicing while that of LAP did not. The voicing usually ceased before the end of the C_1 closure.

The average difference in [æ] duration between LAB and LAP was 98 ms (33 %) in monosyllables and 55 ms (27 %) in disyllables. For C_1 closure, the difference in duration was −25 ms (30%) in monosyllables and −16 ms (21%) in disyllables. The C/V ratios in mono- and disyllables, respectively, were 0.28 and 0.39 for LAB, and 0.55 and 0.65 for

TABLE I. Comparison of acoustic segment durations (ms) in monosyllables and in disyllables, averaged across talkers and tokens.

	[l]	[æ]	C_1 closure	C_1 burst	C_2 closure	Total closure
Monosyllables						
LAB	82	295	83	117		
LAP	83	197	108	136		
Mean	83	246	96	126		
Disyllables						
LAB-	78	201	78	14	84	176
LAP-	77	146	94	21	117	232
Mean	78	173	86	17	101	204

LAP. This comparison shows that the addition of a second syllable increased the C/V ratio, which thus was not invariant.

A third comparison relevant to the perceptual experiments concerns the effect of the voicing of C_2 (GOAT vs COAT) on acoustic properties of the preceding syllable. Table II lists the relevant data, averaged over all other factors. It is evident that there was very little effect: Only the [æ] vowel was slightly shorter preceding COAT than preceding GOAT—a small difference that was almost unreasonably consistent across talkers, $F(1,3) = 358.9, p = 0.0003$. In addition, the C_2 closure seemed to be affected by C_2 voicing, being longer for GOAT than for COAT. However, this difference, which runs counter to the common finding of longer closures for voiceless than for voiced stops, was exhibited by only two talkers and hence was nonsignificant. Westbury's (1977) observation that the *total* closure was shorter in voiced–voiceless than in voiceless–voiced C_1C_2 sequences is thus confirmed by the present data, even though closure durations exhibited large individual differences.

To summarize: "vowel duration" in LAB/LAP is substantially reduced when a second syllable is added; it is shorter in LAP than in LAB; and it is slightly shorter preceding GOAT than preceding COAT. C_1 closure duration tends to be shorter for LAB than for LAP, and it is shortened somewhat when a second syllable is added. C_2 closure may be longer preceding GOAT than preceding COAT. C_1 release bursts are substantially reduced in intersyllabic position as compared to absolute final position.

II. PERCEPTION EXPERIMENTS

The perceptual studies were carried out in two stages. An initial five-part study (experiments 1–5) was followed by a two-part replication experiment conducted several years later with a new set of stimuli (experiments 6 and 7).

A. General method: Experiments 1–5

1. Subjects

Nine paid student volunteers served as subjects. They were all native speakers of American English and reported having no speech or hearing problems.

2. Stimuli

Selected utterances of one female talker (CG) were used for stimulus construction. A continuum from LAP to LAB was constructed from the first syllable of a representative token of LABGOAT. The original duration of that syllable (not including the closure interval) was 320 ms. A seven-member continuum, not including the original syllable, was constructed by deleting pitch pulses from the interior of the [æ] portion. The initial [l] portion, approximately 62 ms long, was left undisturbed. The members of the LAP/LAB continuum had vowel durations ranging from 118–238 ms in 20-ms steps. A second LAP/LAB continuum, used only in experiment 1, was constructed from the first syllable of a good token of LAPCOAT whose original duration was 226 ms (51 ms for [l]), by either deleting or duplicating pitch pulses in the [æ] portion. The vowel durations of the seven members of that continuum ranged from 134 to 255 ms in 20-ms steps.[2] These disyllable-derived stimuli were acceptable as monosyllables with a neutral intonation, whereas stimuli fashioned from LAB or LAP produced in isolation would have been unacceptable in the context of a disyllabic word because of their falling intonation. A strongly falling intonation contour would also have made construction of a syllable duration continuum problematic.

In some conditions, the stimuli from the LAP/LAB continuum were followed by one of two C_1 release bursts. These bursts and the surrounding closure interval were derived from tokens of LABCOAT and LAPGOAT, respectively. Any closure voicing present was replaced with silence, so as to eliminate a potentially overriding (Lisker, 1981) nontemporal cue. The C_1 closure durations were 65 and 88 ms, respectively, the release burst durations were 9 and 13 ms, and the total closure durations were 170 and 243 ms. These durations were representative of those observed in the sample of utterances recorded and analyzed (see Table I). Note that closure duration and origin of release burst were confounded.

A good token of GOAT, with a voice onset time (VOT) of 22 ms and a total duration of 473 ms (including the final [t] release burst) was excerpted from LABGOAT. Since it was desirable to use tokens of GOAT and COAT that differed only in their initial VOT, the initial 66 ms of the second syllable of LABCOAT, representing the aperiodic portion (i.e., the VOT) of COAT, was substituted for the initial 66 ms of GOAT to yield an acceptable token of COAT.[3]

The stimuli for each experiment were recorded on audio tape in five randomized blocks with intertrial intervals of 2.5 s.

3. Procedure

Each subject participated in two sessions. In the first session, the stimulus tapes representing experiments 1, 2, and one additional test (see footnote 3) were presented. In the second session, experiments 5, 3, and 4 were administered, always in that order. Subjects listened over TDH-39 earphones in a quiet room. Their task was to identify in writing the final stop consonant of the first syllable ("b" or "p") and,

TABLE II. Comparison of average acoustic segment durations (ms) as a function of following GOAT or COAT in disyllables.

	[l]	[æ]	C_1 closure	C_1 burst	C_2 closure	Total closure
-GOAT	78	177	87	18	110	215
-COAT	78	170	85	17	91	193

when a second syllable followed, its initial stop consonant as well ("g" or "c"), even when it was constant (except for experiment 4).

B. Experiment 1

The purpose of this first test was twofold: to assess the influence of a final release burst on the LAP/LAB distinction in monosyllabic stimuli, and to compare perception of two LAP/LAB continua, one derived from an original utterance of LAB-, the other from LAP-. Several earlier investigators have noted that it is easier to change an originally voiced syllable-final stop consonant into a voiceless one by manipulating vowel and/or silent closure duration than vice versa (Price and Lisker, 1979; Hogan and Rozsypal, 1980). At the very least, a "trading relation" between vowel duration and differential vowel offset cues was expected in experiment 1. As to the perceptual contribution of the C_1 closure and release burst, previous investigations (e.g., Malécot, 1958; Wolf, 1978; Hillenbrand *et al.*, 1984) employing release bursts appropriate for utterance-final stops generally found only small effects. Although the disyllable-derived release bursts in the present stimuli were acoustically weaker, the relative ambiguity created by varying vowel duration was expected to enhance any effects of secondary voicing cues.

1. Method

For the seven stimuli from each LAP/LAB continuum, a final release burst (with associated C_1 closure) was either present or absent and, if present, derived from either LAB- or LAP-. Stimuli from a GOAT/COAT continuum (see footnote 3) were interspersed as fillers.

2. Results and discussion

The results are shown in Fig. 1. It can be seen, first, that the average percentage of "b" responses increased as vowel duration increased. Thus, vowel duration was an effective cue to the voiced–voiceless distinction, as intended. Second, "b" responses were much more frequent to the LAB-derived stimuli than to the LAP-derived stimuli, which reflects additional important cues presumably located at the offset of the periodic stimulus portion. This difference was highly significant in a repeated-measures analysis of variance, $F(1,8) = 33.1$, $p < 0.001$. Third, the presence of a release burst (and of an associated C_1 closure interval) did make a difference, $F(2,16) = 11.9$, $p < 0.001$. The effect was generally one of reducing "b" responses, even when the burst and closure derived from LAB-, at least in the case of the LAP-derived continuum. A separate analysis of variance was conducted on stimuli with bursts only. Bursts derived from LAP- led to fewer "b" responses than did bursts derived from LAB-, $F(1,8) = 8.8$, $p < 0.02$, and this difference was more pronounced for the LAB-derived continuum, $F(1,8) = 5.8$, $p < 0.05$ for the interaction.

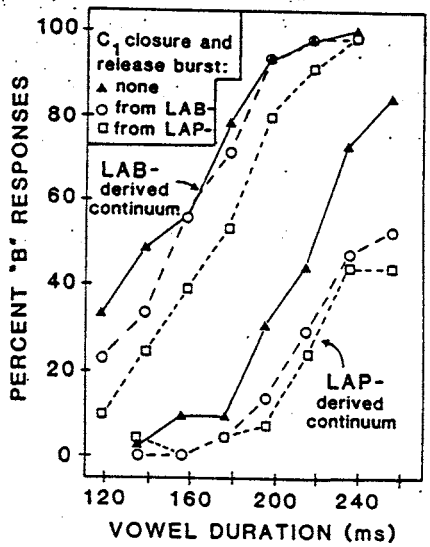

FIG. 1. Effects of vowel offset cues (LAB- vs LAP-derived continuum) and C_1 closure/release burst on the voiced–voiceless distinction for syllable-final stops along a vowel duration continuum (experiment 1).

These results confirm the relative salience of vowel duration as a voicing cue. Although, within the range of durations used here, a complete change of perceived category was not achieved in either continuum, vowel duration was about equally effective in changing LAB to LAP, and LAP to LAB. This is in contrast to some earlier studies that have found it difficult to change final voiceless stops into voiced ones (Price and Lisker, 1979; Hogan and Rozsypal, 1980). The lack of such an asymmetry in the present stimuli may be due to their having been produced in the context of a disyllabic word, which perhaps made the vowel offset cues somewhat less pronounced. Still, they were quite strong, being "worth" roughly 80 ms of vowel duration at the point of maximal ambiguity.

The finding that release bursts reduced "b" responses regardless of the burst's origin may be attributed to the absence of closure voicing: Presence of a release burst defined a closure interval that was always silent and thus more appropriate for "p" than for "b." The differential effect of LAB- and LAP-derived bursts may have been due either to properties of the release bursts themselves or to C_1 closure duration (or both). Why this effect was more pronounced with the LAB-derived continuum is not clear.

C. Experiment 2

The purpose of experiment 2 was fourfold. First, the LAP/LAB stimuli were now presented in a disyllabic context (i.e., followed by GOAT or COAT), and a shortening of the absolute vowel duration necessary to cue the LAP/LAB distinction was expected relative to experiment 1, by analogy to the findings of Nooteboom (1973) and Nooteboom and Doodeman (1980). Second, experiment 2 investigated the perceptual contribution of (total) closure duration in the disyllabic context. Third, the effect of presence versus absence of a C_1 release burst was also studied in this new context. Finally, possible perceptual contrast effects due to the voicing category of the initial consonant of the second syllable (GOAT or COAT) were assessed.

1. Methods

Only the LAB-derived LAP/LAB continuum was used. These syllables were followed by one of four closure intervals

and by either GOAT or COAT. Two of the closure intervals were those also used in experiment 1, which contained a release burst derived from either a voiced or a voiceless syllable-final stop. In contrast to experiment 1, however, where only the C_1 closure was defined (65 or 88 ms), here the C_2 closure was defined as well by the onset of the GOAT or COAT syllable. Two additional conditions resulted from substituting silence for the release bursts so that the closure interval consisted of either 170 or 243 ms of pure silence.

2. Results and discussion

Analysis of subjects' responses revealed no influence of the GOAT/COAT contrast on the LAP/LAB distinction, $F(1,8) = 1.7$. Therefore, Fig. 2 shows the results collapsed over this factor, as a function of total closure duration and presence versus absence of a release burst. It is evident from the figure and from the statistical analysis that the release burst had no systematic effect, $F(1,8) = 1.5$. However, total closure duration did have an influence: Fewer "b" responses were obtained with the longer closure duration, $F(1,8) = 20.9$ $p < 0.002$. Finally, contrary to expectations, the LAP/LAB boundaries were located at about the same point as in experiment 1, revealing no influence of the addition of a second syllable.

The absence of this expected contextual effect must be interpreted with caution because of the different stimulus ensembles used in experiments 1 and 2. In experiment 1 the inclusion of LAP-derived stimuli in the test sequence may have had a contrastive effect that pushed the boundary for the LAB-derived stimuli toward shorter vowel durations. That this was the case is suggested by the results of experiments 4 and 6, which directly compared LAP/LAB stimuli in isolation and in disyllabic context, and obtained a reliable difference (see below).

The absence of an effect of GOAT vs COAT on perception of the LAP/LAB contrast indirectly supports Repp's (1983a) conclusion that there are no perceptual contrast effects between syllable-final and syllable-initial stop consonants (see also Ades, 1974; Samuel *et al.*, 1984). What seemed like a contrast effect (for place of articulation) in Repp's study was ultimately attributed to perceptual information conveyed by closure duration. The present acoustic measurements showed that GOAT vs COAT had only a negligible influence on the duration of the preceding closure, so the absence of any perceptual effect on the LAP/LAB distinction is consistent with speech production. Incidentally, the absence of any response preference for the real word LAB-COAT over the disyllabic pseudowords suggests that semantic biases played no role in the present experiment.

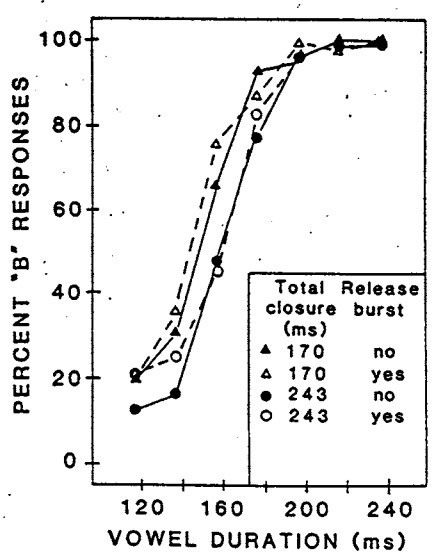

FIG. 2. Effects of total closure duration and release burst on the voiced–voiceless distinction for syllable-final stops in disyllables (experiment 2).

The absence of any effect due to the release burst was somewhat surprising in view of the fairly large effect obtained in experiment 1. While that effect could have been due to either C_1 closure duration or properties of the release bursts themselves, neither variable was effective in the disyllabic context. One possible explanation is that the following syllable had a masking effect on the weak release burst, making it difficult to detect (cf. Henderson and Repp, 1982).

The only significant effect obtained in experiment 2, that of total closure duration, is consistent with the acoustic measurements which showed total closure duration to be longer following a voiceless syllable-final stop.

D. Experiment 3

Experiment 3 investigated further the potential role of the release burst in disyllabic context by manipulating its position within the closure interval.

1. Methods

Stimuli from the LAB-derived continuum were always followed by COAT. The single release burst used was the one originally taken from LAPCOAT. The silent intervals surrounding this release burst were modified as follows: The total closure duration was made either short (120 ms) or long (200 ms), and the release burst was placed 40 or 80 ms (in the short interval) or 40, 80, 120, or 160 ms (in the long interval) after the beginning of the closure, defining corresponding C_1 closure durations.

2. Results and discussion

The effect of total closure duration was again obtained, $F(1,8) = 12.2, p < 0.01$; the results resembled those obtained in experiment 2 (see Fig. 2). Effects of release burst position (i.e., C_1 closure duration), on the other hand, were small and apparent only in the longer closure interval: $F(3,24) = 3.7$, $p < 0.03$, in a separate analysis. The effect was not monotonic: "b" responses decreased slightly as C_1 closure duration increased from 40 to 80 to 120 ms—which is in the expected direction—but increased again for a C_1 closure of 160 ms. Perhaps, this very late-occurring release burst effectively suggested a shortening of the total closure interval. Alternatively, the burst may have been masked by the onset of the second syllable in that condition, which restored a "long C_1 closure" to a neutral value.

E. Experiment 4

In this test, isolated LAP/LAB stimuli were directly compared with disyllabic stimuli, either with or without release bursts. Thus, the issue of whether LAP/LAB syllables from a vowel duration continuum exhibit a shorter category boundary in disyllabic context than in isolation was re-examined. Recall that the experiment 1 versus experiment 2 comparison yielded a negative result, but this was attributed to possible stimulus range effects. In addition, a possible influence of the rate of production of the second syllable on the perception of the LAP/LAB contrast was investigated (cf. Port and Dalby, 1982).

1. Methods

The LAB-derived continuum was used in conjunction with the LAB-derived closure interval (total duration 170 ms) and release burst, which was either present or absent. There were two versions of the second syllable: the GOAT used previously, which had been produced in a disyllabic context and was 473 ms in total duration, and another token of GOAT from the same speaker, which had been produced in isolation and measured 610 ms. In this test, the subjects identified only the syllable-final stop.

2. Results and discussion

Presence versus absence of a release burst made no significant difference, $F(1,8) = 1.5$. In hindsight, this is not surprising in view of the fact that the burst used happened to be the one that had little effect in experiment 1 (see the two leftmost functions in Fig. 1). The average data did suggest an effect in the expected direction for isolated LAP/LAB syllables, but the relevant interaction was not nearly significant, $F(1,8) = 1.2$.

Collapsing over this factor, Fig. 3 compares the labeling functions for LAP/LAB syllables in isolation and when followed by either version of GOAT. In the disyllables, there were somewhat more "b" responses when the long GOAT followed than when the short GOAT followed. This effect was very small but reached significance, $F(1,8) = 7.8$, $p < 0.03$. A similar effect of the rate of production of the second syllable on the DIGGER–DICKER distinction was reported by Port and Dalby (1982), although in their stimuli closure duration, not vowel duration, was the primary voicing cue. In addition, a shift in the boundary for isolated syllables relative to disyllables can be seen in Fig. 3, $F(1,8) = 9.5$, $p < 0.02$. This is the hypothesized effect of syllabic context, which failed to emerge in a comparison of experiments 1 and 2. Thus, the suspicion that this earlier comparison was invalid because of differences in stimulus ensemble tends to be confirmed by the present data. Another replication of this context effect was sought in experiment 6.

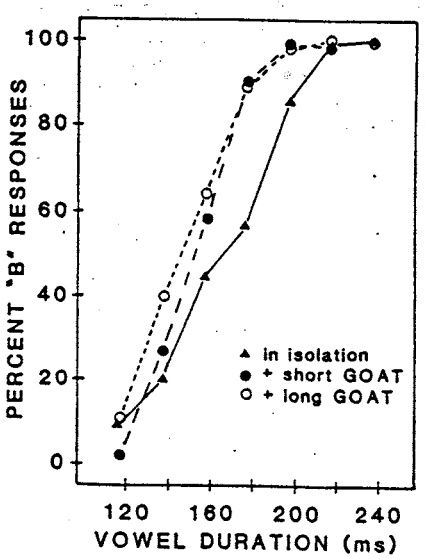

FIG. 3. Effect of adding a second syllable, and of the rate of production of that syllable, on the voiced–voiceless distinction for syllable final stops (experiment 4).

F. Experiment 5

One result has emerged clearly from experiments 2 and 3: In disyllables, the number of "b" responses decreases as total closure duration increases. Presumably, this indicates that phonological voicing information is conveyed by closure duration, in agreement with the acoustic measurements. Experiment 4 also suggests that the LAP/LAB boundary for isolated syllables is indeed at a longer vowel duration than that for the same syllables in disyllabic context. This difference may be attributed to a perceptual compensation for the expected shortening of the first syllable in disyllabic context (cf. Nooteboom, 1973). Note, however, that the direction of the boundary shift when a second syllable is added (viz., the increase in "b" responses) is the same as results from a shortening of the closure duration in disyllables. Thus, it is conceivable that the effects of closure duration and of syllabic context are one and the same, reflecting temporal proximity of following context.

The precise time course of the change in the LAP/LAB boundary in disyllables as a function of a wide range of closure durations may provide relevant information. Certainly, as the closure duration is increased to very long values, the influence of the second syllable on the LAP/LAB boundary should cease, and the boundary should equal that for isolated monosyllables. Experiment 5 sought to determine the temporal separation (closure duration) at which this asymptote is reached, as well as the shape of the function relating the LAP/LAB boundary to closure duration. If there is only a single factor involved—temporal proximity of following context—this function should be monotonically increasing until the asymptote is reached. On the other hand, if there are two factors—closure duration acting as a voicing cue and presence/absence of syllabic context making an independent contribution—then closure durations typical of voiceless stops (i.e., around 230 ms, cf. Table I) should lead to a relative *decrease* in "b" responses counteracting the effect of syllabic context, which increases "b" responses. Thus, depending on the relative strengths of the two opposing effects, the function may either be nonmonotonic, or have an early asymptote, or exhibit a change in slope around the point where the cue value of closure duration changes polarity (i.e., around 200 ms).

The subjects in experiment 5 were also asked to judge on each trial whether they thought the two syllables formed a single compound (pseudo-) word or whether they sounded like two unrelated monosyllables. It was of interest to determine whether the "one word"–"two words" boundary (1/2 boundary, for short) would coincide with the intersyllabic temporal separation (i.e., closure duration) at which the LAP/LAB boundary function reached its asymptote. Such a finding might suggest a top-down influence on phonetic perception, or at least a common factor influencing both types of judgment.

1. Methods

The stimuli from the LAB-derived LAP/LAB continuum were followed by the standard GOAT at each of eight temporal separations (closure intervals) ranging from 150 to 500 ms in 50-ms steps. The closure intervals were completely silent; release bursts were not included in this test.

In addition to identifying both stop consonants, subjects were asked to indicate for each two-syllable sequence whether it sounded like a single disyllabic word (LABGOAT or LAPGOAT) or like two unrelated monosyllabic words (LAB, GOAT or LAP, GOAT). They indicated the latter judgment by placing a comma between the two consonant responses ("b,g" or "p,g"). The results of one subject were discarded because she gave no "p" responses. (Reasons unknown.)

2. Results and discussion

The average category boundary on the LAP/LAB continuum was determined by linear interpolation between the two data points straddling the 50% crossover, separately for each closure duration. The solid line in Fig. 4 shows this vowel duration boundary as a function of closure duration. It is evident that the boundary shifted to longer vowel durations (i.e., "b" responses decreased) as closure duration increased from 150 to 250 ms. The overall effect of closure duration was highly significant, $F(7,49) = 8.6$, $p < 0.0001$. However, the boundary remained fixed at closure durations beyond 250 ms, $F(5,35) = 1.4$. This is just slightly beyond the typical closure duration following voiceless stops. No nonmonotonic trend is evident.

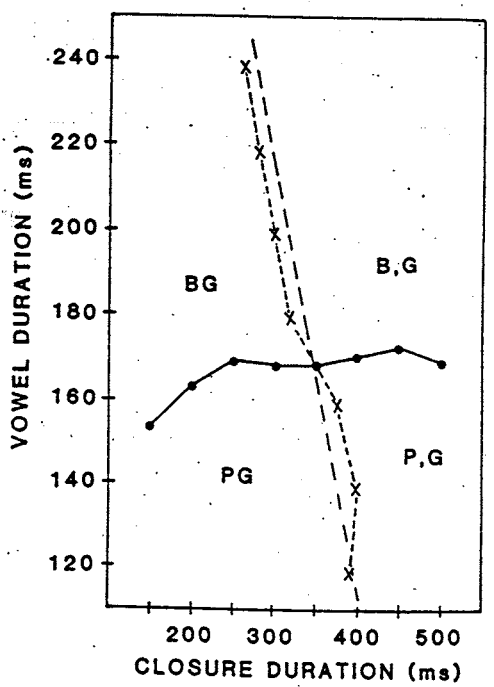

FIG. 4. The voiced–voiceless boundary (solid line) and the 1/2 boundary (dotted line) as a function of closure duration and vowel duration in disyllables (experiment 5). The dashed line represents a slope of −1, i.e., a constant sum of vowel and closure durations.

"Two words" responses increased monotonically as closure duration increased, as expected. The 1/2 boundary, expressed in terms of closure duration, was calculated separately for each vowel duration and is shown as the dotted line in Fig. 4. This boundary decreased strongly as vowel duration increased, $F(6,42) = 30.4$, $p < 0.0001$, except at the shortest vowel durations. The slope of this function is not far from −1 (a dashed line with this slope is drawn in Fig. 4); in other words, the sum of vowel and closure durations at the 1/2 boundary tended to be constant. Thus, it appears that the subjects based their judgments not on the silent intersyllable interval but on the interval between the onsets of the first and second syllables. Note also that the boundary continued to decrease at the longest vowel durations where the syllable-final consonant was almost uniformly labeled "b." Therefore, these judgments did not seem to be contingent on identification of the stop consonant, although the steepest slope of the 1/2 boundary function did occur in the region of the LAP/LAB boundary. In addition, it may be noted that the 1/2 boundary function intersects the LAP/LAB boundary function at 350 ms of closure silence, i.e., 100 ms beyond the point at which closure duration loses its effectiveness as a voicing cue. Thus, whatever process is responsible for the effect of closure duration on voicing judgments, does not seem to be a direct consequence or a common determinant of perceiving the two syllables as part of a single word.

These results give no reason to consider the effect of adding a second syllable as different, in principle, from the effect of shortening the closure interval in a disyllable. Indeed, it appears that adding a second syllable has an effect only when the resulting closure interval is sufficiently short. Relative to LAP produced in isolation, shortening of the vowel in LAP was observed even when a closure interval averaging 232 ms intervened before the second syllable (cf. Table I). Yet, the perceptual effect of adding a syllable after that long a closure duration was almost nil (cf. Fig. 4).[4] This may mean that closure duration should be viewed as an asymmetric cue: Short closures are a cue for the category "voiced," but long closures are devoid of any perceptual cue value. That is, a stop never sounds "more voiceless" in disyllabic context than in isolation. Alternatively, a tendency to hear LAP at closure durations characteristic of voiceless stops (200–250 ms) may have been canceled by an opposing tendency to hear LAB because of temporal recalibration (i.e., a subjective stretching of the vowel) in the presence of a second syllable. Although this two-process model is not implausible, a single-process explanation must be preferred on grounds of parsimony. A summary of this argument is presented schematically in Fig. 5. (However, see Sec. III.)

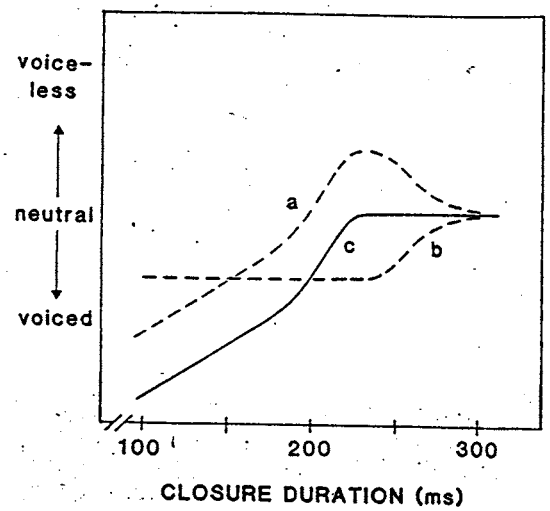

FIG. 5. Schematic illustration of a two-process model of the perceptual effect of closure duration. Closure durations of less than 200 ms are assumed to cue voicedness, whereas closure durations of about 200–300 ms are assumed to cue voicelessness (function a). At the same time, a constant temporal compensation due to the presence of a second syllable is assumed to occur, as long as the closure duration does not exceed about 250 ms (function b). The resultant (function c) shows only an increase in voiced percepts because functions a and b cancel beyond 230 ms or so.

G. Summary of experiments 1–5

In summary, these studies show:

(1) that addition of a second syllable shifts the LAP/LAB boundary toward shorter vowel durations, as long as the temporal separation (closure) is less than about 250 ms;

(2) that total closure duration (in the range below 250 ms) is a cue to the LAP/LAB distinction, with shorter closures leading to more "b" responses;

(3) that the effect of syllabic context is not a direct consequence of hearing the two syllables as part of a single word, and that it may indeed be identical with (2);

(4) that judgments of "two words" increase almost linearly with the duration of the first syllable and thus seem to rest on the perceived separation of syllable onsets;

(5) and that C_1 release burst and properties of the second syllable (VOT, overall duration) play at best a minor role in the perception of the LAP/LAB distinction in disyllabic context.

Experiments 6 and 7 attempted to replicate findings (1)–(4) with a new set of stimuli and a new group of subjects.

H. General methods: Experiments 6 and 7

1. Subjects

Sixteen undergraduate students enrolled in an introductory psychology course at the University of Connecticut participated in the experiment for course credit. All subjects were native speakers of American English with no history of hearing impairment.

2. Stimuli

Three representative disyllabic utterances of the male talker (DW), digitized at 10 kHz, served as bases for the LAB, GOAT, and COAT stimuli used in the present experiments. The LAB stimulus was excerpted from a good LABCOAT. Tokens of LABGOAT and a second LABCOAT yielded the GOAT and COAT stimuli.

A LAP/LAB continuum was constructed by successively deleting every other pitch pulse from the vocalic region of the LAB stimulus. The first deleted pitch pulse began 55 ms into the syllable (following the formant transitions for /l/) and the last, one pitch pulse prior to closure for /b/. In all, eight pitch pulses were excised yielding a series of nine stimuli. The longest stimulus was the original LAB (220 ms); the shortest ("LAP") was 135 ms in duration. The vowel durations thus ranged from 80 to 165 ms across the LAP/LAB continuum. Note that these durations are considerably shorter than those employed in experiments 1–5, reflecting differences in the speaking rates of talkers DW and CG.

When GOAT or COAT was appended to the LAP/LAB stimuli to form disyllables, the closure was completely silent; no release bursts of the syllable-final stop consonant were included in experiments 6 and 7. The duration of the COAT stimulus (487 ms) was somewhat greater than that of the GOAT stimulus (439 ms), although this difference was located mainly in the final release burst.

3. Procedure

Two stimulus tapes (one for each experiment) were prepared with intertrial intervals of 2.5 s and presented to subjects binaurally over TDH-39 headphones. Both tapes were heard during the 75-min session, with experiment 7 always following experiment 6.

I. Experiment 6

In this study, voicing judgments for disyllables were compared with those for LAP/LAB monosyllables when presented in a separate test and when included in the disyllable test. We also wished to replicate the effect of closure duration on the voicing boundary, using (silent) closure durations that were appropriate for the present, shorter stimuli. In addition, the possible influence of the natural GOAT and COAT on voicing judgments was reassessed.

1. Method

Two different stimulus sequences were presented. In the first, only the nine stimuli from the monosyllabic LAP/LAB series were included. Following five repetitions of the endpoint stimuli, subjects listened to ten randomized blocks of the nine stimuli. In the second sequence, each block included two occurrences of the monosyllabic stimuli interspersed among single occurrences of stimuli from four disyllabic continua. The disyllabic stimuli were constructed by appending the GOAT or COAT stimulus to each member of the LAP/LAB series following either a 120- or a 170-ms silent interval. These intervals corresponded to speaker DW's average closure durations in his utterances of LAB- and LAP- disyllables, respectively. There were five blocks of 54 stimuli. Appropriate responses to the monosyllabic stimuli were "b" and "p." For the disyllables, subjects responded with "bg," "pg," "bc," or "pc" depending on whether LABGOAT, LAPGOAT, LABCOAT, or LAPCOAT was heard.

2. Results and discussion

In the disyllables, there was a small but significant effect of the identity of the following syllable, GOAT vs COAT, $F(1,15) = 14.9$, $p < 0.002$. Voiced responses were more frequent preceding COAT, which is consistent with three alternative explanations: (1) If closure durations were longer preceding COAT in production, then listeners' tacit knowledge of that regularity might lead them to shift their perceptual criterion in favor of voiced responses in that context. (2) The effect may be due to second syllable duration, COAT being longer than GOAT in the present experiment. (3) The effect may represent response contrast between the voicing categories of C_1 and C_2. Considering that (1) is not supported by our acoustic measurements (see Table II) and that (3) was not obtained in experiment 2, the most likely explanation seems to be (2), in accordance with the effect of second-syllable duration obtained in experiment 4.

Figure 6 presents the results for monosyllables (dashed lines) and for disyllables collapsed over the GOAT/COAT factor (solid lines). As expected, increasing the closure duration in disyllables had the effect of shifting the voicing boundary toward longer vowel durations. Significantly fewer "b" responses were made to disyllables with long closures than to those with short closures, $F(1,15) = 21.7$, $p < 0.0003$. It is also evident that subjects gave significantly fewer "b" responses to the monosyllables than to the disyllables, $F(1,15) = 73.0$, $p < 0.0001$. Thus, these results replicate for the present set of stimuli the finding (experiment 4) that the LAP/LAB boundary is shifted toward shorter vowel durations in disyllabic context.

Figure 6 further shows that subjects gave fewer "b" responses to the monosyllables that were interspersed among the disyllables than to those that were presented alone, $F(1,15) = 13.2$, $p = 0.003$. This represents a stimulus range effect: Because the syllable-final stops sounded relatively more voiced in disyllables, they presumably sounded relatively more voiceless in the interspersed monosyllables. The finding of such an effect lends further credence to the earlier argument proffered in connection with experiments 1 and 2, where the apparent absence of an effect of syllabic context was attributed to the presence of stimulus range effects.

J. Experiment 7

This experiment replicated experiment 5 using the new set of stimuli.

1. Method

The GOAT stimulus was appended to the stimuli from the LAP/LAB series at each of eight temporal separations (closure durations) ranging from 100 to 450 ms in 50-ms steps. These stimuli were recorded in five randomized blocks. Subjects were asked to decide on each trial whether they heard /b/ or /p/, and whether they heard one two-syllable word or two one-syllable words, using the responses B1, P1, B2, or P2.

2. Results and discussion

As in experiment 5, the average category boundary on the LAP/LAB continuum was determined for each closure duration by means of linear interpolation. These boundaries are plotted as a function of closure duration in Fig. 7 (solid line). As expected, the overall effect of closure duration was highly significant, $F(7,105) = 50.4$, $p < 0.0001$: Increasing closure duration shifted the LAP/LAB boundary toward longer vowel durations. The effect leveled off around closure durations of 200–250 ms. But, in contrast to experiment 5, there was a small increase in the boundary even beyond 300 ms, $F(3,45) = 3.7$, $p < 0.02$. Again, the boundary function is monotonic, with an asymptote close to the boundary for monosyllables in experiment 6.

The 1/2 boundary was calculated separately for each vowel duration and is plotted as a function of vowel duration in Fig. 7 (dotted line). As in experiment 5, it is evident that

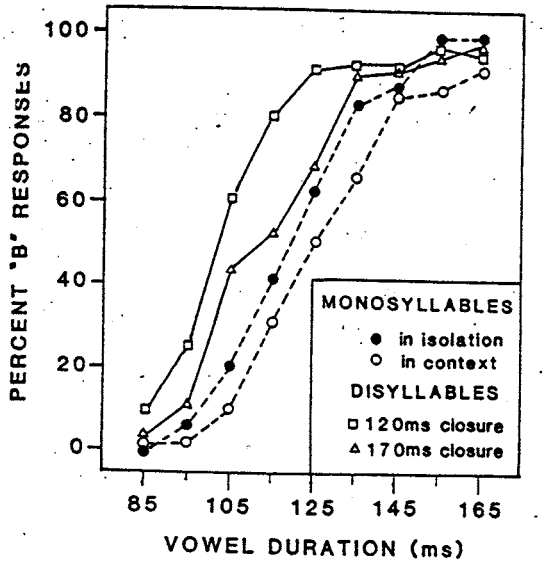

FIG. 6. Percent voiced responses as a function of vowel duration for isolated monosyllables, monosyllables interspersed among disyllables, and disyllables with two closure durations (experiment 6).

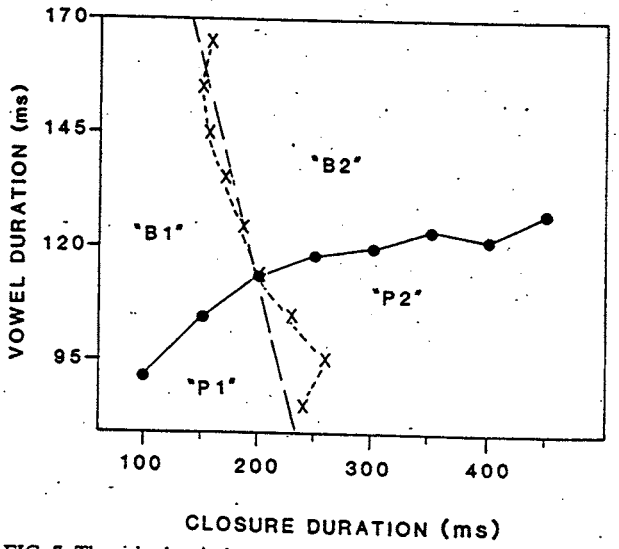

FIG. 7. The voiced–voiceless boundary (solid line) and the 1/2 boundary (dotted line) in disyllables as a joint function of closure duration and vowel duration (experiment 7). The dashed line represents a slope of -1, i.e., a constant sum of vowel and closure durations.

the boundary decreases sharply as vowel duration increases, $F(8,120) = 29.7$, $p < 0.0001$ (except perhaps at the shortest and longest vowel durations). Again, the slope of the function is close to 1, indicating that the sum of vowel and closure duration at the 1/2 boundary tends to be constant. Thus, the data confirm that these judgments were based on the interval between the onsets of the first and second syllables.

Because of the shorter syllable durations, the subjects in experiment 7 were probably more inclined to consider the two syllables as separate words than were the subjects in experiment 5.[5] Thus, the "one word"/"two words" boundary function intersects the voicing boundary function at 200 ms of closure duration (versus 350 ms in experiment 5). While this would be consistent with the hypothesis that syllabic context (or closure duration) has its effect contingent on perception of the two syllables as part of the same utterance, this hypothesis was rejected on the basis of the results of experiment 5. That is, the coincidence of the 1/2 boundary with the leveling off of the syllabic context (or closure duration) effect may be just that, a coincidence. At any rate, the data of experiment 7 are again consistent with a single-process explanation of the effect of closure duration: Short closures increase voiced responses.

K. Summary of experiments 6 and 7

The results of these replication experiments affirm all the major conclusions of experiments 1–5, as summarized above. There are two minor discrepancies: the absence of a stable asymptote of the LAP/LAB boundary function, and of a clear dissociation of voicing and 1/2 judgments. The earlier data were clearer in these regards, but the replication must nevertheless be considered useful in view of the stability of the major findings across different stimulus materials and subject groups.

III. GENERAL DISCUSSION

The principal aim of the present series of studies was to demonstrate two effects in the perception of the voicing category of syllable-final stop consonants, one being due to the addition of a second syllable beginning with a different stop, and the other reflecting the perceptual contribution of the closure duration cue in disyllables. An effect of closure duration was consistently obtained: The shorter the closure, the more likely subjects were to report a voiced stop consonant. This is in agreement with our measurements of closure durations in natural speech and suggests that listeners have incorporated tacit knowledge about these temporal regularities into their perceptual criteria for the voiced–voiceless distinction. By the same token, one should expect that this knowledge includes the fact, well-known from earlier speech production studies and substantiated by our acoustic measurements, that a syllable contracts when a second syllable is added to it (or, equivalently, that a syllable is lengthened in utterance-final position). Nooteboom (1973) has demonstrated such perceptual compensation in a task requiring judgments of phonological vowel length. The present data are consistent with such a temporal compensation mechanism which operates in addition to an independent perceptual effect of closure duration as a voicing cue (see Fig. 5).

Although the results are compatible with such a two-process model, they do not provide compelling evidence in its favor. A more parsimonious interpretation of the results, at least when considered in isolation from other findings in speech perception research, is that there is only a single effect, that of closure duration. That effect, moreover, is unidirectional: A final stop consonant sounds increasingly "voiced" as closure duration decreases, but even at closure durations that are optimal for voiceless stops in disyllables subjects do not give more voiceless responses than they give to monosyllables which lack any closure duration information. In other words, the closure duration cue apparently contributes only to the perception of stops as voiced, not as voiceless. This is not unreasonable: Even though "voicing" may be used as an abstract cover term for a variety of acoustic manifestations of a phonological distinction, it may also be understood more narrowly as designating the common acoustic feature of "presence of low-frequency energy" within a certain time span (Stevens *et al.*, 1985). Absence of low-frequency energy is the neutral state; that is, voicing is an acoustically "marked" feature. The problem in applying this view to the present data lies in the finding that it made little difference whether the second syllable began with a voiced or a voiceless (aspirated) stop. If the decisive factor was presence of low-frequency energy within a certain interval following the offset of the first syllable, aspiration following the closure should not have increased voiced responses as much as did a voiced (actually, weakly aspirated) signal.

A specific auditory mechanism which has been discussed in connection with speech is backward recognition masking (e.g., Massaro, 1975). Although the ineffectiveness of C_1 release bursts in disyllables may be due to auditory backward masking, the perceptual effect of decreasing closure duration is not compatible with such an explanation: Masking of either the vowel offset cues (which favored voiced percepts, since the stimuli were derived from an original utterance of LAB) or of the perceived duration of the first syllable (leading to a reduction in subjective vowel duration—see Massaro and Idson, 1978) should have decreased, not increased voiced responses. It seems, therefore, that proximity of following context had its effect without interfering with or altering those auditory properties of the first syllable that fed into phonetic decisions.

These hypotheses surely do not exhaust the possible mechanisms that may underlie a unidimensional, unidirectional effect of closure duration. The failure of two specific accounts, however, raises doubt about whether the isolated parsimony of a single-process model is indeed preferable to a two-process model, particularly one that ties in with a multitude of related observations suggesting that listeners make phonetic decisions in accord with criteria which reflect the phonetic regularities of the language (Nooteboom, 1973; Nooteboom and Doodeman, 1980; Repp, 1982, 1983c).

A comment is in order concerning the relationship between vowel duration and closure duration at the voiced/voiceless boundary. For single intervocalic post-stressed stops, as in DIGGER/DICKER (Port and Dalby, 1982), the

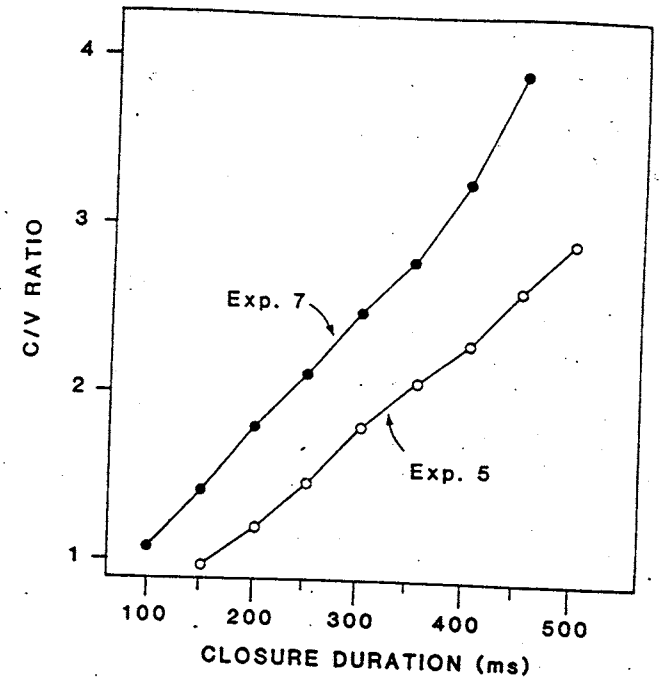

FIG. 8. The average C/V ratio as a function of closure duration (experiments 5 and 7). The ratios were computed from the average data shown in Figs. 4 and 7, respectively.

ratio of closure to vowel durations (C/V) at the voicing boundary remains approximately constant as one of the two temporal variables is manipulated. Port and Dalby varied vowel duration while determining the boundary on a closure duration continuum; we manipulated closure duration while determining the boundary on a vowel duration continuum. In the context of two-stop sequences, closure duration is a much less salient voicing cue than it is for single intervocalic stops. Because of the longer closure durations, larger absolute C/V ratios were to be expected; the question was whether they would remain constant in the region where closure duration influenced the boundary on the vowel duration continuum. As can be seen in Fig. 8, the answer is negative. The average C/V ratios from both experiments 5 and 7 increase as a nearly linear function of closure duration over the whole range. Thus, a constant-ratio rule does not hold for these stimuli; such a rule may be restricted to the specific utterance types considered by Port and Dalby (1982).

Another type of constancy was found in the present data, however: The sum of vowel duration and closure duration at the 1/2 boundary was approximately constant. That is, subjects based their "one word"/"two words" judgments on the onset-to-onset interval between the two syllables and not on the separation (the silent closure) between them. Perception of the syllable-final stop as voiced or voiceless seemed to be independent of these timing judgments. This is in agreement with the recent findings of Miller *et al.* (1984), who showed that perception of a particular temporally cued phonetic contrast was independent of explicit judgments of perceived speaking rate. It appears that the speech signal supports a variety of independent judgments which do not interact, though they may combine at higher levels of organization (cf. Ganong, 1980; Massaro and Cohen, 1983b).

On the whole, then, the present results are consistent with the general notion that human listeners behave as if they knew all the detailed acoustic consequences of articulation, including context-conditioned and position-specific variation. The perceptual effects of various acoustic cues can almost always be rationalized by reference to the systematic patterns that emerge in the acoustic analysis of speech, although, considering the many factors that play a role in perception, a precise prediction of experimental results from acoustic regularities is rarely possible. The future development of a more economic description of speech in terms of dynamic articulatory processes may ease the burden on the perception theorist.

ACKNOWLEDGMENTS

This research was supported by NICHD Grant HD-01994 and BRS Grant RR-05596 to Haskins Laboratories. Results from experiments 6 and 7 were reported at the 107th Meeting of the Acoustical Society of America in Norfolk, VA, May 1984.

[1]Authors such as Klatt (1973) and Port (1981) took the midpoint of the formant transitions in spectrograms to be the acoustic liquid-vowel boundary. The oscillographic criterion used here allots most of the transition portion to the vowel, which seems justified on the basis of perceptual data (Raphael *et al.*, 1980).

[2]The syllable durations were similar across the two continua, as intended, except for a small difference caused by the discreteness of the pitch pulses. The difference in vowel durations is mainly a consequence of the difference in [l] duration. The decision to use vowel duration, rather than syllable duration, in the presentation of results was made at a rather late stage.

[3]In addition, a GOAT-COAT continuum was constructed by substituting aspiration noise from COAT for successive pitch periods in GOAT (see the appendix in Ganong, 1980, for a description of this procedure). This continuum was used in another test administered to the present subjects at the end of the first session. That condition investigated the influence of preceding LAB or LAP, closure duration, and C_1 release bursts on perception of the GOAT/COAT distinction. There were no systematic effects of any of these variables; thus, VOT appeared to be the only salient cue to the GOAT/COAT distinction in these stimuli.

[4]The assumption here is that the constant asymptotic vowel duration boundary of about 170 ms matches that for isolated LAP/LAB syllables. This is supported by a comparison with the results of experiment 4 (cf. Fig. 3).

[5]Note that the 1/2 boundary did not just bisect the range of closure durations but was clearly to the left of the center of the stimulus range. This indicates that 1/2 judgments were not arbitrary and rested on some preestablished internal criteria.

Ades, A. E. (1974). "How phonetic is selective adaptation? Experiments on syllable position and vowel environment," Percept. Psychophys. 16, 61–66.

Barry, W. J. (1979). "Complex encoding in word-final voiced and voiceless stops," Phonetica 36, 361–372.

Ganong, W. F., III. (1980). "Phonetic categorization in auditory word perception," J. Exp. Psychol.: Human Percept. Perform. 6, 110–125.

Gruenenfelder, T. M., and Pisoni, D. B. (1980). "Fundamental frequency as a cue to post-vocalic consonantal voicing: Some data from speech perception and production," Percept. Psychophys. 28, 514–520.

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

Hillenbrand, J., Ingrisano, D. R., Smith, B. L., and Flege, J. E. (1984). "Perception of the voiced–voiceless contrast in syllable-final stops," J. Acoust. Soc. Am. 76, 18–26.

Hogan, J. T., and Roszypal, A. J. (1980). "Evaluation of vowel duration as a cue for the voicing distinction in the following word-final consonant," J.

Acoust. Soc. Am. 67, 1764–1771.
House, A. S., and Fairbanks, G. (1953). "The influence of consonant environment upon the secondary acoustical characteristics of vowels," J. Acoust. Soc. Am. 25, 105–133.
Klatt, D. H. (1973). "Interaction between two factors that influence vowel duration," J. Acoust. Soc. Am. 54, 1102–1104.
Lehiste, I. (1972). "Timing of utterances and linguistic boundaries," J. Acoust. Soc. Am. 51, 2018–2024.
Lehiste, I. (1976). "Influence of fundamental frequency pattern on the perception of duration," J. Phonet. 4, 113–117.
Lisker, L. (1957). "Closure duration and the intervocalic voiced–voiceless distinction in English," Language 33, 42–49.
Lisker, L. (1978). "*Rapid* vs *Rabid*: A catalog of acoustic features that may cue the distinction," Haskins Lab. Stat. Rep. Speech Res. SR-54, 127–132.
Lisker, L. (1981). "On generalizing the *Rabid-Rapid* distinction based on silent gap duration," Haskins Lab. Stat. Rep. Speech Res., SR-65, 251–259.
Malécot, A. (1958). "The role of releases in the identification of released final stops," Language 34, 370–380.
Massaro, D. W. (1975). "Perceptual images, processing times, and perceptual units in speech perception," in *Understanding Language: An Information-Processing Analysis of Speech Perception, Reading, and Psycholinguistics*, edited by D. W. Massaro (Academic, New York), pp. 125–150.
Massaro, D. W., and Cohen, M. M. (1983a). "Consonant/vowel ratio: An improbable cue in speech," Percept. Psychophys. 33, 501–505.
Massaro, D. W., and Cohen, M. M. (1983b). "Phonological context in speech perception," Percept. Psychophys. 34, 338–348.
Massaro, D. W., and Idson, W. L. (1978). "Temporal course of perceived vowel duration," J. Speech Hear. Res. 21, 37–55.
Miller, J. L. (1981). "The effect of speaking rate on segmental distinctions: Acoustic variation and perceptual compensation," in *Perspectives in the Study of Speech*, edited by P. D. Eimas and J. L. Miller (Erlbaum, Hillsdale, NJ).
Miller, J. L., Aibel, I. L., and Green, K. (1984). "On the nature of rate-dependent processing during phonetic perception," Percept. Psychophys. 35, 5–15.
Nooteboom, S. G. (1973). "The perceptual reality of some prosodic durations," J. Phonet. 1, 25–45.
Nooteboom, S. G., and Doodeman, C. J. N. (1980). "Production and perception of vowel length in spoken sentences," J. Acoust. Soc. Am. 67, 276–287.
Peterson, G. E., and Lehiste, I. (1960). "Duration of syllable nuclei in English," J. Acoust. Soc. Am. 32, 693–703.
Port, R. F. (1981). "Linguistic timing factors in combination," J. Acoust. Soc. Am. 69, 262–274.
Port, R. F., and Dalby, J. (1982). "Consonant/vowel ratio as a cue for voicing in English," Percept. Psychophys. 32, 141–152.
Price, P. J., and Lisker, L. (1979). "(/b/→/p/) but ~(/p/→/b/)," in *Speech Communications Papers: 97th Meeting of the Acoustical Society of America*, edited by J. Wolf and D. Klatt (Acoustical Society of America, New York), pp. 49–52.
Raphael, L. J. (1972). "Preceding vowel duration as a cue to the perception of voicing in word-final consonants in American English," J. Acoust. Soc. Am. 51, 1296–1303.
Raphael, L. J. (1981). "Durations and contexts as cues to word-final cognate opposition in English," Phonetica 69, 1462–1464.
Raphael, L. J., Dorman, M. F., and Liberman, A. M. (1980). "On defining the vowel duration that cues voicing in final position," Lang. Speech 23, 297–308.
Repp, B. H. (1982). "Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception," Psychol. Bull. 92, 81–110.
Repp. B. H. (1983a). "Bidirectional contrast effects in the perception of VC–CV sequences," Percept. Psychophys. 33, 147–155.
Repp, B. H. (1983b). "Coarticulation in sequences of two nonhomorganic stop consonants: Perceptual and acoustic evidence," J. Acoust. Soc. Am. 74, 420–427.
Repp, B. H. (1983c). "Trading relations among acoustic cues in speech perception: Speech-specific but not special," Haskins Lab. Stat. Rep. Speech Res. SR-76, 129–132.
Samuel, A. G., Kat, D., and Tartter, V. C. (1984). "Which syllable does an intervocalic stop belong to? A selective adaptation study," J. Acoust. Soc. Am. 76, 1652–1663.
Stevens, K. N., Keyser, S. J., and Kawasaki, H. (1985). "Toward a phonetic and phonological theory of redundant features," in *Invariance and Variability of Speech Processes*, edited by J. S. Perkell and D. H. Klatt (Erlbaum, Hillsdale, NJ) (in press).
Wang, W. S.-Y. (1959). "Transition and release as perceptual cues for final plosives," J. Speech Hear. Res. 2, 66–73.
Wardrip-Fruin, C. (1982). "On the status of phonetic categories: Preceding vowel duration as a cue to voicing in final stop consonants," J. Acoust. Soc. Am. 71, 187–195.
Westbury, J. R. (1977). "Temporal control of medial stop consonant clusters in English," paper presented at the 93rd meeting of the Acoustical Society of America, State College, Pennsylvania, J. Acoust. Soc. Am. Suppl. 1 61, S91.
Wolf, C. G. (1978). "Voicing cues in English final stops," J. Phonet. 6, 299–309.