Reprinted from: The Journal of the Acoustical Society of America

9.5, 9.7, 9.4

Received 4 August 1971

Preceding Vowel Duration as a Cue to the Perception of the Voicing Characteristic of Word-Final Consonants in American English

LAWRENCE J. RAPHAEL*

Herbert H. Lehman College of the City University of New York, Bronx, New York 10468

A number of studies in the literature have stated that the duration of a vowel is a significant cue to the voicing characteristic of the consonant that follows it. The present study investigated the effect of varying preceding vowel duration upon the perception of word-final stops, fricatives, and clusters in synthetic speech. A variety of minimal CVC(C) pairs was synthesized and the vowel of each was varied over a range of values derived from durations found in real-speech samples. It was found that, regardless of the cues for voicing or voicelessness used in the synthesis of the final consonant or cluster, listeners perceived the final segments as voiceless when they were preceded by vowels of short duration and as voiced when they were preceded by vowels of long duration. Discrimination tests revealed that when the voicing characteristic is cued by vowel duration, perception is continuous rather than categorical.

INTRODUCTION

The question of the extent to which vowel duration conditions the perception of the voicing characteristic of syllable-final consonants has been investigated by several writers and experimenters. Daniel Jones, for instance, has noted that in the speech of some speakers of English "words like heed and heat . . . are distinguished solely by the length of the vowel."[1] Experimental work done at Haskins Laboratories on synthetic velar stops in final position, stops in which voicing cues had been neutralized, showed that preceding vowels longer than 200 msec caused listeners to hear /ɛg/. Vowels shorter than 200 msec produced listener judgments of /ɛk/.[2]

Further experimental evidence supplied by Denes indicated that for word-final /s/ and /z/ "the duration of the vowels and of the final consonants have a definite and consistent influence on the perception of 'voicing'." Using real-speech samples, Denes first physically interchanged the final sounds in the tape recordings of the words *use*/jus/ and (to) *use*/juz/, shortening the duration of the /s/ and lengthening that of the /z/. Each transposed sound was then heard by listeners as its cognate, as judged by the identification of the word in which it was contained. A second experiment varied the durations of a preceding synthetic vowel (after synthetic /j/) and a word-final, real speech /s/. Listener judgments indicated that the durations varied were interdependent: ". . . the perception of 'voicing' of the final consonant increases as the ratio of the duration of final consonant to preceding vowel decreases." The strength of the preceding vowel duration as a cue to voicing is indicated by the fact that the final /s/, regardless of its duration, "if produced in isolation, always sounds voiceless, although the same sound when arranged as part of a word may be heard as a voiced sound."[3]

Work done by this writer on human speech revealed that shortening the duration of vowels preceding final voiced stops and fricatives caused them to be perceived as voiceless.[4]

Finally, Malécot has pointed out that the "duration of vowels before final consonants is both a powerful and sufficient acoustic cue . . ." for differentiating between the classes of those cognate pairs which have thus far been tested.[5]

The experiments reported below were designed to extend the work of Denes and others by testing the effectiveness of vowel duration as a cue to the voicing characteristic of a variety of word-final stops, fricatives, and clusters. It was further hoped that discrimination tests of the stimuli used would shed some more light on the perception of speech *vis-à-vis* the motor theory of speech perception.

The question of the mode of listeners' perceptions of voicing that is cued by vowel durational differences is of particular interest. Previous research has indicated that synthetic stop consonants are perceived categorically with regard to changes in place and voicing[6,7] and that changes in vowel perception, expecially in isolation, are somewhat more continuous with regard to quality and duration.[8–10] In the experiment described below, however, the cue to the voicing characteristic of a final consonant lies not within the articulatory period of the consonant itself, but within the duration of the preceding vowel. If, in fact, the changes in duration are continuously perceived, then the changes in consonant perception that they cue should be perceived in a similar manner, as predicted by the motor theory.[11]

I. PERCEPTUAL TESTS

A. Procedures

The experiment reported here was designed to determine the perceptual effect of varying the steady-state duration of vowels before a variety of synthetic word-final cognate stops and fricatives in minimal pairs, as well as before word-final clusters of the types stop plus stop, stop plus fricative, and fricative plus stop. Table I specifies the oppositions tested and the corpus of words used.

The Haskins Laboratories' Pattern Playback was employed to generate the stimuli. The same general scheme was followed for each of the series of stimuli generated. Figure 1 shows the types of patterns used to produce words ending in simple stops. These were in the form of a three-formant steady-state vowel with 50-msec transitions appropriate to the place and voicing characteristics of the initial and final consonants. A series of stimuli of the type shown at the top of Fig. 1(a), with steady-state vowel durations varying from 150 to 350 msec, was prepared and recorded. This was called the "voiced" series. The range of vowel durations used for any specific series of stimuli was determined from investigation of real-speech samples of words spoken in isolation. After the recording of the voiced series, each member of the series was converted to a "voiceless" stimulus by eliminating the final 50-msec F1 transition. This produced stimuli similar to the one at the bottom of Fig. 1(a). The new, voiceless series was then recorded.

A similar procedure was followed in synthesizing the final fricatives and clusters. That is [see Figs. 1(b)–1(e)], the same basic stimuli were used for words ending in both voiced and voiceless sounds. The voiced series was recorded first, and then, after appropriate adjustments had been made in formant transitions and cutbacks, friction durations, silent closure intervals, and release characteristics, the voiceless series was recorded.

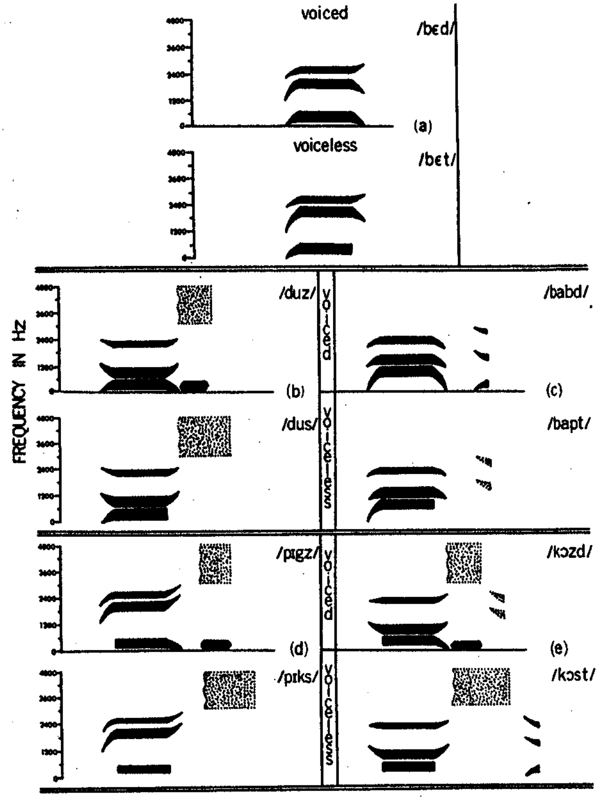

FIG. 1. Examples of patterns used to synthesize the voiced and voiceless stimuli series in the perceptual tests.

The stimuli of each series were randomized and played to 25 undergraduate phonetics students at Hunter College of the City University of New York. There were 20 females and five males in the group. All had normal hearing. The subjects were asked to respond in a forced-choice format, labeling which member of a minimal pair of words they heard for each stimulus played.

The subjects heard the stimuli over a loudspeaker in the sound-treated speech science laboratory at

TABLE I. Corpus and oppositions tested.

	Opposition tested	Minimal pairs
Stops	/p–b/	gape–Gabe
	/t–d/	bet–bed
	/k–g/	Burke–Berg
Fricatives	/f–v/	Duff–dove
	/θ–ð/	teeth–teethe
	/s–z/	deuce–dues
	/ʃ–ʒ/	cash–cas(ual)
Clusters	/pt–bd/	bopped–bobbed
	/st–zd/	cost–caused
	/ps–bz/	cops–cobs
	/kt–gd/	tucked–tugged
	/ts–dz/	tights–tides
	/ks–gz/	picks–pigs

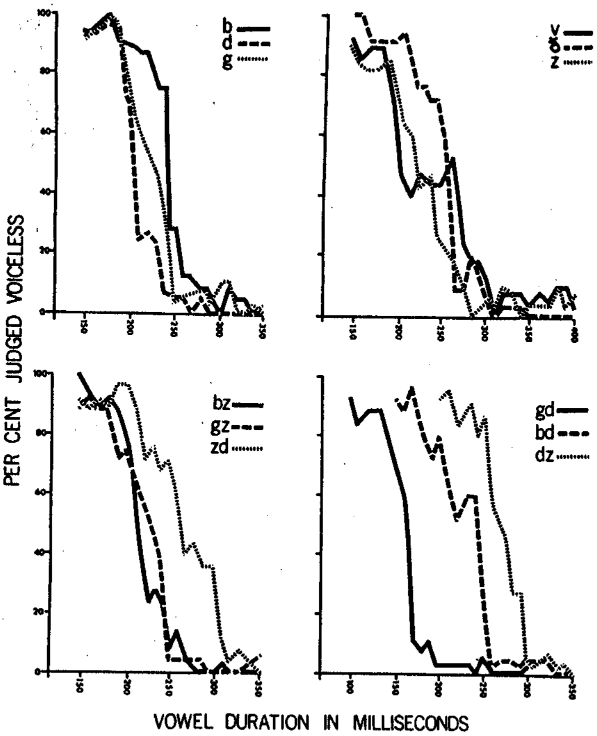

Fig. 2. Perception of the voiced stimuli series as a function of preceding vowel duration.

Hunter College. The tapes were played on an Ampex AG-500 machine at 7½ ips.

B. Results and Discussion

1. Perceptual Changes for All Classes of Sounds

The results of the labeling tests revealed that, with one exception and regardless of the voicing cues used in their synthesis, all final consonants and clusters were perceived as voiceless when preceded by vowels of short duration and as voiced when preceded by vowels of long duration. That is, a final consonant or cluster synthesized with cues appropriate for voicing was perceived as voiceless when the vowel preceding it was of short duration, and as voiced when the preceding vowel was of long duration (Fig. 2). A final cluster or consonant synthesized with cues for voicelessness was perceived in precisely the same way (Fig. 3).

The exception to this result was found in the /ʃ/–/ʒ/ opposition (Fig. 4). Although there was a general tendency for the subjects to perceive these sounds as voiced as the preceding vowel duration increased, the over-all picture was one of confusion, rather than of change, in perception. The listeners were never fully convinced as a group that any stimulus in the series belonged to one class rather than another.

There are at least two possible explanations for this exception. First, the /ʃ/–/ʒ/ opposition is one of extremely low functional yield in English. Thus it is possible that the subjects were not used to listening for it and so lacked the experience necessary for making the labeling decisions. Second, the low yield of the opposition left few options for minimal test pairs, and these few generally included rare or somewhat artificial words. The pair used in this experiment, *cash*/kæʃ/–*cas*/kæʒ/ (short for *cas*ual) certainly fits into the latter, artificial category. It seems safe to assume, therefore, that because of these factors the cue of vowel duration, rather than completely offsetting the cues to voicing or voicelessness during the friction portion of the consonants, tended merely to make the stimuli ambiguous.

Since the indications are that the unique results for /ʃ/ and /ʒ/ are due partly to the structure of the language and partly to the choice of test items, rather than to a failure of the cue being tested, the data for these sounds will not be considered further here, either in comparison with other final fricatives, or in comparing one manner class with another.

2. Comparison of Perceptual Changes among the Classes of Sounds

The general picture, thus, is one of preceding vowel duration as a sufficient and, *for the stimuli used here*, a necessary cue to the voicing characteristic of word-final consonants and consonant clusters. It is clearly, though,

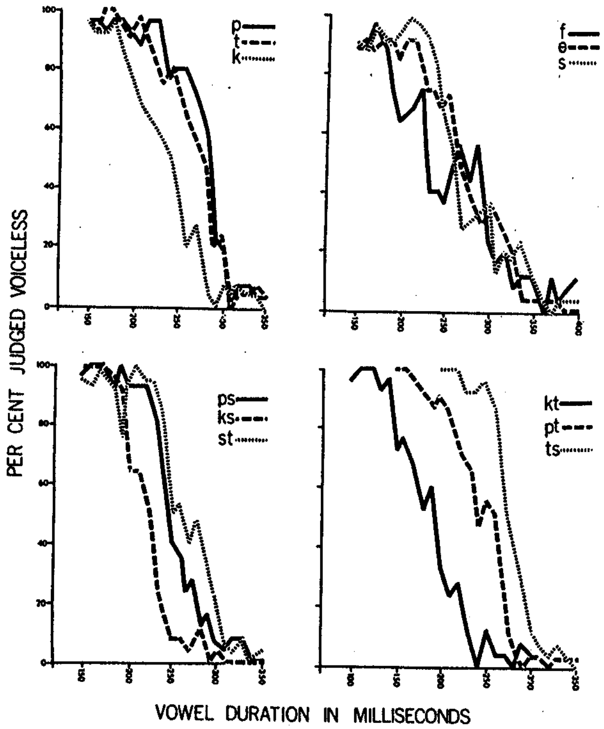

Fig. 3. Perception of the voiceless stimuli series as a function of preceding vowel duration.

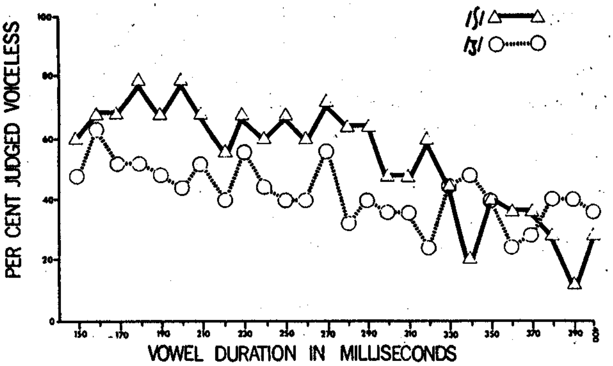

FIG. 4. Perception of synthetic /ʃ/ and /ʒ/ as a function of preceding vowel duration.

not the only cue. The presence of the cues for voicing or for voicelessness during the final segment of the words tested does affect perception. Figure 5 demonstrates this by the direct comparison of the labeling curves for both members of the minimal pairs tested. The generally earlier descent of the curves for the voiced stimuli as compared with those of the voiceless stimuli indicates that stimuli synthesized with cues for voicing in the final segment were perceived as voiced following vowels of shorter duration than those synthesized with cues for voicelessness.

Although this result is found generally, there are some differences between the classes of sounds tested. These differences are manifested by a comparison of those parts of the labeling curves over which the perception changes most radically. This portion of each labeling curve, hereafter called the *critical vowel duration*, is to be taken as the range of stimuli over which perceptual judgments change from 80% to 20% agreement of voicelessness. Figure 6 illustrates the critical vowel duration for the perception of *tucked*.

Three points during the critical vowel duration may be taken as indicative of the differences between the effects of vowel duration on the perception of the various stimuli classes:

(1) The beginning of the critical vowel duration.
(2) The (first) 50% crossover.
(3) The end of the critical vowel duration.

Table II displays each of these points for each of the contrasts tested. The figures in this table show that there is regularly an earlier descent of the voiced than of the voiceless stimuli curves for both the stops and fricatives, although the curves for the latter are, on the average, less separated than are those for the former. The clusters show the least separation, and in two cases, *picks–pigs* and *cost–caused*, there are at least two points where the curves overlap. Although the case for the clusters is thus somewhat ambiguous, there is a clear indication that the presence or absence of voicing cues during the closure period of the final consonants does have some value, although such cue value seems clearly secondary in importance to that of vowel duration.

Table II also reveals that the absolute values for the onset of critical vowel duration vary among the test words. This variation may be ascribed to the particular vowel present in a given stimulus. Within each class of consonants tested the smallest values of the onset of critical vowel duration are those for vowels which are generally described as short,[12] monophthongal, or lax in real speech. The largest values are found in those words which contain long, diphthongized, tense vowels. It seems reasonable to assume that lower stimulus values would begin to register perceptual change for short vowels than for long vowels.

3. *Relative Strength of the Cue of Vowel Duration Preceding Stops, Fricatives, and Clusters*

Data derived from the labeling curves indicates that preceding vowel duration is not equally effective as a voicing cue before each of the consonant types tested. Two interrelated measures support this indication. The first of these is the critical vowel duration. Table III reveals that this duration is consistently longer for the vowels preceding fricatives than for those preceding either stops or clusters. On the average, the critical

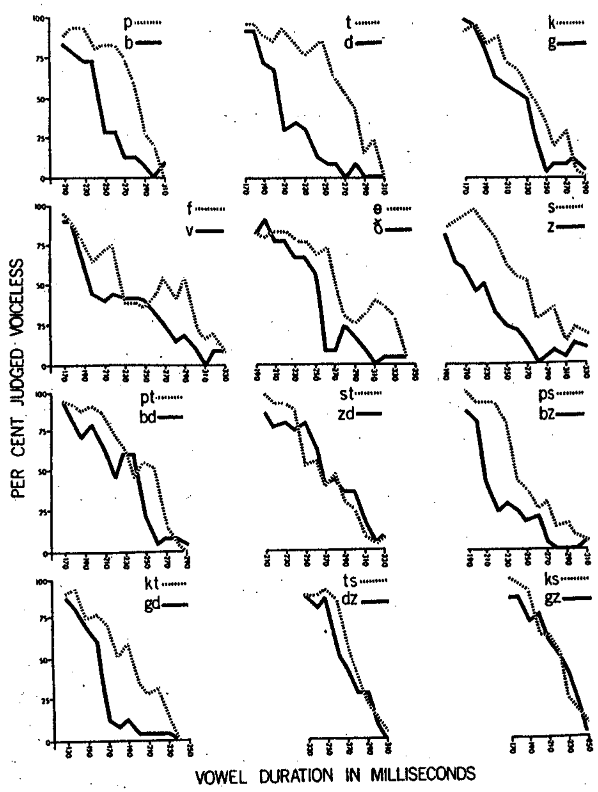

FIG. 5. Comparison of labeling curves for cognate sounds.

TABLE II. Critical vowel durations, points of measurement (msec).

Opposition tested	Beginning point: Voiceless stimulus	Beginning point: Voiced stimulus	First 50% crossover: Voiceless stimulus	First 50% crossover: Voiced stimulus	End point: Voiceless stimulus	End point: Voiced stimulus
Stops						
/p–b/	265	215	282	245	300	265
/t–d/	252	187	275	205	289	235
/k–g/	202	190	243	223	260	242
Average difference	42.3		42.3		35.7	
Fricatives						
/f–v/	186	183	228	198	305	275
/ʃ–ʒ/	228	208	269	251	330	258
/s–z/	243	190	271	217	308	260
Average difference	25.5		34.0		50.0	
Clusters						
/pt–bd/	214	200	239	218	269	252
/st–zd/	243	215	265	265	303	307
/ps–bz/	230	200	238	218	275	243
/kt–gd/	148	140	194	162	226	169
/ts–dz/	262	252	273	263	294	294
/ks–gz/	194	185	221	223	235	240
Average difference	16.5		13.5		17.7	

durations of vowels before fricatives is 88% longer than those before stops and 70% longer than those before clusters.

The slopes of the labeling curves during the critical vowel duration (which are, of course, a function of that duration) present the same evidence in another form. Table IV shows that, although there is one case of overlap between the slopes in the stop and fricative categories, the averages for these categories are clearly separated, with the average for the clusters much more closely resembling that of the stops than that of the fricatives.

One might infer from these results that the greater duration of the fricatives relative to the stops causes the voicing cue present during fricative articulation to be more powerful than that during stop articulation. Thus the range of durations over which the vowel must be varied to overcome the effect of voicing during consonant closure in order to cause a change in perception should be greater before fricatives than before stops. Further, the presence of the stop in each of the final clusters tested seems to condition both the vowel duration and the rapidity of the perceptual change as manifested by the critical vowel duration. This appears to be true even for those clusters containing a fricative and independently of the ordering of the stop and fricative in the cluster. Table V presents data for real speech which reveals that vowel durations before final clusters of all three types tested more closely resemble that of final stops than that of final fricatives.

It can be concluded, then, that identical variations in preceding vowel duration will cause a greater and more rapid change in the perception of voicing for word-final stops and clusters than for fricatives.

II. DISCRIMINATION TESTS

A. Procedures

In order to determine the mode of perception used by the listeners, a series of discrimination tests, using the stimuli of the labeling experiment, was administered. Earlier studies employing this type of test have revealed that where perception is categorical there is a close fit between the discrimination function predicted from the labeling curves and that obtained from the discrimination test itself. In those cases where perception was continuous across a range of varying stimuli, the curves did not fit: The obtained discrimination function was at a considerably higher level than the predicted function.[13]

In the discrimination experiment reported here, one member of each class of the synthetic test consonants was selected: a final stop in the word *Berg*/bɝg/; a final fricative in the word *dues*/duz/; and a final cluster in the word *picks*/pɪks/. These sounds were chosen from their respective categories because the perception of them in the labeling tests changed over the smallest number of stimuli. It was felt that categorical perception, if it were present, would be most likely to occur in those stimuli for which perception changed over the fewest stimulus steps.

The stimuli were presented in an "oddball" format in which the listeners heard a triad of stimuli, one of which was different from the other two. Subjects were asked to identify the odd member of each triad. Each stimulus was matched in a triad with the stimulus

TABLE III. Critical vowel duration (msec).

Opposition tested	Voiceless stimulus series		Voiced stimulus series
/p–b/	35		50
/t–d/	37		48
/k–g/	58		52
av: each series	43.3		50.0
av: both series		46.6	
/f–v/	118.4		92
/θ–ð/	102		50
/s–z/	65		95
av: each series	95.1		79.0
av: both series		87.1	
/pt–bd/	55		52
/st–zd/	60		82
/ps–bz/	45		43
/kt–gd/	78		29
/ts–dz/	32		42
/ks–gz/	41		55
av: each series	51.8		50.5
av: both series		51.2	

which differed from it by one, two, three, and seven stimulus steps. Finally, each odd stimulus was heard once in each position of a triad, so that three triads were prepared for each oddball stimulus for each of the stimulus steps tested.

The seven-step test was prepared in order to compare discrimination across phoneme boundaries with that within phoneme boundaries (as determined from the labeling tests). Seven steps were necessary because of the extended range of stimuli which separated the phonemic labels applied by the listeners (see Figs. 2 and 3). In each of the seven-step tests, one stimulus was from the range labeled voiceless, and two were from the voiced range. This allowed for one comparison across and one within the phoneme boundary. In the seven-step test for *dues*/duz/, there were enough stimuli from the long end of the range to add another comparison within the voiced phoneme category. There were, however, no comparisons for any items within the voiceless category in the seven-step test.

B. Results

Figure 7 shows the results of the one-, two-, and three-step tests. Because the results of these tests contain no interphonemic comparisons, they are not as revealing as the seven-step test results. They do, however, indicate a generally higher level of discrimination than would be expected from the predicted curves. This is particularly true in the case of the final stop and fricative tested, especially as the step-function increases from one to three.

More interesting results may be seen in the seven-step test (Fig. 8), which allowed for inter- and intraphonemic comparisons. In all three categories tested the obtained discrimination function lies at extremely high values, both for inter- and intraphonemic comparisons. On the other hand, the predicted levels, while high in the interphonemic comparisons, still lie considerably below the obtained functions. Further, the predicted levels for the intraphonemic comparisons are virtually at the chance level, while the obtained functions for the same comparisons are at or near the 90% level.

Thus it appears that listeners can discriminate far better than they can label, both across and, more importantly, within phoneme boundaries, and that when the voiced–voiceless opposition is cued by the duration of the preceding vowel, perception is more nearly continuous than categorical.

TABLE IV. Slope of critical vowel duration.

Voiceless stimulus	Slope		Voiced stimulus
Stops			
gape	1.715	1.200	Gabe
bet	1.740	1.240	bed
Burke	1.070	1.150	Berg
av: Voiceless stops	1.508	1.197	av: Voiced stops
Fricatives			
Duff	0.507	0.645	dove
teeth	0.585	1.200	teethe
deuce	0.945	0.858	dues
av: Voiceless fricatives	0.697	0.901	av: Voiced fricatives
Clusters			
bopped	1.110	1.150	bobbed
cost	1.330	0.655	caused
cops	1.385	1.330	cobs
tucked	0.765	2.070	tugged
tights	1.915	1.460	tides
picks	1.460	1.090	pigs
av: Voiceless clusters	1.328	1.293	av: Voiced clusters

Averages: All stops: 1.353 All fricatives: 0.790 All clusters: 1.310

TABLE V. Average vowel durations (real speech).

Vowel precedes:	Milliseconds
Stop+stop	
voiceless	163
voiced	242
Stop+fricative	
voiceless	150
voiced	219
Fricative+stop	
voiceless	157
voiced	240
All clusters	
voiceless	156
voiced	230
All stops	
voiceless	164
voiced	316
All fricatives	
voiceless	210
voiced	362

III. SUMMARY AND CONCLUSION

The results of the experiments described above lead to the following conclusions:

(1) Preceding vowel duration is a sufficient (*and for the types of stimuli employed here*, a necessary) cue to the perception of the voicing characteristic of a word-final stop, fricative, or cluster.

(2) The presence of voicing during the closure period of a final consonant or cluster does have some cue value, although it is minor compared to that of vowel duration.

(3) The cue of preceding vowel duration is more effective before stops and clusters than before fricatives.

(4) The perception cued by the preceding vowel duration is continuous, rather than categorical.

The strength of preceding vowel duration as a cue is not surprising, especially if one contrasts it with the nature of other potential cues to the voicing characteristics of final consonants: Systematic variation in vowel duration is the most consistently present of all the cues. Examples of voiced consonants with little or

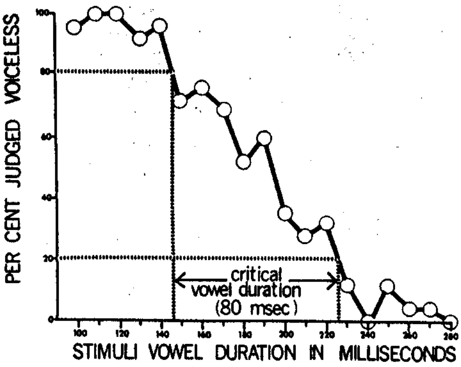

FIG. 6. Specification of critical vowel duration in a labeling curve (stimulus: /tʌkt/).

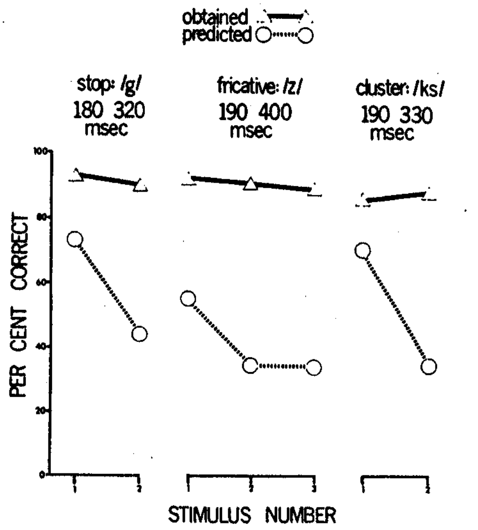

FIG. 8. Discrimination functions for seven-step tests.

no voicing during consonant closure are not uncommon, especially if such consonants are followed in context by voiceless sounds. Similarly, voiceless consonants which evidence vocal pulsing throughout most or all of their closure period are not uncommon, especially when such consonants are followed in context by voiced sounds.[14] The cue of stop release is often absent in American English,[15] and fricative duration differences, although consistently present, may be accompanied by contradictory vocal pulsing cues, depending on the context of the utterance.[16] Since the reliability of these cues is not high, it might be expected that the more consistently present cue of preceding vowel duration would attain primacy in the perceptual categorization scheme of listeners. In a sense, what we may have here is a sort of natural selection among the cues, with the one best equipped to survive the effects of context becoming the most significant cue of all.

The association of the voiced–voiceless opposition with the cue of vowel duration appears to bring about a situation in which continuous rather than categorical perception occurs. Such a situation differs from most others already investigated, in which the cues to the opposition are such that changes in perception are more nearly categorical.

There is, then, something more here than just another case in which different cues to the same distinction are in operation in different phonetic contexts. In this case the cue itself is perceived differently from other cues to the same opposition in other contexts. Further, the time of occurrence of the cue does not fit into the usual pattern of expectation. The phonetic features which are said to distinguish one class of sounds from another are almost universally assumed to be present during, and as a result of, the articulation of the sound(s) in question. For the stimuli tested here, however, the cue for voicing does not coincide with the articulation of the sounds. Rather, it occurs during a preceding vowel segment. We have here an example of Lisker's point that "a single linguistic segment may be identified on the basis of cues contained in more than one acoustic segment, and . . . [that] a single acoustic segment may provide information for the identification of more than one linguistic segment." One could further hold with Lisker that "each acoustic segment might be said to supply cues to a single linguistic segment, while any

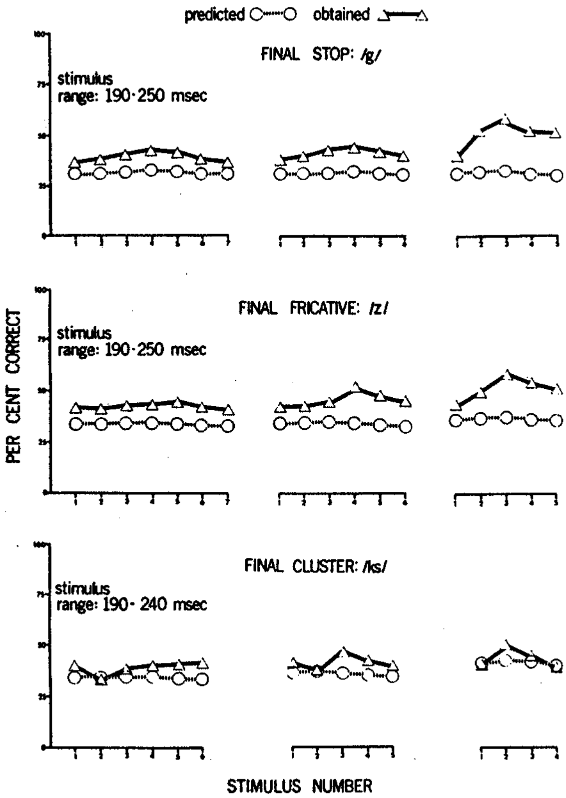

FIG. 7. Discrimination functions for one-, two-, and three-step tests.

features it contains which have cue value for some other linguistic segment could be considered 'automatic' in the neighborhood of the acoustic segment or segments having a recognized relation to this other linguistic segment."[17] Specifically, the acoustic segment manifested by formants here supplies cues to the single linguistic segment identified as the vowel. One of the other features it contains, its duration, has a cue value for another linguistic segment, the following consonant, with which it has a "recognized relation."

ACKNOWLEDGMENTS

The author is extremely indebted to Dr. Franklin S. Cooper and to the staff of Haskins Laboratories. Without their assistance and the use of the Laboratories' facilities this study would not have been possible. The author further wishes to express his gratitude to Professors Arthur J. Bronstein, Katherine S. Harris, Arthur S. Abramson, Leigh Lisker, and Louis G. Gerstman, all of whom commented on and suggested improvements in the experimental procedures and the manuscript.

* Also at Haskins Laboratories, New Haven, Conn. 06510.

[1] Daniel Jones, *The Phoneme: Its Nature and Use* (Heffer, Cambridge, England, 1950), p. 121.
[2] Haskins Laboratories Quart. Prog. Reps., No. 21, Appendix 5, 1–2 (1956).
[3] P. Denes, J. Acoust. Soc. Amer. 27, 761–764 (1955).
[4] L. J. Raphael, "Preceding Vowel Duration as a Cue to the Perception of Voicing of American English Consonants in Word-Final Position," City Univ. New York, unpublished doctoral dissertation, 1970, 112–124.
[5] A. Malécot, J. Acoust. Soc. Amer. 47, 1588–1592 (1970).
[6] A. M. Liberman, K. S. Harris, H. S. Hoffman, and B. C. Griffith, J. Exp. Psych. 54, 358–368 (1957).
[7] A. M. Liberman, K. S. Harris, J. O. Kinney, and H. Lane, J. Exp. Psych. 61, 379–388 (1961).
[8] D. B. Fry, A. S. Abramson, P. D. Eimas, and A. M. Liberman, Language and Speech 5, 171–189 (1962).
[9] J. Bastian and A. S. Abramson, J. Acoust. Soc. Amer. 34, 743(A) (1962).
[10] K. N. Stevens, A. M. Liberman, M. Studdert-Kennedy, and S. E. G. Öhman, Language and Speech 12, 1–23 (1969).
[11] A. M. Liberman, F. S. Cooper, D. P. Shankweiler, and M. Studdert-Kennedy, Psych. Rev. 74, 431–461 (1967).
[12] G. E. Peterson and I. Lehiste, J. Acoust. Soc. Amer. 32, 693–703 (1960).
[13] Ref. 6, 7, 8, and 10.
[14] L. Lisker, A. S. Abramson, F. S. Cooper, and M. H. Schvey, J. Acoust. Soc. Amer. 45, 1544–1546 (1969).
[15] H. Rositzke, Amer. Speech 18, 39–42 (1943).
[16] Ref. 14.
[17] L. Lisker, Language 33, 370–374 (1957).