Consonant/vowel ratio as a cue for voicing in English

ROBERT F. PORT and JONATHAN DALBY
Department of Linguistics, Indiana University, Bloomington, Indiana

Several experiments investigate voicing judgments in minimal pairs like *rabid-rapid* when the duration of the first vowel and the medial stop are varied factorially and other cues for voicing remain ambiguous. In Experiments 1 and 2, in which synthetic labial and velar-stop voicing pairs are investigated, the perceptual boundary along a continuum of silent consonant durations varies in constant proportion to increases in the duration of the preceding vocalic interval. In Experiment 3, it is shown that speaking tempo external to the test word has far smaller effects on a closure duration boundary for voicing than does the tempo within the test word. Experiment 4 shows that, even within the word, it is primarily the preceding vowel that accounts for changes in the consonant duration effects. Furthermore, in Experiments 3 and 4, the effects of timing outside the vowel-consonant interval are independent of the duration of that interval itself. These findings suggest that consonant/vowel ratio serves as a primary acoustic cue for English voicing in syllable-final position and imply that this ratio possibly is directly extracted from the speech signal.

Twenty-five years ago, Denes (1955) reported an interesting phenomenon regarding the perception of phonological voicing by English listeners. He found that a voiceless final fricative in a synthetically constructed monosyllable could be heard as having either the plus or minus value of the English voicing feature, depending on the relative duration of the vowel and final fricative itself for words like *his* and *hiss*. Despite variations in the overall word length, the fricative was heard as a voiceless /s/ when it was about the same duration as the vowel that preceded it or longer, and it was heard as a voiced /z/ when the fricative was much shorter than the vowel. Until quite recently (see Cole & Cooper, 1975), Denes' paper has been cited primarily as evidence for the perceptual effectiveness of the two *individual* parameters of vowel duration and postvocalic consonant duration as cues to the voicing of the consonant,

rather than for the far more general result of a possible ratio cue for voicing in English in syllable-final position.[1]

Recently, interest in this question has been revived by research dealing with three general areas. One issue is the relationship among different cues for a single phonological or phonetic feature. It is well established that vowel duration is an important correlate of changes in syllable-final voicing in English (Klatt, 1976; Peterson & Lehiste, 1960; Raphael, 1972), as is the duration of the consonant constriction itself (Lisker, 1957; Port, 1979, 1981a). Yet it has often been noted that such timing cues are easily dominated by nontemporal cues such as closure voicing and aspiration (Port, 1979; Price & Lisker, 1979; Wardrip-Fruin, 1982; Lisker, Note 3, Note 4). That is, even if we restrict attention to isolated words or to sentence-stressed words in carrier sentences, where temporal cues are the strongest (Umeda, 1975), the presence of glottal pulsing during a final obstruent forces perception of a voiced consonant in English, no matter what manipulations are performed on the timing. This can be interpreted as implying the presence of a hierarchic relationship between the temporal cues for voicing with respect to other acoustic cues that are manifestations of opening and closing the glottis. The hierarchy is such that timing will control the perception of syllable-final voicing in English only when other cues are ambiguous (Barry, 1979; Hogan & Roszypal, 1980; Wardrip-Fruin, 1982; Lisker, Note 3, Note 4). The fact that vowel and obstruent timing play a secondary role as cues for voicing does not, however, reduce the intrinsic theoretical interest of a possible "acous-

tic cue" for a phonological feature that can be represented as the ratio between two prominent acoustic intervals.

The second area of research in which the inverse relation of vowel and consonant durations for the voicing contrast has arisen is in studies of the timing of speech production. In a study of timing in English and English-like words, Port (1981a) found that the temporal effect of a change in the voicing of a syllable-final stop could be expressed as a ratio between the vowel duration and the duration of the following stop closure. Such an effect can be stated as an invariant across changes in the number of syllables in the word (e.g., from *dib-dip* to *dibber-dipper*) and in the identity of the vowel itself (e.g., from *dib-dip* to *deeb-deep*). If an invariant expressible as a temporal ratio can be found in speech production, then the role of such an invariant should be explored more carefully in speech perception.

The third area of research that bears on the issue of consonant-vowel (C/V) ratio as a voicing cue is that of perceptual studies of the effects of speaking tempo. Numerous studies that employ the tempo of a carrier sentence as a variable have demonstrated that the interpretation of temporal cues to segmental contrasts is influenced by the tempo of the context (Miller & Liberman, 1979; Pickett & Decker, 1960; Port, 1979; Summerfield, 1975; Fitch, Note 5), and generally indicate that perception compensates for changes in the temporal patterns of speech produced at different tempos. Miller (1981) surveyed many such studies and concluded that there could not be a rate-independent set of cues for certain phonetic distinctions, such as voicing, and that listeners could not interpret these tempo-dependent cues without "processing them" in relation to the rate of speech at which they were produced. Thus, she concluded that an evaluation of speaking tempo must be made in order for temporal cues (like consonant closure duration for voicing) to be correctly interpreted. On the other hand, there is a paradoxical problem here, in that just those intervals in speech that serve as temporal cues for segmental features also serve as the cues for the perception of tempo itself (Huggins, 1978; Klatt, 1976; Miller & Grosjean, 1979). This would seem to imply a perceptual system containing a loop such that vowel durations, let us say, cue both tempo and segmental features and the tempo judgment itself then plays a role in feature judgments.

The concept of C/V ratio as a cue for the voicing feature becomes important here, since it relies far less on such a loop. If the main temporal cue for voicing were just the ratio of the vowel duration to the consonant duration, then there would be no need for processing in which a perceptual judgment about the rate of speech mediated the evaluation of a temporal cue. The cue itself would be very largely independent of speaking rate and would not require special processing to take tempo judgments into account (Port, 1981b; Summerfield, Note 6).

The following set of experiments employed both synthetic and natural speech to investigate the role of vowel/consonant ratio as a voicing cue in English for intervocalic stops in pairs like *rabid-rapid* and *buggy-Bucky*. In each experiment, the overall duration of the stressed vowel plus its following stop closure was varied and the duration of either the stressed vowel or the consonant constriction was altered in small steps. In Experiment 1, a set of synthetic words, *dipper* and *dibber*, was constructed from five different durations of *dib* plus a continuum of *b-p* closure durations. The words with a longer *dib* syllable sounded simply like slower tempo productions of *dibber* or *dipper*. Throughout this paper, the interval we measured from spectrograms as "vowel duration" (from initial consonant release to closure for the following consonants) is also described as a CVC (e.g., *dib*), since that is the perceptual value of such an interval when heard in isolation.

EXPERIMENT 1: LABIAL STOPS

Method

Subjects. Sixteen native English-speaking American students at Indiana University served as subjects in the experiment. They were paid for their participation.

Stimuli. The synthetic word *dibber*, as displayed in the upper panel of Figure 1, was constructed on the Klatt waveform synthesizer (Klatt, 1977) in the Speech Perception Laboratory at Indiana University (Kewley-Port, Note 7). Parameter values were based on those suggested by Klatt (Note 8) for consonant-vowel syllables and modified by examination of spectrograms of the word produced by a trained speaker at a moderate tempo. The duration of the first vowel from *d*-burst to apparent closure of the /b/ in the basic synthetic syllable was 140 msec (as measured from a wide-band spectrogram). The *dib* syllable was constructed so that the F1 and the F0 both peaked 55 msec after the burst. By varying the duration of this steady-state point, the syllable was modified in duration in five 30-msec steps from 140 to 260 msec. The *ber* syllable had voicing beginning at the burst and an overall duration of 150 msec. Nine different silent medial-stop closure durations were constructed in 15-msec steps from 20 to 140 msec. The intonation contour rose from 105 Hz to a peak of 120 Hz at the stretched steady-state point, then fell to 113 Hz at the medial stop closure (which was silent), and then fell from 109 to 86 Hz at the end of *ber*. Thus, five "vowel durations" were combined with nine "consonant closure durations" to yield 45 stimuli. Two randomized stimulus tapes containing five tokens each of the test words separated by a 3-sec intertrial interval and a 10-sec pause after every block of 20 stimuli were prepared.

Procedure. The subjects listened to a practice block of 16 trials and then heard the two experimental tapes, which yielded a total of 10 presentations of each test item. The subjects were read a set of instructions asking them to decide for each test word whether they heard the word *dibber* or *dipper* and to check the appropriate box on a response sheet. The stimuli were presented over Koss Pro-4AA headphones driven by a Crown D-60 amplifier and a custom-built panel at a comfortable listening level in the phonetics laboratory. The subjects were given a 5-min rest between the two test tapes. The order of presentation of the two randomizations was balanced across the subjects.

Experiment 1
dibber-dipper

Short dib (140 ms)
50 ms closure

Long dib (260 ms)
50 ms closure

Experiment 2
digger-dicker

Short dig (140 ms)
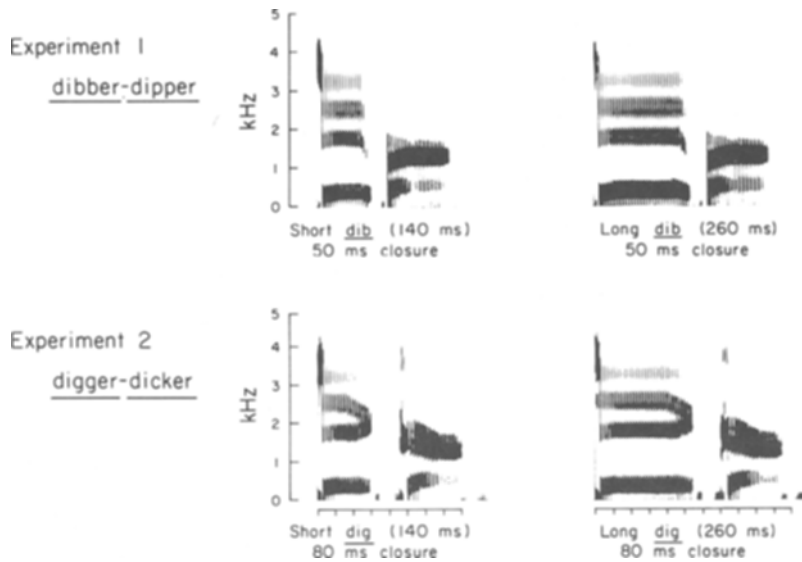80 ms closure

Long dig (260 ms)
80 ms closure

Figure 1. Sound spectrograms of some of the synthetic stimuli employed in Experiments 1 and 2. Those employed in Experiment 4 were modified from those used in Experiment 2.

Results and Discussion

The basic identification result of the experiment are presented in the left-hand panel of Figure 2, which displays the percent identification of the stimuli as *dibber* as a function of the silent medial consonant closure gap for each of the five *dib* syllable durations. It can be seen that the apparent perceptual boundaries (as 50% crossovers) tend to lie farther toward the right as the *dib* duration becomes longer. Ogives were fitted to the identification functions of each subject and Wilcoxon matched-pairs signed-ranks tests (two-tailed) were applied to the 50% crossovers. Results showed that the boundaries shifted to the right (p < .01) for 8 of 10 comparisons between conditions. Two of the adjacent pairs (170 vs. 200 msec and 230 vs. 260 msec), however, were not significantly different.

In order to evaluate the role of C/V ratio for these data, the results were replotted after dividing the medial consonant closure duration by the burst-to-closure duration of the *dib* syllable. Thus, the results plotted as a function of C/V ratio appear in the right-

Duration of dib
• 140 ms
x 170 ms
△ 200 ms
○ 230 ms
□ 260 ms

Percent Identification as dibber

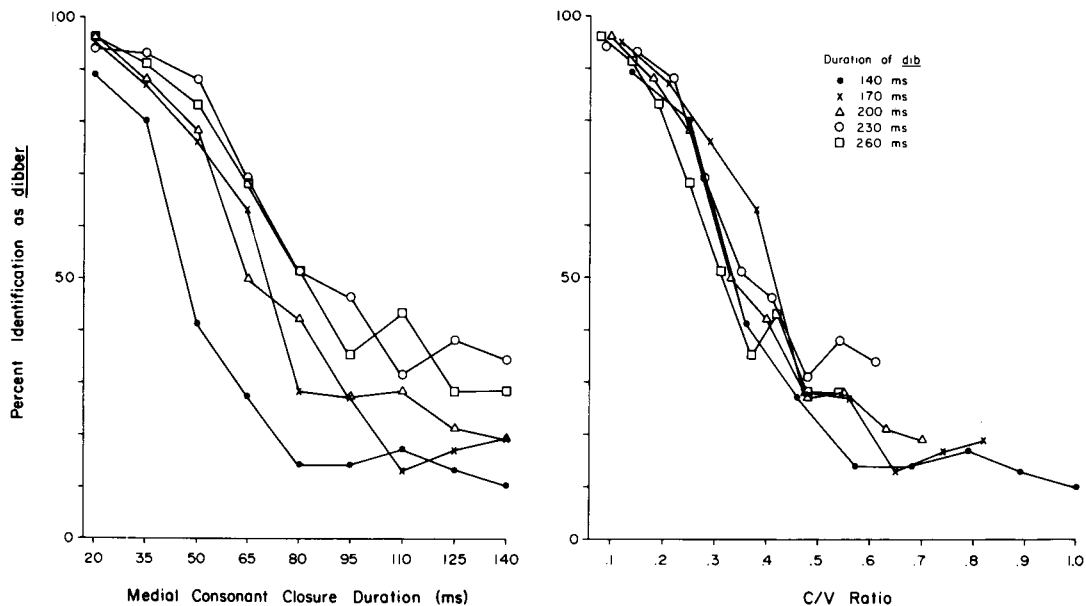Medial Consonant Closure Duration (ms)

C/V Ratio

Figure 2. The identification results of Experiment 1 plotted as a function of medial stop closure duration on the left and as a function of stop closure-to-preceding vocalic interval (C/V ratio) in the right panel.

hand panel of Figure 2. In this display there is far less variation, since all boundary values cluster around a C/V ratio of .35.

A probit analysis was carried out on the non-asymptotic portions of the identification functions in order to more accurately estimate the boundary values using both these descriptions of the results. Probit analysis (Finney, 1971) was used to fit an ogive to each pooled identification function, both with respect to consonant closure duration and C/V ratio. The 50% crossover points for both analyses are presented in Table 1. It can be seen that the cross-over value of consonant closure duration increases for each longer duration of the preceding vowel; however, the crossover values of C/V ratio display no trend at all when compared with the durations of the preceding vowel. This implies that the differences between C/V conditions are simply noise and that C/V ratio is the primary factor employed by the subjects to decide on the voicing value of the medial consonant in this experimental task.

Another way to evaluate the importance of the absolute values of vowel and consonant duration versus the dimensionless ratio of the two intervals is to compare the amount of variance accounted for by the two models using regression techniques. Table 2 shows the results of regression analysis of the number of *dibber* responses in the experiment. When consonant duration and vowel duration are entered separately using stepwise entry of the variables, the multiple R for the combination of the two variables is .793, implying that about 63% of the variance is accounted for. On the other hand, the simple correlation between the *dibber* score and the C/V ratio is .781, which implies that 61% of the variance is accounted for. This means that adding information about the absolute durations of the consonant and vowel permits us to account for only 2% more of the variance in the subjects' responses.

This result clearly replicates for labial stops the observation made by Denes about C/V ratio for /s/ and

Table 2
Regression Analysis of Results of Experiment 1

| | Multiple Regression | | Simple Regression | | |
	Multiple R	R^2		r	r^2
Consonant Duration	.751	.564	C/V Ratio	.781	.610
Vowel Duration	.793	.629			

Note—$p < .001$ in all cases.

/z/. When other features of the context are held constant and yet ambiguous, such as by maintaing a voiceless closure with no significant aspiration for stops, then the relative duration of the vowel to the following obstruent closure is an extremely powerful cue for phonological voicing in English. The absolute durations of the two intervals play only a very small role in determining subjects' judgments.

EXPERIMENT 2: VELAR STOPS

It was important to determine whether the effect of C/V ratio extended to places of articulation other than bilabial. Since the apical stops /d/ and /t/ were neutralized to flaps in the intervocalic poststress position, the C/V ratio concept did not apply. The next experiment, then, investigated the same phenomenon for a matching pair of velar stops.

Method
Subjects. The subjects were 10 Indiana University students who were native speakers of American English and who did not participate in any other experiments in this series. They were paid for their participation.

Stimuli. A synthetic *digger*, displayed in the lower panel of Figure 1, was derived from the *dibber* employed in Experiment 1. The formant transitions and release burst were modified to specify velar place, and, as typically occurs in natural examples, a 20-msec lag in voice onset was inserted at release of the /g/. Again, five durations of the stressed syllable were made, from 140 to 260 msec, and combined with nine different closure durations over the range 35 to 155 msec. Stimulus tapes were constructed as in Experiment 1.

Procedure. The subjects were instructed to identify the test words as *digger* or *dicker* by checking a box on a response sheet. The experiment procedure was the same as Experiment 1.

Results and Discussion
The identification results are presented in the left-hand panel of Figure 3 showing percent identification as *digger* plotted as a function of medial stop closure duration for the five values of *dig* syllable duration. Once again there is an obvious tendency of the 50% crossover to fall at larger values of consonant closure for longer preceding vowels. Each *dig* duration function is significantly distinct (p < .01) from the adjacent function by the Wilcoxon test on individual speakers' fitted-ogive boundaries, except for 230 vs. 260 msec, which is significant only at p < .02.

Table 1
Probit Analysis of Perceptual Boundaries for Experiment 1

| *dib* Duration | NC* | Value of Dependent Variable at 50% Crossover on Probit Scale | |
		CC†	C/V Ratio
140	20-80	50	.36
170	20-110	72	.42
200	20-125	78	.39
230	20-110	88	.39
250	20-125	91	.35

Note—*C/V ratio is defined as the consonant closure duration divided by the* dib *(V₁) duration at the 50% crossover.* *Non-asymptotic closure values (in milliseconds) included in analysis. †Medial consonant closure duration (in milliseconds).
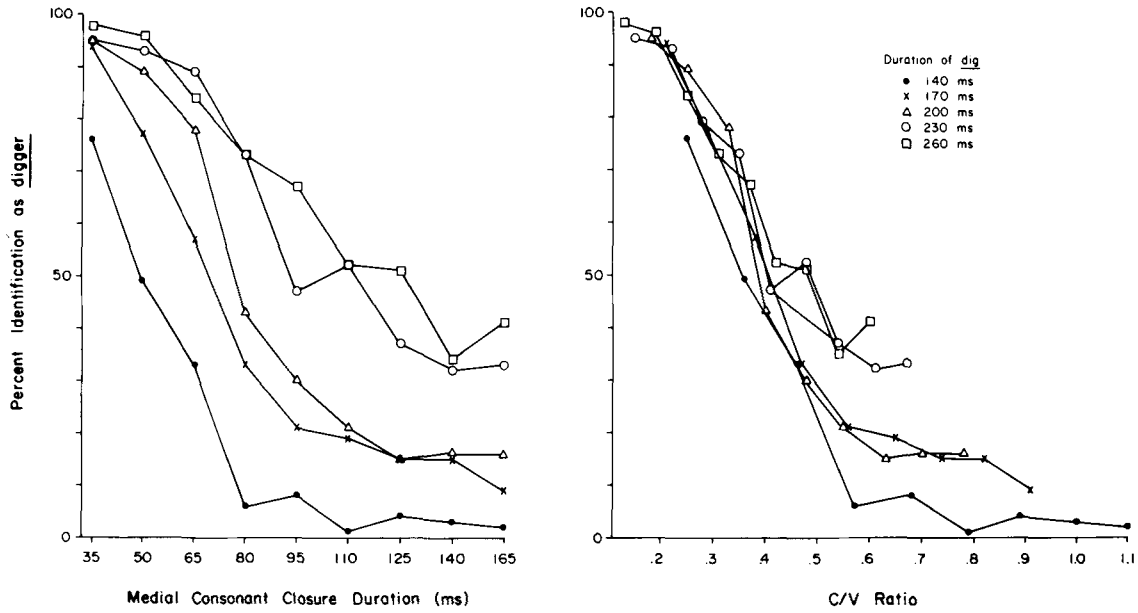
Figure 3. The identification results of Experiment 3 plotted as a function of medial stop closure duration on the left and as a function of stop closure-to-preceding vocalic interval (C/V ratio) in the right-hand panel.

To observe the role of C/V ratio in these results, the closure durations of each stimulus were divided by the duration of the vocalic interval *dig* and re-plotted in the right-hand panel of Figure 3. It can be seen that all the functions lie close together and center around a C/V ratio boundary of .4, a value somewhat larger than the values found for similar stimuli containing labial stops. Crossover boundaries were estimated for both modes of data presentation by using probit analysis on nonasymptotic portions of each function and are presented in Table 3. It can be seen that the boundaries along the C/V ratio continuum are randomly distributed with respect to the *dig* duration but that, along the consonant closure continuum, they increase linearly with respect to the *dig* duration.

Before continuing our discussion of C/V ratio, we should compare these results for velar stops with those for the labial pair in Experiment 1. The perceptual voicing boundary here lies at a C/V ratio of .42, whereas the boundary ratio from Experiment 1 was .36. Unfortunately, this direction of boundary shift does not match the production data. Port (Note 9) had subjects read the words *dibber, dipper, digger,* and *dicker* in carrier sentences. Averaging across the two voicing categories, the mean C/V ratio for each place was .59 for velars and .83 for labials. Thus, these production results should predict a shift for the perceptual boundary opposite in direction to that observed here. Since the measurement criteria for the synthetic stimuli and the production data were quite comparable, the discrepancy resulted

from some other aspect of the synthetic stimuli. For example, if the VOT on the velar stimuli were 0 (as in the labial pair) rather than +20, the boundary would surely have fallen at a smaller C/V ratio value.

In order to evaluate the importance of the absolute values of the consonant and vowel durations with respect to the C/V ratio hypothesis, two regression analyses were performed as in Experiment 1 and are displayed in Table 4. When consonant and vowel durations are entered separately in the multiple regression, a total of 70% of the variance can be accounted for. But when C/V ratio alone is employed, 66% can be accounted for. Thus, addition of information about the absolute values of the two intervals permits us to account for only about 4% more of the variance.

These data clearly show that C/V ratio works as a powerful cue for the voicing of velar stops as well as

Table 3
Probit Analysis of Perceptual Boundaries for Experiment 2

dig Duration	NC*	Value of Dependent Variable at 50% Crossover on Probit Scale	
		CC†	C/V Ratio
140	35-110	65	.48
170	35-125	90	.53
200	35-110	100	.49
230	35-140	123	.55
260	35-140	133	.50

*Nonasymptotic closure values (in milliseconds) included in analysis.   †Medial consonant closure duration (in milliseconds).

Table 4
Regression Analysis of Results of Experiment 2

Multiple Regression			Simple Regression		
	Multiple R	R²		r	r²
Consonant Duration	.689	.475	C/V Ratio	.813	.661
Vowel Duration	.835	.698			

Note—p <.001 in all cases.

labials. The results strongly suggest that the vowel durations and consonant durations are directly compared with each other in each stimulus and that the C/V ratio itself determines the degree of perceptual voicing of the stimuli.

On the other hand, there is also evidence that C/V ratio is *not* completely invariant across changes in speaking tempo (Derr & Massaro, 1980; Port, 1979). Port (1979), for example, found that changes in the tempo of a preceding carrier sentence changed the perceptual boundary along a continuum of closure durations for *rabid* vs. *rapid*. Since the preceding vowel durations were unchanged, the C/V ratios were necessarily different at the perceptual boundaries. He also noted, however, that the changes in C/V ratio boundary were very small (about 10%), whereas the tempo in the external carrier sentence changed by a factor of 2. Are these results contradictory? Possibly not. It might be that the degree of voicing of a stimulus item is determined primarily by the C/V ratio but that the boundary value of C/V ratio is determined by other temporal factors outside the test word. In this case, the C/V ratio boundary should be the *same* for a slow-tempo test word as for a fast-tempo test word (comparable to the stimuli employed in Experiments 1 and 2) when embedded in carrier sentences of any given tempo. This is to say, the effects of tempo external to the test word will be independent of the effect of change in duration of the stressed vowel itself.

EXPERIMENT 3: INTERNAL VS. EXTERNAL TEMPO

In Experiment 3, the effects of changes in the speaking tempo of a test word and of changes in the tempo of the carrier sentence were contrasted. In general, we predicted that both factors would change the perceptual boundary for voicing—if only because tempo closer to the test interval had been shown to have a larger effect on perception of that cue than tempo further away (Summerfield, Note 6). More specifically, if C/V ratio was the cue, then tempo outside the test word would have the same effect on a fast-tempo word that it did on a slow-tempo word. That is, there should have been no interaction between these variables.

Since it was important to test the generality of these effects across vowel contexts and with natural speech, the next experiment employed edited natural speech stimuli and a different pair of test words.

Method

Subjects. Twelve students from Indiana University who were native speakers of American English volunteered to serve as subjects. The listeners were students in the Department of Linguistics.

Stimuli. A trained speaker recorded a casual production of the sentence "I'm trying to say *rabid* to you" at both a fast tempo and a slow tempo. After digitization and 5-kHz filtering of these natural utterances, using the PCM system at Haskins Laboratories, stimuli were prepared under computer control as illustrated in Figure 4. The test word *rabid* was excised from each sentence by cutting at the end of the initial /r/ constriction and in the middle of the final /d/ closure. Then within each word, cuts were made immediately after closure for the medial /b/ and just before release. The glottal pulsing during the closure was deleted and a continuum was prepared by inserting silent gaps in each test word varying from 50 to 200 msec in steps of 10 msec. With the short stop closures, both test words sounded like *rabid*, and with the long closures, both sounded like *rapid*. Individual members of each of these continua were then embedded in both the fast and the slow carrier sentences.
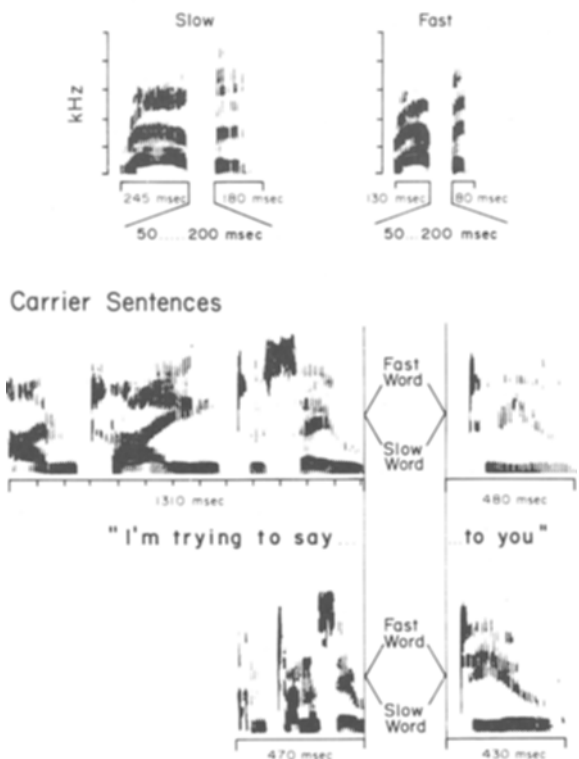

Figure 4. Sound spectrograms illustrating construction of the natural speech stimuli employed in Experiment 3. Note that the fast tempo production of the carrier sentence involved considerable phonetic reduction.

Procedure. A listening tape containing 10 tokens of each of these 64 stimuli in random order was prepared and was presented to the subjects after they had heard a 10-item practice tape. The listeners were instructed to check on a response sheet whether they heard the word *rabid* or *rapid* in each sentence.

Results and Discussion

The results of these identifications are presented in Figure 5, in which the percent identification of each stimulus as *rabid*, that is, as containing a medial voiced stop, is plotted as a function of the closure duration of the stop. All four conditions differed from each other at p < .01 by the Wilcoxon matched-pairs signed-ranks test on subject boundaries based on fitted ogives. In the fast-word/fast-sentence condition, stop closures longer than about 70 msec were heard predominantly as /p/, while in the slow-word/slow-sentence condition, the /b-p/ boundary was more than twice as long at 165 msec. Starting with the slow-word/slow-sentence condition, it was found that changing just the carrier sentence from slow to fast—that is, from the open circles to the open triangles—moved the boundary about 30 msec, or 20%, toward shorter values. If, however, just the test word was changed from slow to fast and the carrier sentence was left the same (that is, from open circles to filled circles in the figure), the boundary shifted by 65 msec. This amounted to a shortening of the voicing boundary by about 40%—far larger than the effect of a change in the carrier sentence alone. Changing just the carrier sentence moved the boundary significantly, but changing just the test word moved it considerably more. These results imply that timing in the two syllables adjacent to the stop closure was considerably more important in determining

the durational boundary for medial stop voicing than was the speaking tempo in a surrounding carrier sentence.

In order to evaluate the constancy of C/V ratio here, these identifications results were replotted to the right in Figure 5 as a function of C/V ratio. For this purpose, sound spectrograms were obtained and measurements were made of the interval from the point of steepest rising F3 slope in the transition from the /r/ into /æ/ up to the point of apparent closure for the /b/ for both fast and slow test words (found to be 110 and 200 msec, respectively), and each of these values was divided into each of the respective closure durations employed. It can be seen that, in the fast carrier sentence, the fast and slow test words (in filled and open triangles) cross the 50% boundary at the ratio of .56 (shown by probit analysis). Similarly, in the slow carrier sentence, the identification functions for the two test words (represented by filled and open circles) have perceptual boundaries very near each other (at .68 and .75, respectively). Thus, for a given tempo of carrier sentence, the voicing boundary along a continuum of *ratios* of stop closure to the preceding vowel interval tends to be relatively constant. Since inspection suggests that the two test words may not have the same C/V boundary in the slow condition, the independence of the internal and external tempo variables was evaluated with analysis of variance on the number of /b/ responses. Because consonant duration clearly has a large effect on the scores, consonant duration was treated as a covariate to provide the strongest test of the hypothesis of independence. Of course, both internal tempo [F(1,2) = 625, p < .001] and external tempo [F(1,2) = 73.5, p < .001] are highly significant, but their interaction
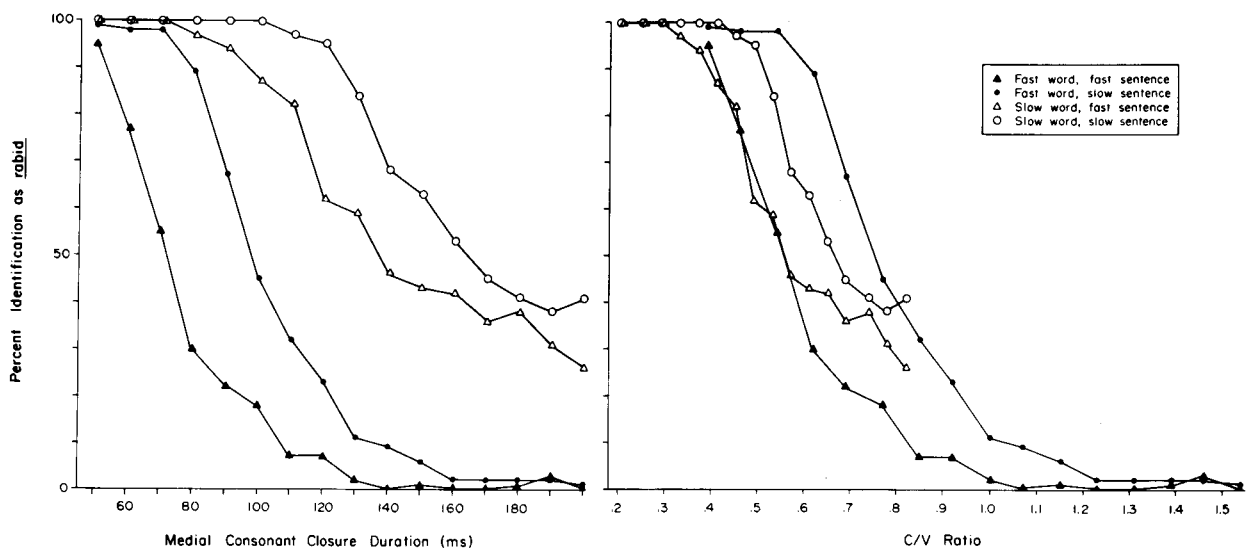

Figure 5. The identification results of Experiment 3 plotted as a function of medial stop closure duration on the left and C/V ratio on the right.

is not [F(1,2) = 3.03, p = n.s.]. This implies that the C/V ratio boundary is the same for both test word tempos.

These findings confirm that timing both within the word and external to it has perceptual effects on the closure duration cue for voicing. For the values employed in these two natural sounding utterances, the tempo external to the test word has far smaller effects than the tempo within the test word. We predicted on the basis of Experiments 1 and 2 (as well as Denes, 1955, and Port, 1979) that, if C/V ratio is invariant across changes in local or within-word tempo, then any effects of word-external tempo on the C/V ratio boundary should be the same for both tempos of test word. This prediction is supported by the apparent independence of the effects of word-external and word-internal tempo. It is reasonable to propose, then, that across the normal range of vowel and consonant durations, C/V ratio itself acts as the primary temporal cue for the postvocalic obstruent voicing, even though it is one which may be dominated by nontemporal ones. The tempo of the carrier sentence surrounding the test word may have some influence on the temporal cue, but that influence is most economically stated as a modification of the C/V ratio cue itself, rather than as a modification of a consonant duration cue for voicing.

The purpose of the experiment was to explore perception with natural production of the test sentences, but this use of natural speech stimuli permits a possible criticism of our interpretation of the experiment. Although we have proposed C/V ratio as the only relevant property of these artificially modified test words for specifying the voicing of the medial stop, we have not controlled for possible effects of the large difference in duration of the second vowel, or *bid* syllable, in these stimuli (as was seen in Figure 4). One might argue that the absence of an interaction between carrier sentence and the test word is in some way due to effects of the following vowel. In Experiment 4, this was directly addressed.

EXPERIMENT 4: EFFECT OF FOLLOWING VOWEL

It is likely a priori that the duration of the second syllable in these test words has some effect on the consonant closure boundary for voicing, but how strong is the effect? It has been shown that if a carrier sentence preceding a target word is divided into three segments, the tempo of a segment closest to the target word will have the strongest effect on a voice-onset time boundary (Summerfield, Note 6). On the basis of this proximity principle, we might expect that, since the following vowel is just as close to the consonant closure cue interval as the preceding vowel, it should have just as strong an effect in changing the

consonant closure boundary for voicing. This would be coherent with the hypothesis that tempo always mediates voicing judgments that are based essentially on consonant closure duration. The well-known effects of the preceding vowel duration (Raphael, 1972) on voicing perception might then be interpreted as demonstrating that vowel duration is fundamentally a *local tempo* cue that serves to adjust the perceptual judgment of the consonant closure duration for voicing. By this hypothesis, there should be nothing special about the preceding vowel versus the following vowel in evaluating local speaking tempo. Thus, they should have fairly similar effects on the consonant closure boundary for voicing. The only factor that might make the preceding vowel more salient is that it is stressed rather than unstressed.

If, on the other hand, the fundamental temporal cue for voicing in this syllabic position is C/V ratio, then the only way the following vowel could have any effect is as a cue for speaking tempo—the same way that word-external tempo affects C/V ratio. Thus, we would predict that the effect of the second syllable duration should be quite small and, furthermore, that, like word-external tempo, it should be independent of the effects of the duration of the first syllable.

Method

Subjects. Fourteen native American English-speaking Indiana University students were paid to serve as subjects. None of them served in other experiments in this series.

Stimuli. The stimulus items were constructed from the synthetic *digger* employed in Experiment 2. Two *ger* syllables, or V_2s, were produced, with durations of 120 and 220 msec. Two *dig* syllables, or V_1s, were employed with the shortest (140 msec) and the second longest (230 msec) durations from Experiment 2. The four combinations of short and long V_1 with the short and long V_2 were combined with nine different silent stop-closure intervals varying in 15-msec steps from 20 to 140 msec, yielding a total of 36 stimuli. The stimulus tapes and practice tapes were prepared as in earlier experiments.

Procedure. The subjects were instructed to identify the test words as either *digger* or *dicker* by checking a box on a response sheet. Other aspects of the procedure were identical to Experiments 1 and 2.

Results and Discussion

The basic results are presented in the left-hand panel of Figure 6. It can be seen that both variables, the duration of the preceding *dig* syllable (or V_1) and the duration of the following *ger* syllable (V_2), have some effect on the apparent perceptual boundary between /g/ and /k/ (Wilcoxon tests performed as before indicate that all conditions are distinct, p < .01, except for the two conditions with short *dig*, which are distinct at p < .02.) The effect of the preceding syllable on the consonant closure boundary [F(1,2) = 125.0, p < .001, with the consonant duration as a covariate] is obviously considerably larger than the effect of the following *ger* syllable
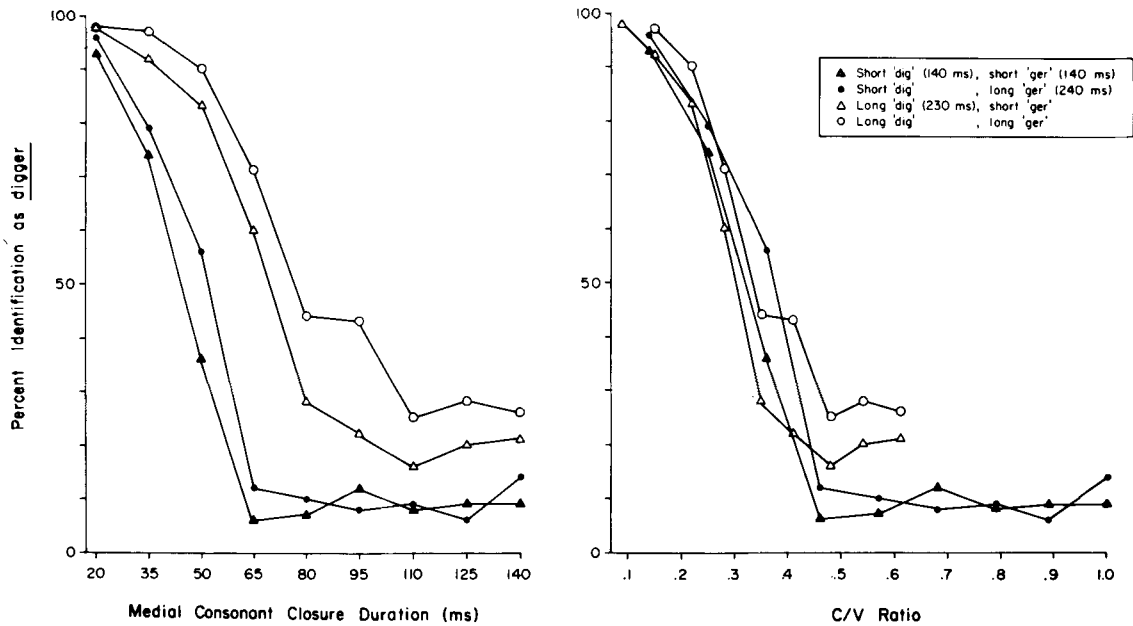
Figure 6. The identification results of Experiment 4 plotted on the left as a function of medial stop closure duration and on the right as a function of C/V ratio.

duration $[F(1,2) = 10.3$, p < .001, with consonant duration as a covariate]. When the results are replotted in terms of C/V_1 ratio, however, as in the right-hand panel of Figure 6, it appears that the two conditions having the same V_1 but different V_2s lie nearly on top of each other, and analysis of variance reveals no interaction between V_1 and V_2 $[F(1,2) = 1.31$, p = n.s., with consonant duration as a covariate]. Thus, the following syllable duration appears to have an effect only on the C/V_1 ratio itself.

This can be more easily seen in terms of the estimated perceptual boundary obtained from probit analysis as in Table 5. The four distinct boundaries along the consonant closure continuum appear toward the left of the table, while on the right it can be seen that the two conditions with the long V_2 and with the short V_1 have essentially identical boundaries when evaluated in terms of C/V ratio. It is ap-

Table 5
Probit Analysis of Perceptual Boundaries for Experiment 4

V_1 Duration	Boundary in Terms of Consonant Duration		Boundary in Terms of C/V Ratio	
	Long V_2	Short V_2	Long V_2	Short V_2
Long	87	73	.38	.32
Short	52	45	.37	.32

Note—For stimuli with a long V_1, the closure durations from 20 to 110 msec of closure duration were judged nonasymptotic and included in the analysis. For short-V_1/long-V_2 stimuli, values from 20 to 95 msec of closure duration were included. For short-V_1/short-V_2 stimuli, values from 20 to 80 msec of closure duration were included. All durations are given in milliseconds.

parent, then, that a change in the duration of the following syllable has a *different kind of effect* on the voicing of a medial stop from the effect of changes in the preceding vocalic interval. A change in the following vowel changes the C/V ratio itself by a small amount. This effect of the following syllable duration is clearly independent of the preceding vowel effect.

A multiple regression analysis of the number of /g/ responses was conducted using two models. In the first model, consonant duration, first vowel duration, and second vowel duration were entered, using a stepwise procedure. The results are shown on Table 6. It can be seen that consonant duration alone accounts for about 53% of the variance and the addition of V_1 duration accounts for 9% more. The duration of V_2, however, adds only an additional 1%. In the second model, C/V_1 ratio was computed for each stimulus, and this variable was extracted along with the duration of the second syllable. As can be seen in Table 6, C/V_1 ratio again accounts for almost as much variance as vowel duration and consonant duration combined (58% vs. 62%) and yet contains no information about the absolute durations of the vowel and consonant.

The results clearly demonstrate that the preceding vowel has a far stronger effect on the consonant closure boundary than does the following vowel. This implies that the preceding vowel plays a special and privileged role in relation to voicing judgments. The independence of the effect of the second vowel, while not necessarily irreconcilable with a model of tempo mediation, still provides strong support for the view

Table 6
Multiple Regression Analysis of Experiment 4

	R	R²
Model 1		
Consonant Duration	.725	.526
V₁ Duration	.787	.619
V₂ Duration	.792	.627
Model 2		
C/V₁ Ratio	.760	.578
V₂ Duration	.765	.585

Note—p <.001 in all cases. In Model 1, the consonant closure duration, preceding vocalic interval and following vocalic interval are separately entered. In Model 2, C/V ratio is treated as an elementary variable along with the duration of the following vocalic interval.

that the second syllable serves primarily to slightly modify the C/V ratio cue for the voicing of the stop. Furthermore, the multiple regression analyses indicate for these data, as well, that the use of ratio of stressed vowel duration to the duration of the following consonant accounts for nearly as much of the variance as the use of the absolute values of those dimensions. In this situation, the conclusion seems unavoidable that, for a given contextual situation, there is a certain ratio between the vowel and consonant that serves to differentiate voiced values from voiceless values.

GENERAL DISCUSSION

The results of these experiments bear on several general issues in phonetics and speech perception. The first issue has to do with the nature of the voicing contrast in English. A bewildering variety of acoustic cues has been shown to affect voicing judgments (see Abramson, 1977), and various proposals have been made for unifying these in terms of a single underlying contrast, such as the action of opening vs. not opening the glottis (Lisker & Abramson, 1971) and so forth (Slis & Cohen, 1969; Stevens & Blumstein, 1981). The concept of C/V ratio as an independent voicing cue resembles these proposals in that it serves as a way to *reduce* the number of independent parameters from the speech wave that must be measured and combined by the perceptual system. The view that vowel duration and consonant duration are separate and independent cues requires that both intervals be measured in terms of some absolute temporal units that would be formally comparable to milliseconds. But since C/V ratio is relatively stable as a cue for voicing despite radical changes in overall duration, it is tempting to suppose that the nervous system has some way of measuring and using C/V ratio directly without requiring "prior" measurement of each interval in absolute terms (such as in milliseconds). In this case, the nervous system would be extracting the most phonet-

ically relevant information. Of course, C/V ratio is not perfectly invariant, so it must be subject to biasing due to surrounding tempo and to the place of articulation and manner of the consonant. In addition, this cue still must be integrated with others, such as closure voicing and aspiration, which, in this context, are weighted far more strongly than C/V ratio itself as cues for the voicing feature.

To illustrate the importance of the nontemporal cues, we may compare these perception results with earlier production data on timing in the very test words used in Experiments 1, 2, and 4 (Port, 1981a, Note 9). First, the productions of *dibber, dipper, digger,* and *dicker* tended to have shorter vowels, because these productions were medial in the carrier sentence (in Port, Note 9) rather than utterance-final like our perception test words. Thus, the durations of the first vowels were around 70-80 msec, in contrast with 140 msec for the shorter first vowels in our test words. Second, the C/V ratios observed in the production data tended to be larger; that is, the consonants were longer relative to the vowels for a given voicing category. For example, the C/V ratios for *digger* and *dicker* in sentence-medial position were .48 and .72, respectively (Port, Note 9, Tables 3 and 4), thus predicting a perceptual boundary ratio lying between them at about .6. Our observed boundary for the synthetic stimuli of Experiment 2 was about .42. This difference in the C/V ratio boundary probably reflects the absence of voicing during the closure of /g/. The *dibber-dipper* pair offers a more extreme example, since the produced ratios (Port, Note 9, Tables 3 and 4) were .77 and 1.2, respectively—far higher than the perceptual boundary for our synthetic stimuli of .36. Apparently, the absence of closure voicing during the stop made them far more "voiceless" in quality. The important point, however, is that this increased "voiceless quality" of our synthetic stimuli is still consistently manifested as a shift toward smaller values of the C/V ratio boundary for the feature rather than, for example, being tied just to the consonant closure itself. Of course, the production data themselves also demonstrate strong independent evidence that C/V ratio is directly manipulated by speakers in accordance with values of the voicing contrast (see Port, 1981a).

It should also be noted that English is not unique in employing C/V ratio as a constraint on speech production and as a cue in speech perception. A very similar phenomenon has been demonstrated for several other Germanic languages in production (Bannert, 1976; Elert, 1964; Garnes, 1973; Kohler, Note 10; Kohler & Künzel, Note 11) and in perception (Bannert, 1976; Kohler, 1979). Kohler's (1979, Note 10) results on the role of the ratio of V to C as a voicing cue in German are very similar to the results for English. In most of the other cases, the contrast

between VC: and V:C is at least partially independent of the voicing contrast but is restricted to stressed syllables.[2] The extent to which such simple temporal relations play a role in languages outside the Germanic group is not known.

The proposal has been made that listeners must make a judgment about speaking tempo *before* evaluating, say, the closure duration cue for voicing (Miller, 1981; Miller & Grosjean, 1979; Miller & Liberman, 1979). The results of these experiments by no means rule out such a model, but they do seem to make it less plausible and less necessary. To sustain such a model over the C/V ratio model, it would be necessary to account for why the *preceding* vowel plays such a unique role as a cue for local tempo. The preceding vowel alone accounts for nearly all of the context-sensitive change in the consonant-closure duration cue for voicing. Of course, C/V ratio does in fact change somewhat with changes in tempo, so there must exist some mechanism for achieving these changes due to factors like speaking tempo, the presence of closure voicing, and aspiration, as well as the place and manner of the consonant. Nevertheless, the most basic point demonstrated by our results is that, however this adjustment takes place, it is an adjustment of C/V ratio, not of vowel or consonant duration cues independently.

The model that is proposed here for the temporal correlates of the voicing contrast in syllable-final position also has important implications for the general theory of phonetics (Port, 1981b). Normally, the phonetic space is assumed to contain *segments* primarily, each of which has some "inherent duration" (Klatt, 1976; Lehiste, 1970). Other timing effects are to be accounted for with rules of "temporal implementation" (Klatt, 1976). But when the production and perception of speech are viewed as processes that deal with segment-sized units, an important phonetic structure is missed; a "linguistically significant generalization" is overlooked (Chomsky & Halle, 1968). C/V ratio in English behaves like an elemental phonetic unit that is identified with the voicing or tensity contrast in syllable-final and ambisyllabic positions.

This interpretation of the C/V ratio effects found in production and perception is coherent with the general view that linguistic features have a variety of abstract and relatively invariant acoustic correlates that speakers encode into sound and that hearers listen for when perceiving speech (Jakobson, Fant, & Halle, 1952; Stevens & Blumstein, 1981). These acoustic correlates, like the linguistic features themselves, are relatively independent of each other. But, unlike features, they are not arranged ordinally in time and may instead overlap or be defined in strictly temporal terms (Fowler, 1980; Fowler, Rubin, Remez, & Turvey, 1980). From this point of view there is no basis for the implicit assumption that

"vowel durations" and "consonant durations" are somehow more concrete or more fundamental as perceptual units than the relative duration of the two of them. C/V ratio, as associated with contrasts in voicing, provides a constraint on speech production that is invariant across a variety of contextual changes and also specifies an acoustic pattern that can be directly extracted from the speech signal during speech perception without prior determination of speaking rate.

REFERENCE NOTES

1. Port, R. F. *Effects of word-internal vs. word-external tempo on the voicing boundary for medial stop closure* (Status Report on Speech Research SR-55/56; DTIC No. AD-A065575). New Haven, Conn: Haskins Laboratories, 1978.

2. Port, R. F. *Temporal aspects of the English voicing contrast: Implications for a theory of phonetics.* Invited address to the American Association of Phonetic Sciences, Los Angeles, November 1980.

3. Lisker, L. *On buzzing the English /b/* (Status Report on Speech Research SR-55/56; DTIC No. AD-A065575). New Haven, Conn: Haskins Laboratories, 1978.

4. Lisker, L. *Closure hiatus: Cue to voicing, manner and place of consonant occlusion* (Status Report on Speech Research SR-53; DTIC No. AD-A055853). New Haven, Conn: Haskins Laboratories, 1978.

5. Fitch, H. *Distinguishing temporal information for speaking rate from temporal information for intervocalic stop consonant voicing* (Status Report on Speech Research SR-65). New Haven, Conn: Haskins Laboratories, 1981.

6. Summerfield, A. Q. *Timing in phonetic perception: Extrinsic or intrinsic?* (Arbeitsberichte 12). Kiel: Institute of Phonetics, University of Kiel, 1979.

7. Kewley-Port, D. KLTXEC: *Executive program to implement the KLATT software speech synthesizer* (Research on Speech Perception Report 5, pp. 327-346). Bloomington, Ind: Department of Psychology, Indiana University, 1978.

8. Klatt, D. H. *Analysis and synthesis of consonant-vowel syllables in English.* Book in preparation, 1982.

9. Port, R. F. *Influence of speaking tempo on duration of stressed vowel and medial stop in English trochee words.* Bloomington, Ind: Indiana University Linguistics Club, 1976.

10. Kohler, K. J. *The production of plosives.* (Arbeitsberichte 8). Kiel: Institute of Phonetics, University of Kiel, 1977.

11. Kohler, K. J., & Künzel, H. J. *The temporal organization of closing-opening movements for sequences of vowels and plosives in German.* (Arbeitsberichte 10). Kiel: Institute of Phonetics, University of Kiel, 1978.

REFERENCES

ABRAMSON, A. S. Laryngeal timing in consonant distinctions. *Phonetica*, 1977, **34**, 295-303.

BANNERT, R. *Mittelbairische Phonologie auf akustischer und perzeptorischer Grundlage.* Lund, Sweden: Gleerup-Fink, 1976.

BARRY, W. Complex encoding in word-final voiced and voiceless stops. *Phonetica*, 1979, **36**, 361-372.

CHOMSKY, N., & HALLE, M. *Sound pattern of English.* New York: Harper & Row, 1968.

COLE, R., & COOPER, W. E. Perception of voicing in English affricates and fricatives. *Journal of the Acoustical Society of America*, 1975, **58**, 1280-1287.

DENES, P. Effect of duration on the perception of voicing. *Journal of the Acoustical Society of America*, 1955, **27**, 761-764.

DERR, M., & MASSARO, D. The contribution of vowel duration, F₀ contour and frication duration as cues to the /juz-jus/ distinction. *Perception & Psychophysics*, 1980, 27, 51-59.

ELERT, C.-C. *Phonological studies of quantity in Swedish*. Stockholm: Almqvist and Wiksell, 1964.

FINNEY, D. J. *Probit analysis* (3rd ed.). Cambridge: Cambridge University Press, 1971.

FOWLER, C. Coarticulation and theories of extrinsic timing. *Journal of Phonetics*, 1980, 8, 113-133.

FOWLER, C., RUBIN, P., REMEZ, R., & TURVEY, M. Implications for speech production of a general theory of action. In B. Butterworth (Ed.), *Language production*. New York: Academic Press, 1980.

GARNES, S. Phonetic evidence supporting a phonological analysis. *Journal of Phonetics*, 1973, 1, 273-284.

HOGAN, J. T., & ROSZYPAL, A. J. Evaluation of vowel duration as a cue for the voicing distinction in the following word-final consonant. *Journal of the Acoustical Society of America*, 1980, 67, 1764-1771.

HUGGINS, A. W. F. Speech timing and intelligibility. In J. Requin (Ed.), *Attention and performance VII*. Hillsdale, N.J: Erlbaum, 1978.

JAKOBSON, R., FANT, G., & HALLE, M. *Preliminaries to speech analysis*. Cambridge, Mass: M.I.T. Press, 1952.

KLATT, D. H. The linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America*, 1976, 59, 1208-1221.

KLATT, D. H. Cascade/parallel terminal analog speech synthesizer and a strategy for consonant-vowel synthesis. *Journal of the Acoustical Society of America*, 1977, 61, S68(A).

KOHLER, K. Dimensions in the perception of fortis and lenis plosives. *Phonetica*, 1979, 36, 332-343.

LEHISTE, I. *Suprasegmentals*. Cambridge, Mass: M.I.T. Press, 1970.

LISKER, L. Closure duration and the intervocalic voiced-voiceless distinction in English. *Language*, 1957, 33, 42-49.

LISKER, L., & ABRAMSON, A. Distinctive features and laryngeal control. *Language*, 1971, 47, 767-785.

MILLER, J. L. The effect of speaking rate on segmental distinctions: Acoustic variation and perceptual compensation. In P. D. Eimas & J. L. Miller (Eds.), *Perspectives in the study of speech*. Hillsdale, N.J: Erlbaum, 1981.

MILLER, J. L., & GROSJEAN, F. Further studies of the influence of sentence rate on the closure duration cue for voicing. In J. Wolf & D. Klatt (Eds.), *Speech communication papers: 97th Meeting of the Acoustical Society of America*. New York: Acoustical Society of America, 1979.

MILLER, J. L., & LIBERMAN, A. M. Some effects of later occurring information on the perception of stop consonant and semivowel. *Perception & Psychophysics*, 1979, 25, 457-465.

PETERSON, G., & LEHISTE, I. Duration of syllabic nuclei in English. *Journal of the Acoustical Society of America*, 1960, 32, 693-703.

PICKETT, J. M., & DECKER, L. Time factors in perception of a double consonant. *Language and Speech*, 1960, 3, 11-17.

PORT, R. The influence of tempo on stop closure duration as a cue for voicing and place. *Journal of Phonetics*, 1979, 7, 45-56.

PORT, R. F. Linguistic timing factors in combination. *Journal of the Acoustical Society of America*, 1981, 69, 262-274. (a)

PORT, R. F. On the structure of the phonetic space with special reference to speech timing. *Lingua*, 1981, 55, 181-219. (b)

PRICE, P., & LISKER, L. (/b/ → /p/) but not (/p/ → /b/). In J. Wolf & D. Klatt (Eds.), *Speech communication papers: 97th Meeting of the Acoustical Society of America*. New York: Acoustical Society of America, 1979.

RAPHAEL, L. Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English. *Journal of the Acoustical Society of America*, 1972, 51, 1296-1303.

SLIS, I. H., & COHEN, A. On the complex regulating the voiced-voiceless distinction I. *Language and Speech*, 1969, 1, 80-102.

STEVENS, K. N., & BLUMSTEIN, S. The search for invariant acoustic correlates of phonetic features. In P. Eimas & J. Miller (Eds.) *Perspectives in the study of speech*. Hillsdale, N.J: Erlbaum, 1981.

SUMMERFIELD, A. Q. How a full account of segmental perception depends on prosody and vice versa. In A. Cohen & S. G. Nooteboom (Eds.), *Structure and process in speech perception*. New York: Springer-Verlag, 1975.

UMEDA, N. Vowel duration in American English. *Journal of the Acoustical Society of America*, 1975, 58, 434-445.

WARDRIP-FRUIN, C. On the status of phonetic cues to phonetic categories: Preceding vowel duration as a cue to voicing in final stop consonants. *Journal of the Acoustical Society of America*, 1982, 71, 187-195.

NOTES

1. In a replication and extension of the Denes experiment using the same lexical pair—/jus-juz/—Derr and Massaro (1979) found similar results. They prefer, however, to emphasize the extent to which there is a change in C/V ratio, in both Denes's results and their own. For example, Derr and Massaro orthogonally varied durations of the vocalic and fricative portions to create test words whose total duration varied over a range of more than 2:1, yet the C/V ratio boundary between /jus/ and /juz/ varied only from about .4 with the shortest fricative duration to about .6 with the longest fricative duration (computed for their stimuli with level F0 contours when the total vocalic interval is compared with the total fricative duration). This is not a large ratio change for such a great modification of total duration.

2. The notation V:C means a long vowel followed by a short consonant, and VC: means a short vowel following by a long consonant. In Swedish, for example, stressed short vowels occur only before long consonants (as in *backa* [bak:ə] "to back up"), while long vowels occur only before short consonants (as in *baka* [bɔ:kə] "to bake").