70-24,483

RAPHAEL, Lawrence Joel, 1938-
PRECEDING VOWEL DURATION AS A CUE TO THE PERCEPTION OF VOICING OF AMERICAN ENGLISH CONSONANTS IN WORD-FINAL POSITION.

The City University of New York, Ph.D., 1970
Speech

University Microfilms, A XEROX Company, Ann Arbor, Michigan

PRECEDING VOWEL DURATION AS A CUE TO THE PERCEPTION OF VOICING OF AMERICAN ENGLISH CONSONANTS IN WORD-FINAL POSITION

by

LAWRENCE J. RAPHAEL

A dissertation submitted to the
Graduate Faculty in Speech in partial
fulfillment of the requirements for the
degree of Doctor of Philosophy,
The City University of New York.

1970

This manuscript has been read and accepted
for the Graduate Faculty in Speech in
satisfaction of the dissertation requirement
of Doctor of Philosophy.

April 8, 1970

Chairman of Examining Committee

Executive Officer

The City University of New York

Acknowledgements

Any author, student or professional, who devotes a considerable amount of time and effort to a project is bound to owe a debt of gratitude to a number of people who have helped him along the way. This writer is no exception. What follows, then, are the words of thanks; the feelings, too, are intended, although the words do not adequately convey them.

First, I would like to thank my chief sponsor and adviser, Professor Arthur J. Bronstein, whose instruction and inspiration since my undergraduate days have influenced more than just the contents of this dissertation. He once said that he never doubted that I would finish. Because he never doubted it, neither did I.

I am also greatly indebted to the other members of my advisory committee, Dr. Katherine S. Harris and Professor Louis J. Gerstman. Their knowledge and confidence were invaluable to me in surmounting the many problems I encountered in my research and writing.

A special vote of thanks must go to Professor Arthur S. Abramson of the University of Connecticut, to whom I can trace both my interest in experimental phonetics, and, indeed, the idea which grew into this dissertation.

Further, I would like to express my appreciation to Dr. Franklin S. Cooper, director of Haskins Laboratories,

and to his staff. Without the facilities of the Laboratory and the assistance of those employed there, the research reported below would have been impossible. I am particularly indebted to Miss Agnes McKeon and Mr. David Zeichner whose special skills and kind attention were welcome on many occasions.

I am grateful to a number of people whose challenging questions, thoughtful advice and personal research played a major role in the development of this dissertation. They include Professors Leigh Lisker, John B. Newman, CJ Stevens, and Dr. Tom Gay.

To my fellow students and colleagues, Mr. Albert Abrams, Miss Patricia Dooley, Mrs. Jill Giottino, Mr. Irving Maston, Mr. David Pisoni, Mrs. Lorraine Russell, and Mrs. Eleanor Sass, I also express thanks. In the course of many conversations with them several previously unformed ideas took shape and ultimately found their way into my work.

Finally, and by no means least, I thank my wife, Carolyn. Her inspiration, assistance, understanding and quiet perseverance were the final, essential components of my efforts. If this dissertation was written for anyone, it was written for her.

Lawrence J. Raphael

New York City
February, 1970

TABLE OF CONTENTS

ACKNOWLEDGEMENTS iii

LIST OF TABLES vi

LIST OF ILLUSTRATIONS viii

Chapter

I. SOME THEORIES OF THE PSYCHOLOGICAL AND ACOUSTIC BASES FOR THE OPPOSITION OF MEMBERS OF COGNATE PAIRS IN AMERICAN ENGLISH 1

II. CUES TO THE CLASSIFICATION OF THE MEMBERS OF COGNATE PAIRS IN WORD-FINAL POSITION: SPECTROGRAPHIC ANALYSIS AND TAPE-CUTTING EXPERIMENTS 75

III. VOWEL DURATION AS A CUE TO THE PERCEPTUAL SEPARATION OF WORD-FINAL COGNATE SOUNDS: SYNTHETIC-SPEECH EXPERIMENTS 127

IV. SUMMARY AND IMPLICATIONS 189

APPENDIX 194

BIBLIOGRAPHY 211

AUTOBIOGRAPHICAL STATEMENT 220

LIST OF TABLES

Table Page

I-1 English Cognate Pairs 1

I-2 Comparison of Data for Variation in Vowel Duration Before 'Voiced' and Before 'Voiceless' Stops (All Durational Values Given in Milleseconds) 22

II-1 Spectrographic Data for Stops 86

II-2 Spectrographic Data for Fricatives 87

II-3 Data for Minimal Pairs Ending in /p-b/ . . . 88

II-4 Data for Minimal Pairs Ending in /t-d/ . . . 89

II-5 Data for Minimal Pairs Ending in /k-g/ . . . 90

II-6 Data for Minimal Pairs Ending in /f-v/ . . . 91

II-7 Data for Minimal Pairs Ending in /θ-ð/ . . . 92

II-8 Data for Minimal Pairs Ending in /s-z/ . . . 93

II-9 Data for Minimal Pairs Ending in /ʃ-ʒ/ . . . 94

III-1 Test Corpus and Oppositions Tested 140

III-2 Specifications of Stimuli Used to Synthesize Stop-Final Syllables 141

III-3 Specifications of Stimuli Used to Synthesize Fricative-Final Syllables 142

III-4 Specifications of Stimuli Used to Synthesize Syllables Ending in Stop-Plus-Stop Clusters 144

III-5 Specifications of Stimuli Used to Synthesize Syllables Ending in Stop-Plus-Fricative Clusters 145

III-6 Specifications of Stimuli Used to Synthesize Syllables Ending in Fricative-Plus-Stop Clusters 146

III-7 Analysis of Frequency of Occurrence of Stops Used to Initate Test-Stimuli 148

III-8 Critical Vowel Durations: Points of Measurement 175

III-9 Critical Vowel Durations 179

III-10 Slope of Critical Vowel Duration 180

III-11 Average Vowel Durations 182

LIST OF ILLUSTRATIONS

Figure Page

2-1 Voice Onset Time: Stops 76

2-2 Voice Offset Time: Stops 78

2-3 Voice Offset Time: Fricatives 80

2-4 Points Marked on Spectrograms 83

2-5 Spectrographic Measurements 84

2-6 Vowel Duration Before Stops 96

2-7 Vowel Duration Before Fricatives 97

2-8 Voicing During Closure: Stops 98

2-9 Voicing During Closure: Fricatives 99

2-10 Total Voicing Duration: Vowel + Stop 100

2-11 Total Voicing Duration: Vowel + Fricative . . . 101

2-12a Voicing Offset Time: Stops 102

2-12b Voicing Offset Time: Fricatives 103

2-13a Closure Duration: Stops 104

2-13b Closure Duration: Fricatives 105

2-14, 15, 16 Perception of Final /b,d,g/ As /p,t,k/ As a Function of Voicing Offset Time 111

2-17 Perception of Final /b/ As /p/ As a Function of Preceding Vowel Duration 114

2-18 Perception of Final /d/ As /t/ As a Function of Preceding Vowel Duration 115

2-19 Perception of Final /g/ As /k/ As a Function of Preceding Vowel Duration 116

2-20 Perception of Final /v/ As /f/ As a Function of Preceding Vowel Duration 118

2-21 Perception of Final /θ/ As /ð/ As a Function of Preceding Vowel Duration 119

2-22 Perception of Final /z/ As /s/ As a Function of Preceding Vowel Duration 120

2-23 Perception of Final /ʃ/ As /ʒ/ As a Function of Preceding Vowel Duration 121

3-1 Sample Stimuli: Stop-Final Syllables 131

3-2 Sample Stimuli: Fricative-Final Syllables . . . 133

3-3 Sample Stimuli: Stop + Stop-Final Syllables . . 135

3-4 Sample Stimuli: Stop + Fricative-Final Syllables 136

3-5 Sample Stimuli: Fricative + Stop-Final Syllables 138

3-6 Analysis of Frequency of Occurrence of Vowels Used in Test Stimuli 147

3-7 Perception of Final Stops: Voiceless Stimuli . 151

3-8 Perception of Final Stops: Voiced Stimuli . . . 152

3-9 Perception of Final Fricatives: Voiceless Stimuli 153

3-10 Perception of Final Fricatives: Voiced Stimuli 154

3-11a Perception of Final Clusters: Voiceless Stimuli 155

3-11b Perception of Final Clusters: Voiceless Stimuli 156

3-12a Perception of Final Clusters: Voiced Stimuli . 157

3-12b Perception of Final Clusters: Voiced Stimuli . 158

3-13 Perception of Final /ʃ-ʒ/ 159

3-14 Discrimination Test: Final Stop: /g/ 165

3-15 Discrimination Test: Final Fricative: /z/ . . . 166

3-16 Discrimination Test: Final Cluster: /ks/ . . . 167

3-17 Seven-Step Discrimination Tests 169

3-18 Comparison of Perceptual Change: Voiced vs. Voiceless Stops 171

3-19 Comparison of Perceptual Change: Voiced vs. Voiceless Fricatives 172

3-20 Comparison of Perceptual Change: Voiced vs. Voiceless Clusters 173

3-21 Comparison of Perceptual Change: Voiced vs. Voiceless Clusters 174

3-22 Comparison of Critical Vowel Durations: Stops and Fricatives 184

CHAPTER I

SOME THEORIES OF THE PHYSIOLOGICAL AND ACOUSTIC BASES FOR THE OPPOSITION OF MEMBERS OF COGNATE PAIRS IN AMERICAN ENGLISH

Within the phonemic inventory of American English there are several pairs of sounds which are termed cognates by many linguists and phoneticians. There is general agreement that the members of each cognate pair are characterized by the identical manner of articulation (e.g. stop) and by the identical place of articulation (e.g. apico-alveolar).[1] These cognate pairs may be designated as follows:[2]

TABLE I-1

ENGLISH COGNATE PAIRS

	Stop	Fricative
Bilabial	/p-b/	
Labio-Dental		/f-v/
Lingua-Dental		/θ-ð/
Apico-Alveolar	/t-d/	/s-z/
Alveolar-Palatal		/ʃ-ʒ/
Dorso-Velar	/k-g/	

There is considerably less agreement about the characterization of the feature(s) which distinguishes one member of a cognate pair from the other. Many linguists have sought one distinctive feature which would separate the members of all of the cognate pairs, whatever their phonetic realizations, into two groups: /ptkfθsʃ/ vs. /bdgvðzʒ/. The two features most commonly proposed as performing this function are the voiced/voiceless opposition and the fortis/lenis opposition.

On the other hand, some investigators, doubting or denying the existence of any generally distinguishing feature, have proposed a series of individual phonetic features which, taken individually, or in various combinations, are said to underly the opposition of the members of the pairs of sounds in a variety of phonetic environments. Among these features are the nature of the burst and aspiration of stops, the duration of closure of stops, the duration of the friction of fricatives, and the duration of sounds preceding the members of both stop and fricative cognates.[3]

The Voiced/Voiceless Opposition

In its most elementary manifestation, the voiced/voiceless opposition is caused by the presence or absence of periodic vibrations of the vocal folds during the articulation of a speech sound.[4] Thus the vocal folds may be described as vibrating during the articulation of such sounds as /b,g,v,z/ and at rest during the articulation of such sounds as /p,k,f,s/.

A variety of impressionistic tests is recommended to demonstrate this assertion. C. K. Thomas suggests that one touch the tip of one's larynx lightly while saying the words <u>fail</u> and <u>veil</u> to demonstrate the voiceless/voiced contrast between /f/ and /v/.[5] Daniel Jones remarks that when "breathed [voiceless] and voiced sounds are pronounced while the ears are stopped a loud buzzing sound is heard in the latter case but not in the former," and adds that "It is possible to sing tunes on voiced sounds, but not on breathed [voiceless] ones."[6] The validity of such tests is, of course, questionable, particularly with regard to stops such as English /b,d,g/, where tactile sensitivity may not be acute enough along the time dimension to distinguish the voicing during closure (if there is any) from that of the following vowel.[7]

It is, in fact, the category of stops which supplies much of the basis of the argument which denies that the voiced/voiceless opposition (in the form so far considered) effectively separates the members of cognate pairs in English. Spectrographic evidence reveals that in initial position before a stressed vowel, both /p,t,k/ and /b,d,g/ are often produced without voicing during their closure stage.[8] Thus, according to the elementary definition of voicing, both series of stops are voiceless in English in at least one very common phonetic context.

Spectrographic investigation further indicates that there is often audible voicing present in the closure for /p,t,k/ when these stops are preceded by voiced sounds in running speech.[9] Finally, there is evidence that cognate stops in whispered (i.e. voiceless) speech can usually be distinguished.[10]

There is some indication that members of the fricative series /v,ð,z,ʒ/ may be articulated without accompanying vocal fold vibrations upon occasions. Transillumination studies reveal that members of these series are often produced with an open glottis, and, less frequently, without vocal pulsing.[11]

It appears, then, for fricatives as well as for stops, that the presence or absence of periodic vocal fold vibrations is not always a certain indication that a sound will be perceived as belonging to one phonemic category rather than another. That is, the voiced/voiceless contrast does not serve to separate the perception of stop and fricative cognates in all their phonetic manifestations.

The Fortis/Lenis Opposition

An alternative contrast often said to perform the above function successfully is the fortis/lenis opposition. Most linguists use the terms fortis and lenis to refer to either or both of two parameters:

1. The degree of oral breath pressure during the articulation of a sound.
2. The tension of the muscles involved in articulation.[12]

In this scheme sounds such as /p,k,f,s/ are said to be characterized by relatively greater muscular tension and high oral breath pressure and so are labelled fortis. Sounds such as /b,g,v,z/ are said to be characterized by relatively lax articulation and low oral breath pressure and so are termed lenis.

Many linguists hold that in English the fortis/lenis opposition is more fundamental than the voiced/voiceless opposition. Jakobson, Halle and Fant comment that "In languages lacking an autonomous opposition of voiced and voiceless consonants, the latter is either used as a mere concomitant of the opposition of lax and tense consonants, as in English, or oral consonants are normally voiceless . . ."[13] Stetson, in much the same manner, states that the pressure difference underlying the fortis/lenis opposition "is more fundamental than the voicing of consonants, and persists after the voicing distinction is lost, i.e. after both surd (fortis) and sonant (lenis) are voiced . . ."[14]

Although earlier statements of this theory were often based on impressionistic evidence, some experimental work in recent years has supported the validity of the opposition in a variety of phonetic environments. Black found the mean oral air pressure during the articulation of /p,t,f,s/ greater than that during the articulation of /b,d,v,z/. This finding held for sounds in initial, intervocalic and final positions.[15] Stetson found that vertical movements of the larynx caused pressure differences in the oral cavity and that these pressure differences were in agreement with the assignment of /p,t,k,f,θ,s,ʃ/ to the fortis category, and of their cognates to the lenis group.[16]

Malécot found that oral air pressure measurements taken by means of a nasally inserted catheter matched his subjects' "kinesthetic impression of <u>force of articulation</u> in a general way and . . . [fit] the <u>lenis-fortis</u> theory perfectly."[17] In a later study, using improved mechanical devices for measuring pressure, Malécot found that there were significant differences in peak pressure, time integral of instantaneous pressure and total duration of pressure between classes of stops in intervocalic, pre-stress position.[18] Further, Malécot determined that the subjects of his experiment were sensitive to differences in peak pressure and time integral of pressure that were "well within the range of values observed real speech."[19]

Some recent studies, however, by Malécot, Lisker and others, have shown that there is no consistent physical basis for a fortis/lenis distinction in English, either in terms of breath pressure or of articulatory energy.[20] Measuring the breath pressure component of the opposition in stops, Lisker found "questionable the ready acceptance of the view that a difference in supraglottal air pressure is a thoroughly reliable index of the /ptk/-/bdg/ contrast, and that such a pressure difference constitutes the evidence for the existence of a fortis-lenis dimension independent of and more fundamental than one of voicing."[21] Lisker measured peak pressure, the time integral of instantaneous pressure, and the duration of pressure buildup and decay for stops in initial, medial-pre-stress, medial-post-stress, and final positions.[22] No one of these measures of pressure effectively separated the stops in each of the positions investigated, although all of the measures were effective in final position and one or more measures were operative in each position.[23]

It may be noted that Lisker and Malécot differ in their results for medial-pre-stress position in regard to peak pressure. Both, however, found the time integral of instantaneous pressure to separate stop categories in this position,[24] and Lisker, in a later study, found total duration of pressure to be a reliable indicator of categorization.[25]

Lisker has pointed out that the fact that initial stops of both classes show no significant difference in pressure profiles is of special import: Since initial stops of both classes are often produced without voicing during the closure period while pressure is increasing, "the absence of any difference between the two . . . does contradict the view that a difference in pressure invariably marks the contrasts between /ptk/ and /bdg/."[26]

Further, in questioning the consistency with which phoneticians apply the terms <u>fortis</u> and <u>lenis</u>, Lisker notes that a medial post-stress stop, such as the /p/ in copper has been characterized in the literature as being lenis. Since /p/ is elsewhere referred to as fortis, Lisker concludes that "we must reckon with the fact that there is some disagreement as to the existence of a fortis-lenis contrast between /ptk/ and /bdg/ in just that position where the two series (alveolars excepted) differ most [according to studies done at Haskins Laboratories] in pressure. Conversely, it is where phoneticians seem most in agreement that there is a fortis-lenis difference, namely in initial position, that no pressure difference may be found." Lisker concludes that "if we wish to continue to say that the initial stops differ in tension, we must reject the [air] pressure definition of the fortis-lenis difference and search for some other physical interpretation of this dimension."[27]

The second of the fortis/lenis measures, muscular tension of the articulators, was investigated by Malécot. Measuring pressure duration, peak pressure, and pressure impulse of the muscles of opposing articulators, for labial and alveolar stops and nasals, he found that "American speakers do not articulate consonants in such a way as to yield values [of articulatory pressure] sufficiently distinctive to be used proprioceptively in identifying phonetic class." Malécot concludes that motor commands to the articulators (for the sounds he studied) contain no information regarding tension, but merely that an occlusion be made at a particular place in the mouth.[28]

Force of muscular tension during articulation has also been measured in two independently conducted studies which employed electromyographic techniques. Measuring contractions of the orbicularis oris muscle during the articulation of oral and nasal labial stops, in initial and final syllable position, a research group at Haskins Laboratories found that although some differences in tension exist between /p/ and /b/, they "are small at best and non-existent for some subjects. The differences observed in these experiments could not serve as the basis for a workable phonemic distinction based on muscular tenseness or laxness."[29]

Fromkin, in a separate study, found such a great overlap between the peak amplitudes of the EMG traces for /p/ and /b/ in initial and final positions that she concluded "the gestures at the lips were virtually identical" for both sounds. She did, however, find a greater duration of muscular tension for /b/ than for /p/ in every utterance tested.[30] This finding contradicts the fortis/lenis theory. In the data supplied by the principal subject of Fromkin's experiment, "initial /b/ shows both greater amplitude and greater duration of muscular activity than initial /p/, while final /p/ shows greater amplitude but shorter duration than final /b/." On the basis of these results Fromkin concludes, in agreement with the Haskins group, that "a feature other than the tense-lax feature must differentiate these two phonemes [i.e. /p/ and /b/]

in English."[31,32]

It seems, then, in English, that both the voiced/voiceless and the fortis/lenis oppositions, if interpreted in their most basic and fundamental senses, do not possess the generality necessary to separate the members of cognate pairs into two classes. That is, there is at least one actual contrasting phonetic manifestation of each cognate pair that neither of the oppositions adequately explains.

Acoustic Cues to Cognate Differentiation

Phoneticians, both in attempts to extend the nature of the oppositions described above, or in recognition of their inadequacy, have investigated a number of phonetic features that characterize the allophone of one phoneme in a given phonetic environment, but not the allophone of its cognate in the same environment. It is likely that those who view these features as reflexes of the voiced/voiceless or fortis/lenis oppositions recognize that these oppositions must be substantially modified from their fundamental forms in order to serve as theoretical bases which adequately account for each of the several acoustic events that might be ascribed to them.

One of the features which has been studied is aspiration. It has long been noted that aspiration accompanies the release of English /ptk/ (but not of /bdg/) when these sounds occur initially in a syllable before a stressed vowel.[33,34]

Two independently conducted tape-cutting experiments are widely interpreted as demonstrating the presence of aspiration as cue to /ptk/ and its absence as a cue to /bdg/. In both of the studies it was found that when the /s/-friction was removed from the beginnings of words such as *spill*, *still*, *skill*, native speakers of English identified them as *bill*, *dill*, *gill*, rather than as words initiated by /ptk/. Since the stops in these words are often described as voiceless and unaspirated, and since they were heard as "voiced" stops by the listeners, the experimenters concluded that "aspiration is a more dominant cue than voicing in the perceptual separation of these two classes of stops."[35]

Hultzén argues that stops after /s/ in initial clusters should be interpreted phonemically as /bdg/. He uses tape-cutting experiments as evidence, and refers to analogous cases where "the segment after /s/ is a recognized lenis, elsewhere voiced, which has no recognized fortis, usually voiceless counterpart, [e.g. /sm/, /sn/, /sw/, /sj/, /sl/ . . .]."[36]

Lisker objects to Hultzén's assumption that voiced consonants such as /m,n,w,j,l/ in post-/s/ position are lenis and voiceless. Referring to work by Sholes, Lisker indicates that air pressure measurements for stops after /s/ do not vary significantly from those for initial /ptk/. Further, "spectrographic evidence also contradicts Hultzen in his consignment of the second members of such clusters as sp, sl, and tr all to a single voiceless lenis category, inasmuch as there is no acoustic basis for claiming that they are equally devoiced in that position." Lisker conjectures that Hultzén may have mistaken a "reduction in the duration of the second consonant" in such clusters as /sm/, /sn/, /sl/, for devoicing.[37]

Lisker further rejects Hultzén's contention, pointing out that although the tape-cutting experiments indicate that post-/s/ stops "can be distinguished from initial /ptk/, they certainly cannot be taken to force the decision that . . . [they are] /bdg/." That is, the stops beginning the words labelled as bill, dill, gill, are not initial stops. Rather, they are acoustic segments which have been displaced from their actual context. Thus, comparing them with initial /ptk/ and /bdg/ may lead listeners to make a phonemic identification that in fact "conflict[s] with an assignment required by the phonological facts of the language." Lisker notes that when "the /p/ in words like rapid copper, etc. were presented as initial stops for judgments . . . they were by and large called /b/. Since English contains such pairs as rapid-rabid it is very clear that the results of this experiment cannot be phonologically decisive."[38]

Some phoneticians feel that aspiration is simply a reflex of the pressure differences which characterize the fortis/lenis opposition. If it is, it is a reflex that is absent in several phonetic contexts: /ptk/ are often unaspirated (i.e. unreleased) in English in absolute final position,[39] and preceding other stops, especially their cognates.[40] For those contexts where /ptk/ are clearly distinguishable from and in opposition to /bdg/, in spite of the lack of aspiration,[41] one must assume either that the fortis/lenis opposition is inoperative, or that it manifests itself in other ways (see below: vowel duration preceding final consonants).

Other investigators tend to view aspiration as a reflex of the voiced/voiceless opposition, although their interpretation of the opposition is different from that described above. This interpretation will be dealt with at length in a later section of this study.

Much the same line of argument that has been applied to the cue value of the aspiration of stops may also be applied to the burst-release of /ptk/.[42] As with aspiration, some linguists tend to view burst pressure or intensity as a reflection of fortisness or lenisness. Halle, Hughes and Radley, making acoustic measurements of stop consonants, noted that burst pressure is the essential cue that separates /ptk/ from /bdg/:[43] "lax stops . . . show a significant drop in level in the high frequencies [of the burst]. This high-frequency loss is a consequence of the lower pressure associated with the production of lax stops and is therefore a crucial cue for this class of stops."[44] However, Halle et al. found that using an electronic gating device to remove the bursts of final-position stop consonants[45] caused little confusion between members of cognate pairs, "because of the very marked difference in duration of vowels before tense and before lax stops. . . ."[46,47] There was, however, some significant confusion in judging syllables originally ending in /ptk/ which were thought to be "open" (i.e. unchecked) when heard with the burst removed.[48] This would seem to indicate that the burst is a cue for the presence or absence of /ptk/, but not necessarily a cue opposing those stops to their cognates.

Malécot and Wang report similar results in their studies of plosives in final position. Malécot states that "suppression of the releases leaves the voiced stops practically unaltered, but results in a drastic reduction of the manner [i.e. voicing] identification of the voiceless stops."[49] Wang explains that since American speakers tend to release final /ptk/ more frequently than final /bdg/, the releases of the former set are more important cues for identification than are the releases for the latter set.[50]

Both Wang and Malécot found that replacing the releases of final /ptk/ with those of /bdg/ generally tended to shift the perception of the final consonants from the former to the latter set, although there was no significant shift when final /bdg/ releases were replaced with those of /ptk/. These results were explained by the fact that there was voicing during half the closure for both classes of stops. Thus, for example, when a /p/ release was substituted for a /b/ release, the voicing during closure cued the original perception as /b/, whereas a /b/ release substituted for a /p/ release merely reinforced the perception of /b/ cued by the voicing during closure.[51,52] This is further evidence of the relatively greater importance of the release of /ptk/ as a cue for identification, as opposed to the release of /bdg/.

As in the case of aspiration, it must be noted that burst-release is not a universally present reflex of a fortis/lenis opposition, if, in fact, it _can_ be ascribed to differences in breath pressure and articulatory pressure. That is, there are cases where stops are unreleased in English and in which, therefore, no differentiation is possible on the basis of the acoustic nature of the burst. It is interesting to note that these unreleased allophones occur in phonetic contexts where Lisker found that there is a pressure difference between the two classes of stops (see above, p. 7).[53] Further, when stops in these contexts are released, it does not appear that the pressure difference always cues a distinction between them.

Another phonetic feature, studied for both stops and fricatives, is duration. In the case of stops the duration measured is that of the closure or 'hold' period. For fricatives, the duration of the friction in the acoustic signal is considered relevant. Both of these types of duration have generally been found to be relatively longer for the /p,t,k,f,θ,s,ʃ, / series than for their cognates, in those contexts where differences are found.[54]

Lisker, using varying lengths of blank tape to simulate the closure portion of intervocalic bilabial and velar stops, found that "there exists a closure duration, perhaps about 75 msecs for trochees in isolation, such that all stops of a briefer duration are heard as voiced by American subjects; and that there is another closure duration, 130 msecs . . . such that stops of this duration and longer are assigned to the voiceless category."[55]

In a similar study, using spectrographic analysis, Sharf found a significant difference between the closure durations for /p/ and /b/, and for /k/ and /g/ in intervocalic post-stress position. No such difference, however, was found between /t/ and /d/. This was expected, since the /t/-/d/ opposition is effectively neutralized in medial post-stress position by many, if not most, speakers of American English in such words as <u>latter</u>-<u>ladder</u>.[56]

These results received further confirmation in a later study in which Lisker measured closure duration by noting the duration of the rise and fall of intraoral pressure assumed to accompany stop closure. The data were virtually identical with those derived from spectrographic measurement. Lisker found, however, that there was only a negligible difference in closure duration between the two categories of stops in other than intervocalic post-stress position.[57,58,59] He does, however, note another type of difference between the two classes of stops: The closure duration of intervocalic /bdg/ is much more sensitive to variation in stress on the preceding and following syllables than is that of /ptk/. In fact, the difference between /bdg/ closure duration following stressed vs. unstressed syllables is as significant as "the difference between voiced and voiceless categories in the one position in which it is found."[60]

In a later experiment, employing spectrographic measurement, Lisker found the relation of stop closure duration of /ptk/ to /bdg/ to be the reverse of that for post-stress intervocalic position: /bdg/, separated from a preceding /s/ by a word boundary as in It's Bill's (/s # b,d,g/), evidence a longer closure duration than do /ptk/ in a similar context such as It's pills (/s # p,t,k/). The "neutralized" stop in expressions of the type It's still (/s # sp,st,sk/) and It spills (/ # sp,st,sk/) evidence closure durations that do not seem to differ significantly from the closure durations found for stops in contexts such as It's pills (/s # p,t,k/).[61]

Lisker concludes from his investigations that stop closure duration is unreliable as an indication of the fortis/lenis opposition because it fails to separate the categories of stops in a consistent manner. The fortis/lenis opposition must thus be rejected as a useful theory if it is maintained that "the two stop categories differ in force of articulation only where they differ in duration."[62]

Investigation of the duration of friction for

/f,θ,s,ʃ/ indicates that it is relatively greater than is the duration for /v,ð,z,ʒ/.[63] The perceptual significance of friction duration has been demonstrated by Denes in an experiment on the words (the) *use* /jus/ and (to) *use* /juz/. By interchanging the final consonants and adjusting their durations so that /s/ was shortened by two-thirds and /z/ lengthened by a factor of three, Denes caused listeners to reverse their judgments of the originally recorded words: /jus/ was heard as /juz/, and vice-versa.[64] In the same experiment, varying durations of /s/ friction were coupled with a preceding vowel of varying duration. The results of listeners' judgments indicated that ". . . the duration of the vowels and of the final consonants have a definite and consistent influence on the perception of 'voicing.'" The durations which were varied were found to be interdependent, with the effect that "the perception of 'voicing' of the final consonant increases as the ratio of the durations of final consonant to preceding vowel decreases."[65]

Two experiments by Sholes, using synthesized /s/, in final position, produced results similar to Denes'. In the first study, preceding vowel (syllable) duration and friction duration were co-varied along with formant tapering and total voice duration. Friction duration was found to be the most important of the variables tested, although preceding vowel (syllable) duration also proved crucial as a cue for most of the listeners who judged the final /s/ as /z/: ". . . shorter total length [of the preceding vowel-syllable] tends to disfavor /z/, whereas longer, or at least medium total length tends to favor it. In this, total length works at odds with friction length, which favors /z/ when shorter, but rules it out when longer."[66] Further analysis of the data indicated that a constant friction duration of 110 msec was most appropriate for /z/, whereas the most unanimous judgments of /s/ were "for some (sufficiently long) value of friction length."[67] This finding contrasts with Denes' that /s/-/z/ identification depends on the "ratio of friction and vowel durations." However, Denes' use of very short syllables in which the optimal friction duration for /z/ (as determined by Sholes) was not always reached, may explain the discrepancy. In any event, both authors clearly indicate that "non-spectral characteristics, such as durational relations, play a by-no-means insignificant role in our perception" of speech sounds as 'voiced' or 'voiceless.'[68]

Both Denes and Sholes comment on the cue value of the intensity ratio between the preceding vowel and the consonant. Denes notes that "the effect of intensity is greatest . . . where the duration ratios [of vowel to consonant] change over from the 'voiced' to the 'unvoiced' values. . . . [T]he final consonant is more likely to be heard as a 'voiced' sound as its intensity becomes less and less as compared with the vowel."[69]

Sholes, by synthesizing vowel formants with varying degrees of tapering, was able to adjust the intensity ratios between final consonant and preceding vowel. He found that two of the four subjects required a tapered formant in order to hear /z/, and that none required a tapered formant to hear /s/.[70] The first of these results, for /z/, seems, in part, to contradict Denes' findings. That is, tapering the vowel formant tends to lower the intensity of the vowel relative to the consonant. This, according to Denes, should make /s/ a more likely response, rather than /z/.

An oft-noted fact of (American) English pronunciation is that vowels preceding consonants of the class /b,d,g,v,ð,z,ʒ/ are of greater duration than those preceding the class /p,t,k,f,θ,s,ʃ/.[71] Frequently the difference is associated with or ascribed to the voiced/voiceless opposition: "The same vowel if stressed, is longer when final or before a voiced consonant than it is before a voiceless consonant."[72] The fortis/lenis classification of the consonant is also offered as an alternative basis for the systematic variation in vowel duration: ". . . the voicing characteristic (or the effort) associated with . . . the following consonant . . . changes the average duration of vowels markedly."[73] Differences in vowel duration have been most closely studied for stops in final position in isolated words.[74] Absolute measurements of duration vary greatly from vowel to vowel, since intrinsically shorter vowels such as /I,ɛ,ə,U/ when lengthened before a consonant of the class /bdg/ will not be as long as intrinsically longer vowels such as /i,e^{I},æ,u/ in the same context.[75] One must thus view absolute values of duration as fully significant only with relation to a single vowel in a single phonetic environment, especially since a vowel such as /i/ may, in the context /b_t/ be shorter than a vowel such as /I/ in the context /b_d/. Per cent differences, ratios, and ranges of absolute values will thus be more revealing than any single absolute value or mean of absolute values. Table I-2 presents some of the measurements and estimations of vowel duration reported in the literature. Figures are presented as originally stated by the writers with the addition of ratios, per cent differences and ranges of variation as determined, when possible, by this writer from the data. All durations have been converted to milliseconds for the purposes of comparison.

TABLE I-2

COMPARISON OF DATA FOR VARIATION IN VOWEL DURATION BEFORE 'VOICED' AND BEFORE 'VOICELESS' STOPS (ALL DURATIONAL VALUES GIVEN IN MILLESECONDS)

	VOWEL DURATION BEFORE ALL STOPS						RANGES OF VOWEL DURATION BY PLACE OF ARTICULATION OF FOLLOWING STOP					
Author	% Difference	Ratio	Range /ptk/	/bdg/	Average /ptk/	/bdg/	/p/	/b/	/t/	/d/	/k/	/g/
Heffner[76]	20	4:5	140-270	190-330			140-250	190-280	150-270	200-330	140-250	270-300
Rositzke[77]	50	1:2										
Peterson[78]	33	2:3	138-243	168-318	197	297	138-188	203-307	147-210	206-318	145-200	243-314
Stetson[79]	60	1:2.3							60 (Avg.)	140 (Avg.)		
Gimson[80]	50	1:2										
Belasco[81]	40	3:5	140-270	260-290								

Results similar to those recorded in Table I-2 have been found for vowel duration preceding medial stops. Lisker found an average difference of 25 msec between the shorter vowel before /p/ and the longer vowel before /b/ for trochees of the type staple-stable, rupee-ruby.[82] Lorge notes the difference in vowel duration before intervocalic post-stress /t/ and /d/,[83] as does Sharf, whose spectrographic measurements reveal a 9 msec difference. This contrasts with the 30 msec difference found before intervocalic post-stress labial and velar stops.[84] This discrepancy might well be expected, however, since, as has been noted above, the /t/-/d/ opposition in words such as bitter-bidder, latter-ladder and metal-medal is often neutralized in American English.[85] It is clear, therefore, that the difference in vowel duration before final stops is preserved in the case of intervocalic[86] post-stress stops, although it is greatly diminished, especially before alveolar /t/ and /d/.[87,88]

Vowels preceding fricatives show a similar variation in duration. Peterson and Lehiste found an average vowel duration (for intrinsically long and short vowels taken together) of 228 msec before fricatives of the class /f,θ,s,ʃ/, and of 376 msec before fricatives such as /v,ð,z,ʒ/.[89] The corresponding ranges of variation were 192-278 msec and 231-410 msec for an approximate ratio of 2:3 and a per cent difference of 33.[90]

Although the differences in vowel duration recorded above have been noted for at least a century,[91] it is only recently that their cue value has been investigated. Daniel Jones notes that, "In some varieties of English it would appear that no difference is made between the sounds of final d and t, both being pronounced d̥. In the speech of those who pronounce in this way words like heed and heat . . . are distinguished solely by the length of the vowel. H. J. Uldall . . . considers this pronunciation common in America, but the point does not seem to have received much attention from American phoneticians."[92] Lisker and Abramson concur, adding that "The single differentiating feature [for stops in final position] that all descriptions agree is regularly present is the greater length of vowels followed by /bdg/, but this is usually treated as a matter of vowel allophones."[93]

The view of the difference in vowel duration before consonants of different classes as non-distinctive is put forth by several American phoneticians. Thomas states that "Some of these differences are, of course, so slight that we usually disregard them."[94] Wise comments that "Since changes of length . . . take place in the utterance of English-speaking people automatically and without significance, it is of little value to the layman even to know of the changes. . . ."[95] Heffner and Locke, writing specifically of the durational difference of vowels before final /t-d/, doubt if it is long enough to be perceptually significant: "This difference is of the order of .04 to .06

seconds, which is 'perceptually' very short indeed. . . ."[96]

It is not unreasonable, however, to assume that the duration of a preceding sound might have perceptual significance as a cue to the nature of the sound following it. As Lisker suggests, there may be "a one-many relationship [which] exists between linguistic segments and acoustic elements specified by a subset of the features contributing to speech perception." Thus "each acoustic segment might be said to supply cues to a single linguistic segment, while any 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." In sum, "a single linguistic segment may be identified on the basis of cues contained in more than one acoustic segment, and . . . a single acoustic segment may provide information for the identification of more than one linguistic segment."[97]

Some recent work, often based on experimental evidence derived from newly developed and improved measuring devices, tends to ascribe relatively great perceptual significance to vowel duration. Jakobson and Halle suggest that "The relative duration of the consonant and the antecedent phoneme may remain for certain contextual or optional variants of tense and lax consonants the chief or even the only cue to their distinction."[98] Pilch notes that the durational difference in vowels "before voiced rather than before voiceless stops . . . has . . . in certain dialects, acquired phonemic significance by virtue of being maintained even when /d/ and /t/ coalesced under [ɾ] between stressed and unstressed vowels as in *latter* /læɾɚ/ ≠ *ladder*, /læːɾɚ/, and similarly in *writer* . . . *rider*, *pouter* . . . *powder*. . . ."[99] Lorge's study revealed that Ss, when identifying their own and other Ss' recordings of potentially minimal pairs with intervocalic post-stressed /t/ or /d/ (e.g. *latter*-*ladder*), tended to make more errors "on the words containing short vowels than on those containing long ones. . . . This imbalance might possibly have occurred because the distinction between /t/ and /d/, when lost in the intervocalic stop may be preserved in the length of the preceding stressed vowel, and diphthongs allow for greater variation in length than do short vowels."[100]

Denes' experiment, noted above (see pp. 18-19), indicated that vowel duration can be a sufficient cue to the identification of a word-final fricative as /s/ or as /z/: [s] preceded by a sufficiently long vowel was identified as /z/.[101]

Related experiments, employing only synthetic speech, and attempting to synthesize final /z/ using hiss (aperiodic vibrations), revealed that vowel duration or total syllable duration can be a significant and sometimes essential cue, although not as important as friction duration.[102,103]

Other experiments, on synthesized final stops, indicated that a shorter vowel duration (225 msec) enhanced the perception of /ptk/, while /bdg/ were more readily identified when the preceding vowel was lengthened (to 300 msec).[104] More specifically, "A syllable having a vowel duration of 200 msec. was so synthesized that it could be heard equally well as /ɛk/ or /ɛg/ (Previously identified voicing cues . . . were . . . at neutral values). The duration of the /ɛ/ portion was then extended and contracted in steps of 50 msec., so that the final range of vowel durations to be tested was from 100 to 400 msec." Vowels longer than 200 msec produced listener judgments of /ɛg/; vowels shorter than 200 msec caused listeners to hear /ɛk/.[105]

In a study employing tape-cutting and filtering techniques, Noll found that shortening the vowels before 'voiced' fricatives, although it did not cause a change in perception, did result "in a slight impression of voicelessness," when the voice bar was simultaneously filtered out.[106]

In testing final "voiced" fricatives, Noll found that "The perception of post-vocalic /ʃ/ and /s/ as voiced consonants seems to be especially affected by increasing the duration of the preceding vowel, particularly if this is combined with an addition of a voice bar. . . . The post-vocalic /f/ and /θ/, on the other hand, do not seem to be nearly so much affected by the vowel duration factor. Thus vowel duration would seem to have some effect on the perception of voicing in post-vocalic consonants [fricatives], but the evidence of this study would indicate that this effect varies with different phonemes and is secondary as compared to the presence or absence of a periodic component in the consonant spectrum."[107]

Estimates of the perceptual cue value of the duration of vowels in re the sounds that follow them bear a close relationship to the question as to whether these durational differences are physiologically conditioned or learned. If the differences were found to be a function of physiology, that is, if vowels must be longer before voiced sounds than before voiceless sounds because of factors inherent in the human vocal apparatus and its control mechanisms, then one might confidently expect that vowel duration would turn out to be a highly significant perceptual cue, since a vowel of a given duration would <u>always</u> be associated with consonants of a given class.[108] On the other hand, physiological conditioning is not crucial to the importance of vowel duration as a cue in any given language: Theoretically, durational differences could be learned and just as consistently associated with different classes of consonants as if they had been physiologically conditioned.

Perhaps the most popular argument for the case of physiological conditioning is that identical relationships between vowel duration and the nature of the following consonant have been found in a number of languages: e.g. English, Norwegian, Icelandic, German, French, Italian, and Hungarian. "The existence of the lengthening effect . . . in so many languages supports the assumption that it has a physiological background."[109]

Peterson and Lehiste, however, on the basis of cross-linguistic investigation, maintain that the conditioning of vowel duration in monosyllables is a particular feature of English and not a "general principle applicable to all languages induced by physiological factors."[110] House's rejection of physiological conditioning is based on the idea that whereas vowels preceding fortis sounds should have a greater duration than those preceding lenis sounds, the opposite obtains in English.[111]

Other writers, however, disagree sharply with House's position. Delattre points out, contrary to House's estimate of the fortis/lenis effect on vowel duration, that "We have always made the opposite hypothesis, namely that the <u>anticipation</u> of greater effort would make one shorten the vowel more in <u>bit</u> than in <u>bid</u>."[112]

Despite some experimentally-based, though ambiguous and frequently unreliable evidence presented by Stetson,[113] as well as some speculation concerning neuro-physiology by Heffner and Locke,[114] and by House and Fairbanks,[115] there is very little in the way of empirical or theoretical evidence relating to the physiological events underlying the phenomenon of vowel lengthening in English. Further theorizing and speculation should await such evidence.

Timing Relationships: Articulation and Glottal States

Among the most recent theories explaining the perceived differences between members of cognate phonemes is that proposed by Abramson and Lisker. This theory seeks the generality of the fortis/lenis and of the voiced/voiceless oppositions. In fact, one may easily view the theory as an extension and adaptation of the voiced/voiceless opposition which was presented at the beginning of this chapter. Using an idea presented by Fant, Lisker and Abramson have suggested that differences between cognates may be explained in terms of the timing relationships between activity at the glottis and supraglottal articulatory movements.[116] Since the glottis is one of the primary foci of attention in this theory, the relation to earlier concepts of voicing is apparent.

The development of the new voicing hypothesis derives in part from the early work done in speech synthesis at Haskins Laboratories. The presence or absence of voicing, originally assumed to be the primary cue differentiating members of stop cognate pairs,[117] was shown, in a series of experiments, to be of less importance than the nature of the first formant transition and "time interval between the burst and transition."[118] These latter cues are both evident from spectrographic analysis of real speech. Gimson notes that "there is likely to be a marked rising bend of F1 of the adjacent vowel in the case of /b,d,g/ which is not as marked in the case of /p,t,k/."[119] Fischer-Jørgenson states that the beginning of the [vowel] transition after /ptk/ is concealed by "the voiceless pause [containing aspiration] between the explosion and the vowel. . . ."[120]

Delattre, reviewing the findings of the Haskins group, indicates that it is the presence of the first-formant transition which "seems to contribute very strongly to the presence of voicing and its absence [along with aspiration in the second- and third-formant transitions] to voicelessness."[121]

Since the transition of F1 is a significant cue to voicing, the cutback (or removal) of F1 may be viewed as a (further)[122] delay in the onset of voicing. Thus it is possible to "define the amount or degree of voicing of a stop [in initial position] as the duration of the time interval by which the onset of periodic pulsing either precedes or follows release."[123] If the two classes of stops display distinctly different ranges of values of voicing onset time (hereafter, VOT) it is then possible, in a sense, to reinterpret the difference between them as one of voicing. Further, the presence or absence of aspiration, often taken as a sign of the fortis/lenis opposition, and "coordinate with voicing, is then regarded simply as the automatic concomitant of a large delay in voice onset."[124]

Abercrombie explains this last point in Elements of General Phonetics: "Finally we must consider the coordination of the states of the glottis with the different phases of a stop. . . . If voicing does not set in until after phase iii [release] is completed, the plosion itself will be voiceless, and the stop is said to be voiceless aspirated. If, however, voicing starts simultaneously with the beginning of phase iii, then the plosion is voiced and the stop is termed voiceless unaspirated. . . . A voiced stop, a voicless unaspirated stop and a voiceless aspirated stop differ, then, according to the point at which voicing sets in. . . ."[125]

Lisker and Abramson, in a cross-language study using spectrographic analysis, sought to determine if, in fact, the VOT measure would serve to distinguish classes of stops in initial position. Their data for four English speakers, speaking isolated words, revealed that in every case of initial /bdg/ there was an earlier onset of voicing of the following vowel than there was in the case of the cognate class /ptk./[126]

The authors conclude that their findings can be explained by "certain inferences about glottal behavior. . . . It is to be understood that this is not merely the obvious question of whether or not there is laryngeal vibration. Instead, it seems evident that a fairly complicated acoustic output is dependent upon the relatively simple matter of varying the area of the glottis. If the speaker closes the glottis down enough for phonation, he does not directly 'command' the vocal folds to vibrate; rather, he makes the necessary muscular adjustments that set the conditions for vibration when sufficient airflow is supplied."[127] In sum, Lisker and Abramson propose that "glottal mechanisms" are the basis for explaining most, if not all, of the phonetic features which distinguish stop classes in various phonetic contexts.

A later study sought to determine what effect running speech (context) would have on VOT in stops. In general, it was found that VOT values were reduced in running speech as compared with the pronunciation of isolated words. Another effect was that the VOT values for members of cognate pairs were found to overlap in context.[128] Lisker and Abramson maintain that the failure of the VOT measure to separate the two stop categories in context means either that there is "some general 'blurring' of distinctiveness that may be inevitable at . . . higher articulation rates . . . where linguistic constraints will take up any 'slack' in intelligibility . . ." or that there are "other acoustic features, whose contribution to intelligibility may be redundant for deliberate speech but indispensable at higher rates of transmission."[129]

In seeking to explain the cause of the overlap in VOT values for cognate sounds, the authors point to several contextual factors and experimental artifacts. One factor contributing to the overlap is stress. The VOT values for /ptk/ are greater when these sounds appear before stressed vowels than before unstressed vowels. The reverse is sometimes true for /bdg/, where values are slightly greater before unstressed, rather than stressed, vowels, in those cases where there is no voicing lead.[130]

A complicating factor in measuring VOT, especially for /bdg/ before unstressed vowels, is that in a voiced context, vocal pulsing usually continues from the preceding sound through the stop closure.[131] Thus overlap between stop categories appears to be greater than it is, since only very few "occurrences of /bdg/ . . . could be measured for VOT," and most of them overlapped with /ptk/ values. If, however, these cases of continuous voicing are taken as negative VOT, then there is little overlap between stop categories.[132]

The /ptk/ series also provides measurement difficulties before unstressed vowels. These sounds, in a context of voicing, often have audible "edge vibrations" of the vocal folds during the articulatory closure period. Lisker and Abramson view edge vibrations as the continuing "laryngeal oscillations of a preceding voiced environment," which occur "for a while even after the glottis has begun to open for a voiceless stop."[133] These vibrations, however, drop "from 2 to 20 db. below auditory threshold" after the release of the stop.[134] Since these cases of audibly discontinuous voicing are not taken into account when one measures VOT from the stop release, the overlap between categories is again made to appear greater than it might otherwise seem.[135]

Taking these factors into consideration, the authors conclude that there does seem to be "a very close match with respect to the relations between stress and VOT,"[136] and that when the effects of stress are taken into account it can still be maintained that "/ptk/ and /bdg/ are characterized by significantly different distributions of VOT values."[137] In isolated words before stressed or unstressed vowels, VOT completely separates the two stop categories. In words in sentences, although there is an overlap in VOT values, especially before unstressed vowels, "not only are the mean VOT values greater for /ptk/ than for /bdg/ in the same context, but . . . no contextual variant of /ptk/ has a mean value as small as that of any variant of /bdg/."[138]

The fact that there is an overlap, however, indicates that there is possibly "no single acoustic feature by which the category membership of any given stop may be determined independently of context. Instead, there is reason to believe that /ptk/ are invariably characterized by a gesture of laryngeal opening, and that such a gesture does not accompany the articulations of /bdg/."[139]

The inference concerning the states of the glottis during stop production is based, in part, on early transillumination experiments done at Haskins Laboratories. Although the data for these experiments is somewhat restricted with regard to number of subjects (only one) and utterances analyzed, there were examples of "all the English stop and fricative consonants" in initial and medial positions.[140] The experimenters found that "in the case of the stops, the voiced and voiceless categories are almost perfectly sorted on the basis of presence versus absence of glottal opening or interruption of pulsing." As in the context experiment, the authors found that /ptk/, especially before an unstressed vowel, occasionally showed either no glottal opening during articulatory closure, and/or no break in phonation.[141] Before stressed vowels /ptk/ was produced 96% of the time with an open glottis; before unstressed vowels, 84% of the time. /bdg/, on the other hand, was produced 94% of the time with the glottis closed.[142]

Other experimental findings support the correlation between VOT and stop classes, as well as inferences about glottal states. Lisker's study of stops after initial /s/ (see above, p. 17) indicated that VOT "sharply" separated utterances of the type "It's pills" (/s#p,t,k/) from those of the types "It spills" (/#sp,st,sk/), "It's skill" (/s#sp,st,sk/), and "It's Bill's" (/s#b,d,g/): The "mean value of 55 msec [for "It's pills"] is appropriate for stops of the kind called voiceless and aspirated."[143] If, of course, VOT were crucial to perception, by virtue of the fact that it conditioned the presence or absence (or degree) of aspiration after the stop release, we would expect from these results that the stops in initial /sp, st,sk/ combinations, not having appropriate VOT values for the set /ptk/, should be perceived as /bdg/. The two independent tape-cutting experiments cited above (see pp. 11-12) confirm this hypothesis. That is, when the /s/-friction is removed from before the stops in words such as <u>spill</u>, <u>still</u>, and <u>skill</u>, they are perceived as <u>bill</u>, <u>dill</u>

and <u>gill</u> by speakers of English.[144,145]

Electromyographic evidence also lends support to the effectiveness of the VOT measure in separating stop categories. Although no differences in muscular tension could be found between initial and medial /p/ and /b/, experimenters felt that the "difference between the bilabial stops can be accounted for on the basis of the behavior of the glottis. During the series recorded here [e.g. /əpik, əbik, pikə, bikə, dæpl, dæmbl, dæmpl,/ etc.] the vocal cords were always vibrating at the time the stop occlusion was released for voiced sounds and never were vibrating at the time of release of voiceless sounds."[146]

Simiarly, Fromkin found an average difference of 45 msec between the peak of muscular activity for initial /b/ and /p/ and the onset of voicing, the shorter duration being between /b/ and the following vowel.[147] For /p/ and /b/ in final position, Fromkin's data indicate that the offset of voicing occurs simultaneously with labial closure for final /p/, but continues for some time during the closure for final /b/.[148]

Experiments employing synthetic speech and designed to confirm the perceptual significance of VOT have been performed. Abramson and Lisker synthesized a vowel preceded by an alveolar stop. The VOT was varied from a lead of 150 msec to a lag of 150, as measured from the release. The synthesized stimuli consisted of a burst, preceded by a voicing bar of varying length in those cases where there was voicing lead, or silence when there was no lead, followed by an F1 with a 50 msec transition that was replaced by hiss as it was cut back in the series of stimuli with voicing lag. Attempts were made to keep the burst-hiss intensities at a ratio equally appropriate to both /t/ and /d/.[149] When the stimuli were heard by speakers of English it was found that the phoneme boundary between initial /t/ and /d/ was at +25 msec, a value which "coincides neatly with the boundary between the two non-overlapping ranges of VOT values for productions" which had been determined in an earlier experiment.[150]

Lorge, using the same synthetic stimuli described above, in another experiment, sought to compare perception with production for specific subjects. Her fifteen subjects' data gave a /t/-/d/ phoneme boundary "between 30 and 40 milliseconds of voicing lag," slightly longer than that found by Abramson and Lisker. The phoneme boundary was sharp, "/t/ responses going from 33% at 30 msec. [of voicing lag] to 89% at 40 msec."[151] Four subjects were then chosen, two with "100% consistency in identifying sounds up to a certain point [specifically between 30 and 40 msec of voicing lag] as /d/ and beyond that point as /t/," and two whose responses were less consistent and whose phoneme boundaries were different from the group mean.[152] The subjects chosen produced initial /t/ and /d/ in monosyllabic words, and their productions were analyzed. It was found that the subjects whose perceptual phoneme boundaries were

100% consistent produced utterances in which VOT values clearly fell into two categories on either side of the perceptual boundary. The subjects with less consistent perceptual responses produced utterances in which VOT values did not fall consistently on the appropriate side of phoneme boundaries as determined for the group of 15 subjects. These subjects tended to produce some /t/'s having VOT values identical with those of the synthetic stimuli which they heard as /d/.[153]

There has been very little experimental work done on the applicability of VOT and of glottal states to fricatives. In general it may be noted that the presence or absence of a voicing bar and of the first formant transition have perceptual significance for fricatives as for stops, and have often been used as cues in speech synthesis. Delattre found, however, that the most effective cue in synthesizing voiced fricatives was the continuation of low intensity formants through the articulatory closure period.[154]

Kanter and West present a theory of aspiration in final fricatives which bears a distinct similarity to theories based on VOT and glottal states. They point out that in "English the transition from a vowel sound to a surd fricative is usually accomplished by opening the glottis at the instant at which the tongue quits the vowel position and begins to move in the direction of the position it will occupy for the fricative. Thus in the interval between the cessation of the voiced tone and the beginning of the fricative sound air is streaming out without resistance [i.e. aspiration is present]. . . . If the timing at the glottis is altered in such a way that the voice continues until the fricative sound starts, there is a typical unaspirated surd fricative, not standard in English. . . . If one continues the voice on into the fricative, he affects the voiced fricative usual in English. . . ."[155]

MacNeilage, however, found that for /f/ in absolute final position, or as the first member of a final cluster, the cessation of voicing corresponded with the onset of friction,[156] a state of affairs which Kanter and West, as noted above, feel should produce a non-standard English sound.[157] MacNeilage presents no data for final /v/ or any other fricatives of either category and thus no comparison between classes is possible.

The early transillumination experiments indicate that both classes of fricatives are produced with an open glottis, and differentiation between classes depends mainly upon whether or not the vocal folds are in vibration.[158] It is pointed out that the "high incidence of glottal opening for the voiced fricatives as compared with the voiced stops is not unreasonable if we suppose that a well-formed fricative requires audible turbulence, and that the airflow needed for this is most easily supplied when the glottis is partially open. . . . [I]t is almost exclusively the fricatives with higher levels of noise intensity, primarily /z/ and /ž/, that show opening of the glottis as the normal accompaniment of the oral constriction."[159] There is, at present, little significant data available on the timing relationships between the onsets of articulatory activity and glottal states in the case of fricatives.

Conclusion

In a sense, this chapter has come full circle in its attempt to trace the theory and experimentation that seek to explain the physical bases which underlie the perceived difference between two classes of (American) English sounds. The sense in which this is true is that glottal activity is again being considered by several investigators as the dominant factor contributing to the opposition which has been variously labelled as fortis/lenis, voiced/voiceless, surd/sonant, etc. The search for a general explanation, of whatever form, clearly reveals that although such a theory may be simply stated, its manifestation in human behavior is indeed complex: It must account for a large number of grossly different acoustic events which are in themselves audible cues to the opposition in different phonetic environments. That is, the general underlying difference, which may not itself be "audible" in the usual sense of the word (e.g. one cannot "hear" timing relationships per se between glottal states and articulation), must be shown to give rise to those numerous acoustic features which are perceived by listeners.

Nor is it abundantly clear that such a general explanation will be found, or at least that it can be verified. It does seem, at this time, that simple notions about presence or absence of voicing, or aspiration, and differences in articulatory tension or breath pressure have been discredited as general theories. Yet such differences do exist and certainly bear perceptual significance in one phonetic context or another. Whether or not such differences may, in turn, be related to differences between glottal states and their timing in relation to articulatory events remains, for the most part, to be proven. Initial experimental work is promising in this regard, but it may well be that phoneticians and linguists will have eventually to fall back on a selection from a large set of acoustic-articulatory cues to explain the opposition of /p,t,k,f, θ,s,ʃ/ to /b,d,g,v,ð z,ʒ/ in any given context in which such an opposition may appear. Such cues have been summarized in the following table (I-3).

TABLE I-3

SUMMARY OF ACOUSTIC AND ARTICULATORY FEATURES ASSOCIATED WITH OPPOSING MEMBERS OF COGNATE CLASSES OF STOPS AND FRICATIVES

Acoustic Articulatory Feature	Prevocalic	Inter-, Postvocalic Pre-Stress	Inter-, Postvocalic Post-Stress	Post-Vocalic, Word-Final
Vocal Pulsing	Traditional: Present for /bdgvðzʒ/. Absent for /ptkfθsʃ/. Often absent in stops during closure stage when preceding a stressed vowel and preceded by silence or a voiceless sound. /ptk/ may have voicing during closure when preceded by voiced sounds.	Intervocalic /vðzʒ/, especially /zʒ/, often produced with open glottis, and occasionally with no vocal fold vibration.	Intervocalic /vðzʒ/, especially /zʒ/, often produced with open, vibrating glottis.	Conflicting evidence: 1. /ptkfθsʃ/ usually produced with no vocal pulsing; /bdgvðzʒ/ usually produced with some vocal pulsing. 2. Voicing found during half of closure duration for both /ptk/ and /bdg/.

TABLE I-3..Continued

Acoustic Articulatory Feature	Prevocalic	Inter-, Postvocalic Pre-Stress	Inter-, Postvocalic Post-Stress	Post-Vocalic, Word-Final
Intra-Oral Pressure	Conflicting Evidence: Earlier data: Higher values for /ptk/; Later data: no difference between stop cognates.	Conflicting Evidence: Earlier data: High peak pressure, time integral of instantaneous pressure, for intervocalic /ptkfθsʃ/. Later data: difference only in time integral of instantaneous pressure--higher for /ptk/. No difference found between /ptk/ after /s/ and initial /ptk/.	Higher values for peak pressure, duration of onset of pressure, and time integral of instantaneous pressure for inter-vocalic /ptk/ than for /bdg/.	Values of all measures higher for /ptk/ than for /bdg/.
Articulatory Tension	EMG studies reveal no difference between initial /p/ and /b/.			EMG studies reveal no differences between /p/ and /b/ in terms of muscular activity.

TABLE I-3..Continued

Acoustic Articulatory Feature	Prevocalic	Inter-, Postvocalic Pre-Stress	Inter-, Postvocalic Post-Stress	Post-Vocalic Word-Final
Aspiration	Present for /ptk/; Absent for /bdg/, except before a stressed vowel after /s/.	Present for intervocalic /ptk/; Absent for /bdg/. Absent for /ptk/ after /s/.*		
Burst-Release	Higher intensity for /ptk/; Lower for /bdg/.			More intense burst for /ptk/ than for /bdg/. Burst may be absent for both /ptk/ and /bdg/, but more commonly present for /ptk/, for which it is a cue.* Absence of burst may lead to confusion between identification of cognates.*

*Cue value demonstrated by synthetic experiments, manipulation of real speech, or both.

TABLE I-3..Continued

Acoustic Articulatory Feature	Prevocalic	Inter-, Postvocalic Pre-Stress	Inter-, Postvocalic Post-Stress	Post-Vocalic, Word-Final
Duration	No durational differences found, either spectrographically or in terms of duration of intra-oral pressure rise and fall.	No difference between /ptk/ and /bdg/ after an /s/ in the same word. Longer closure duration for /bdg/ after an /s/ ending a preceding word.	Longer closure duration for /pk/ than for /bg/. Evidence from spectrography and from duration of rise and fall of intra-oral pressure.*	Friction duration greater for /fθsʃ/ than for /vðzʒ/.*
Intensity Ratio Between Vowel and Consonant	Some indication of greater ratio, consonant to vowel, for initial /ptkfθsʃ/.			Conflicting evidence: 1. Lower ratio--vowel to consonant before /s/.* 2. Lower ratio--vowel to consonant before /z/.*

---

*Cue value demonstrated by synthetic experiments, manipulation of real speech, or both.

TABLE I-3..Continued

Acoustic Articulatory Feature	Prevocalic	Inter-, Postvocalic Pre-Stress	Inter-, Postvocalic Post-Stress	Post-Vocalic Word-Final
Vowel Duration			Greater before /bg/ than before /pk/. Greater before /d/ than before /t/, although difference is less than before labials and velars.	Greater before /bdg vðzʒ/ than before /ptkfθsʃ/. Cue value demonstrated before /s-z/.*
F1 transition	Full transition after /bdg/; Attenuated transition after /ptk/.* Same F1 relationship found between fricative cognates.	Full transition after /bdg/; Attenuated transition after /ptk/. Same F1 relationship found between fricative cognates.		Full transition before /bdg/; Attenuated transition before /ptk/.

*Cue value demonstrated by synthetic experiments, manipulation of real speech, or both.

TABLE I-3..Continued

Acoustic Articulatory Feature	Prevocalic	Inter-, Postvocalic Pre-Stress	Inter-, Postvocalic Post-Stress	Post-Vocalic Word-Final
Voice Onset (Offset) Time	VOT lag for /ptk/; 0 VOT or voicing lead (positive VOT) for /bdg/ in initial position in isolated words.* EMG data: Vocal folds vibrating at /bdg/ release; not vibrating at /ptk/ release. Peak of muscular activity closer to voice onset for /b/ than for /p/.	Greater mean VOT values for /ptk/ than for /bdg/, but values overlap in running speech.		EMG Data: Labial closure for /p/ simultaneous with offset of voicing. Offset of voicing after labial closure for /b/.

*Cue value demonstrated by synthetic experiments, manipulation of real speech, or both.

Footnotes - Chapter I

[1] Arthur J. Bronstein, The Pronunciation of American English (New York: Appleton-Century-Crofts, 1960), pp. 70, 83, 85.

[2] The affricates, /tʃ/ and /dʒ/, which are often included among the cognate pairs, have been omitted here because of the lack of agreement as to their classification in English. That is, some linguists hold that the affricates are to be regarded as unit sounds, others that they are sequences of sounds.

[3] It must be added, however, that many of these features are interpreted by linguists as manifestations of the fortis/lenis or of the voiced/voiceless oppositions. The discussion of these features below will thus relate them to their appropriate theories as well as examine them as independent cues.

[4] Claude E. Kantner and Robert West, Phonetics, revised edition (New York: Harper and Brothers, 1960), p. 159. See also Claude M. Wise, Introduction to Phonetics (Englewood Cliffs, New Jersey: Prentice-Hall Inc., 1957), p. 62; David Abercrombie, Elements of General Phonetics (Chicago: Aldine Publishing Company, 1967), p. 27; Roman Jakobson, D. Gunnar Fant, and Morris Halle, Preliminaries to Speech Analysis (Cambridge, Mass.: MIT Press, 1967), p. 26.

[5] The Phonetics of American English, 2nd edition (New York: Ronald Press, 1958), p. 34. Wise recommends the same sort of test, Introduction, p. 64.

[6] The Pronunciation of English, 4th edition (Cambridge, Cambridge University Press, 1966), p. 34. N.B. Although this study deals with American English, there are many features that AE shares with British Received Pronunciation and other dialects. Thus the work of British phoneticians will occasionally be used in these pages when their statements and findings are in agreement with or relevant to American English.

[7] Wise, Introduction, p. 64. Wise notes this shortcoming for the "ear-ringing" test as well.

[8] Leigh Lisker and Arthur S. Abramson, "A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements," Word, 20 (December, 1964), 384.

[9] Leigh Lisker and Arthur S. Abramson, "Some Effects of Context on Voice Onset Time in English Stops," Language and Speech, 10 (January-March, 1967), 8.

[10] Alvin M. Liberman, Pierre C. Delattre, and Franklin S. Cooper, "Some Cues for the Distinction Between Voiced and Voiceless Stops in Initial Position," Language and Speech, 1 (July-September, 1958), 153. Kantner and West hold that fricatives, unlike stops, cannot be distinguished in whispered speech. Phonetics, p. 208.

[11] Leigh Lisker, Arthur S. Abramson, Franklin S. Cooper, and Malcom H. Schvey, "Transillumination of the Larynx in Running Speech," Haskins Laboratories Status Reports on Speech Research, 5-6 (November-June, 1965-6), 4.4. A. C. Gimson indicates that initial and especially final

/vðzʒ/ in British English may be completely voiceless. An Introduction to the Pronunciation of English (London: Edward Arnold, Ltd., 1962), p. 174.

[12] Kantner and West, Phonetics, p. 60.

[13] Preliminaries, p. 26. The terms tense and lax are taken here to be synonymous with fortis and lenis respectively, as they apply to consonants. There are some differences between the interpretation of the former and the latter pairs of terms, but such differences do not seem to be distinctive enough to this writer to alter the essential meanings of statements in which the terms are taken as synonyms. Jakobson, Halle and Fant define tense phonemes in terms of the "greater distinctness and pressure" with which they are articulated, as opposed to the "corresponding Lax phonemes . . . a lower vs. higher air pressure in the cavity behind the main or only source (i.e. . . . behind the point of articulation for the consonants)." "Tenseness and Laxness," reprinted in Jakobson, Halle and Fant, Preliminaries, p. 60.

[14] R. H. Stetson, Motor Phonetics (Amsterdam: North Holland Publishing Company, 1951), p. 50. See also Lee S. Hultzen, "Voiceless Lenis Stops in Prevocalic Clusters," Word, 18 (December, 1962), 308 and passim, and Gimson, Introduction, p. 32.

[15] John W. Black, "The Pressure Component in the Production of Consonants," Journal of Speech and Hearing Disorders, 15 (June, 1950), 208-9.

[16] Motor Phonetics, p. 50.

[17] Andre Malécot, "An Experimental Study of Force of Articulation," Studia Linguistica, 9, No. 1 (1955), 42. It is not quite clear exactly what Malécot's subjects are responding to when they give their impressionistic responses of effort needed to produce sounds. That is, they may be responding to intra-oral pressure differences, and/or articulatory tensions, and/or some other factors associated with the production of the sounds in question.

[18] "The Effectiveness of Intra-Oral Air-Pressure-Pulse Parameters in Distinguishing Between Stop Cognates," Phonetica, 14, No. 2 (1966), 69. It should be noted, however, that Malécot ascribes these pressure differences to variations in glottal posture during the production of the two classes of sounds. Ibid., p. 77.

[19] Ibid., p. 76.

[20] Many of the differences between the results of the earlier and later studies may be explained by the use of more sensitive and sophisticated measuring and recording devices in recent years.

[21] Leigh Lisker, "Supraglottal Air Pressure in the Production of English Stops," Haskins Laboratories Status Reports on Speech Research, 4 (November, 1965), 3.12. Lisker made his measurements with a pressure transducer, attached to a catheter, which was inserted through the nose and suspended in the pharynx above the glottis.

[22] Ibid., p. 3.5

[23] Ibid., p. 3.12.

[24] Ibid.

[25] "Measuring Stop Closure Duration from Intraoral Pressure Records," Haskins Laboratories Status Report on Speech Research, 7-8 (July-December, 1966), 5.3.

[26] "On Hultzén's 'Voiceless Lenis Stops in Prevocalic Clusters,'" Word, 19 (December, 1963), 384.

[27] Ibid.

[28] "Mechanical Pressure as an Index of 'Force of Articulation,'" Phonetica, 14, No. 3 (1966), 178.

[29] Katherine S. Harris, Gloria F. Lysaught, and Malcom M. Schvey, "Some Aspects of the Production of Oral and Nasal Labial Stops," Language and Speech, 8 (July-September, 1965), 146.

[30] Victoria A. Fromkin, "Some Phonetic Specifications of Linguistic Units: An Electromyographic Investigation," Working Papers in Phonetics, UCLA, 3 (1965), 166.

[31] Ibid., pp. 168-69.

[32] Fischer-Jørgensen, in her analysis of Danish stops reaches the same conclusion: "There is no physiological evidence . . . for a stronger tension of ptk. Muscular feeling goes in the opposite direction, particularly for t-d, and palatograms show identity of contact, or slightly more contact for d and g." "Acoustic Analysis of Stop Consonants," Miscellanea Phonetica, 2, Nos. 1-3 (1954), 45.

[33] Bronstein, Pronunciation, pp. 68, 70. Kantner and West, Phonetics, p. 275. C. K. Thomas, Phonetics, p. 157. Jones, Pronunciation, pp. 69-76.

[34] Fischer-Jørgensen notes the same thing for Danish stops. "Acoustic Analysis," p. 44.

[35] J. A. Reeds and W. S-Y. Wang, "The Perception of Stops after s," Phonetica, 6, No. 1 (1961), 80-1. See also John Lotz, Arthur S. Abramson, Louis J. Gerstman, Frances Ingemann, and William J. Nemser, "The Perception of English Stops by Speakers of English, Spanish, Hungarian, and Thai: A Tape-Cutting Experiment," Language and Speech, 3 (April-June, 1960), 77.

[36] "Voiceless Lenis Stops," pp. 310-11.

[37] "On Hultzén's," p. 386.

[38] *Ibid*, pp. 386-87.

[39] Bronstein, Pronunciation, pp. 70, 72, 77.

[40] *Ibid*., pp. 71, 76, 78.

[41] e.g. [bɛt-] /bet/ vs. [bed-] /bed/; (put your) [bɛt- daʊn] /bet daʊn/ vs. [bed- daʊn] /bed daʊn/ (the horses).

[42] Most writers use "release" in final position to refer to both burst and following aspiration in the case of /ptk/. For /bdg/ the term generally refers to the burst plus the embryonic voiced or voiceless (aspirated) vowel transitions following the burst.

[43] "Acoustic Properties of Stop Consonants," Journal of the Acoustical Society of America, 29 (January, 1957), 107.

[44] *Ibid*., p. 108. The experimenters also found the lax /bdg/ bursts to be more intense than the tense /ptk/

bursts: 8 db vs. 5 db, respectively. This finding would seem to be at odds with the air pressure relationships that Halle et al. posit between tense and lax stops. He does not comment on this apparent contradiction.

[45] It should be noted that Halle et al. do not say whether or not the stop closure was also removed by the gating, although their description so suggests. This appears to weaken their conclusions seriously, since they state earlier in the same paper (p. 107) that voicing during closure can distinguish voiced from voiceless stops.

[46] Halle at al., "Acoustic Properties," p. 114.

[47] The experimental findings of Halle et al. and of others discussed below, are taken here as a rejection of the claim made by Rositzke that non-release of final stops, since it "produces no discernable sound, . . . reduces the distinction between [cognate stops] . . . to a point where these sounds . . . are often impossible to discriminate with certainty." "The Articulation of Final Stops in General American Speech," American Speech, 18 (February, 1943), 39.

[48] Halle et al., "Acoustic Properties," p. 115.

[49] "The Role of Releases in the Identification of Released Final Stops," Language, 34 (July-September, 1958), 379.

[50] "Transition and Release as Perceptual Cues for Final Plosives," Journal of Speech and Hearing Research, 2 (March, 1959), 71.

[51] Malécot, "Role of Releases," pp. 379-80. Malécot does not indicate that there was voicing during closure as Wang does. He did, however, cut the closure halfway between its beginning and the release, so if there were voicing during the first part of the closure, that is precisely the part that would have been left.

[52] Wang, "Transition and Release," p. 70. This would seem to indicate that voicing during closure is a more important cue than burst release. Further, it should be noted that Wang does not indicate whether the "release" he is using is simply the burst, or the burst plus aspiration (for final /p/) or low intensity vowel transitions (for final /b/). In either case the results might be expected to be the same. What is not clear, however, from Wang's description, is why, given the importance ascribed to closure voicing as a perceptual cue, the substitution of a /ptk/ release for one of a /bdg/ did not reverse perception. Such a substitution, all other things being equal, should create exactly the cues which Wang describes as being present in the sounds originally perceived as /ptk/. In fact, if closure voicing is as significant as Wang states, one may well wonder why listeners perceived the <u>un</u>altered final /ptk/ set as they did. Two explanations are suggested here. One is that all other things are not equal with respect to closure voicing: i.e. there is a greater duration of voicing during /bdg/ than during /ptk/, although Wang does not state that this is so. The other answer might be that some other cue is

operating here, namely vowel duration. For example, it might be that the long vowel before /bdg/ is a more important cue than the release of the consonant (for as Wang says, release of final /bdg/ is less common than that of /ptk/), whereas the short vowel before /ptk/ is not as significant a cue as the release of these consonants. Thus when the release is changed to one that is inappropriate, the perception changes also.

[53] "Supraglottal Air Pressure," p. 3.12.

[54] Pierre C. Delattre, "Acoustic Cues in Speech: First Report," trans. by George Sholes from the article appearing in Phonetica, 2 (1958), 108-118, 226-251. Translated version distributed in mimeographed form by Haskins Laboratories, pp. 29-30.

[55] "Closure Duration and the Intervocalic Voiced-Voiceless Distinction in English," Language, 33 (January-March, 1957), 49. Gimson notes the same closure duration relations for medial stops in British English. Introduction, p. 149.

[56] "Duration of Post-Stress Intervocalic Stops and Preceding Vowels," Language and Speech, 5 (January-March, 1962), 29.

[57] "Measuring Stop Closure Duration from Intraoral Pressure Records," Haskins Laboratories Status Report on Speech Research, 7-8 (July-December, 1966), 5.3.

[58] Spectrographic measurements made by the writer confirm that in final position there is no consistent

difference in the closure duration for the two sets of stops.

[59] Danish stops, according to Fischer-Jørgensen, do show a difference in closure duration in initial position. This difference is the reverse of the one for intervocalic post-stress stops in English. That is, the closure for /ptk/ is shorter than for /bdg/. "Acoustic Analysis of Stop Consonants," Miscellanea Phonetica, 2, Nos. 1-3 (1954), 45.

[60] "Measuring Stop Closure," p. 5.3.

[61] "The English Stops After /s/ at Word Boundary: A Three-Way Contrast," 11th Final Report on Speech Research and Instrumentation, Haskins Laboratories (March, 1965), pp. 5-7, 9.

[62] Ibid.

[63] John Douglas Noll, "The Perceptual Significance of Certain Acoustical Correlates of Consonant Voicing Contrasts," (unpublished Ph.D. dissertation, University of Iowa, 1960), p. 27. See also Pierre C. Delattre, Alvin M. Liberman, and Franklin S. Cooper, "Formant Transitions and Loci as Acoustic Correlates of Place of Articulation in American Fricatives," Studia Linguistica, 16, No. 2 (1964), p. 109; and Peter Denes, "Effect of Duration on the Perception of Voicing," Journal of the Acoustical Society of America, 27 (July, 1955), 761.

[64] "Effects of Duration," p. 761.

[65] Ibid., p. 763. It is interesting to note that the final /s/ used in the experiment was always heard as /s/ when judged in isolation, regardless of its duration. This

indicates that the contextual cues supplied by the vowel may be of greater significance than Denes indicated in his study.

[66] "Hiss Alone as a Cue for Final /z/ on Voback," *Haskins Laboratories Quarterly Progress Reports*, No. 31 (1959), Appendix 1, pp. A1, 6-7.

[67] "Acoustic Cues for Final /s/ and /z/," *Haskins Laboratories Quarterly Progress Reports*, No. 31 (1959), Appendix 3, p. A3. Sholes does not report any specific friction duration as being most appropriate for /s/. It is understood, however, that the friction durations appropriate for /s/ are greater than the 110 msec. which characterizes the highest percentage of /z/ responses.

[68] *Ibid.*, pp. A3-4.

[69] "Effect of Duration," p. 763.

[70] "Hiss Alone," p. A1-7. See also Louis J. Gerstman, "Cues for Distinguishing Among Fricatives, Affricates, and Stop Consonants," (unpublished Ph.D. dissertation, New York University, 1957). Gerstman concludes that for fricatives and affricates in initial position, the intensity ratio of consonant to vowel is likely to be a significant cue to cognate opposition. He finds that /v ð z ʒ dʒ/ have a lower intensity in relation to the following vowel than do /f θ s ʃ tʃ/.

[71] Pilch notes that durational differences have had a changing significance in American English. Prior to the 19th century "the lengthening of the short vowels as it

occurred before *s*, *f*, *p*, *r* was phonemic, i.e. it meant their transfer to the level of contrasting long phonemes," the oppositions at that time being of the nature /i/ ≠ /i:/. When this phonemic lengthening ceased to be distinctive, that is, when the current qualitative oppositions of the nature /I/ ≠ /i/ came into being, "the lengthening of the short vowels . . . before voiced consonants and voiceless spirants was only of allophonic significance, i.e. it produced no phonemic changes, but merely variant pronunciations of the same phoneme as in . . . [bI·d] vs. [bIt]." H. Pilch, "The Rise of the American English Vowel Pattern," *Word*, 11 (April, 1955), 57.

One should note, however, that this does not exclude the possibility that there were durational variations within the ranges of the phonemically short and long vowels prior to the 19th century. That is, there may have been a long variant of /i/ before "voiced" consonants, but not so long as to overlap with the range of durations of /i:/; and there may have been a short variant of /i:/ before "voiceless" consonants, but not so short as to overlap with the range of durations of /i/. The probability of such durational differences relates directly to the question as to whether or not they are physiologically conditioned (see above, pp. 28-29).

[72] John S. Kenyon, *American Pronunciation*, 10th edition (Ann Arbor, Michigan: George Wahr Publishing Co., 1951), pp. 41, 45, 121-31. See also Arthur S. House, "On

Vowel Duration in English," Journal of the Acoustical Society of America, 33 (September, 1961), 1175; Roe-Merrill S. Heffner, General Phonetics (Madison, Wisconsin: University of Wisconsin Press, 1964), p. 209; Thomas, Phonetics, p. 165; Wise, Introduction, pp. 117-18. Durational differences are also noted for British English: Abercrombie, Elements, p. 81; Gimson, Introduction, pp. 147, 174.

[73] House, "Vowel Duration," p. 1176. See also W. N. Locke and R-M. S. Heffner, "Notes on the Length of Vowels II," American Speech, 15 (February, 1940), 74-75, and for British English, Gimson, Introduction, p. 90. Belasco also ascribes the difference to the force of articulation of the following consonant, although his measurements were made for French utterances: "Influence of Force of Articulation of Consonants on Vowel Duration," Journal of the Acoustical Society of America, 25 (September, 1953), 1016.

[74] See especially the series of articles authored and co-authored by R-M. S. Heffner: "Notes on the Length of Vowels (I-VI)," appearing in American Speech from 1937-1943.

[75] The notion of intrinsically short and intrinsically long vowels has been put forth by many authors, notably by Peterson and Lehiste for American English: "Duration of Syllable Nuclei in English," Journal of the Acoustical Society of America, 32 (June, 1960), 701-03; by Kantner and West, Phonetics, p. 69; and for British English by Daniel Jones, Pronunciation, p. 137; and by Abercrombie, Elements, p. 81. Heffner and Rositzke, on

the other hand, can find no evidence for dividing vowels into a long and a short series. "The most that may be said is that in each of these series the vowels [I], [ɛ], [U], and [ʌ] head the list of brevity. . . . That is to say, the relative shortness of these four vowels is real only within each type of phonetic context and cannot be a characteristic of the vowels only. . . . The length of a vowel in English is not determined by the nature of the vowel but only by its phonetic environment. . . ." Heffner, <u>General Phonetics</u>, p. 210. Heffner bases this conclusion on his studies of vowel duration before /t/ and /d/, in which he found no "absolute sequence of vowel durational values which might be in any way structural." Although two classes of duration emerge from the data before /d/, no such division is possible before /t/. Heffner and Locke, "Notes (II)," p. 78. Similarly, Rositzke points out that "The classification of GA vowels into 'short' and 'long' must accordingly rest upon the natural grouping of the objective measurements into categories more or less clearly set off from each other, if these terms are to have any meaning. It has been pointed out that no such clearly demarcated groups of durations appear in the pronunciation of the five subjects [used in this study]. If, notwithstanding this overlapping, the GA vowels be grouped into categories according to length, the following divisions based mainly on the durations of the vowels before mediae [voiced stops] may be taken as most nearly correct: Long [i u e o æ ɔ ɑ] Short [I U ʌ ɛ]."

"Vowel-Length in General American Speech," *Language*, 15 (April-June, 1939), 107. Peterson-Lehiste, Heffner and Rositzke used General American pronunciation as a source for their data. The differences between their findings may be ascribed to the more precise techniques of spectrographic measurement available to Peterson and Lehiste, which had not yet been developed when Heffner and Rositzke did their studies.

[76] Heffner, *General Phonetics*, pp. 209-10. The data for /p-b/ and /k-g/ are from Heffner, "Notes on the Length of Vowels (IV)," *American Speech*, 16 (October, 1941), 209.

[77] "Vowel-Length," p. 104.

[78] "Duration," p. 702.

[79] *Motor Phonetics*, p. 58.

[80] *Introduction*, p. 90. Gimson, speaking of British English, presents no data, and the ratio presented in the table is only an approximation: "/i:/ in *beat* is only about half as long as the /i:/ of *bee* or *bead*. . . ." *Ibid*.

[81] "Influence," p. 1015. Belasco's measurements are for French stops preceded by the French vowel /ɛ/. The data are included here for purposes of comparison with that of English.

[82] "Closure Duration," p. 45.

[83] "A Study of the Relationship Between Production and Perception of Initial and Intervocalic /t/ and /d/ in Individual English Speaking Adults," *Haskins Laboratories Status Report on Speech Research*, 9 (January-March, 1967), 3.15.

[84] "Duration of Post-Stress," p. 26.

[85] Oswald, for one, maintains that the label "voiced t" often given to the stop in such words as little, latter, etc. is a misnomer: "Whatever the variants of the sound, whether they be well articulated, or trilled or tapped, or merely flipped, they are all voiced sounds. . . . Now there is nothing about such a sound, phonetically speaking which would lead an objective observer to conclude that it is a kind of [t]." Oswald concludes from his own experimental work and from that of others, that "except when a voiceless stop was used to represent t, the listeners heard otherwise similar words spelled with t and d as homonyms." "Voiced /t/--A Misnomer," American Speech, 18 (February, 1943), 21-25.

[86] The term "intervocalic" here is to be taken both literally (e.g. latter-ladder) and to mean between a vowel and a syllabic consonant (e.g. petal-pedal).

[87] Gimson notes that the difference is also preserved when final or medial stops are preceded by other consonants, "notably /l, m, n/ . . . especially when the consonants themselves are preceded by a short vowel. . . ." Introduction, p. 147.

[88] Sharf points out that differences in vowel quality may occur with consistency before /t/ and /d/ in intervocalic post-stress position. He maintains that in General American speech /aI/ is realized as [ɑI] before voiced stops and as [ʌI] before voiceless stops. He then notes that

listeners consistently hear a "voiced *t*" as /d/ when preceded by [ɑI], but show confusion between /t/ and /d/ when it is preceded by [ʌI]. "Distinctiveness of 'Voiced t' Words," *American Speech*, 35 (May, 1960), 107, 109.

[89] "Duration," p. 700.

[90] *Ibid*., p. 702.

[91] Claes-Christian Elert, *Phonologic Studies of Quantity in Swedish* (Uppsala, Sweden: Almqvist and Wiksells, 1964), p. 134. Elert notes Sweet's recognition of "The lengthening effect of a voiced consonant on the preceding vowel" in 1877.

[92] *The Phoneme: Its Nature and Use* (Cambridge, England: W. Heffer and Sons Ltd., 1950), p. 121. Gimson agrees with Jones about the cue value of vowel or syllable duration, stating that final stop and fricative cognates, in which "the voicing factor is not strongly operative," are differentiated largely "by the length of the syllable which they close." *Introduction*, pp. 147, 174.

[93] "A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements," *Word*, 20 (December, 1964), 385n.

[94] *Phonetics*, pp. 165-66. Thomas notes a similar variation in the duration of consonants: "Thus [n] is slightly longer before the voiced [d] of *bend* [bɛn·d] than before the voiceless [t] of *bent* [bɛnt]." As in the case of the vowels, however, Thomas notes that these variations are so slight that they may usually be disregarded." *Ibid*., p. 167.

[95] _Introduction_, pp. 117-18.

[96] "Notes (II)," p. 78.

[97] Leigh Lisker, "Linguistic Segments, Acoustic Segments, and Synthetic Speech," _Language_, 33 (July-September, 1957), 372.

[98] "Tenseness and Laxness," p. 60.

[99] "The Rise of the American English Vowel Pattern," _Word_, 11 (April, 1955), 85-86.

[100] "A Study of the Relationship," p. 3.15.

[101] "Effect of Duration," p. 761.

[102] Sholes, "Acoustic Cues," p. A3-4.

[103] Sholes, "Hiss Alone," p. A1-1.

[104] "Terminal Stops and Nasals," _Haskins Laboratories Quarterly Progress Report_, No. 13, Appendix 2 (1954), pp. A2-2-3. This article and the one in the following reference are unsigned in the _Progress Report_. The staff at Haskins Laboratories has informed the author, however, that the research reported was carried out by George Sholes.

[105] "Stop Consonants in Final Position," _Haskins Laboratories Quarterly Progress Reports_, No. 21, Appendix 5 (1956), pp. A5-1-2.

[106] "Perceptual Significance," pp. 58-62.

[107] _Ibid_., pp. 64-65.

[108] Assuming, of course, that use were made of the class distinction to begin with. That is, in a language where [s] and [z] were allophonic variants of the same phoneme, vowel duration, even if physiologically

conditioned, would be ignored, just as would the other phonetic features which might distinguish one of these phones from the other in a variety of phonetic environments.

[109] Elert, Quantity in Swedish, p. 134. See also Jones, The Phoneme, pp. 177-78, and Pierre C. Delattre, "Some Factors of Vowel Duration and Their Cross-Linguistic Validity," Journal of the Acoustical Society of America, 34 (August, 1962), 1142-43, and S. A. Zimmerman and S. M. Sapon, "Note on Vowel Duration Seen Cross-Linguistically," Journal of the Acoustical Society of America, 30 (February, 1958), 152-53. Zimmerman and Sapon comment that although the difference exists in Spanish, it is too small to be of perceptual significance.

[110] "Duration," p. 693.

[111] "Vowel Duration," p. 1177. Together with Fairbanks, House goes so far as to say that "attempts to interpret the effect of voicing of the consonant upon vowel duration have thus far been fruitless." "The Influence of Consonant Environment upon the Secondary Acoustical Characteristics of Vowels," Journal of the Acoustical Society of America, 25 (January, 1953), 108.

[112] Delattre's argument in terms of anticipation of greater effort is well taken here, in the light of Lisker's finding that all measures of air pressure difference are operant in final position (see above, p. 7). Support for Delattre's contention is found in a relatively early study of vowel duration by Heffner and Locke: "Mr. Locke's

final [t] articulation, when he began his experimental work, was very weak, and normally not exploded. When he saw this on the recordings, and before we began to take measurements, he became conscious of this fact, and as a result, may have expended more energy on the formation of his final [t]'s than was entirely normal for him. It seems possible that this added energy expended on the articulation of the final [t] should have the effect of shortening the preceding vowel slightly." "Notes (II)," pp. 74-75. See also Belasco's "Influence," p. 1016, and Elert, Quantity in Swedish, pp. 134-35.

[113] Motor Phonetics, pp. 35-36. Stetson's work has, of course, been proven unreliable in many respects in recent years. For example, see Philip Lieberman, Intonation, Perception, and Language (Cambridge, Massachusetts: MIT Press, 1967), pp. 191-93. In this case, Stetson's findings indicate that "The duration of the syllable depends somewhat on the force with which it is uttered. Increased force of utterance involves a greater contraction of the positive muscles of expiration and this requires a longer time for the process of the arresting movement. Sometimes this increase in length shows in the vowel, when the negative chest muscles which arrest the movement come into play during the later part of the vowel."

The ambiguity in these findings arises from the problem of where in the syllable the force of utterance is manifested. If in the vowel, and if one posits a total energy for all syllables that approaches a constant, then one might predict the sort of vowel-consonant relationship proposed by Delattre: That is, the more energy expended on the vowel, the less will be available for the consonant. This interpretation, however, is one which would explain the fortis/lenis differences between final consonants in terms of the preceding vowel, rather than one which would explain durational vowel differences as a function of the nature of the following consonants. On this point, see Daniel Jones, *The Phoneme*, p. 118.

If the syllable force is to be determined from the final consonant, then Stetson seems to be saying that the consonant will be lengthened. It appears, however, that only final fricatives evidence durational differences between members of cognate pairs.

Finally, if forcefulness is to be determined for the syllable as a whole, one might expect to find the type of theoretical relationship posited by House. That is, more forceful (long) vowels would be associated with more forceful (fortis) consonants, and inversely.

[114] "Notes on the Length of Vowels (VI)," *American Speech*, 18 (October, 1943), 211. Heffner and Locke suggest reasonably that the innervation initiating a final consonant occurs sooner when that consonant is voiced than when it is voiceless.

[115] "Influence," p. 108.

[116] "Cross-Language Study," p. 422.

[117] Alvin M. Liberman, Pierre C. Delattre, and Franklin S. Cooper, "The Role of Selected Stimulus Variables in the Perception of Unvoiced Stop Consonants," *American Journal of Psychology*, 65 (October, 1952), 499n.

[118] Franklin S. Cooper, Pierre C. Delattre, Alvin M. Liberman, John M. Borst, and Louis J. Gerstman, "Some Experiments on the Perception of Synthetic Speech Sounds," *Journal of the Acoustical Society of America*, 24 (November, 1952), 600; Liberman et al., "Selected Stimulus Variables," p. 499; Pierre C. Delattre, Alvin M. Liberman, and Franklin S. Cooper, "Acoustic Loci and Transitional Cues for Consonants," *Journal of the Acoustical Society of America*, 27 (July, 1955), 773; Alvin M. Liberman, Pierre C. Delattre, Franklin S. Cooper, and Louis J. Gerstman, "The Role of Consonant-Vowel Transitions in the Perception of the Stop and Nasal Consonants," *Psychological Monographs*, 68, No. 8 (1954), 5; Alvin M. Liberman, Pierre C. Delattre, Louis J. Gerstman, and Franklin S. Cooper, "Tempo of Frequency Change as a Cue for Distinguishing Classes of Speech Sounds," *Journal of Experimental Psychology*, 52 (August, 1956), 128. As some of their titles indicate, not all of the above studies deal primarily with the voiced/voiceless distinction. The pages cited are thus those which contain the pertinent information.

[119] *Introduction*, p. 149.

[120] "Acoustic Analysis," p. 45. Although Fischer-Jørgensen's study deals with Danish stops, she notes that

the similarity between Danish and English stops allows for a meaningful comparison of data from the two languages.

[121] "Acoustic Cues," p. 29.

[122] F1 cutback and VOT are not necessarily identical measures. A sound with no F1 cutback may evidence a gap between release and the onset of voicing. For example, a stop burst may be followed by a period of silence before a relatively unattenuated F1 transition. If such a stop is perceived as voiced, cutting back on the F1 transition might change the perception to voiceless. The cutback in such a signal would thus be equivalent to further delay in voicing onset.

[123] Lisker and Abramson, "Cross-Language Study," p. 387.

[124] Ibid.

[125] Elements, p. 148. Abercrombie's "voiceless unaspirated stop" is, of course, usually referred to in English as a 'voiced' stop. See also Kantner and West, Phonetics, pp. 159, 163, 270-74.

[126] "Cross-Language Study," p. 394. In general, VOT values for English stops clustered around + 10 msec. and + 75 msec. for /bdg/ and /ptk/ respectively. Ibid., p. 403. In general there was little or no overlap in VOT values between /p/ and /b/, /t/ and /d/, /k/ and /g/, or between labials, alveolars and velars. Ibid., p. 399.

VOT values are assigned in reference to the stop release, which is arbitrarily given a value of zero.

". . . thus, measurements of voice onset time before release are stated as negative numbers and called voicing lead, while measurements of voice onset time after the release are stated as positive numbers and called voicing lag." Ibid., p. 389.

[127] Ibid., pp. 414-15.

[128] Lisker and Abramson, "Effect of Context," p. 8.

[129] Ibid., p. 11.

[130] Ibid., p. 16.

[131] Ibid., p. 8.

[132] Ibid., p. 18.

[133] "Cross-Language Study," p. 417. In 1941, Heffner, offering a tentative explanation for the production of "voiced t," described "residual vibrations" of the vocal folds during the phase of glottal opening which are akin to, if not identical with, the "edge vibrations" described by Lisker and Abramson. Heffner notes that in words such as butter "the opening of the [vocal] cords at the occlusion of t has been so quickly followed by the closure of the glottis for the final syllable that the residual vibrations of the vocal bands occupy the very short interval taken up by the occlusion." "An Adjunct to the Graphic Method," American Speech, 16 (February, 1941), 37-38.

[134] Lisker and Abramson, "Effect of Context," p. 8.

[135] Ibid., p. 18.

[136] Ibid.

[137] Ibid., p. 24.

[138] *Ibid.*, pp. 24-25.

[139] *Ibid.*, p. 27.

[140] Lisker, Abramson, *et al.*, "Transillumination," pp. 4.2-4.

[141] *Ibid.*, p. 4.4.

[142] *Ibid.*, p. 4.5.

[143] "English Stops," p. 6.

[144] Lotz, *et al.*, "Perception of English Stops," pp. 71-76.

[145] Reeds and Wang, "Perception of Stops After *s*," pp. 78-81.

[146] Harris, *et al.*, "Aspects of Production," p. 147.

[147] "Some Phonetic Specifications," pp. 137, 147.

[148] *Ibid.*, p. 167.

[149] "Voice Onset Time in Stop Consonants," *Haskins Laboratories Status Reports on Speech Research*, 3 (August, 1965), 1.3-4. See also Arthur S. Abramson, Leigh Lisker, and Franklin S. Cooper, "Laryngeal Activity in Stop Consonants," *Haskins Laboratories Status Reports on Speech Research*, 4 (November, 1965), 6.7, fig. 2, for a schematic diagram of the pattern used in the synthesis.

[150] "Voice Onset Time in Stop Consonants," p. 1.4.

[151] "Study of Intervocalic /t/ and /d/," p. 3.3.

[152] *Ibid.*, pp. 3.3-4.

[153] *Ibid.*, pp. 3.8-11 (figs. 3-6).

[154] Delattre, *et al.*, "Acoustic Cues," pp. 28-29.

[155] *Phonetics*, p. 272.

[156] "Electromyographic and Acoustic Study of the Production of Certain Final Clusters," *Journal of the Acoustical Society of America*, 35 (April, 1963), 462.

[157] Kantner and West do, however, note that such an "unaspirated surd fricative" does occur in "many dialects." *Phonetics*, p. 272.

[158] Lisker and Abramson, "Transillumination," p. 4.5.

[159] *Ibid.*, pp. 4.4-6.

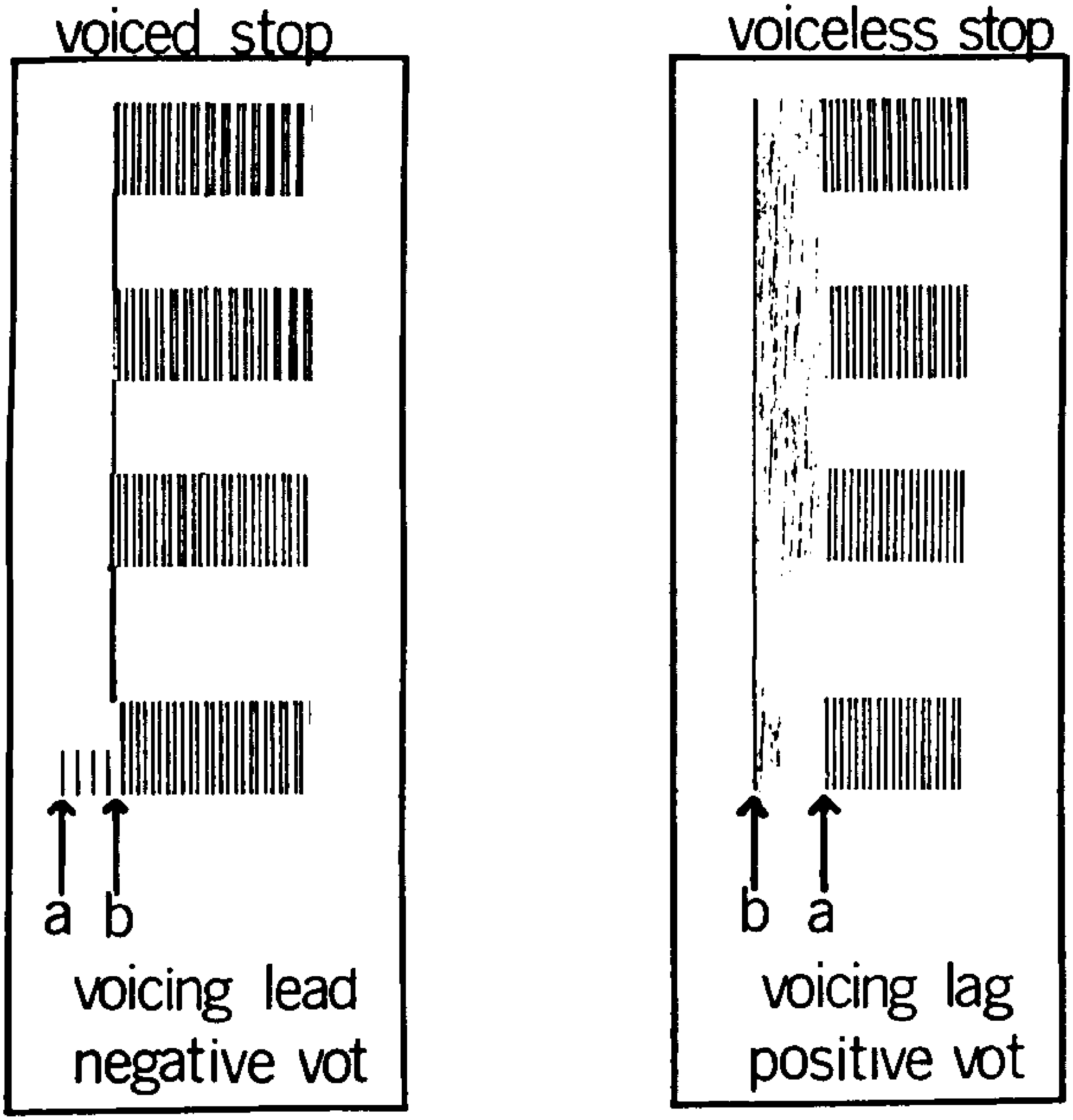
CHAPTER II

CUES TO THE CLASSIFICATION OF THE MEMBERS OF COGNATE PAIRS IN WORD-FINAL POSITION: SPECTROGRAPHIC ANALYSIS AND TAPE-CUTTING EXPERIMENTS

The experiments described in this chapter were designed to determine the relative importance of two potential cues to the categorization of the members of cognate pairs in word final position. The cues investigated were (1) Voicing Offset Time: the final-position analogue of Voicing Onset Time, and (2) the duration of the vowels preceding the final consonants. Two other related potential cues were also investigated: (1) voicing during consonant closure, and (2) total voicing duration in the syllable.

Voicing Onset Time is measurable from the release of an initial stop consonant, with a lag in onset after release being specified by positive values, onset occurring before release being specified by negative values[1] (see Fig. 2-1). An analogous method may be used to determine the Voicing Offset Time for stop consonants in final position: Voicing which ceases before the release of a final stop may be specified by positive values of offset time, and voicing which continues after release may be specified by negative offset values. In those cases where a final stop is

VOICE ONSET TIME: STOPS



point a: onset of voice
point b: release of stop

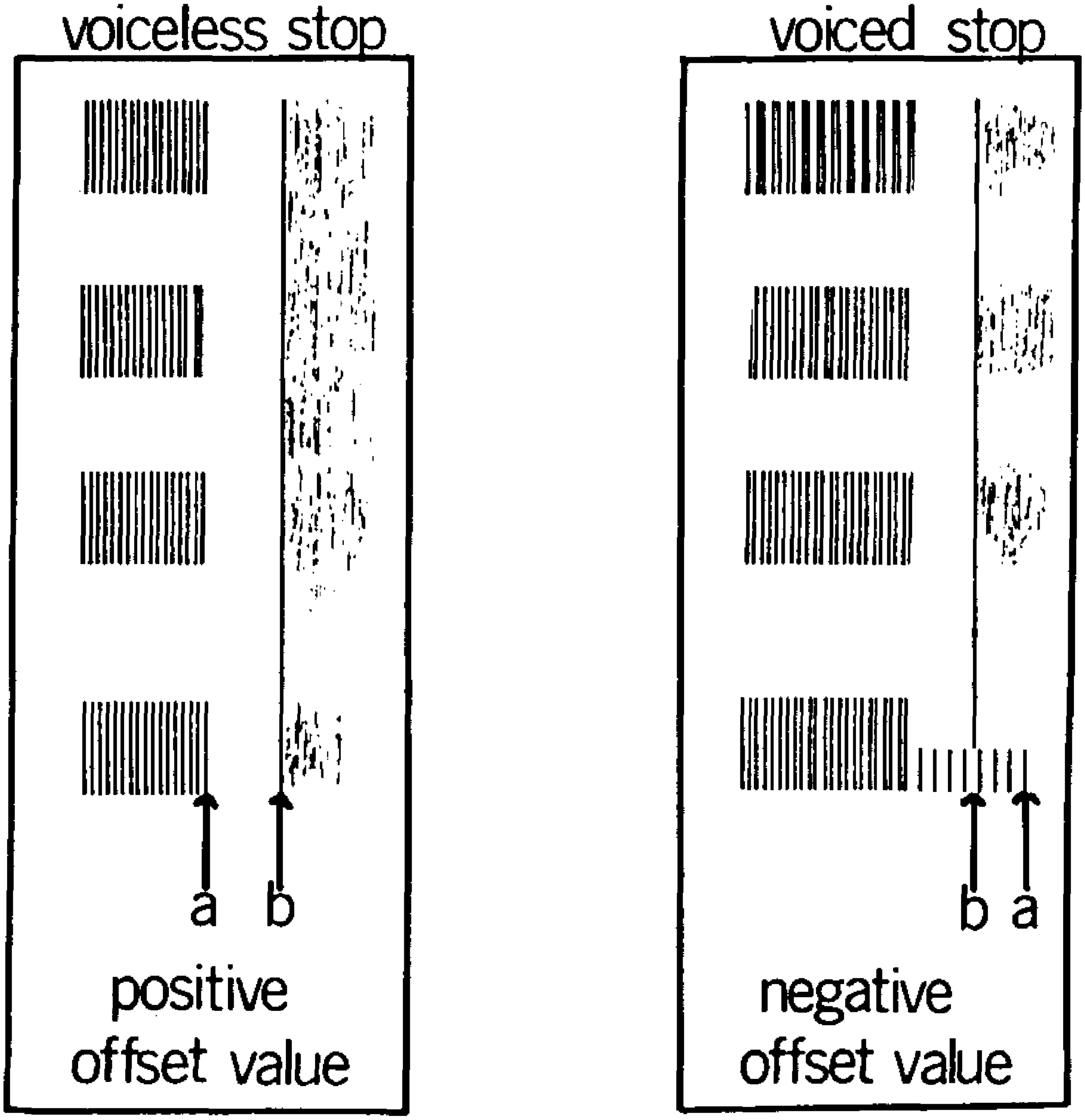
figure 2-1

unreleased, the offset value is assumed to be zero: Since the silence which marks the closure is unbounded, it is reasonable to assume that the consonant is coterminous with the offset of voicing (see Fig. 2-2).

The time at which voicing ceases may be coincidental with the termination of the vowel preceding a final consonant. In spectrograms this point is taken to be that at which (because of the closure for the consonant) there is no trace of formant structure above F1, either as voice or as noise (aspiration). Thus, in the case of stops, voicing during closure (i.e. any presence of voicing, as indicated by the presence of F1 or of a fundamental frequency which extends beyond the termination of the formant structure of the vowel) will be the complement of any positive Voicing Offset value for the total duration of closure. Positive Voicing Offset Time is thus the duration of the period of voicelessness, if any, between the release of the final consonant and the offset of voicing during closure. For example, if the duration from the termination of the vowel to the release of the final consonant (i.e. closure duration) is 100 msec., and if voicing continues for 40 msec. into the closure, then the voicing offset value for the final consonant is + 60 msec.

The procedure used for measuring Voicing Offset Time for fricatives is similar to that used in the case of stops. Positive values are assigned to those utterances where voicing ceases before the end of the friction, and

VOICE OFFSET TIME: STOPS



point a: offset of voice
point b: release of stop

figure 2-2

negative values are assigned in those cases where voicing continues unbroken through and beyond the cessation of friction.[2] The question of non-release is obviously not relevant here since the friction cannot continue indefinitely. The duration of voicing during closure (i.e. voicing during the friction portion of the consonant) is complementary to the positive voicing offset value, for the total duration of closure (see Fig. 2-3).

In general, then, an increase in voicing offset values for a given utterance mandates a decrease of the duration of voicing during closure, and thus in most cases both measures can be determined by a single measurement (once closure duration is known) and can be altered by a single experimental manipulation of the stimulus.

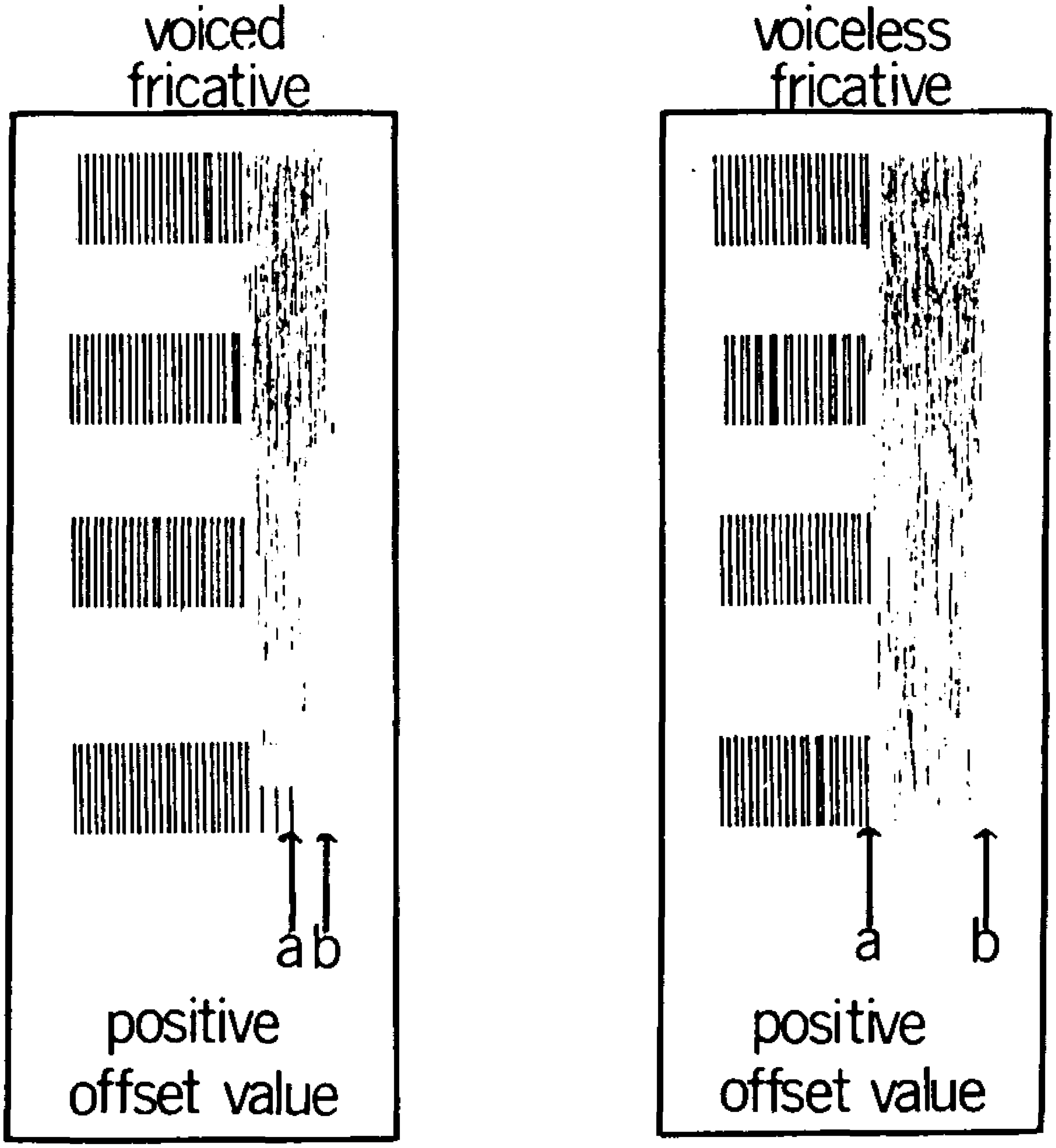
The duration of a vowel preceding a final consonant can be determined spectrographically by measuring from the point where a full formant structure first becomes evident, either as periodic vibration or as noise (aspiration), to the point where such formant structure ceases to be evident.

Experiment I: Spectrographic Measurements of Potential Cues

Procedure

In this first experiment a corpus of test items was developed which consisted of minimal and sub-minimal pairs of English words. The pairs contrasted by virtue of the final-position phonemic oppositions /p-b/, /t-d/, /k-g/,

VOICE OFFSET TIME: FRICATIVES



point a: offset of voice
point b: cessation of friction

figure 2-3

/f-v/, /θ-ð/, /s-z/, /ʃ-ʒ/ (see Appendix 1 for a complete listing of the corpus). The vowels and diphthongs chosen to precede the final consonants were /i, I, e, ɛ, æ, u, U, o, ɔ, a, ʌ, ɝ, aI, aU, ɔI/. Each word was spoken by each of three native male speakers of Standard New York City dialect. The speakers each read from one of three differently randomized lists. For the purposes of this experiment the results drawn from one speaker's recording of the list of words are reported.

Wide band spectrograms were made for each recorded item using a standard Kay Sonograph. A -10 db amplitude display was incorporated into each spectrogram and was used as an aid to detect both the onset of consonant closure and the cessation of voicing. Narrow band spectrograms were also made in those cases where it seemed that an inspection of harmonics might be useful as an aid to determine the presence or absence of voicing.

Four points were marked on the wide band spectrograms:

(1) The onset of the vowel (or diphthong) preceding the final consonant.

(2) The offset of the vowel (or diphthong) preceding the final consonant.

(3) The offset of voicing during the closure period of the final consonant.

(4) The release of the final consonant: i.e. the burst (if present) in the case of stops, and

the cessation of friction in the case of fricatives. (See Fig. 2-4)

The distances between these four points were then measured in millimeters and converted to their equivalents in milliseconds. Figures were rounded to the nearest five milliseconds as an estimate of experimental accuracy. Four principal measures were then computed for each item:

(1) The duration of the vowel preceding the final consonant (point 1 to point 2).

(2) The duration of the voicing during the closure of the final consonant (point 2 to point 3).

(3) The total duration of the voicing during the syllable, discounting the initial consonant (i.e. vowel duration plus duration of voicing during closure: point 1 to point 3).

(4) The Voicing Offset Time (i.e. the duration of the period of voicelessness, if any, between the release of the final consonant and the offset of voicing during closure: point 3 to point 4). (See Fig. 2-5)

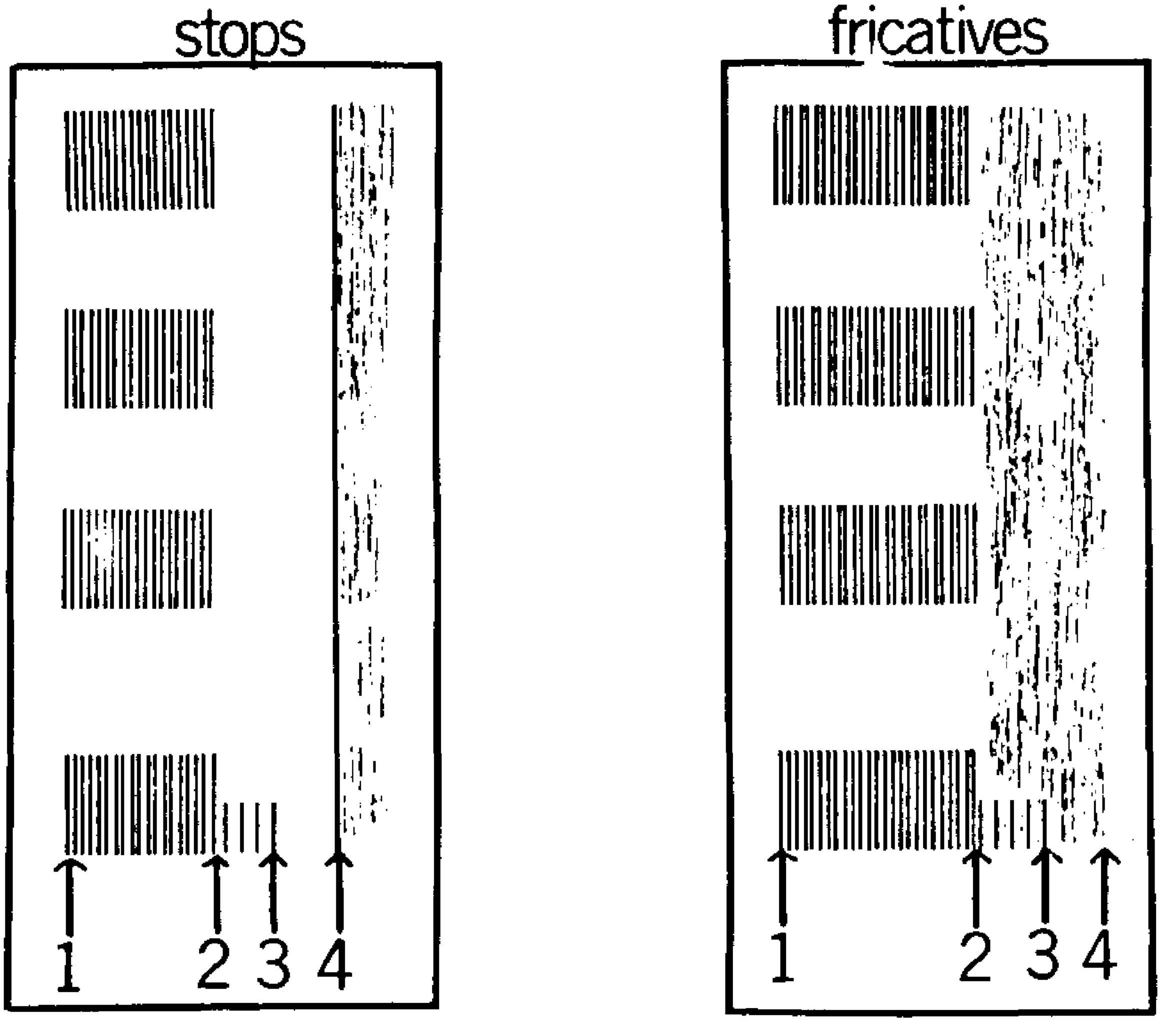
An additional measure was also computed:

(5) The total closure duration (point 2 to point 4).

Results

An inspection of the data reveals that, with few exceptions, all four principal measures correlate with the opposition of the members of cognate pairs: Vowel duration,

POINTS MARKED ON SPECTROGRAMS

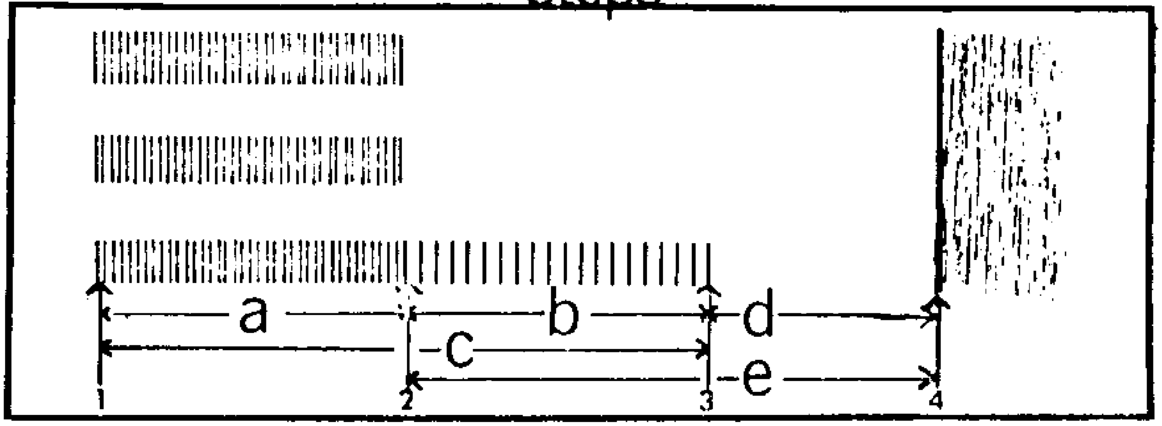


1: onset of vowel
2: offset of vowel
3: offset of voicing
4: consonant release

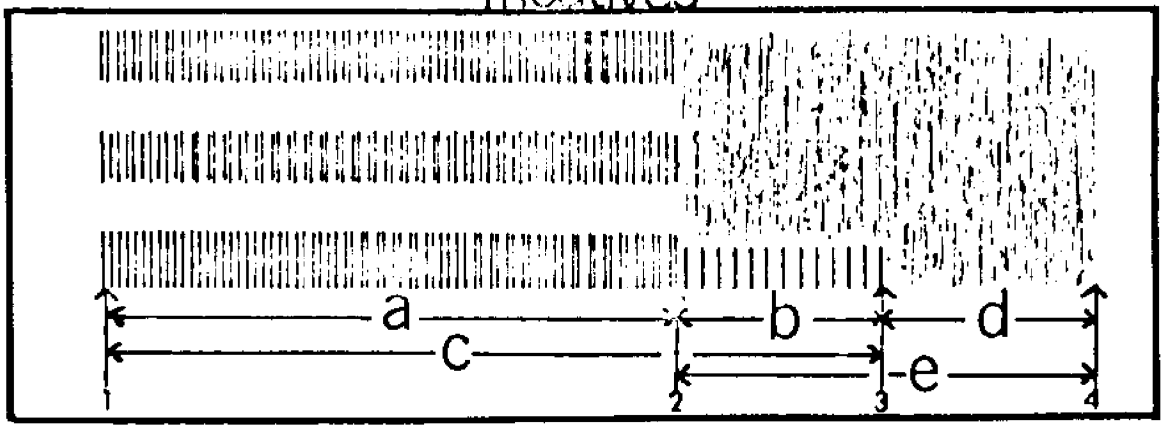
figure 2-4

SPECTROGRAPHIC MEASUREMENTS

stops



fricatives



a: vowel duration 1 ⟷ 2
b: closure voicing duration 2 ⟷ 3
c: total voicing duration 1 ⟷ 3
d: voice offset time 3 ⟷ 4
e: closure duration 2 ⟷ 4

figure 2-5

voicing during closure, and total voicing duration exhibit larger values in syllables ending with /bdgvðzʒ/, than in syllables terminating with /ptkfθsʃ/. Voicing Offset Time, as predicted from its relation to closure voicing, exhibits larger values in syllables ending with /ptkfθsʃ/ than in those ending with /bdgvðzʒ/. (See Tables II-7 and II-8.)

Vowel duration averaged 164.0 msec before /ptk/ and 316.4 msec before /bdg/, a ratio of slightly less than 2:1.[3] For fricatives the average vowel durations were 210.4 msec before /fθsʃ/, and 362.2 msec before /vðzʒ/, a ratio of 1:1.7 (see Tables II-1 and II-2). Average Voicing Offset Time was 57.6 msec before /ptk/ and 23.2 msec before /bdg/, a ratio of 2.5:1. The average Voicing Offset Time before /fθsʃ/ was 212.0 msec, and before /vðzʒ/ 66.1 msec, a ratio of 3.2:1 (see Tables II-1 and II-2).[4]

TABLE II-1

SPECTROGRAPHIC DATA FOR STOPS

	$\frac{p}{b}$		$\frac{t}{d}$		$\frac{k}{g}$		$\frac{ptk}{bdg}$	
	avg.	ratio	avg.	ratio	avg.	ratio	avg.	ratio
Vowel Duration	140.0 293.6	1:2.1	176.0 330.7	1:1.9	157.3 291.4	1:1.9	164.0 316.4	1:1.9
Closure Voicing	38.2 136.8	1:3.6	38.7 115.0	1:2.9	53.2 115.9	1:2.2	42.8 121.8	1:2.8
Total Voicing	204.1 428.6	1:2.2	214.0 445.0	1:2.1	217.3 394.6	1:1.8	212.0 425.1	1:2
Voicing Offset	59.1 20.5	2.9:1	54.3 27.0	2:1	60.5 20.9	2.9:1	57.6 23.2	2.5:1
Closure Duration	113.3 157.5	1:1.4	103.8 135.4	1:1.3	113.6 125.5	1:1.1	109.7 123.9	1:1.1

TABLE II-2

SPECTROGRAPHIC DATA FOR FRICATIVES

	f/v		θ/ð		s/z		fθs*/vðz*	
	avg.	ratio	avg.	ratio	avg.	ratio	avg.	ratio
Vowel Duration	207.8 360.6	1:1.7	183.0 356.0	1:1.9	217.5 365.4	1:1.7	210.4 362.2	1:1.7
Closure Voicing	71.1 80.6	1:1.1	53.0 125.0	1:2.4	34.6 120.8	1:3.5	50.7 108.1	1:2.1
Total Voicing	267.8 353.3	1:1.3	236.0 481.0	1:2	252.1 478.3	1:1.9	257.4 337.2	1:1.3
Voicing Offset	178.3 50.0	3.6:1	205.0 62.0	3.3:1	240.0 75.0	3.2:1	212.0 66.1	3.2:1
Closure Duration	252.7 141.6	1.8:1	258.0 191.0	1.4:1	274.6 195.0	1.4:1	263.8 176.2	1.5:1

*Data for the one minimal pair ending in /ʃ-ʒ/ may be found in Table II-9.

TABLE II-3

DATA FOR MINIMAL PAIRS ENDING IN /p-b/ TIME VALUES FOR EACH PARAMETER ARE GIVEN IN MILLISECONDS

	Vowel Duration	Closure Voicing	Total Voicing	Voicing Offset	Closure Duration
Heap	210	30	240	130	160
Hebe	330	175	505	-55	175
Nip	95	75	175	95	170
Nib	130	150	280	0*	150
Gape	140	40	180	90	130
Gabe	250	135	390	0*	135
Rep	225	85	310	30	115
Reb	230	170	400	-35	170
Lap	200	10	210	0*	10
Lab	355	120	475	-25	120
Mop	165	5	170	0	5
Mob	410	145	555	0*	145
Lope	215	60	275	95	155
Lobe	340	175	515	0	175
Loop	85	50	135	130	180
Lube	350	80	430	80	160
Sup	100	0	100	25	25
Sub	215	150	360	0*	150
Slurp	195	25	220	55	80
Herb	270	80	350	0*	80
Tripe	190	40	230	0*	40
Tribe	350	145	495	-30	145

*Denotes unreleased stop.

TABLE II-4

DATA FOR MINIMAL PAIRS ENDING IN /t-d/ TIME VALUES GIVEN IN MILLISECONDS

	Vowel Duration	Closure Voicing	Total Voicing	Voicing Offset	Closure Duration
Pleat	155	40	195	55	95
Plead	305	135	440	-30	135
Bit	105	45	150	105	150
Bid	185	160	340	-55	160
Rate	190	65	255	55	120
Raid	360	90	450	0	90
Bet	165	0	165	20	20
Bed	260	100	360	-45	100
Hat	245	10	255	25	35
Had	350	120	470	0	120
Pot	125	25	150	0*	25
Pod	315	70	385	0*	70
Sought	190	25	210	100	125
Sawed	320	140	460	-40	140
Moat	210	45	260	100	145
Mowed	355	105	460	-40	105
Put	115	30	145	30	60
Good	180	150	330	-40	150
Suit	175	40	215	100	140
Sued	355	70	425	45	115
Hurt	135	20	150	0*	20
Heard	405	135	540	-40	135
Height	200	30	230	95	125
Hide	390	60	445	0*	60
Hoyt	210	85	290	65	150
Boyd	375	175	550	-85	175
Clout	185	45	230	65	110
Cloud	450	100	550	-30	100
Pallet	235	75	310	0	75
Pallid	355	115	470	0*	115

*Denotes unreleased stop.

TABLE II-5

DATA FOR MINIMAL PAIRS ENDING IN /k-g/ TIME VALUES GIVEN IN MILLISECONDS

	Vowel Duration	Closure Voicing	Total Voicing	Voicing Offset	Closure Duration
Leak	135	65	200	100	165
League	275	135	405	80	215
Wick	260	35	295	35	70
Wig	270	55	325	-25	55
Fake	145	70	285	20	90
Vague	320	145	325	-25	145
Peck	145	100	245	40	140
Peg	320	125	450	-15	125
Sack	130	65	200	75	140
Sag	215	105	320	10	115
Clock	140	65	205	65	130
Clog	335	75	410	-55	75
Hawk	175	0	175	70	70
Dog	310	155	465	-70	155
Broke	150	30	180	100	130
Brogue	330	165	495	-40	165
Duke	185	15	200	65	80
Krug	335	105	440	25	130
Tuck	110	70	180	35	105
Tug	180	105	285	10	115
Burke	155	70	225	60	130
Burg	315	105	420	25	130

TABLE II-6

DATA FOR MINIMAL PAIRS ENDING IN /f-v/
TIME VALUES GIVEN IN MILLISECONDS

	Vowel Duration	Closure Voicing	Total Voicing	Voicing Offset	Closure Duration
Grief	145	80	225	175	255
Grieve	330	80	410	65	145
Syph	165	65	230	145	210
Sieve	270	100	370	35	135
Safe	215	80	295	200	280
Save	385	95	480	70	165
Ref	225	60	285	220	280
Rev	255	175	430	25	200
Half	280	60	340	210	270
Have	380	100	480	45	145
Proof	200	70	270	130	200
Prove	410	90	510	20	110
Duff	200	40	240	215	245
Dove	325	75	400	65	140
Serf	190	115	305	150	265
Serve	400	25	425	80	105
Life	250	70	320	160	230
Live	490	85	575	45	130

TABLE II-7

DATA FOR MINIMAL PAIRS ENDING IN /ð-θ/
TIME VALUES GIVEN IN MILLISECONDS

	Vowel Duration	Closure Voicing	Total Voicing	Voicing Offset	Closure Duration
Teeth	175	30	205	200	230
Teethe	315	105	420	85	110
With	150	85	235	185	270
With	265	170	435	30	200
Loath	180	65	245	210	275
Loathe	455	55	510	55	110
Sooth	160	25	185	200	225
Soothe	435	35	470	50	85
Mouth	250	60	310	230	290
Mouthe	310	260	570	90	350

TABLE II-8

DATA FOR MINIMAL PAIRS ENDING IN /s-z/ TIME VALUES GIVEN IN MILLISECONDS

	Vowel Duration	Closure Voicing	Total Voicing	Voicing Offset	Closure Duration
Peace	150	20	170	235	255
Peas	310	60	375	125	185
Hiss	115	85	200	200	285
His	265	160	425	40	200
Pace	220	10	230	250	260
Pays	360	80	440	95	175
Press	205	20	225	220	240
Pres	320	165	485	45	210
Ass	345	40	385	220	260
As	420	130	550	65	195
Floss	250	15	265	255	270
Flaws	330	120	350	85	205
Dose	225	25	250	255	280
Doze	420	140	560	95	235
Deuce	225	25	250	255	280
Dues	400	125	525	70	195
Bus	140	70	210	225	295
Buzz	225	170	395	60	230
Dice	255	50	305	255	305
Dies	475	75	550	110	185
Boyce	245	40	285	240	280
Boys	470	50	520	110	160
House	235	15	250	270	285
House	390	175	565	0	175

TABLE II-9

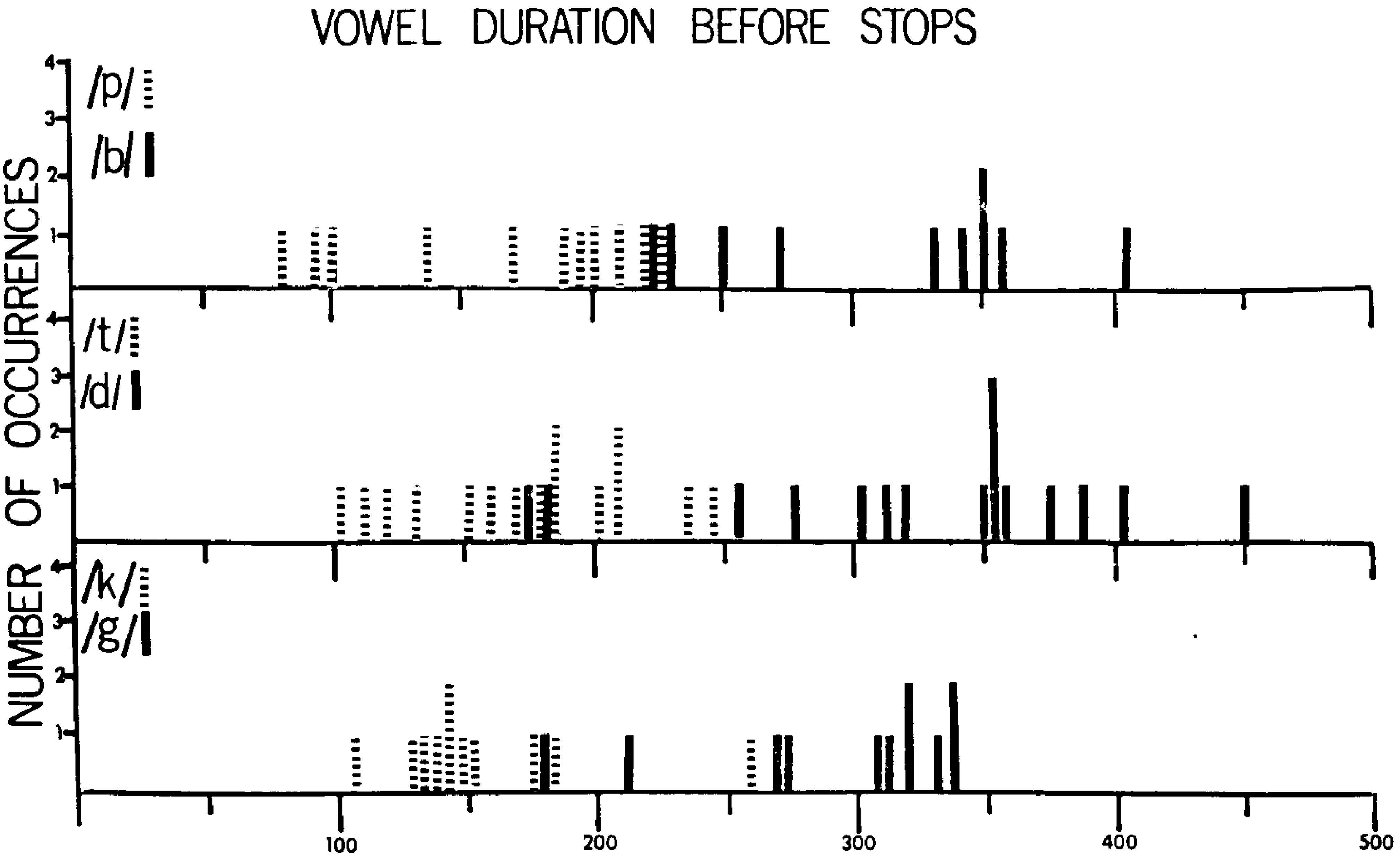
DATA FOR MINIMAL PAIRS ENDING IN /ʃ-ʒ/
TIME VALUES GIVEN IN MILLISECONDS

	Vowel Duration	Closure Voicing	Total Voicing	Voicing Offset	Closure Duration
Cash	285	50	335	215	265
Cas(ual)	370	110	480	125	235

Figures 2.6-2.13 indicate that the two categories of stops and fricatives are generally, though not absolutely separated by all four measures. That is, there is relatively little overlap of values between categories along each of the dimensions measured. More important, with but three exceptions in each one of pairs of words studied, there was a vowel of shorter duration and an earlier cessation of voicing during closure before the sounds usually labelled as "voiceless" than before those usually labelled as "voiced." In the exceptional cases, and in those cases where differences were minimal, the addition of the two measures revealed that for each minimal pair no total voicing duration was greater in a word ending in /ptkfθsʃ/ than in a word ending in /bdgvðzʒ/.

An analogous observation obtains for the Voicing Offset measure. That is, for each minimal pair the values were greater for words ending in /ptkfθsʃ/ than for words ending in /bdgvðzʒ/. This fact is somewhat obscured in the case of the stops which were unreleased sounds; these, assigned a value of zero for the Voicing Offset measure, occasionally had the same value as their cognates (see Tables II-3 and II-4).

It should be noted that eight of the cases of unreleased stops were for the sounds /b,d/, while only four cases of unreleased /p,t/ and none of /k,g/ were found. This finding corroborates the view that voiceless stops are released more frequently than are voiced stops.[5]



VOWEL DURATION BEFORE STOPS

figure 2·6

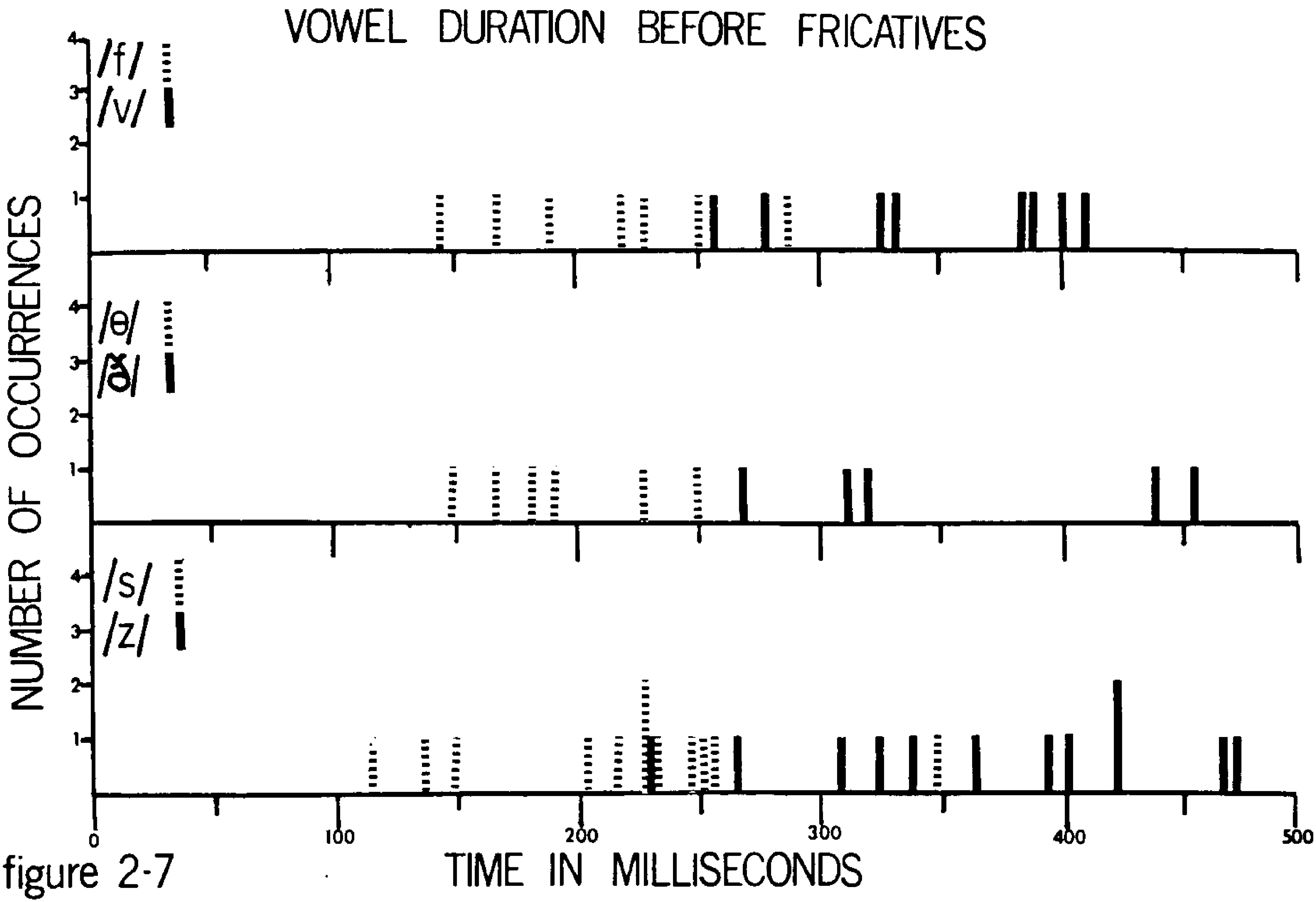

VOWEL DURATION BEFORE FRICATIVES
NUMBER OF OCCURRENCES
/f/
/v/
/θ/
/ð/
/s/
/z/
0
1
2
3
4
100
200
300
400
500
TIME IN MILLISECONDS

figure 2-7

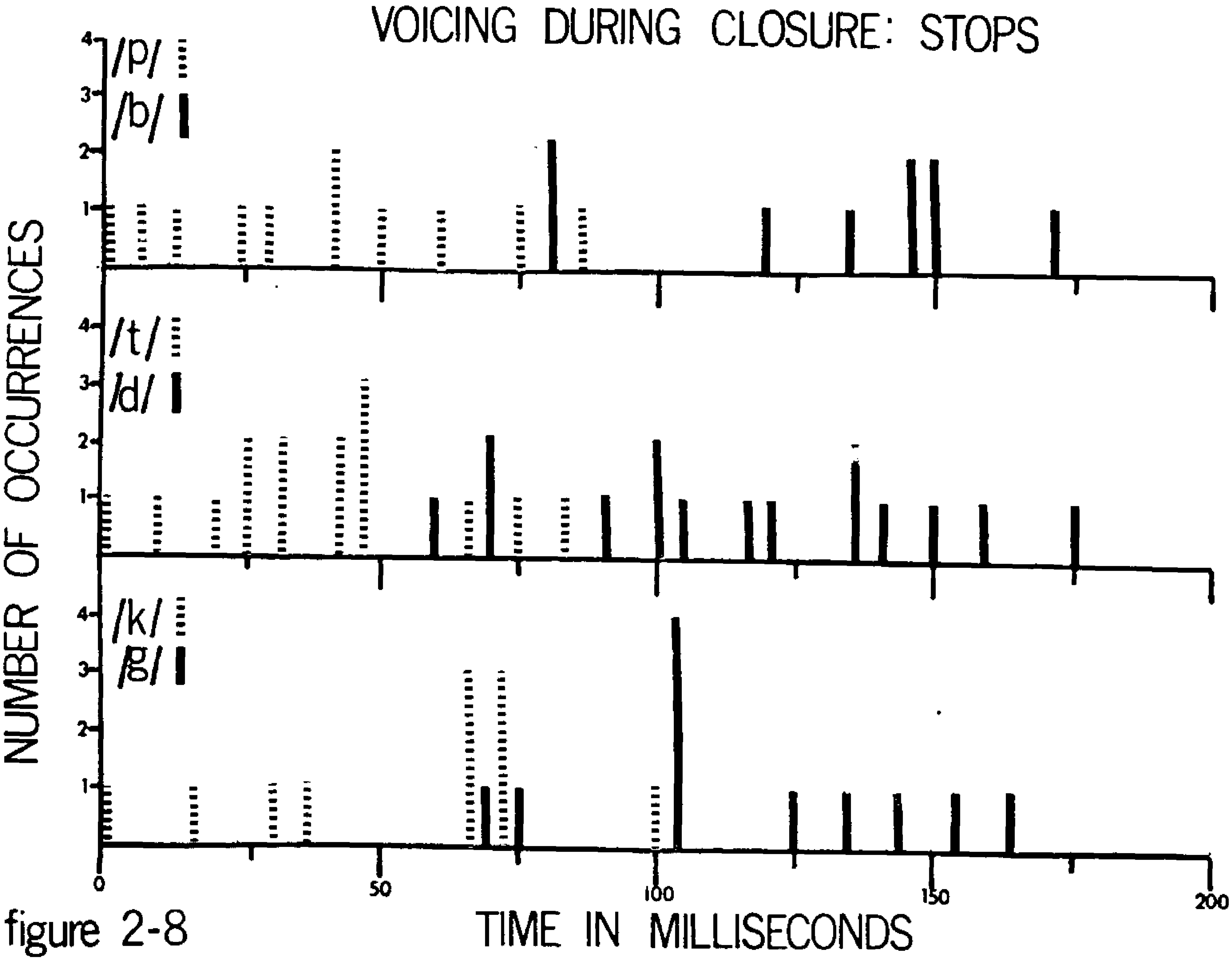

VOICING DURING CLOSURE: STOPS
/p/
/b/
/t/
/d/
/k/
/g/
NUMBER OF OCCURRENCES
1
2
3
4
0
50
100
150
200
TIME IN MILLISECONDS

figure 2-8

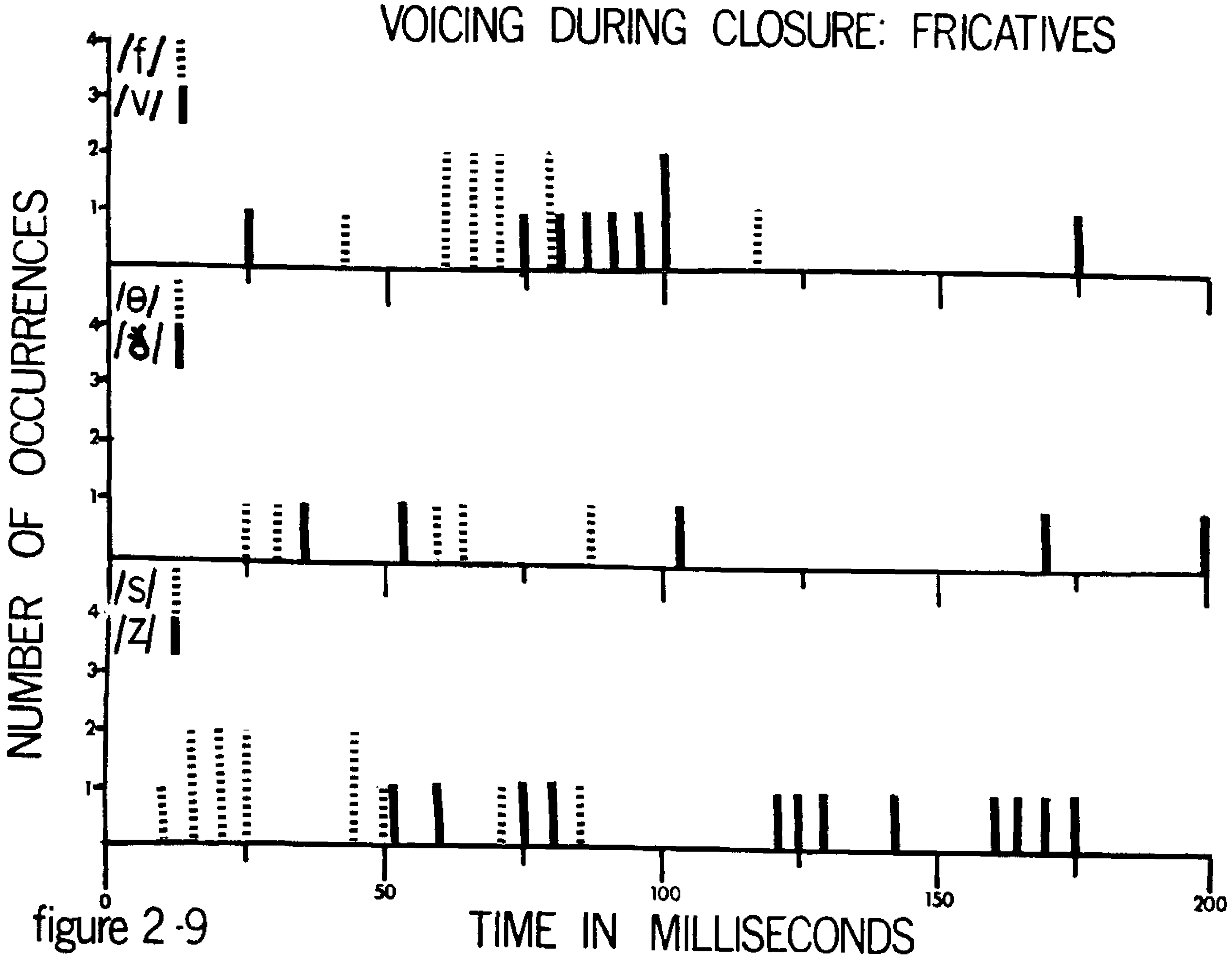

VOICING DURING CLOSURE: FRICATIVES
NUMBER OF OCCURRENCES
/f/
/v/
/θ/
/ð/
/s/
/z/
4
3
2
1
0
50
100
150
200
TIME IN MILLISECONDS

figure 2-9

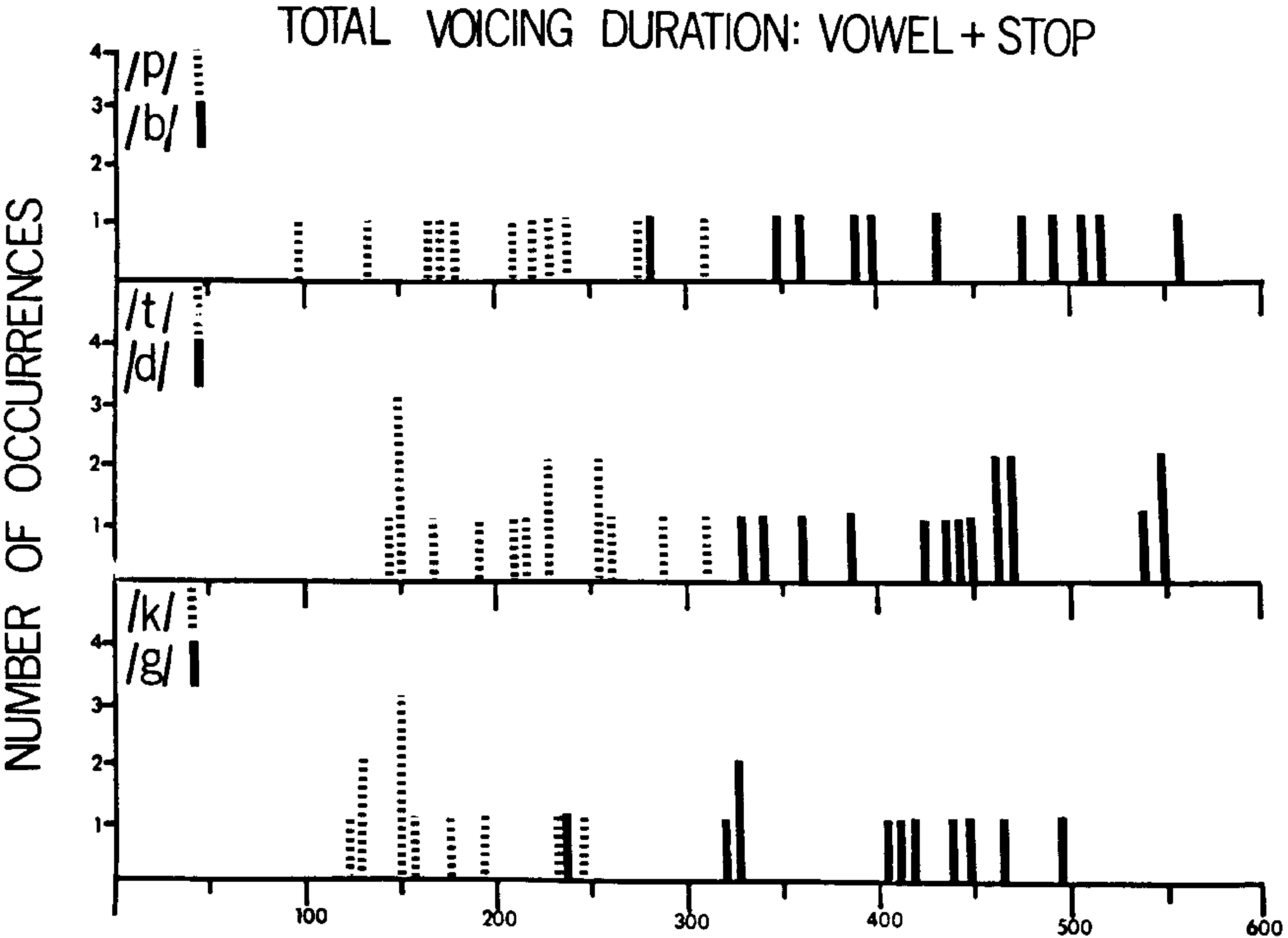

TOTAL VOICING DURATION: VOWEL + STOP
NUMBER OF OCCURRENCES
/p/
/b/
/t/
/d/
/k/
/g/
1
2
3
4
100
200
300
400
500
600
TIME IN MILLISECONDS

figure 2-10

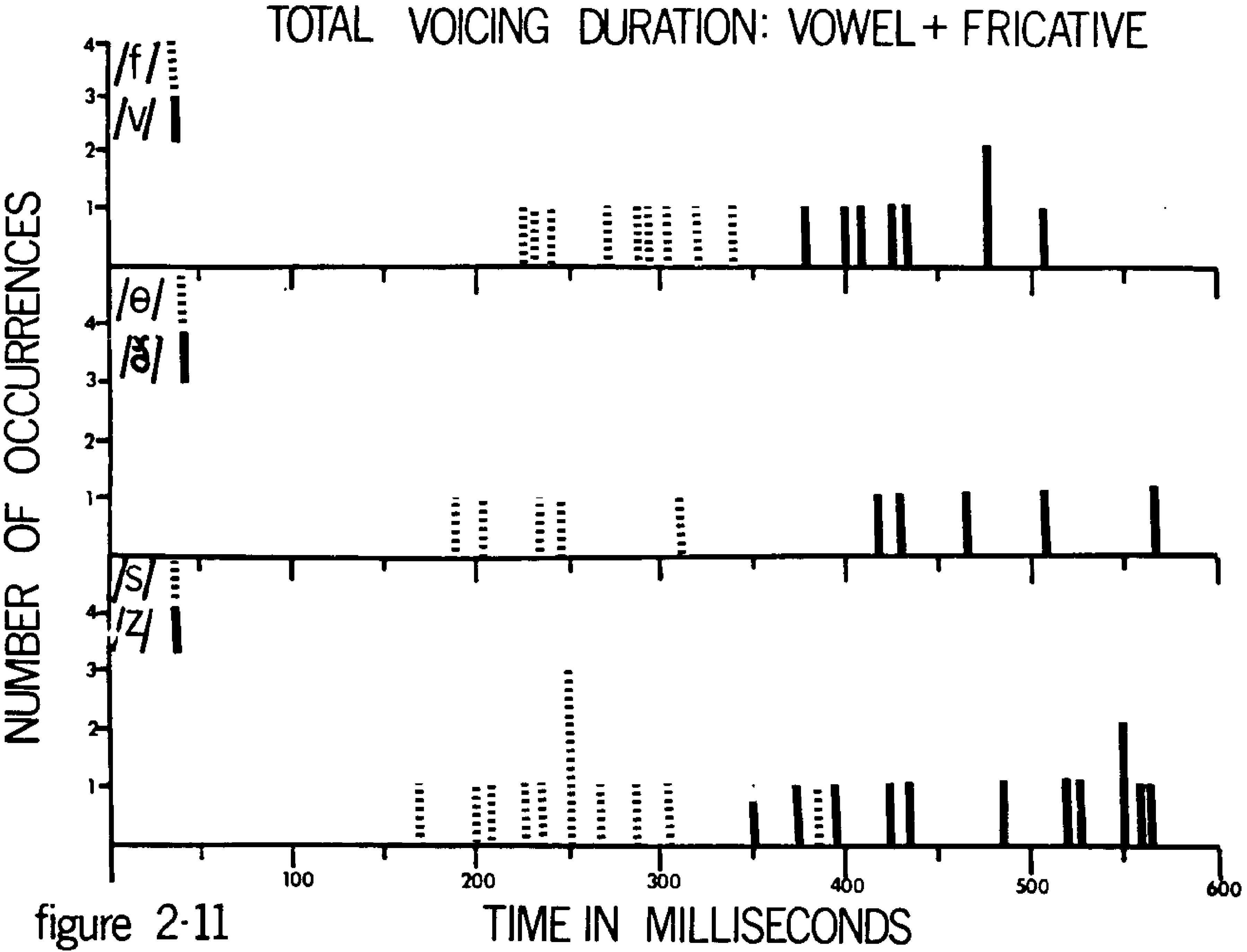

TOTAL VOICING DURATION: VOWEL + FRICATIVE
NUMBER OF OCCURRENCES
/f/
/v/
/θ/
/ð/
/s/
/z/
1
2
3
4
100
200
300
400
500
600
TIME IN MILLISECONDS

figure 2·11

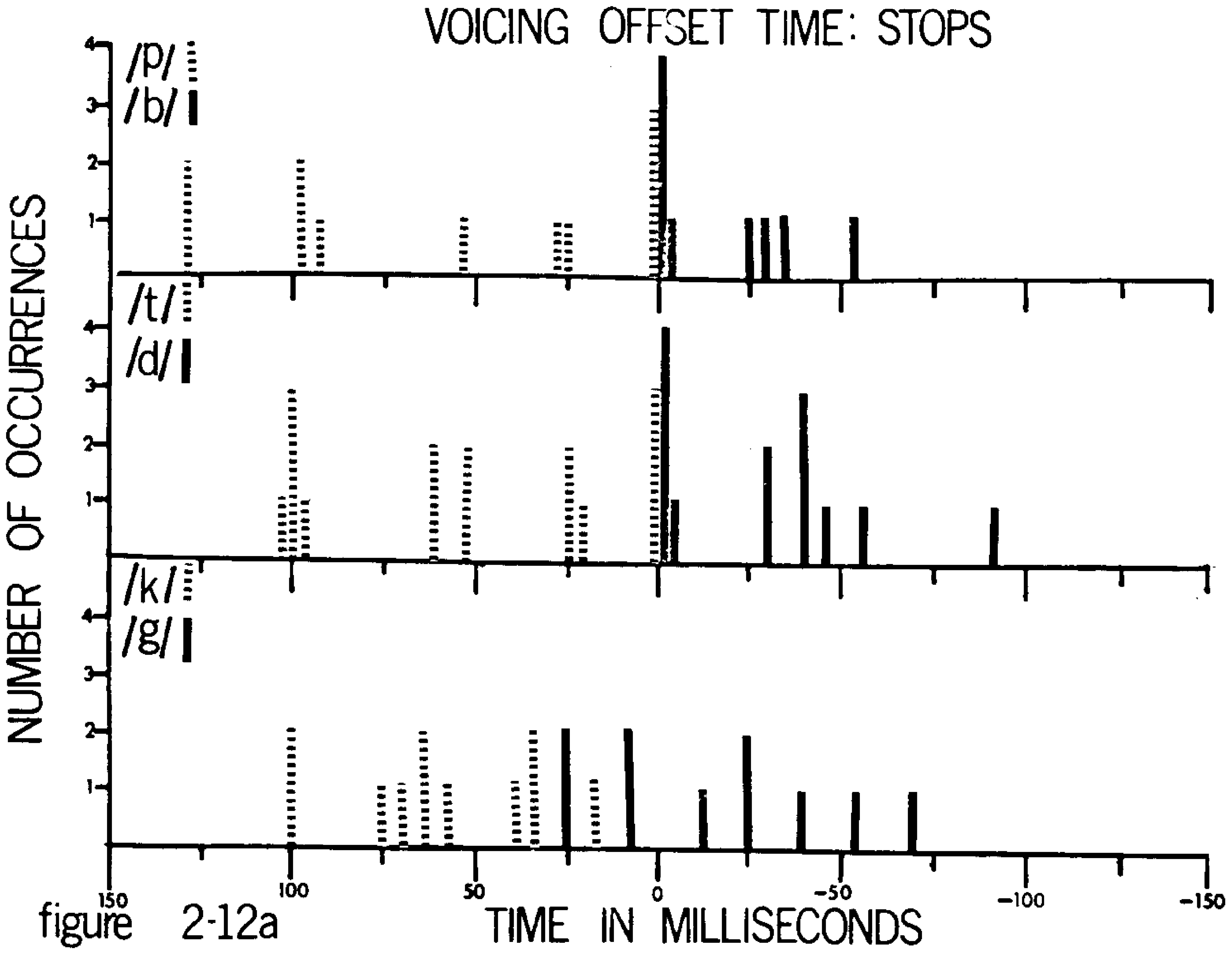

VOICING OFFSET TIME: STOPS
NUMBER OF OCCURRENCES
/p/
/b/
/t/
/d/
/k/
/g/
4
3
2
1
150
100
50
0
-50
-100
-150
TIME IN MILLISECONDS

figure 2-12a

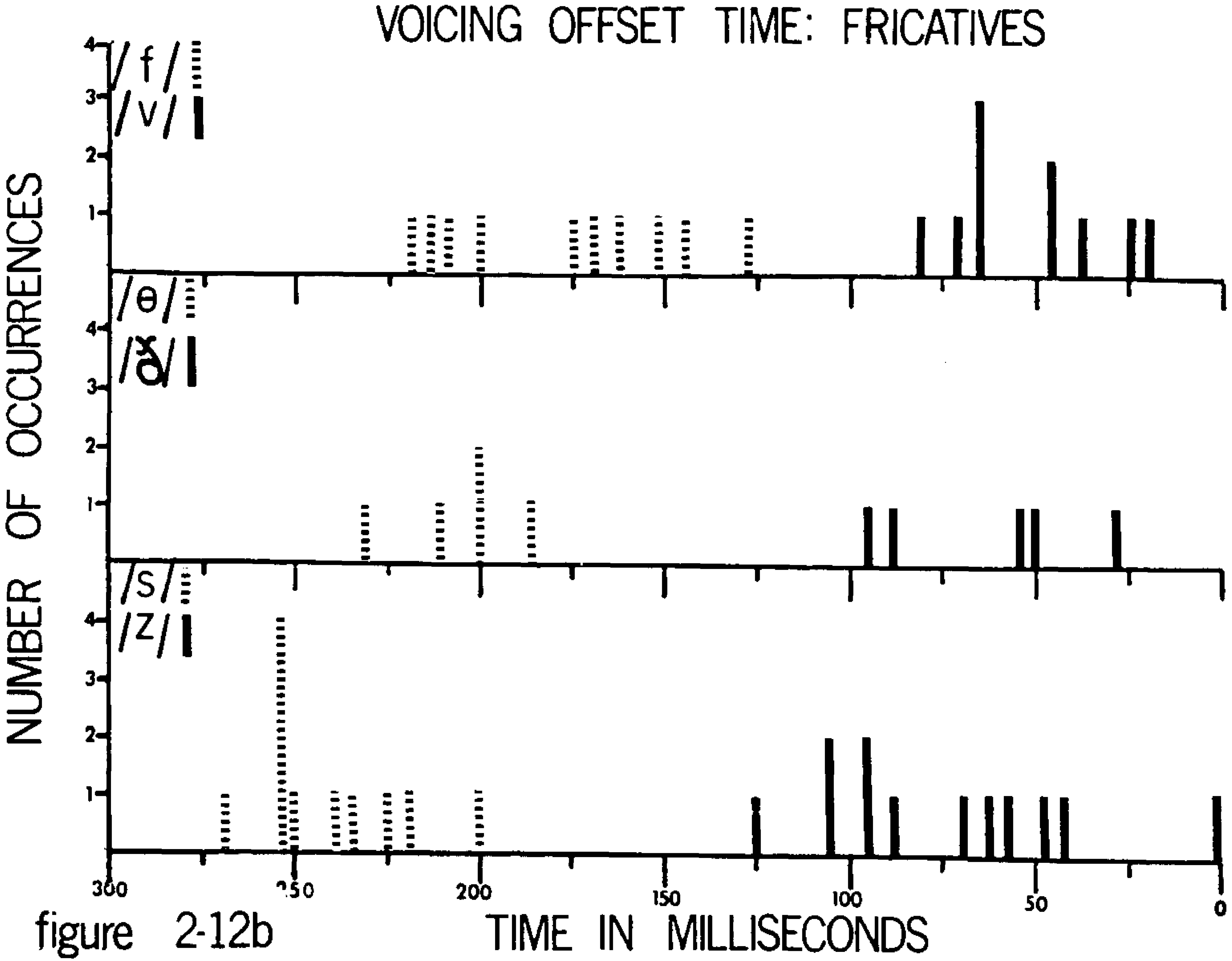

VOICING OFFSET TIME: FRICATIVES
NUMBER OF OCCURRENCES
/f/
/v/
/θ/
/ð/
/s/
/z/
4
3
2
1
300
250
200
150
100
50
0
TIME IN MILLISECONDS

figure 2-12b

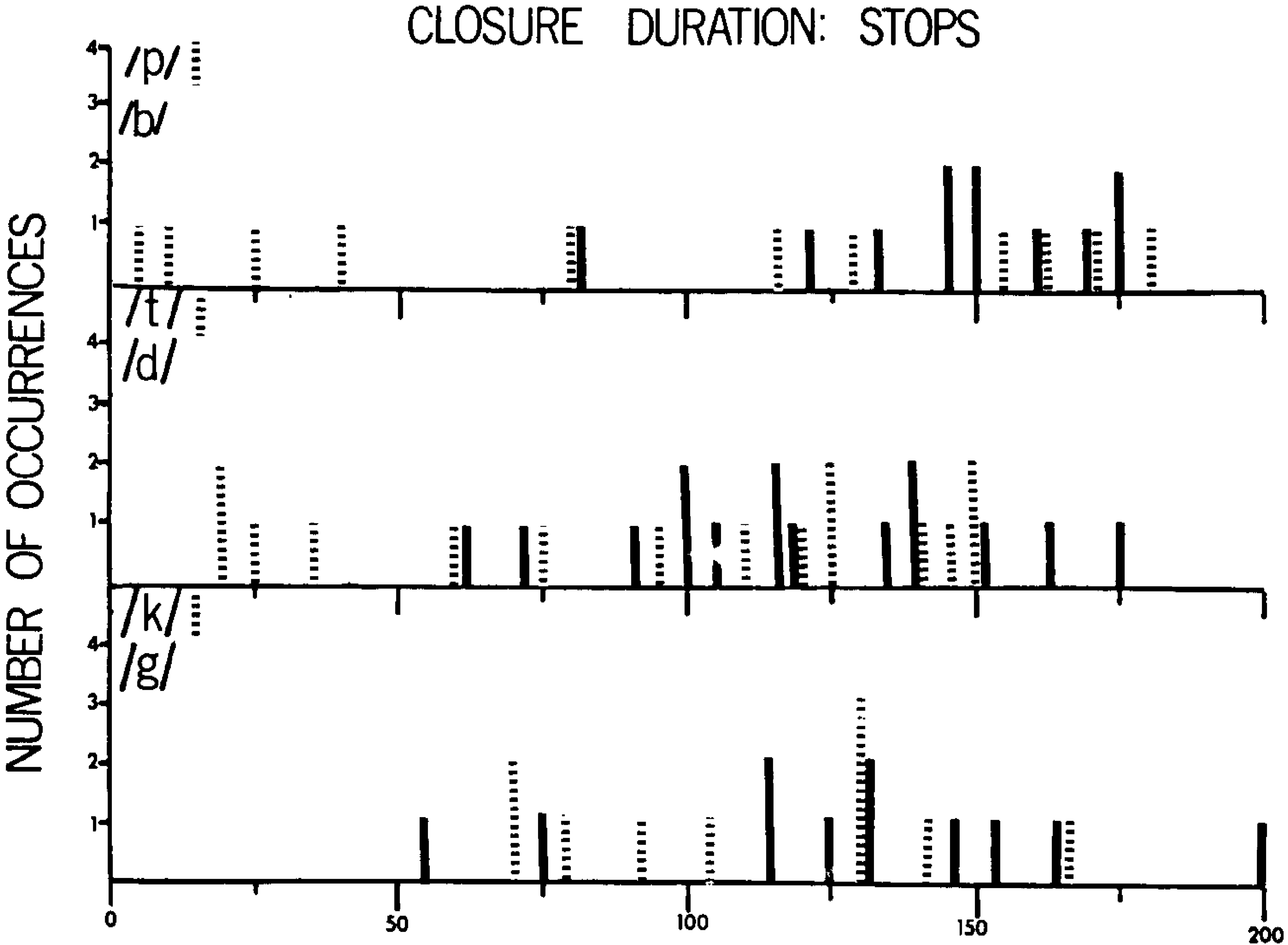
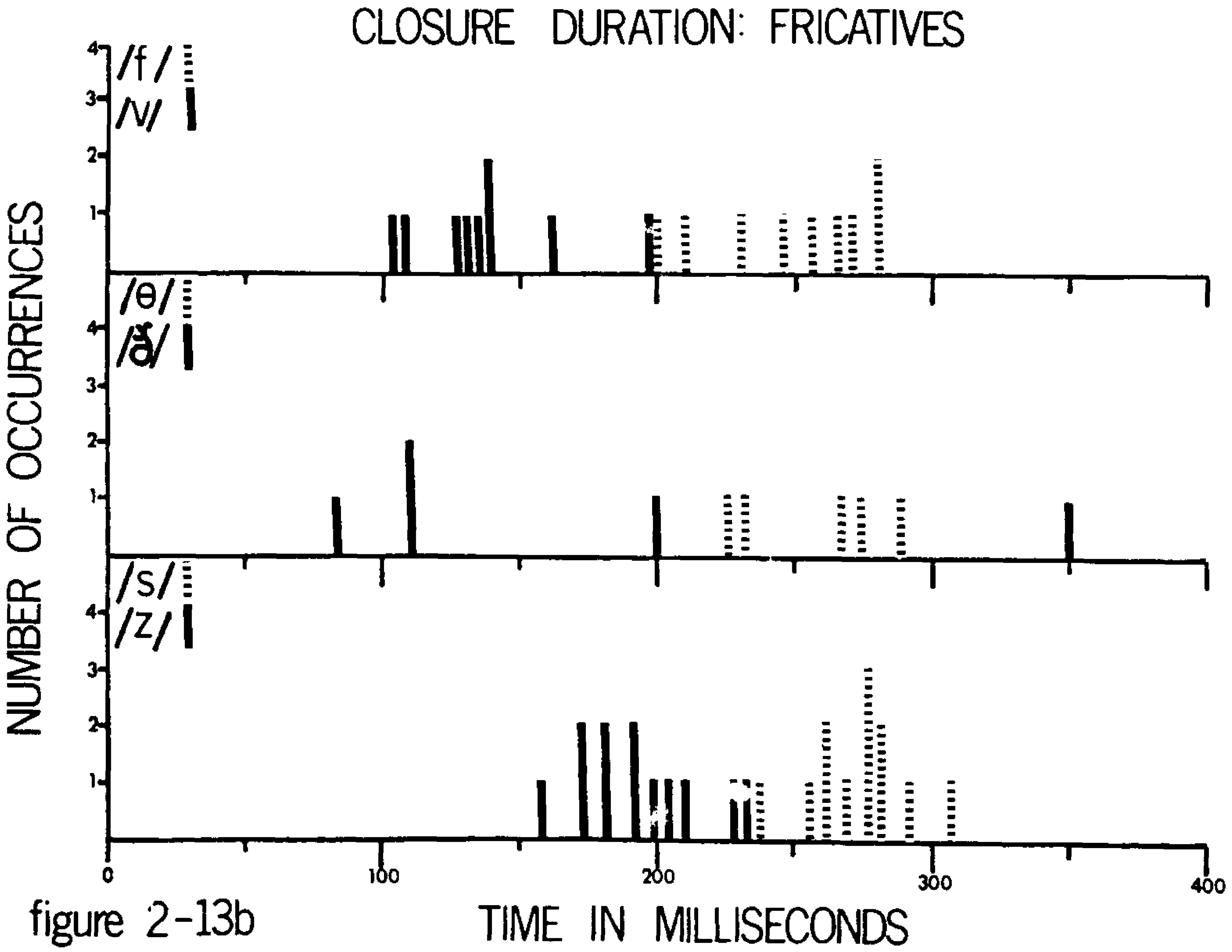

CLOSURE DURATION: STOPS
NUMBER OF OCCURRENCES
/p/
/b/
/t/
/d/
/k/
/g/
4
3
2
1
0
50
100
150
200
TIME IN MILLISECONDS

figure 2-13a



CLOSURE DURATION: FRICATIVES

figure 2-13b

The additional measure of closure duration proved to successfully categorize the fricative consonants into cognate classes. With one exception, the closure durations for /fθs/ exceeded those for /vðz/ in every case by at least 30 msec. The average closure duration for each member of a cognate pair and for each class of cognates as a whole were clearly separated, and there was a minimum of overlap in the absolute values (see figure 2-14 and Tables II-1 and II-2).

Discussion

All of the parameters investigated provide data which effectively categorize cognate sounds. The most effective measure was that of total voicing duration. This measure, in turn, may be viewed as being comprised of two component measures: vowel duration and, alternatively, voicing during closure or Voicing Offset Time (since these last two are, in general, complementary). The fact that these measures allow for categorization of cognate sounds does not indicate that any or all of them are perceptually significant. The correlation between acoustic measurements and linguistic categories does, however, suggest the likelihood that there is a significant cue value to be found among the parameters investigated. The experiments described below were designed to test this likelihood.

Experiment II: Voicing Offset Time/Voicing During Closure as a Cue to the Classification of the Members of Stop Cognates in Word-Final Position: A Tape-Cutting Experiment

This experiment was designed to determine the effectiveness of the measure(s) of Voicing Offset Time/ Closure-Voicing Duration as a cue to the categorization of cognate stop phonemes in word-final position.[6]

Procedure

A typical minimal pair was selected from each of the three categories of stops tested in Experiment I (see Tables II-3, II-4, II-5). All of the syllables, except for the one ending with /p/, had clearly released stops in final position. Each pair (six words in all) was re-recorded many times. These dubbings were made at a tape speed of 15 ips.

A cutting and splicing technique was then employed to remove some portion of the voicing during closure from each member of the pairs ending in /bdg/. The amount of closure voicing removed from each successive dubbing was increased in steps of 10 msec. Blank tape was added to replace the segment of tape with closure voicing that had been removed. The effect of the cuts was thus to decrease, progressively, the duration of the voicing during closure from one stimulus to the next, while at the same time increasing the Voicing Offset value for the final stop.

The cutting and splicing technique, used in this and the following experiment, began by placing an audible click on the tape recording of the original utterances before the dubbings were made. This was done by touching the edge of a magnetized razor blade perpendicularly to the surface of the tape. The click was manifested on a wide band spectrogram by a vertical spike, occurring some distance after the release of the final stop.

The distance from the click to the point of voicing offset was then measured and converted to a time value in milliseconds. This time value was in turn converted to the corresponding length of tape which represented it on the 15 ips dubbings. It was then possible to locate the reference click on each dubbed stimulus by turning the tape reels slowly by hand until the desired point on the tape passed over the recording head. This point was marked and the point at which the voicing during closure ceased was determined by measuring the calculated distance along the tape which corresponded to the time value derived from the spectrogram. The tape was cut at this point and a further length of tape, corresponding to the desired time value of closure voicing to be removed was marked off. Another cut was made at this point. The piece of tape containing voicing during closure was removed and a piece of blank magnetic tape, equal in length to that just removed, was substituted in its place. The several sections of tape were then spliced together to yield a completed stimulus.[7]

The cuts were made on an Editall splicing block having a splicing groove milled at an angle of 80° to the perpendicular. This angle provides a stimulus that does not lose as much of the time dimension as the usual 45° splicing cut. Further, it avoids placing an audible click on the tape which is often heard when a 90° cut is made. The accuracy of each cut was verified by making wide-band spectrograms of each altered stimulus.

The stimuli used in each of the three tests of the experiment consisted of one unaltered utterance ending with one of the stops /bdg/, a series of altered words with progressively greater durations of voicing during closure removed and ending with the same stop as the unaltered stimulus, and an equal number of recordings of the second member of the minimal pair which ended in the cognate of the unaltered stimulus.[8]

The stimuli were then randomized and played to twenty phonetically-naive native English Speakers (standard New York City Dialect). The subjects were asked to judge each item as either one or the other of the minimal pair originally recorded. They were told to make a judgment for each item, even if they were unsure as to which it was. Listeners heard the stimuli as played on an Ampex 301 tape recorder over a loudspeaker in a sound-treated room.

Results

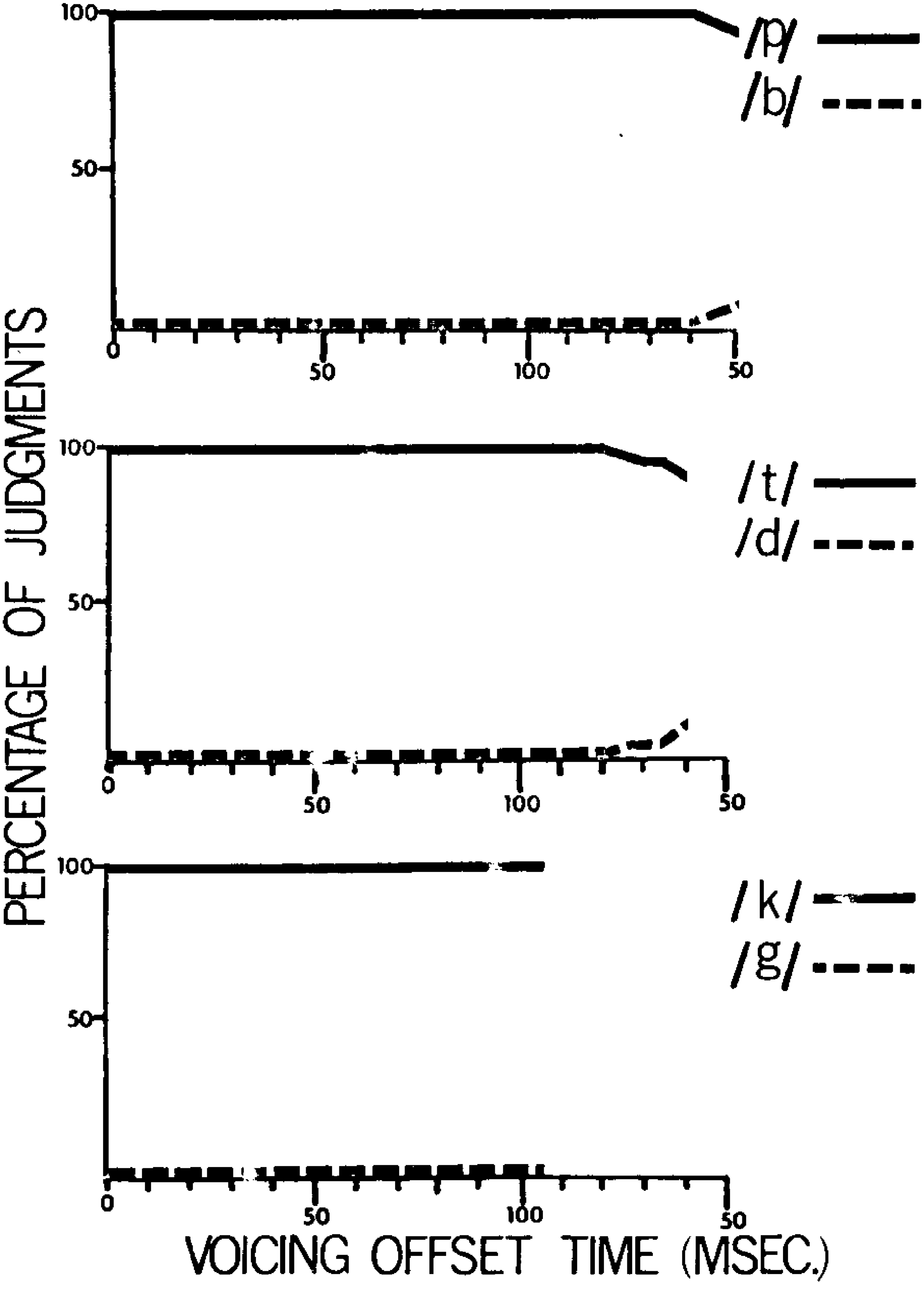
No significant changes in the listeners' perceptual judgments were found for any stimulus. That is, there were no instances of a majority of listeners hearing words originally ending with /b/, /d/, or /g/ as words terminated by /p/, /t/, or /k/, regardless of the amount of voicing during closure that had been removed and the resultant increase in Voicing Offset Time. Similarly, all of the unaltered items which ended in /ptk/ were perceived as originally intended.

Figures 2.14-2.16 reveal the almost total lack of change in perceptual judgments. The final /k-g/ opposition displayed no change whatsoever in percentage of listener identification: All stimuli were heard as originally recorded, whether they were altered or not. The final /t-d/ opposition yielded only six judgments (out of a possible 300) of /d/ as /t/. Two of these judgments were made of stimuli in which the voicing had been cut back 5 msec <u>into</u> the end of the vowel, another for a stimulus in which the voicing had been cut back <u>to</u> the end of vowel, and another of a stimulus in which only 5 msec of voicing during closure remained. The /p-b/ opposition yielded similar results. Of the two judgments (out of a possible 120) of /b/ heard as /p/, both were of items in which the voicing had been cut back 5 msec into the end of the vowel.

Discussion

The measure(s) of Voicing Offset Time/closure voicing duration does not appear to be a significant cue to the perception of final stops as belonging to the class /bdg/ or /ptk/. Any perceptual changes caused by increasing Voicing Offset Time (i.e. removing voicing during closure) are few in number and limited to the stimuli in which

PERCEPTION OF FINAL /b,d,g/ AS /p,t,k/
AS A FUNCTION OF VOICING OFFSET TIME



figures 2-14,15,16

voicing during closure has been virtually eliminated, or in which voicing has in fact been cut back into the vowel. These findings suggest that vowel duration may be a more potent cue to the categorization of word-final stops than the measure tested in this experiment.

Experiment III: Preceding Vowel Duration as a Cue to the Classification of the Members of Stop and Fricative Cognates in Word-Final Position: A Tape-Cutting Experiment

This experiment was designed to determine the effectiveness of preceding vowel duration as a cue to the categorization of word-final stop and fricative phonemes. From the corpus described above (see pp. 80-81 and Appendix 1) one minimal pair from each of the stop and fricative categories of place of articulation was chosen. The tape cutting and splicing procedure previously described (see pp. 108-13) was used, except that instead of measuring from the reference click to the offset of voicing, the distance (and duration) from the reference click to a point in the vowel just before the final transition was calculated and marked off. The cutting and splicing in this experiment removed progressively greater sections of the vowel in ten millisecond steps.

The words chosen, with one exception, all contained what are generally termed "lax" or "checked" vowels. Such vowels display a relatively steady-state formant structure and thus sound more natural when subjected to the cutting and splicing procedure than do "tense" or "free" vowels, whose formant structures often display large changes in frequency from beginning to end. The accuracy of all cuts was again verified by spectrographic analysis.

In this experiment only the stimuli which were originally recorded and perceived to end in /bdgvðzʒ/ were used in the listening tests. That is, the signal heard by the Ss was comprised of an initial consonant (or cluster), a vowel (usually shortened to some extent), and the complete, originally "voiced" consonant, with the transitions leading to it, as spoken by the principal informant.

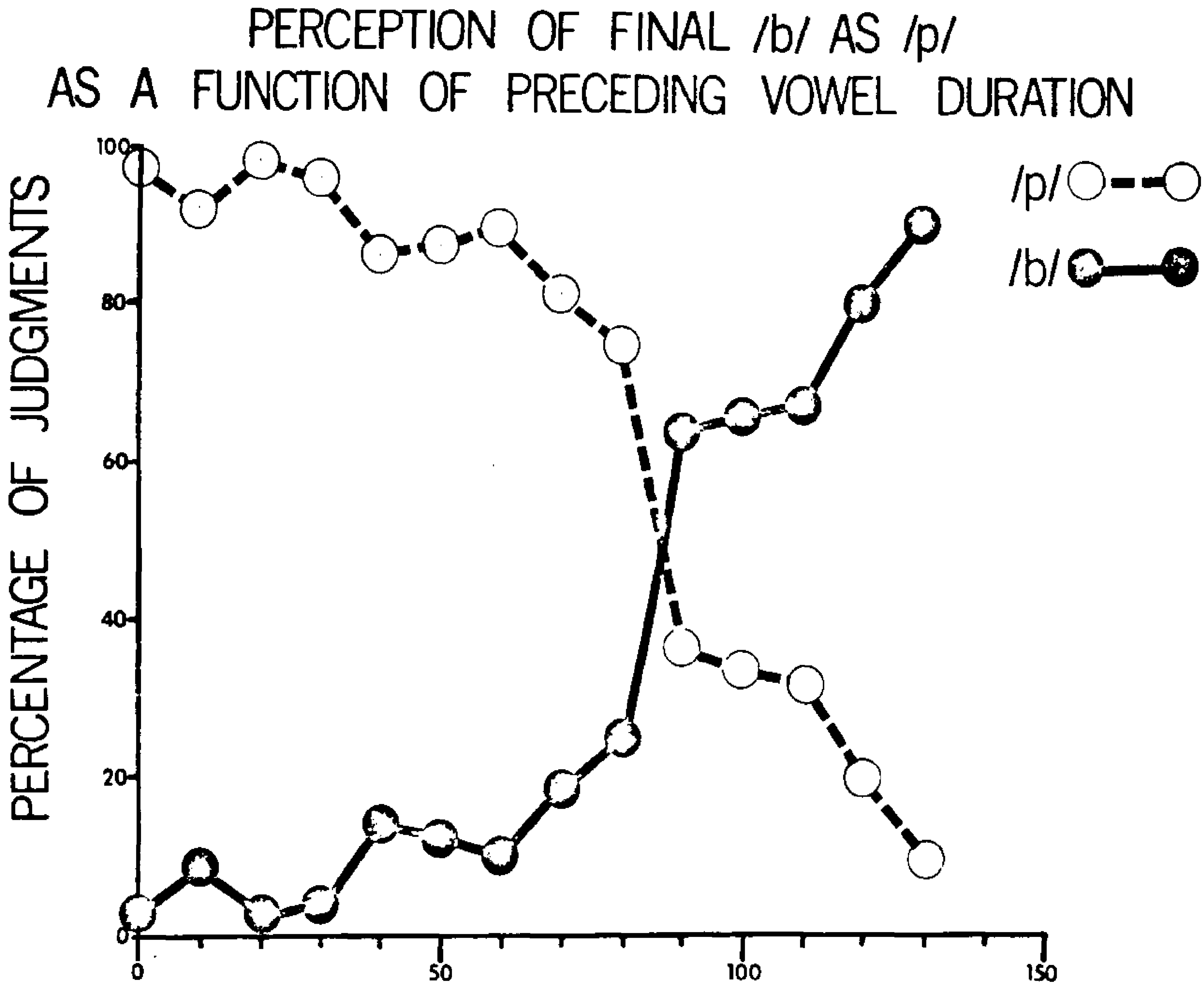
The stimuli were randomized and played over a loudspeaker for 100 phonetically-naive, native speakers of American English (standard New York City Dialect) on an Ampex 301 tape deck in a sound-treated room.

Results

Stops

Figures 2.17-2.19 indicate the results for the stimuli ending in /bdg/. A crossover in listener judgments was obtained for each place category of final stop consonant. In the case of the labials, once the vowel had been shortened by 90 msec, more Ss heard the final stop as /p/ than as /b/, the crossover point being approximately 86 msec. The crossover for /t-d/ was 75 msec, and for /k-g/, 82 msec.

It should be noted that the /t-d/ and /k-g/ judgments reconverge at the extreme limits of vowel reduction, there being a secondary crossover in the case of the



PERCEPTION OF FINAL /b/ AS /p/
AS A FUNCTION OF PRECEDING VOWEL DURATION

figure 2-17

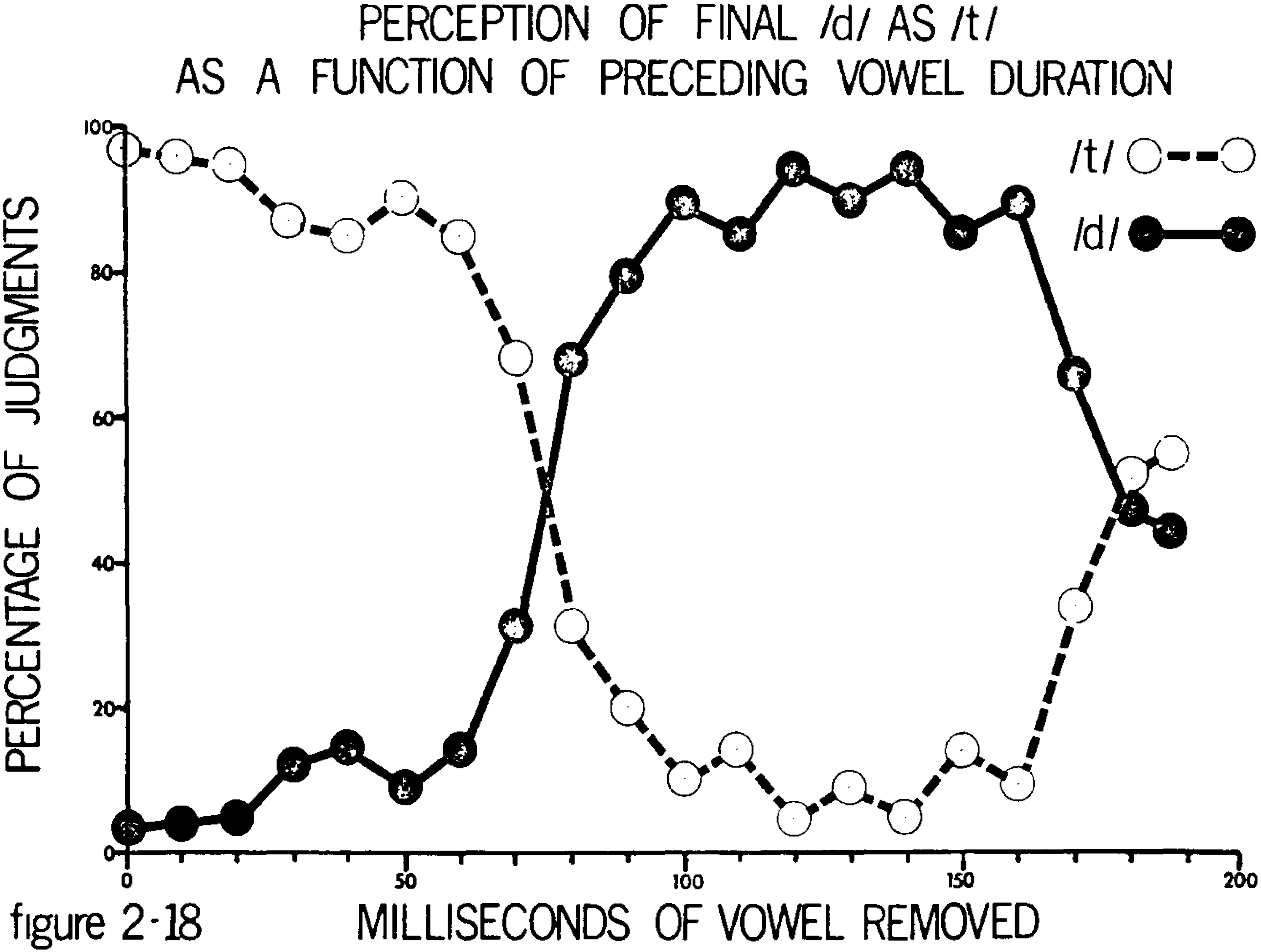

PERCEPTION OF FINAL /d/ AS /t/
AS A FUNCTION OF PRECEDING VOWEL DURATION
PERCENTAGE OF JUDGMENTS
100
80
60
40
20
0
0
50
100
150
200
MILLISECONDS OF VOWEL REMOVED
/t/
/d/

figure 2-18

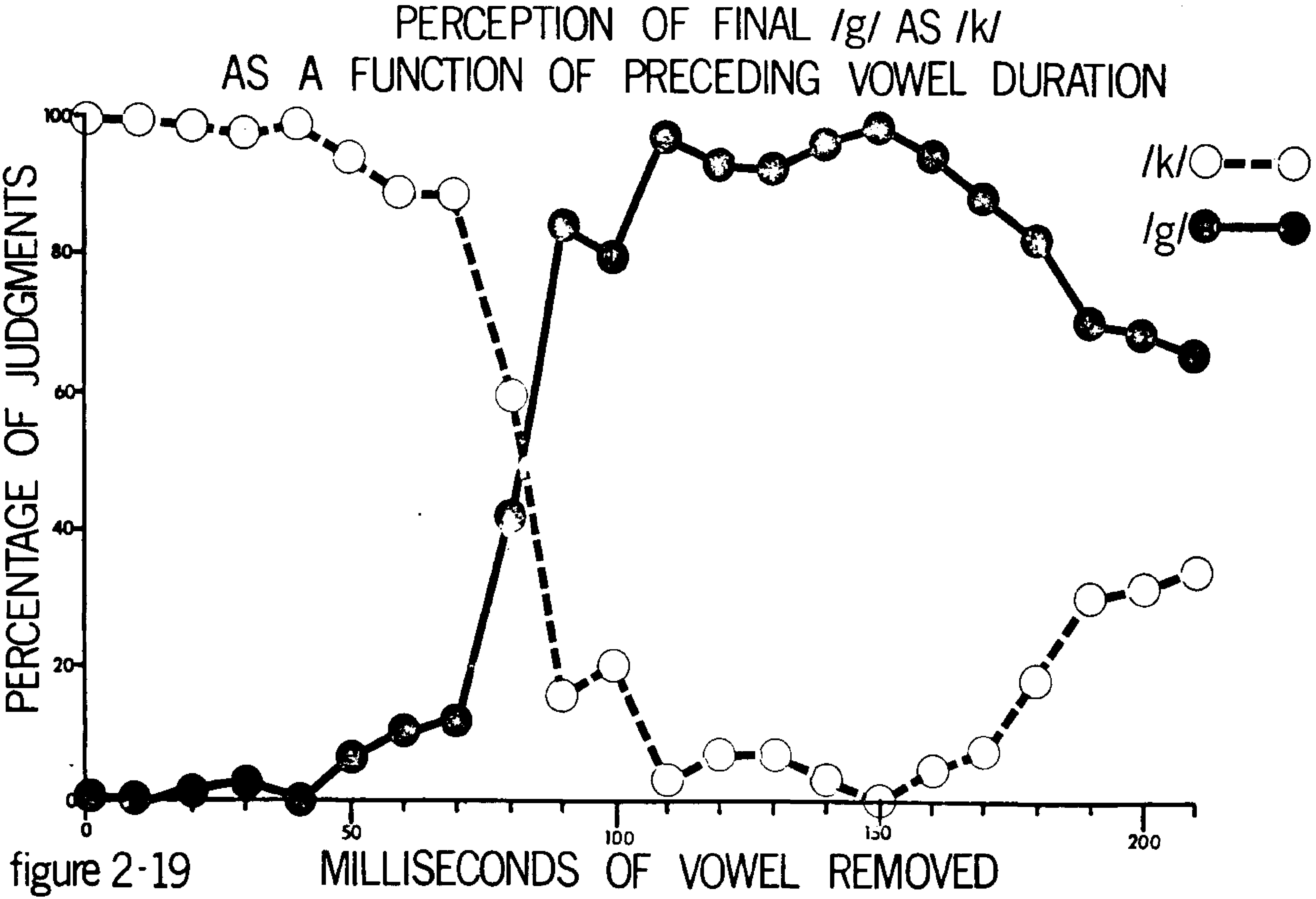

PERCEPTION OF FINAL /g/ AS /k/
AS A FUNCTION OF PRECEDING VOWEL DURATION
PERCENTAGE OF JUDGMENTS
100
80
60
40
20
0
0
50
100
150
200
MILLISECONDS OF VOWEL REMOVED
/k/ ○- - -○
/g/ ●——●

figure 2-19

alveolars. This "end effect" is not evident in the data for /p-b/, where the relatively short vowel (130 msec, as opposed to 185 and 215 msec for the other vowels before /d/ and /g/) may preclude its occurrence.

Fricatives

Figures 2.20-2.23 indicate the results for the stimuli ending in /vðzʒ/. It is clear that the fricative data differs somewhat from that for the stops, although there is a general similarity between the two manner classes.

The /f-v/ opposition is similar to that for the stops, except that there is no "end effect." Further, the crossover in judgments takes place only after a proportionately greater segment of the vowel has been removed than in the case of the stops.

The /ʃ-ʒ/ opposition resembles the stop results most closely. There is both a relatively early crossover of judgments and a reconvergence of judgments, resulting in a secondary crossover, near the extreme limits of the vowel reduction.

The /θ-ð/ and /s-z/ oppositions are the least dramatic in terms of crossover judgments. Such crossovers are sporadic and weak, and reflect, perhaps, an ambivalence of judgments and an ambiguity of stimuli rather than a real change in the perception of the listeners from one category to another.

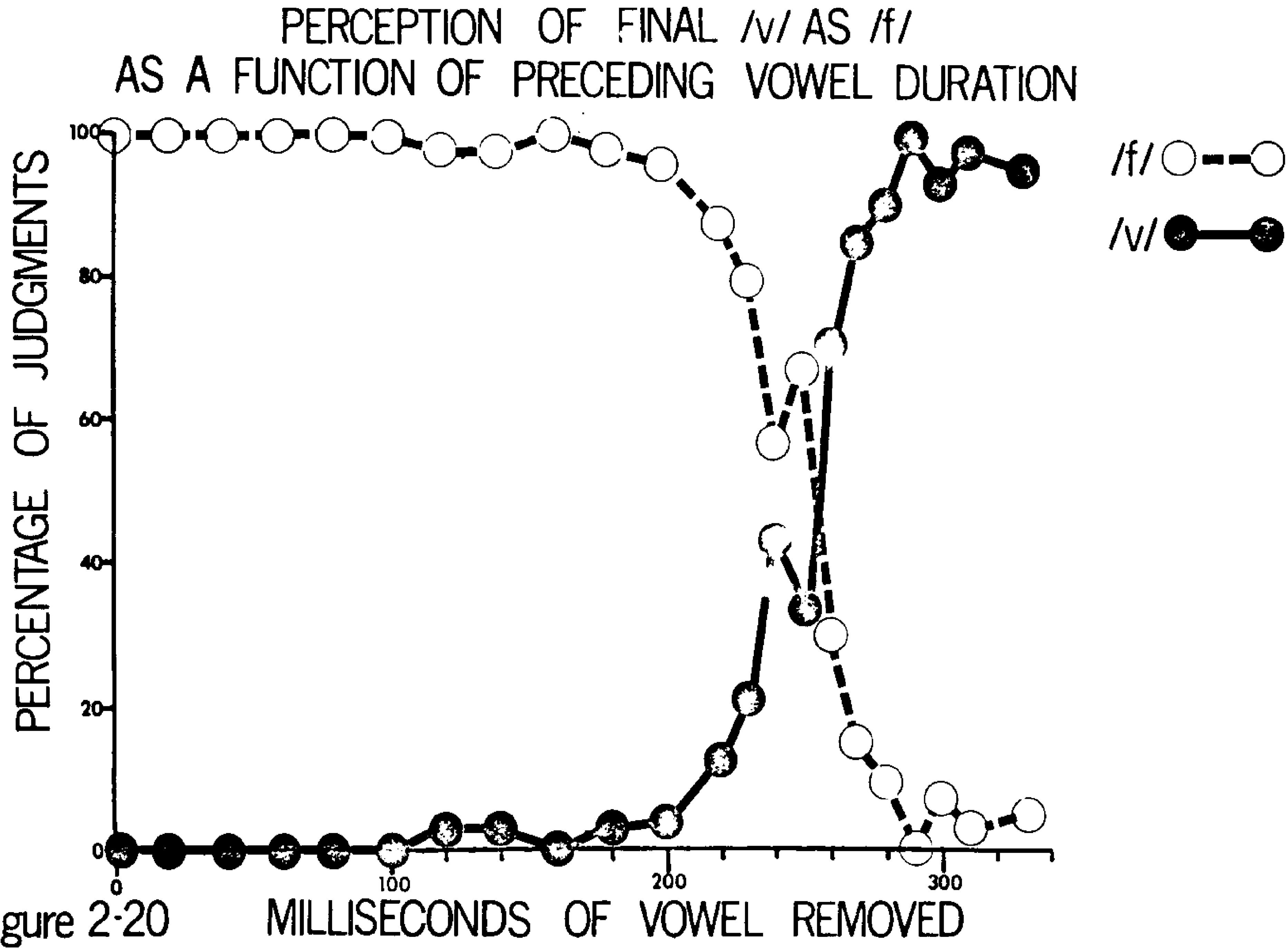

PERCEPTION OF FINAL /v/ AS /f/
AS A FUNCTION OF PRECEDING VOWEL DURATION
PERCENTAGE OF JUDGMENTS
100
80
60
40
20
0
0
100
200
300
MILLISECONDS OF VOWEL REMOVED
/f/
/v/

figure 2-20

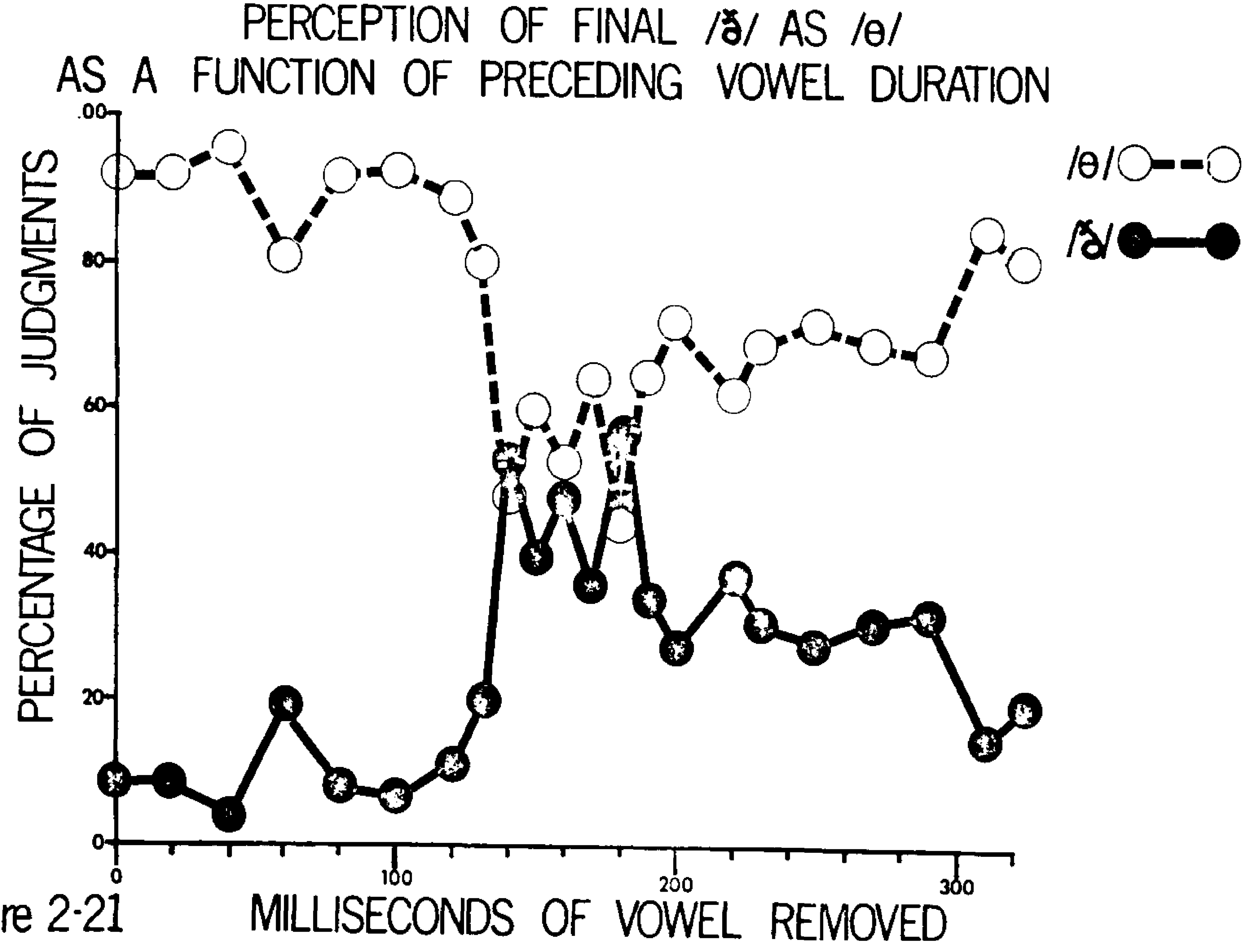

PERCEPTION OF FINAL /ð̌/ AS /θ/
AS A FUNCTION OF PRECEDING VOWEL DURATION
/θ/
/ð̌/
PERCENTAGE OF JUDGMENTS
.00
80
60
40
20
0
0
100
200
300
MILLISECONDS OF VOWEL REMOVED

figure 2-21

PERCEPTION OF FINAL /z/ AS /s/
AS A FUNCTION OF PRECEDING VOWEL DURATION

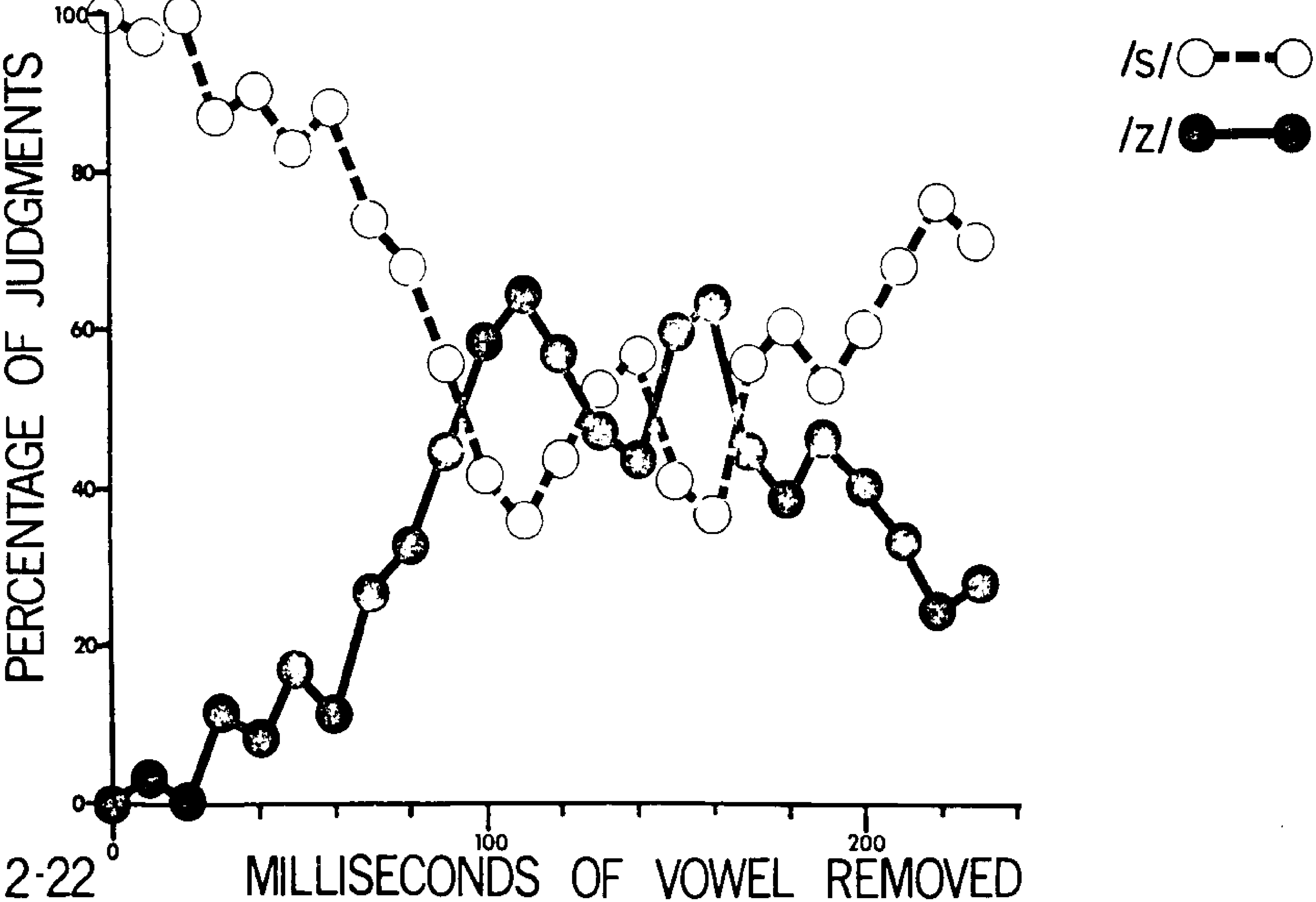


figure 2-22

PERCEPTION OF FINAL /ʒ/ AS /ʃ/
AS A FUNCTION OF PRECEDING VOWEL DURATION

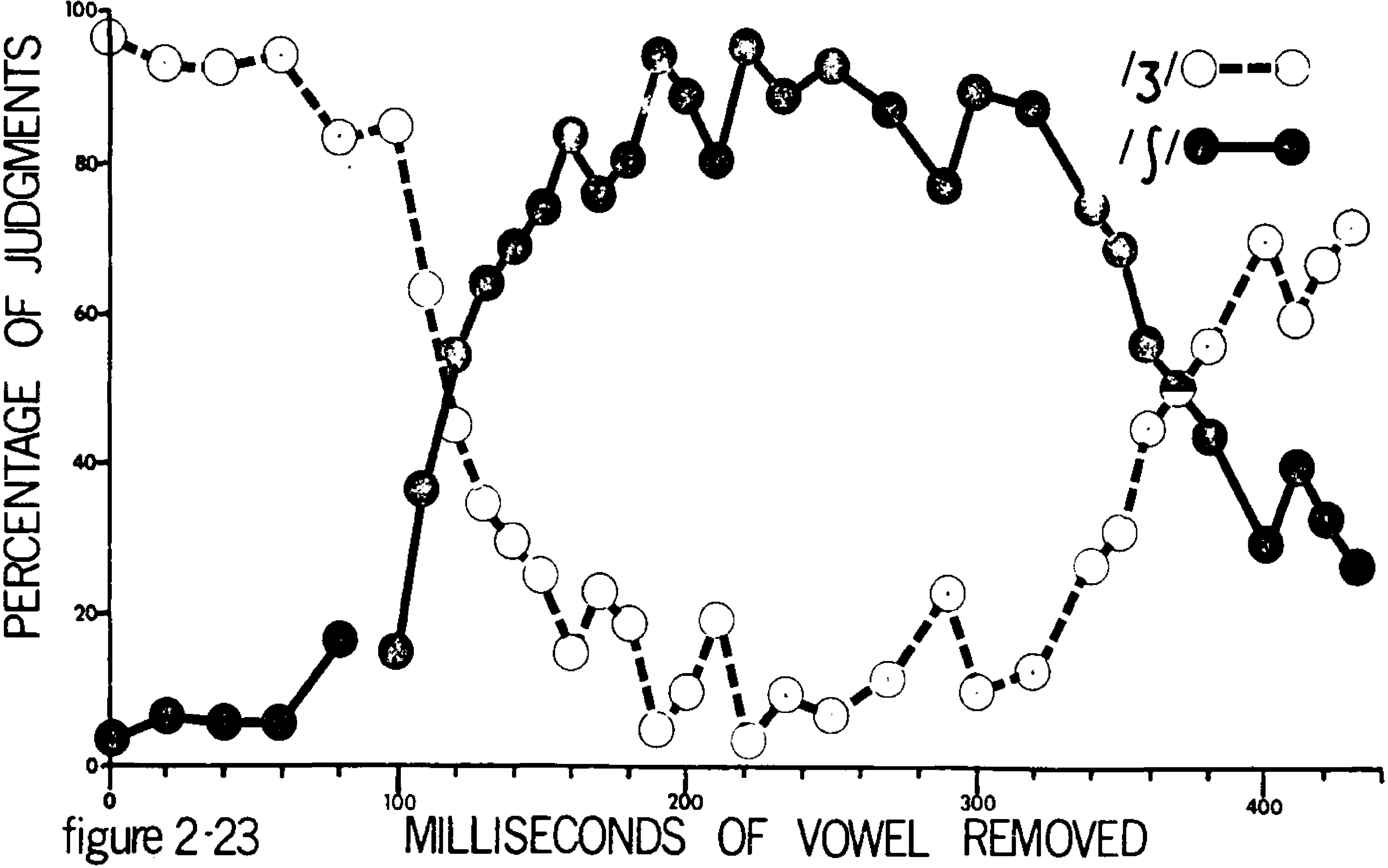


figure 2-23

Discussion

Preceding vowel duration, in general, does prove to be a sufficient cue to the perceptual categorization of the members of cognate phoneme pairs in final position of isolated words. The data indicate that the cue is more effective in the case of stops than in the case of fricatives. Fricatives generally require a proportionately more extensive reduction in preceding vowel duration than do stops, in order for perception to change from one category to another. Further, the members of each of the cognate pairs /θ-ð/ and /s-z/ do not exhibit a significant crossover in percentage of listener judgments. The sounds are merely confused as the vowel is shortened, indicating an uncertainty rather than a change in the perception of the subjects. The remaining fricative pairs and all of the stops, on the other hand, show a definite categorical change in judgment on the part of the listeners.

A possible explanation for the difference between stops and fricatives might be that the longer duration of closure for the fricative manner allows the voicing during closure to play a more prominent role in cueing the cognate opposition than in the case of stops. One might then expect that the preceding vowel would have to be reduced proportionately more in the case of fricatives in order to overcome the relatively stronger cue within the closure and bring about a perceptual change.

The reconvergence and secondary crossover of judgments noted for several of the stops and fricatives tested suggest that the parameter of voicing during closure/Voicing Offset Time may indeed be a cue of some significance. It seems possible that this cue, although a minor one as compared to preceding vowel duration, may assert itself once the vowel has been sufficiently shortened. That is, once the vowel is reduced almost to the point of elimination it may be too brief to be of any value as a cue, and so the listeners may be forced to rely on the secondary cue in making their judgments.

Two other observations support the hypothesis that Voicing Offset Time may be a cue of minor importance. First, when the stop closure in Experiment II was fully eliminated, a cut of 5 to 10 msec into the vowel did produce a slight shift in the percentage (10%) of judgments of the listeners for the /p-b/ and /t-d/ oppositions. In Experiment III, on the other hand, when the stop closure was left intact, such a slight reduction of the vowel caused a smaller change in the percentage of judgments. Second, if voicing during closure were of no significance whatsoever as a cue, one might expect the crossover in judgments to occur at that point where the vowels before /bdgvðzʒ/ had been reduced to the same duration as the vowels preceding /ptkfθsʃ/ in the corresponding member of each minimal pair. In fact, with one exception, the crossovers were attained only after the vowel had been reduced to a point where it was 10 to 140 msec shorter than the vowel preceding the other member of the cognate pair. In the exceptional case of the /θ-ð/ opposition, which showed a confusion rather than a clear change in perception, the first crossover in judgments occurred when the vowel was reduced to a point where its duration was equal to that of the vowel before /θ/.

The results of this experiment, then, indicate the relative importance of preceding vowel duration as a cue to the separation of the members of cognate phoneme pairs when such vowels are reduced before /bdgvðzʒ/. The tape cutting and splicing procedures employed here, however, do not allow for the testing of the reciprocal relation: the lengthening of vowels before /ptkfθsʃ/. In order to determine if such a manipulation of the acoustic signal would cue an analogous change in perception it was necessary to turn to speech synthesis.

Footnotes - Chapter II

[1] Leigh Lisker and Arthur S. Abramson, "A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements," Word, 20 (December, 1964), 387 ff.

[2] In the isolated words studied in these experiments no cases of voicing extending beyond the cessation of friction were found.

[3] See Table I-2, Chapter I, p. 22, for a comparison of these data with those of other investigators.

[4] If the data supplied by the unreleased stops is ignored, the averages for Voice Offset Time before /ptk/ and /bdg/ respectively are 70.8 and 23.0 msec. The ratio is then 3.1:1, which is quite similar to the 3.2:1 ratio found before fricatives. The exclusion of the data for unreleased stops might be justified on the grounds that the arbitrary zero value assigned to Voice Offset Time in these cases might not reflect the fact that closure was, in fact, released, although not audibly, or that some cue to the cessation of closure might not be evident on the spectrograms.

[5] William S-Y. Wang, "Transition and Release as Perceptual Cues for Final Plosives," Journal of Speech and Hearing Research, 2 (March, 1959), 71.

[6] The nature of the experimental procedure used precluded the investigation of Voice Offset Time/voicing during closure for fricatives. In cutting and splicing

the tape it would have been necessary to alter the duration of the friction, a duration which has been shown to exhibit significant cue value for the categorization of cognate fricatives in final position (see above, Chapter I, p. 18). This, of course, would have introduced a second variable to the experiment.

[7] In cases where there was a negative value for Voice Offset Time, a preliminary cut was made removing all voicing following the stop release.

[8] In this experiment, closure duration was essentially unaltered. This writer has performed a similar experiment in which blank tape was not used to replace the voicing during closure that was removed. This procedure progressively eliminates the voicing during closure and shortens the closure duration, but does not affect the Voice Offset Time. The results for that experiment were virtually identical with those reported here for the cue value of Voice Offset Time. That is, neither total closure duration nor the amount of voicing during closure proves to be a significant cue to cognate opposition for word-final stops.

CHAPTER III

VOWEL DURATION AS A CUE TO THE PERCEPTUAL SEPARATION OF WORD-FINAL COGNATE SOUNDS: SYNTHETIC-SPEECH EXPERIMENTS

The purpose of the experiment described in this chapter was to determine the effect of preceding vowel duration upon the perception of final consonants as being one or the other of the members of a cognate pair. This experiment employed synthetic speech and thus differed from the one described in the preceding chapter in that it was possible to explore the effects of varying vowel duration before sounds of the category /ptkfθsʃ/, as well as before the cognates of this category.

Further, this experiment tested the effect of varying vowel duration before word-final clusters of the types: stop-plus-stop; stop-plus-fricative; and fricative-plus-stop.

Using the results of the tape-cutting experiments as a basis, four hypotheses were developed and tested experimentally:

<u>Hypothesis I</u> predicted that final consonants and clusters regardless of the voicing cues used in their synthesis, would be perceived as voiced when preceded by vowels

of long duration and as voiceless when preceded by vowels of short duration.

Hypothesis II predicted that any changes in the perception of final consonants or clusters would be continuous rather than categorical. That is, the Ss in the experiment would be able to discriminate the test stimuli far better than they could label them. Thus, the discrimination functions for the stimuli would be at substantially higher levels, both across and within phoneme boundaries, than would be predicted from the labelling data. This hypothesis rests both on the results of the tape-cutting experiment, and on the assumption that the perceptual classification of the final consonants is in fact conditioned by the duration of the preceding vowels (Hypothesis I). Perceptual changes in vowel duration, as well as quality, have been shown to be continuous, rather than categorical.[1,2] Thus, in the case of final-position, post-vocalic consonants and clusters, perceptual changes will be predicted to be continuous.

Hypothesis III predicted that voicing during consonant closure would be of minor perceptual significance as a cue to the classification of word-final consonants. Once again, this assumption rests on the findings of the tape-cutting experiment, in which it was found that (1) in order to cause a perceptual change, vowels preceding voiced consonants had to be shortened to a point where their durations were less than those which preceded their voiceless cognates, and (2) once the vowel had been so shortened that its duration could no longer serve as a cue, perception tended to revert to the original categorization of the final sound as voiced. In the synthetic experiments it was expected that this cue value would be manifested by a difference between the durations of vowels needed to cause equivalent perceptions of sounds synthesized as cognates. That is, it was expected that a shorter vowel would be needed to cause a final consonant to be perceived as voiceless, when that final consonant had been synthesized with cues for voicing, than when cues for voicelessness had been employed. Conversely, it was expected that a longer vowel would be needed to cause a final consonant to be perceived as voiced, when that final consonant had been synthesized with cues for voicelessness, than when cues for voicing had been employed.

Hypothesis IV predicted that the varying of vowel duration would be more effective in causing perceptual changes before final stops than before final fricatives. That is, it was expected that the results would be similar to those obtained from the tape-cutting experiment in which the variations in vowel duration that caused perceptual changes before stops would be smaller than those before fricatives. Since the durations of vowels preceding clusters in real speech were more like those preceding stops than those preceding fricatives,[3] it was further predicted that the range of variation of vowel duration needed to cause perceptual changes in clusters would be more similar to that needed in the case of the stops than in the case of the fricatives.

Procedure

The synthesizer employed in these experiments was the Haskins Laboratories' Pattern Playback. The Pattern Playback is a device that "generates speech by converting hand-painted spectrograms into sound."[4]

The stimuli produced by the Playback varied with the type(s) of consonants being tested. The final stops were synthesized by using a three-formant, steady-state vowel with 50 msec transitions appropriate to the place and voicing characteristic desired for the (initial and) final consonant(s).[5] A typical stimulus pattern for a syllable ending in a voiced stop is shown in Figure 3-1.

A series of stimuli of the type shown in Figure 3-1 with vowel durations varying from 150 to 350 msec was prepared. The range of vowel durations synthesized was determined from both the experiments reported in Chapter II and from studies of vowel duration reported in the literature reviewed in Chapter I. All final stops were synthesized

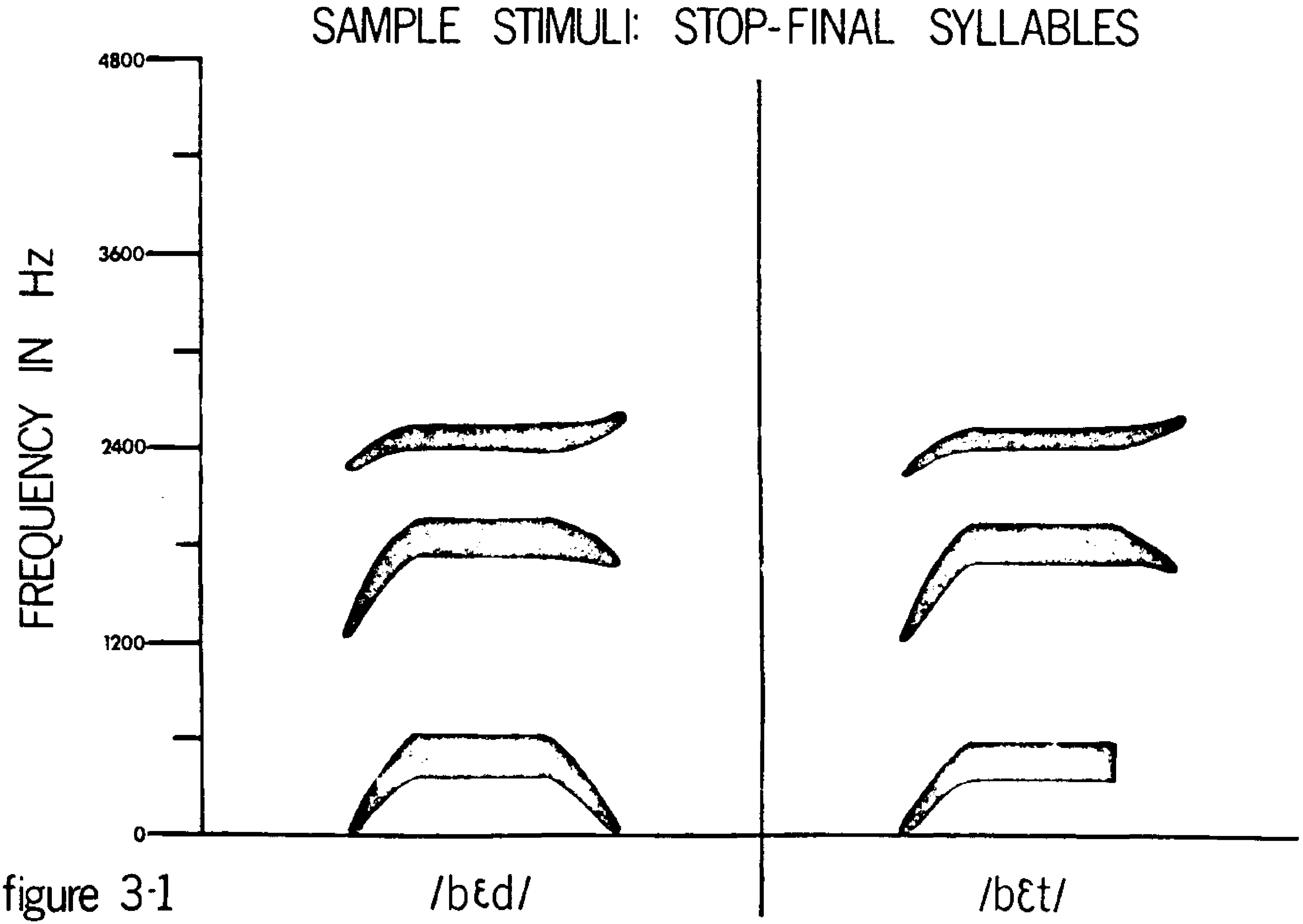

SAMPLE STIMULI: STOP-FINAL SYLLABLES
FREQUENCY IN Hz
4800
3600
2400
1200
0
/bɛd/
/bɛt/

figure 3-1

without burst-releases.[6]

After recording the series of "voiced" stimuli, each member of the series was converted to a "voiceless" stimulus by eliminating the final 50 msec F1 transition. The new voiceless series of stimuli was then recorded (see Figure 3-1).

Final fricatives were synthesized by random stippling in the frequency ranges appropriate to the desired place of articulation as reported in the Haskins literature. Fig. 3-2 shows a pattern used for a stimulus ending in a voiced fricative. The duration of the friction patch was 70 msec and was accompanied by a voicing bar in the lowest harmonic produced by the playback. After the voiced stimuli had been recorded, with vowel durations ranging from 200 to 400 msec, they were converted to stimuli ending in voiceless fricatives by a three-step process: (1) the voicing bar beneath the friction patch was removed; (2) the F1 transition was removed; (3) 80 msec of friction in the appropriate frequency range was added to the friction patch, making its total duration 150 msec (see Fig. 3-2).

Final clusters were synthesized in one of three ways, depending on whether they were comprised of stop-plus-stop, stop-plus-fricative, or fricative-plus-stop. The two syllables ending in voiced stop-plus-stop clusters were synthesized exactly as were those ending in simple stops, with the addition of a 100 msec silent interval after the first stop of the cluster, followed in turn by a release in

SAMPLE STIMULI: FRICATIVE-FINAL SYLLABLES

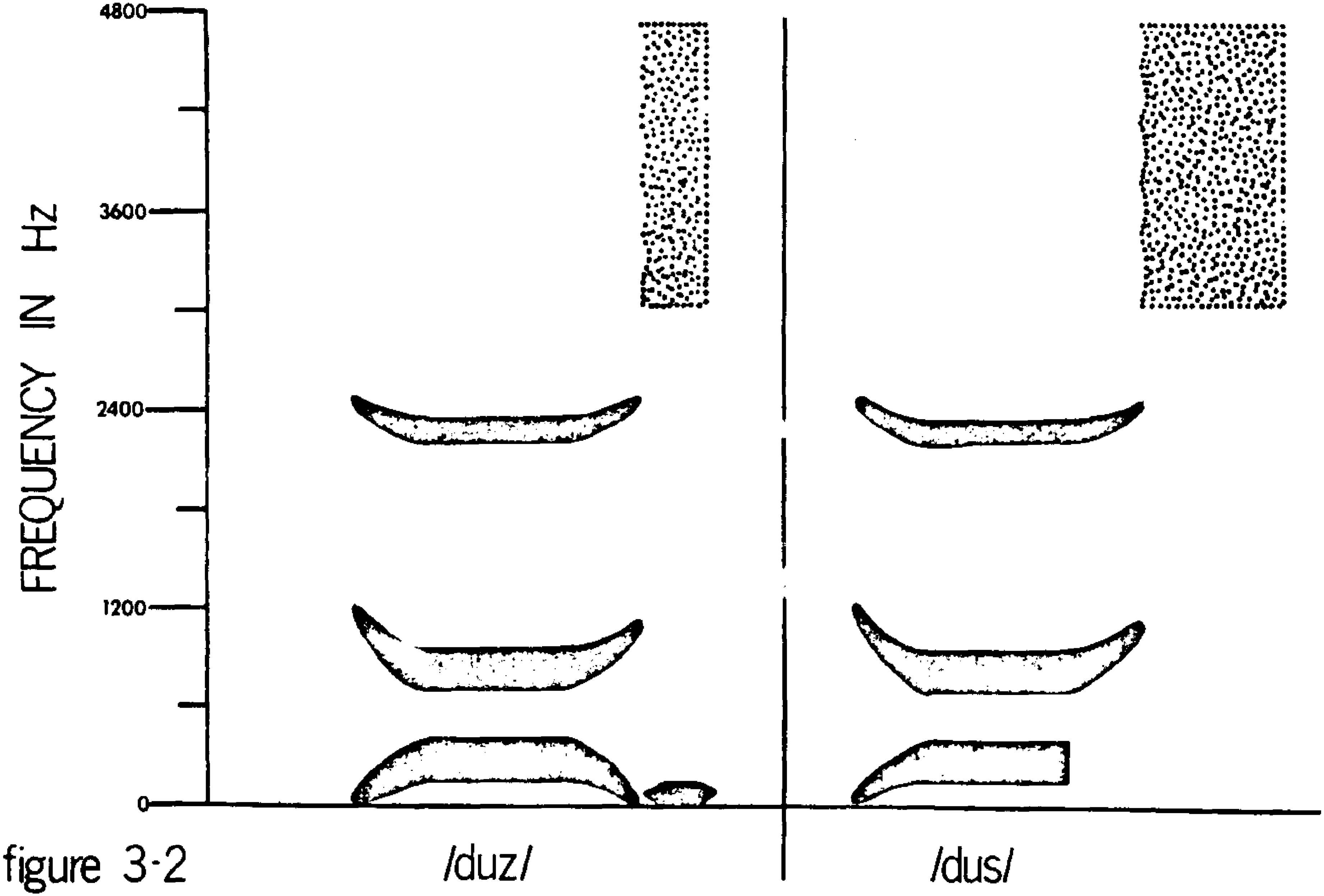


figure 3-2

the form of embryonic transitions of 30 msec duration. These transitions were appropriate to the place of articulation of the second consonant in the cluster and pointed toward a following, non-existent, schwa-like vowel. Figure 3-3 shows the pattern used to synthesize a typical syllable ending in a voiced stop-plus-stop cluster. After these stimuli had been recorded, they were converted to syllables ending in voiceless clusters by: (1) removing the final F1 transition of the vowel of the syllable; (2) removing the F1 transition after the final consonant of the cluster; and (3) changing the final embryonic transitions from solid to dotted lines in order to simulate aspiration (see Figure 3-3).

The three syllables ending with voiced stop-plus-fricative clusters were synthesized as were those ending in simple stops, with the addition of a 100 msec silent gap after the stop, followed by the voice bar and random stippling appropriate for the final fricative as described above in the synthesis of final fricative stimuli (see Figure 3-4). After being recorded, these stimuli were converted to syllables ending in voiceless stop-plus-fricative clusters by (1) removing the 50 msec F1 transition of the stop in the cluster; (2) removing the voicing bar beneath the friction patch; and (3) extending the duration of the friction patch to one appropriate to a voiceless fricative. The friction patch was extended to the right, thus maintaining the silent interval at a duration of 100 msec (see

SAMPLE STIMULI: STOP+STOP-FINAL SYLLABLES

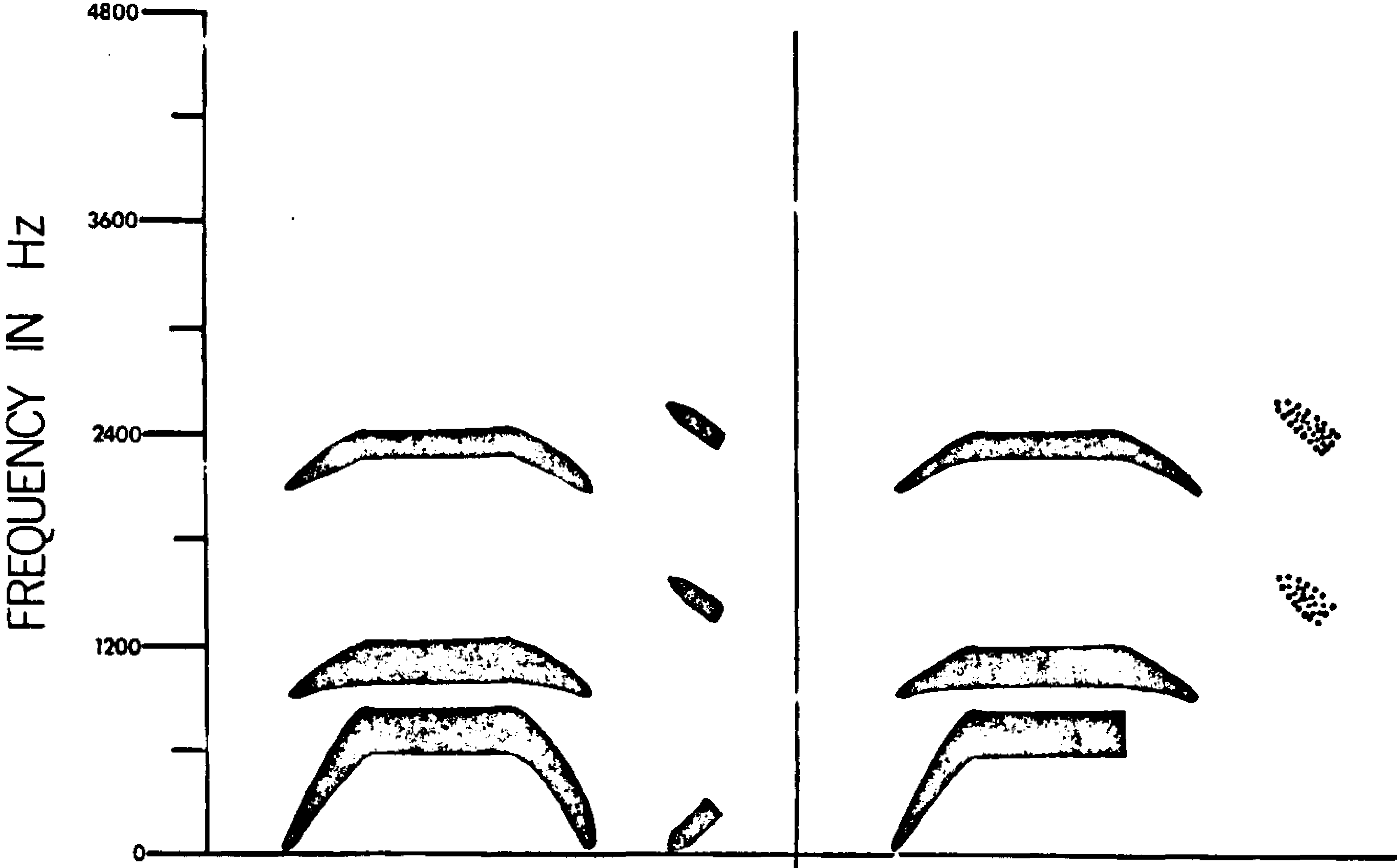


figure 3-3

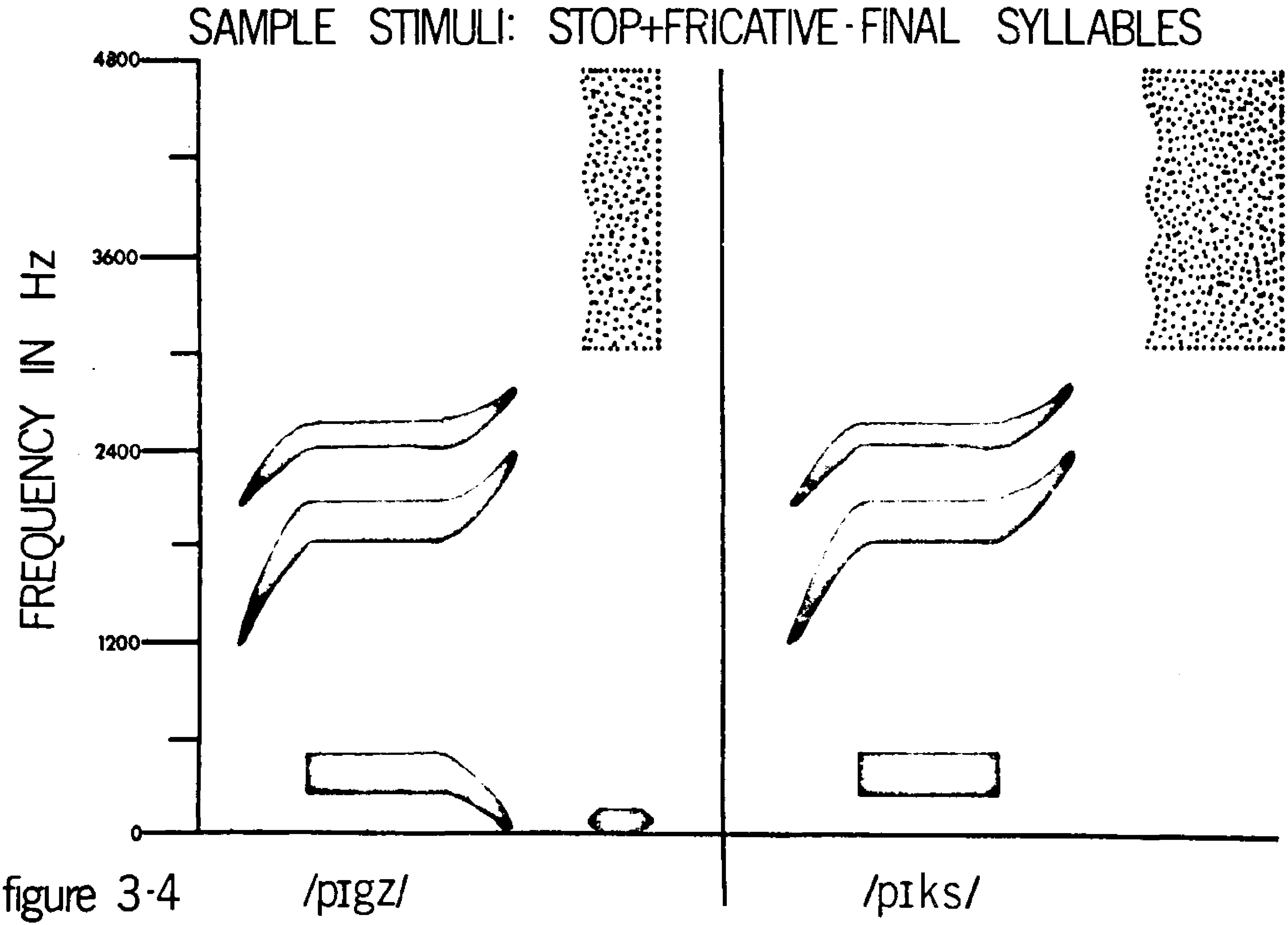

SAMPLE STIMULI: STOP+FRICATIVE-FINAL SYLLABLES

figure 3-4

Figure 3-4).

The single syllable ending in a voiced fricative-plus-stop cluster was synthesized as were those ending with a simple fricative, with the addition of a silent interval of 100 msec followed by the embryonic transitions described above in the synthesis of the final voiced stop-plus-stop clusters (see Figure 3-5). After recording, the conversion to voiceless final clusters was accomplished by: (1) appropriate lengthening of the friction patch; (2) removal of the voice bar beneath the friction patch; (3) removal of the embryonic F1 transition; and (4) changing the solid F2 and F3 transitions to dotted ones to simulate aspiration. In the case of these stimuli the final F2 and F3 transitions were removed and repainted to maintain the duration of the silent interval for the stop closure at 100 msec (see Figure 3-5).

The range of vowel durations synthesized before the final clusters, as determined from the results of spectrographic analysis, varied from one set of stimuli to another. The diphthong in the minimal pair tights-tides ranged from 200 to 400 msec. The lax vowel in the tucked-tugged minimal pair ranged from 100 to 300 msec. The other vowels in the final cluster stimuli were varied from 150 to 350 msec, the same range used before syllables ending in simple stops.

The stimuli chosen were all intended to represent a corpus of real English words. The test corpus and the

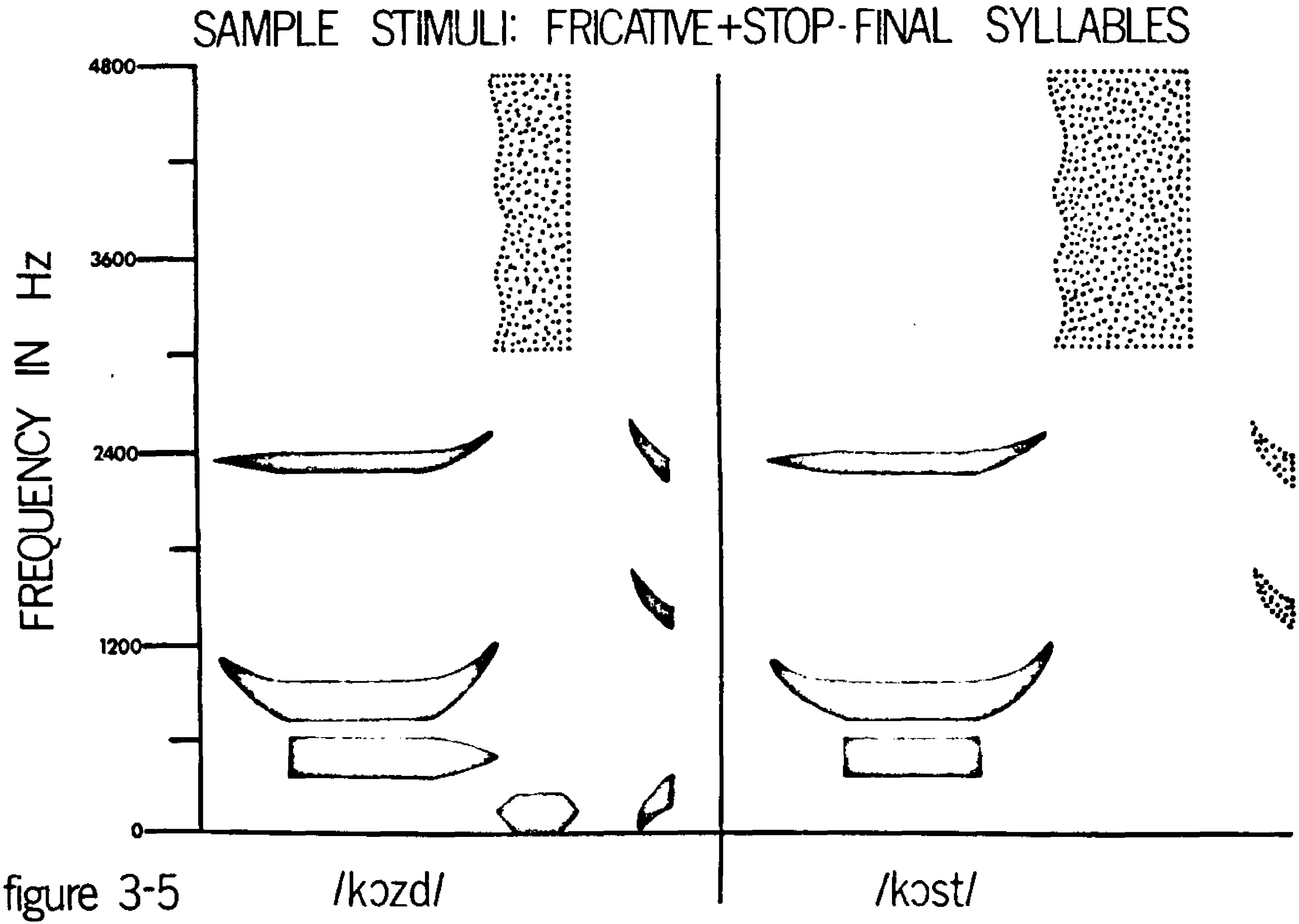

SAMPLE STIMULI: FRICATIVE+STOP-FINAL SYLLABLES
FREQUENCY IN Hz
4800
3600
2400
1200
0
/kɔzd/
/kɔst/

figure 3-5

oppositions tested are shown in Table III-1. The use of stops exclusively as the only initial sounds in the stimuli tended to restrict somewhat the choice of minimal pairs available, but outside of the relatively artificial pair <u>cash</u> /kæʃ/-<u>cas</u>(ual) /kæʒ--/ the subjects reported no difficulty in accepting the stimuli as real words.

Tables III-2 through III-6 give the full specifications of all the stimuli used in this experiment.

An attempt was made to use a representative sample of all vowels from the traditional categories: (1) front-central-back; (2) high-mid-low; (3) tense-lax. Figure 3-6 gives a detailed breakdown of the classification of the vowels used in the synthetic stimuli. In addition to the simple vowels, one diphthong, /aI/, was used. This diphthong was chosen as representative of its class because it appears commonly in minimal pairs.

A similar attempt was made to balance the classes of stops initiating each of the stimuli. Table III-7 shows the classes of stops represented according to place of articulation and voicing characteristic.

Each of the series of stimuli was randomized and placed into the following order on the test tape:

(1) Syllables ending in voiced stops. /bdg/

(2) Syllables ending in voiced fricatives. /vðzʒ/

(3) Syllables ending in voiced clusters. /bd, zd, bz, gd, dz, gz/

(4) Syllables ending in voiceless stops. /ptk/

TABLE III-1

TEST CORPUS AND OPPOSITIONS TESTED

Stops	
Gape-gabe	/p-b/
Bet-bed	/t-d/
Burke-berg	/k-g/
Fricatives	
Duff-dove	/f-v/
Teeth-teethe	/θ-ð/
Deuce-dues	/s-z/
Cash-cas(ual)	/ʃ-ʒ/
Clusters	
Bopped-bobbed	/pt-bd/
Tucked-tugged	/kt-gd/
Cost-caused	/st-zd/
Cops-cobs	/ps-bz/
Tights-tides	/ts-dz/
Picks-pigs	/ks-gz/

TABLE III-2

SPECIFICATION OF STIMULI USED FOR SYNTHETIC STOP-FINAL SYLLABLES[f]

Stimulus	t1[a]	t2	t3	F1	F2	F3	T1[b]	T2	T3
/geb/	0	2640	2760	420[c]	2220	2520	0	1480	2160
/gep/	0	2640	2760	420	2220	2520	-	1480	2160
/bɪd/	0	1320	2280	540	1920	2520	0	1860	2680
/bɪt/	0	1320	2280	540	1920	2520	-	1860	2680
/bɝg/	0	900	1800	600-480[d]	1260-1200[d]	2160-1380[d]	0	2040	+[e]
/bɝk/	0	900	1800	600-480[d]	1260-1200[d]	2160-1380[d]	-	2040	+[e]

[a] t: Transition for initial consonant.

[b] T: Transition for final consonant. t, T values are terminating transition frequencies, not loci.

[c] F values are for the central formant frequency. F1 and F2 were painted with a bandwidth of two harmonics (240 Hz.), F3 with a bandwidth of one harmonic (120 Hz.).

[d] There were 60 msec of steady-state at the second of the two frequencies listed, before the final transition.

[e] A slight positive transition of 10-20 msec was used. Because of the steep positive T2 it was not possible to use a full 50 msec final transition.

[f] The range of vowel durations synthesized before stops was 150-350 msec.

TABLE III-3

SPECIFICATIONS OF STIMULI USED TO SYNTHESIZE FRICATIVE-FINAL SYLLABLES[f]

Stimulus	t1[a]	t2	t3	F1	F2	F3	T1[b]	T2	T3	Friction Range	Friction Duration[c]	120 Voice Bar Below Friction[d]
/dʌv/	0	1560	2580	660[e]	1200	2400	660	900	2100	960 - 2760	70 msec	yes
/dʌf/	0	1560	2580	660	1200	2400	-	900	2100	960 - 2760	150 msec	no
/tið/	-	2100	2880	360	2400	3000	360	1920	2800	960 - 2760	70 msec	yes
/tiθ/	-	2100	2880	360	2400	3000	-	1920	2800	960 - 2760	150 msec	no
/duz/	0	1320	2520	300	840	2280	300	1200	2520	3000 - 4800	70 msec	yes
/dus/	0	1320	2520	300	840	2280	-	1200	2520	3000 - 4800	150 msec	no
/kæʒ/	-	2400	2520	780	1740	2400	780	1680	2580	1920 - 3600	70 msec	yes
/kæʃ/	-	2400	2520	780	1740	2400	-	1680	2580	1920 - 3600	150 msec	no

[a] t: Transition for initial consonant.

[b] T: Transition for final consonant. t, T values are terminating transition frequencies, not loci.

TABLE III-3..Continued

---

[c] The edges of the friction patches were painted unevenly to avoid inducing stop or affricated stop perception.

[d] The voice bar was painted as a formant tapered at both ends.

[e] F values are for the central formant frequency. F1 and F2 were painted with a bandwidth of two harmonics (240 Hz.), F3 with a bandwidth of one harmonic (120 Hz.).

[f] The range of vowel durations synthesized before final fricatives was 150-400 msec.

TABLE III-4

SPECIFICATION OF STIMULI USED TO SYNTHESIZE SYLLABLES ENDING IN STOP-PLUS-STOP CLUSTERS

Stimulus	t1[a]	t2	t3	F1	F2	F3	T1[b]	T2	T3	Closure (Silent) Duration	30 msec T1[c]	30 msec T2	30 msec T3	Range of Vowel Durations
/bæbd/	0	900	2040	720[d]	1080	2400	0	900	2040	100 msec	0 (solid)	1560 (solid)	2520 (solid)	150-350 msec
/bæpt/	0	900	2040	720	1080	2400	-	900	2040	100 msec	-	1560 (dotted)	2520 (dotted)	150-350 msec
/tʌgd/	-	1560	2520	600	1200	2400	0	2160	2400	100 msec	0 (solid)	1560 (solid)	2520 (solid)	100-300 msec
/tʌkt/	-	1560	2520	600	1200	2400	-	2160	2400	100 msec	-	1560 (dotted)	2520 (dotted)	100-300 msec

[a] t: Transition for initial consonant.

[b] T: Transition for first consonant of cluster.

[c] T: Transition after release of second consonant of cluster. t, T, T values are terminating transition frequencies, not loci.

[d] F values are for the central formant frequency. F1 and F2 were painted with a bandwidth of two harmonics (240 Hz.), F3 with a bandwidth of one harmonic (120 Hz.).

TABLE III-5

SPECIFICATIONS OF STIMULI USED TO SYNTHESIZE SYLLABLES ENDING IN STOP-PLUS-FRICATIVE CLUSTERS

Stimulus	t1[a]	t2[a]	t3[a]	F1	F2	F3	T1[b]	T2	T3	Closure (Silent) Duration	Friction Range	Friction Duration	120 Voice Bar Below Friction	Range of Vowel Durations
/kabs/	-	1200	2580	720[c]	1080	2400	0	900	2040	100 msec	3000 - 4800	70 msec	yes	150-350 msec
/kaps/	-	1200	2580	720	1080	2400	-	900	2040	100 msec	3000 - 4800	150 msec	no	150-350 msec
/taɪdz/	-	1400	2560	*760-360	*1080-2040	*2400-2640	0	1920	2640	100 msec	3000 - 4800	70 msec	yes	200-400 msec
/taɪts/	-	1400	2560	*760-360	*1080-2040	*2400-2640	-	1920	2640	100 msec	3000 - 4800	150 msec	no	200-400 msec
/pɪgz/	-	1200	2160	390	1980	2520	0	2400	2760	100 msec	3000 - 4800	70 msec	yes	150-350 msec
/pɪks/	-	1200	2160	390	1980	2520	-	2400	2760	100 msec	3000 - 4800	150 msec	no	150-350 msec

[a] t: Transition for initial consonant.

[b] T: Transition of stop in cluster. t, T values are terminating transition frequencies, not loci.

[c] F values are for the central formant frequency. F1 and F2 were painted with a bandwidth of two harmonics (240 Hz.), F3 with a bandwidth of one harmonic (120 Hz.).

*Starting frequencies of formants of the diphthong /aI/ had a steady state of 50 msec; final frequencies held at steady state for 20 msec.

TABLE III-6

SPECIFICATIONS OF STIMULI USED TO SYNTHESIZE SYLLABLES ENDING IN FRICATIVE-PLUS-STOP CLUSTERS[e]

Stimulus	t1[a]	t2	t3	F1	F2	F3	T1[b]	T2	T3	Friction Range	Friction Duration	120 Voice Bar Below Friction	Closure (Silent) Duration	T1[c]	T2	T3
/kɔzd/	-	1080	2400	600[d]	840	2400	600	1200	2580	3000-4800	75 msec	yes	100 msec	0 (solid)	1560 (solid)	2520 (solid)
/kost/	-	1080	2400	600	840	2400	-	1200	2580	3000-4800	150 msec	no	100 msec	-	1560 (dotted)	2520 (dotted)

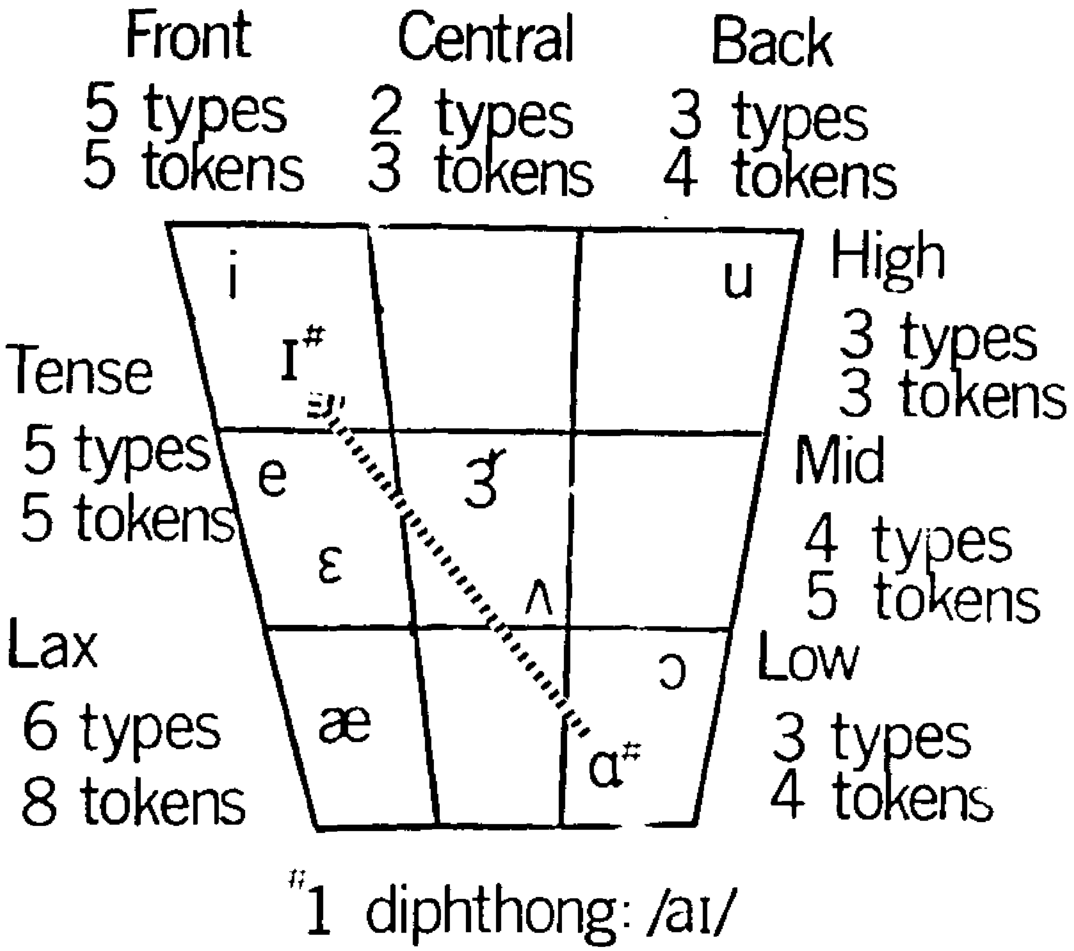
[a] t: Transition for initial consonant.

[b] T: Transition for first consonant of cluster.

[c] T: Transition after release of second consonant of cluster. t, T, T values are terminating frequencies of transitions, not loci.

[d] F values are for the central formant frequency. F1 and F2 were painted with a bandwidth of two harmonics (240 Hz.), F3 with a bandwidth of one harmonic (120 Hz.).

[e] The range of vowel durations synthesized before the fricative-plus-stop clusters was 150-350 msec.



ANALYSIS OF FREQUENCY OF OCCURRENCE OF VOWELS USED IN TEST STIMULI

figure 3-6

TABLE III-7

ANALYSIS OF FREQUENCY OF OCCURRENCE OF STOPS USED TO INITIATE TEST-STIMULI

/p/ - 1	/t/ - 3	/k/ - 3	Voiceless - 7
/b/ - 3	/d/ - 2	/g/ - 1	Voiced - 6
Bilabial - 4	Alveolar - 5	Velar - 4	

(5) Syllables ending in voiceless fricatives./fθsʃ/

(6) Syllables ending in voiceless clusters. /pt, st, ps, kt, ts, ks/

The test tapes were then played to 25 undergraduate students of Hunter College of the City University of New York. The students were enrolled in an advanced course in phonetics. There were 20 females and 5 males, all with normal hearing, in the group. The subjects were told only that they would hear a series of words produced synthetically and that they would have to judge, for each stimulus they heard, whether it was one or the other of the two choices on their response sheets. Sample stimuli were played for the subjects in order to familiarize them with the type of signal produced by the pattern playback. There was a five-minute rest period between the playing of the voiced stimuli and the playing of the voiceless stimuli.

The subjects heard the stimuli on a loudspeaker in the sound-treated speech science laboratory at Hunter College. The tapes were played on an Ampex AG-500 machine at 7-1/2 i.p.s.

Results and Discussion

Hypothesis I

The experimental results confirm Hypothesis I, with one exception. All final consonants and clusters among the stimuli, regardless of the cues used in their synthesis, were perceived as voiceless when preceded by vowels of short duration and voiced when preceded by vowels of long duration.

That is, a final consonant or cluster synthesized with the cues for voicing would be perceived as voiceless when the vowel preceding it was of short duration, and voiced when the preceding vowel was of long duration. A final cluster or consonant synthesized with the cues for voicelessness was perceived in exactly the same way.

Figures 3-7 through 3-12b show the percentage of listener judgments of the consonants and clusters heard as voiceless as a function of the preceding vowel duration.

The exception to this result was found in the case of the /ʃ/-/ʒ/ opposition. As can be seen in Figure 3-13, although there is a general trend on the part of the subjects toward perceiving the stimuli with longer vowels as ending in voiced sounds, the overall picture is one of confusion of perception, rather than change in perception of the type seen in the other consonants and clusters. The subjects never perceived any stimulus synthesized as ending in /ʒ/ more than 64% of the time as ending in /ʃ/, or more than 76% of the time as ending in /ʒ/. Of the stimuli ending in /ʃ/, none was ever perceived more than 80% of the time as ending in /ʃ/, or more than 88% of the time as ending in /ʒ/.

It is true, however, that the peaks in the perception of /ʃ/ occurred near the short end of the vowel duration range, and the peaks of the perceptions of /ʒ/ near the long end of the range. Nevertheless, it is clear that the subjects were never fully convinced as a group that any

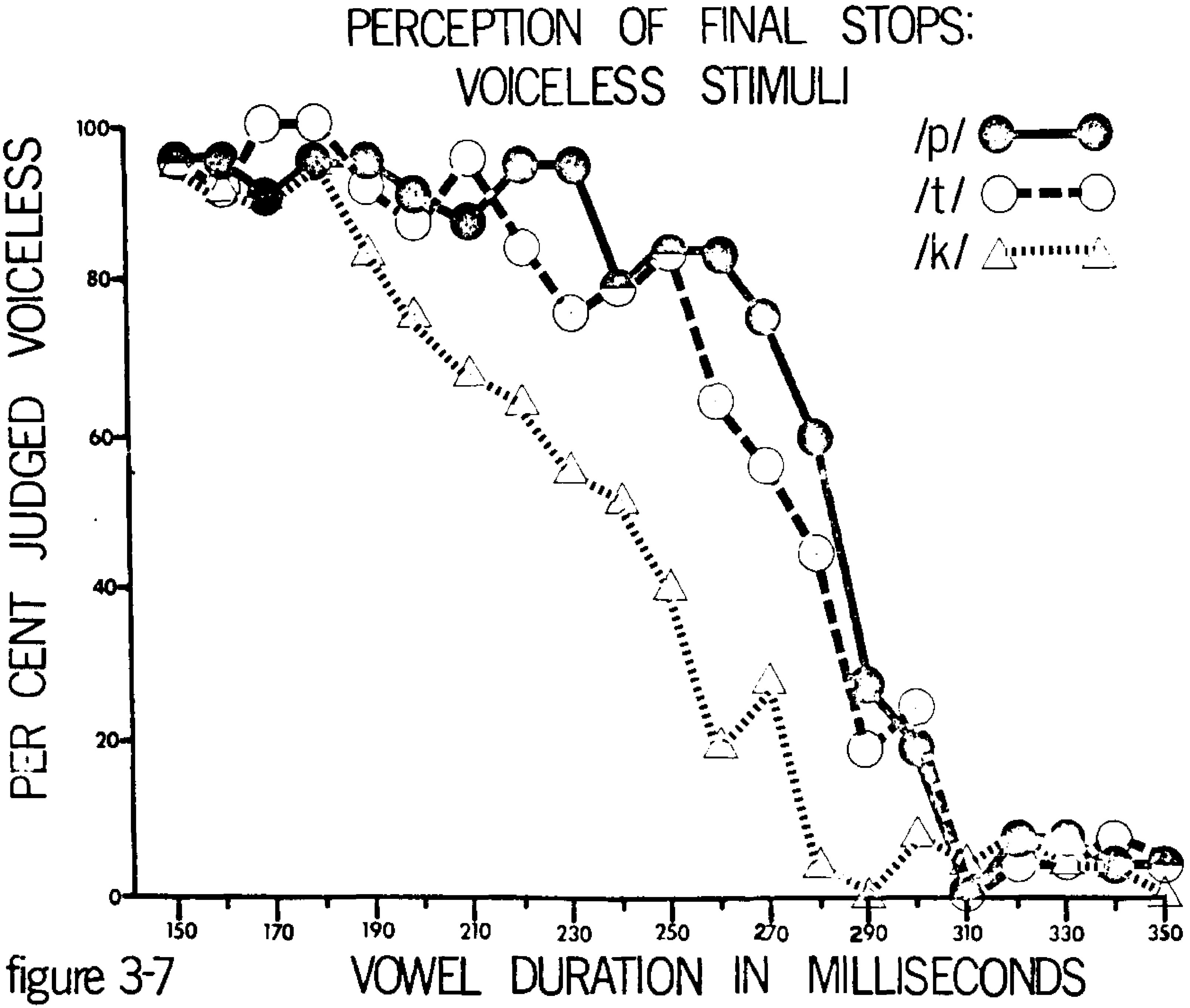
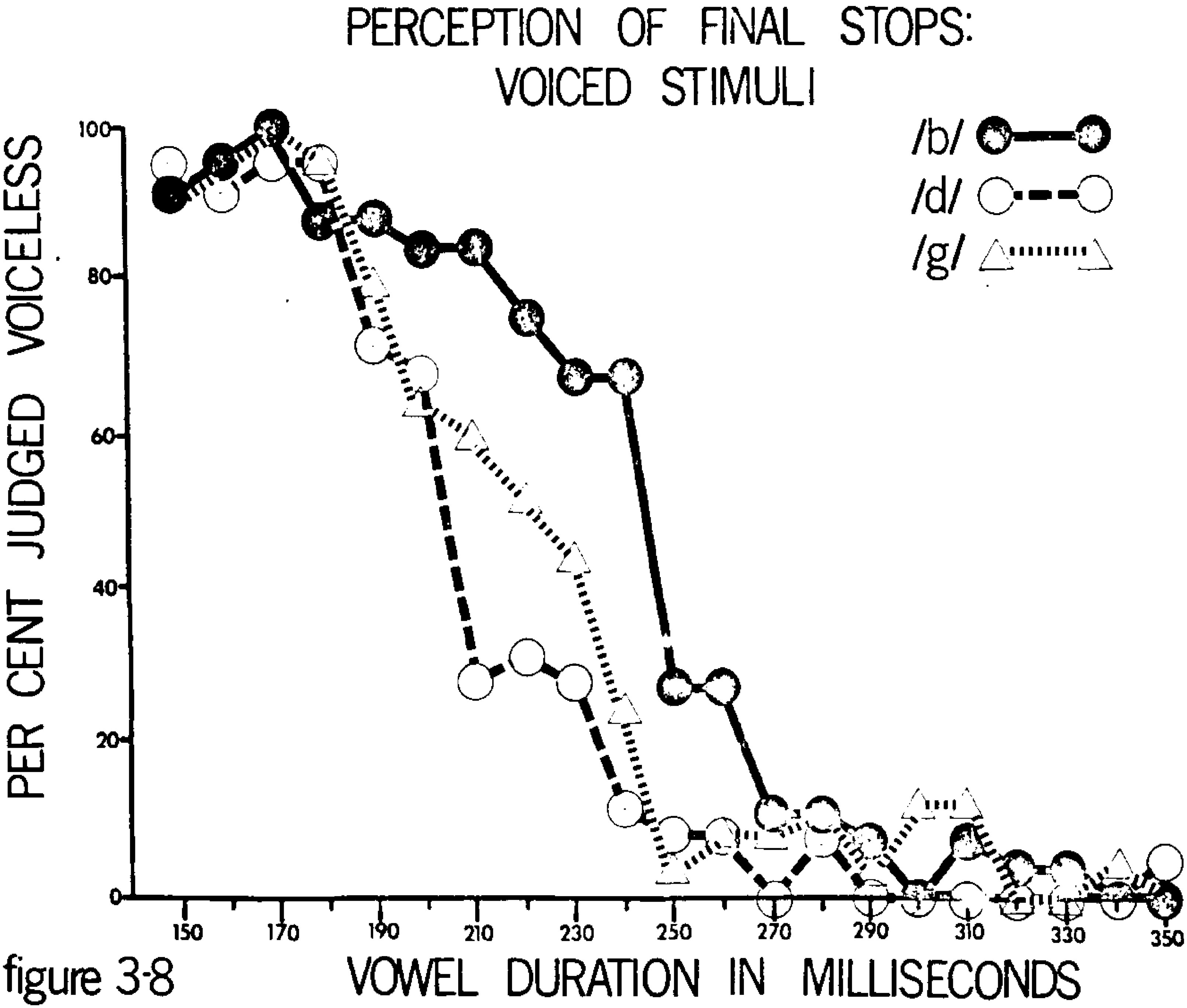

PERCEPTION OF FINAL STOPS:
VOICELESS STIMULI
/p/
/t/
/k/
PER CENT JUDGED VOICELESS
100
80
60
40
20
0
150
170
190
210
230
250
270
290
310
330
350
VOWEL DURATION IN MILLISECONDS

figure 3-7



PERCEPTION OF FINAL STOPS: VOICED STIMULI

figure 3-8

PERCEPTION OF FINAL FRICATIVES
VOICELESS STIMULI

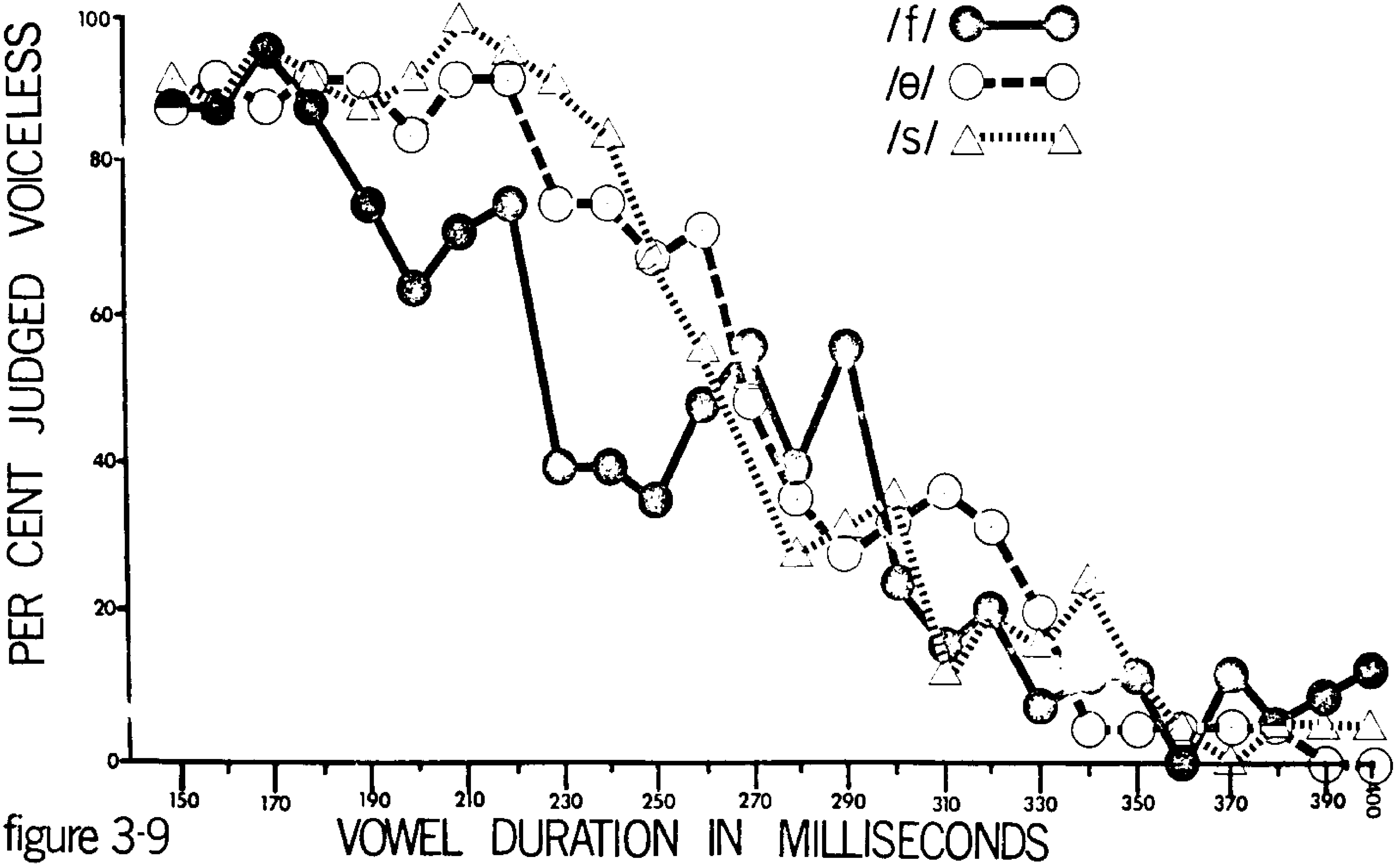


figure 3-9

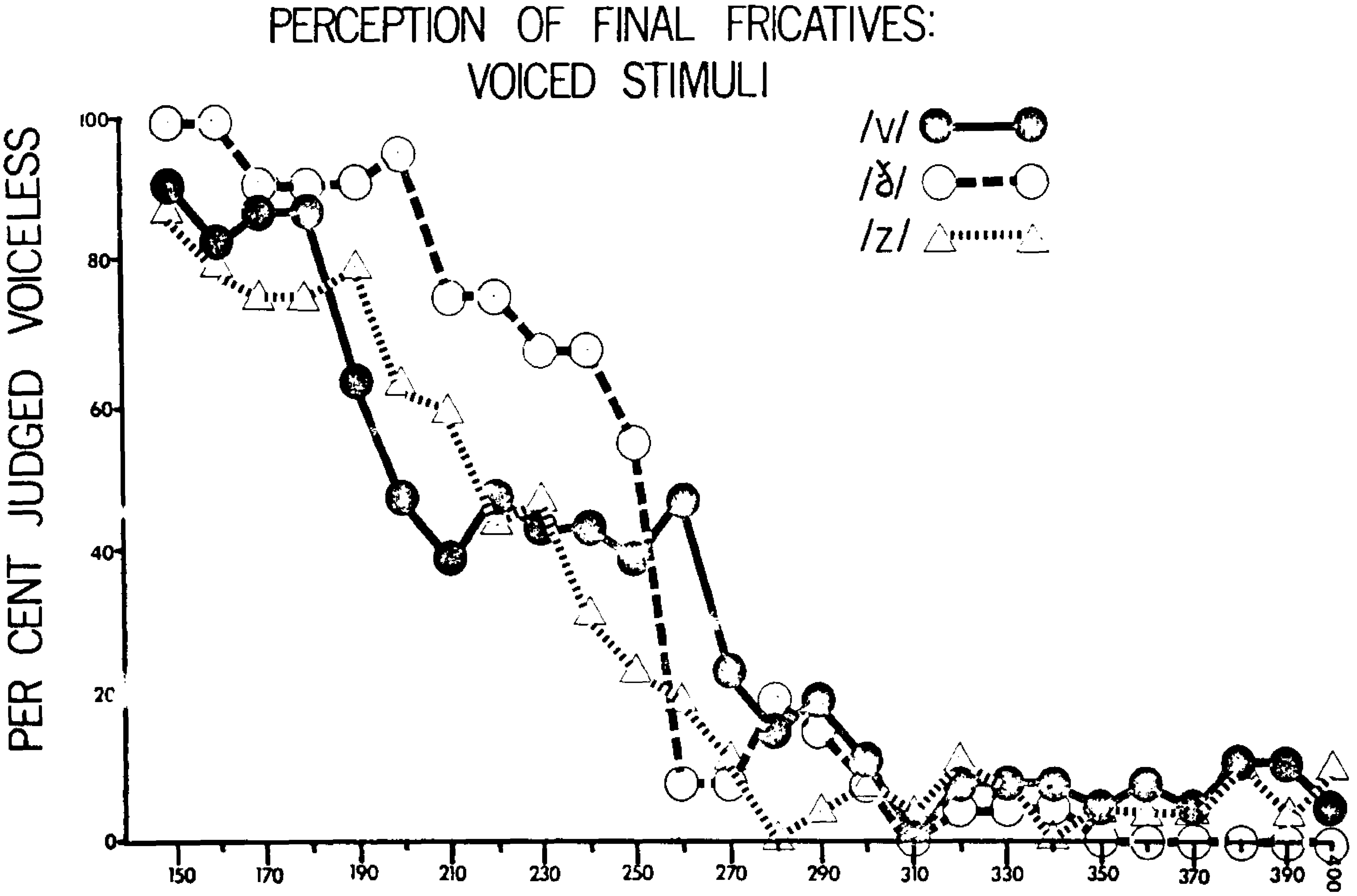

PERCEPTION OF FINAL FRICATIVES:
VOICED STIMULI
/v/
/ð/
/z/
PER CENT JUDGED VOICELESS
100
80
60
40
20
0
150
170
190
210
230
250
270
290
310
330
350
370
390
400
VOWEL DURATION IN MILLISECONDS

figure 3-10

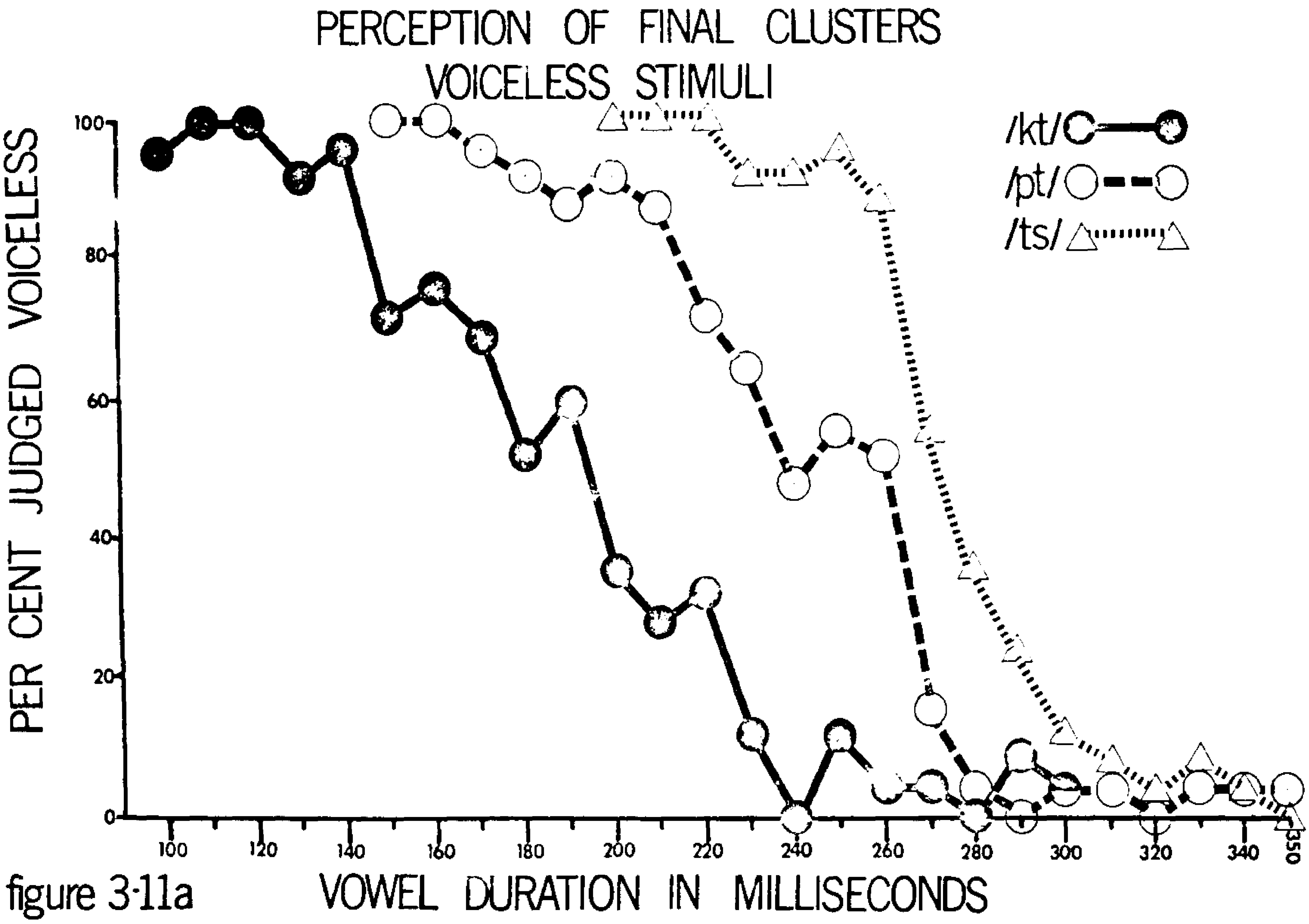

PERCEPTION OF FINAL CLUSTERS
VOICELESS STIMULI
PER CENT JUDGED VOICELESS
0
20
40
60
80
100
VOWEL DURATION IN MILLISECONDS
100
120
140
160
180
200
220
240
260
280
300
320
340
350
/kt/
/pt/
/ts/

figure 3·11a

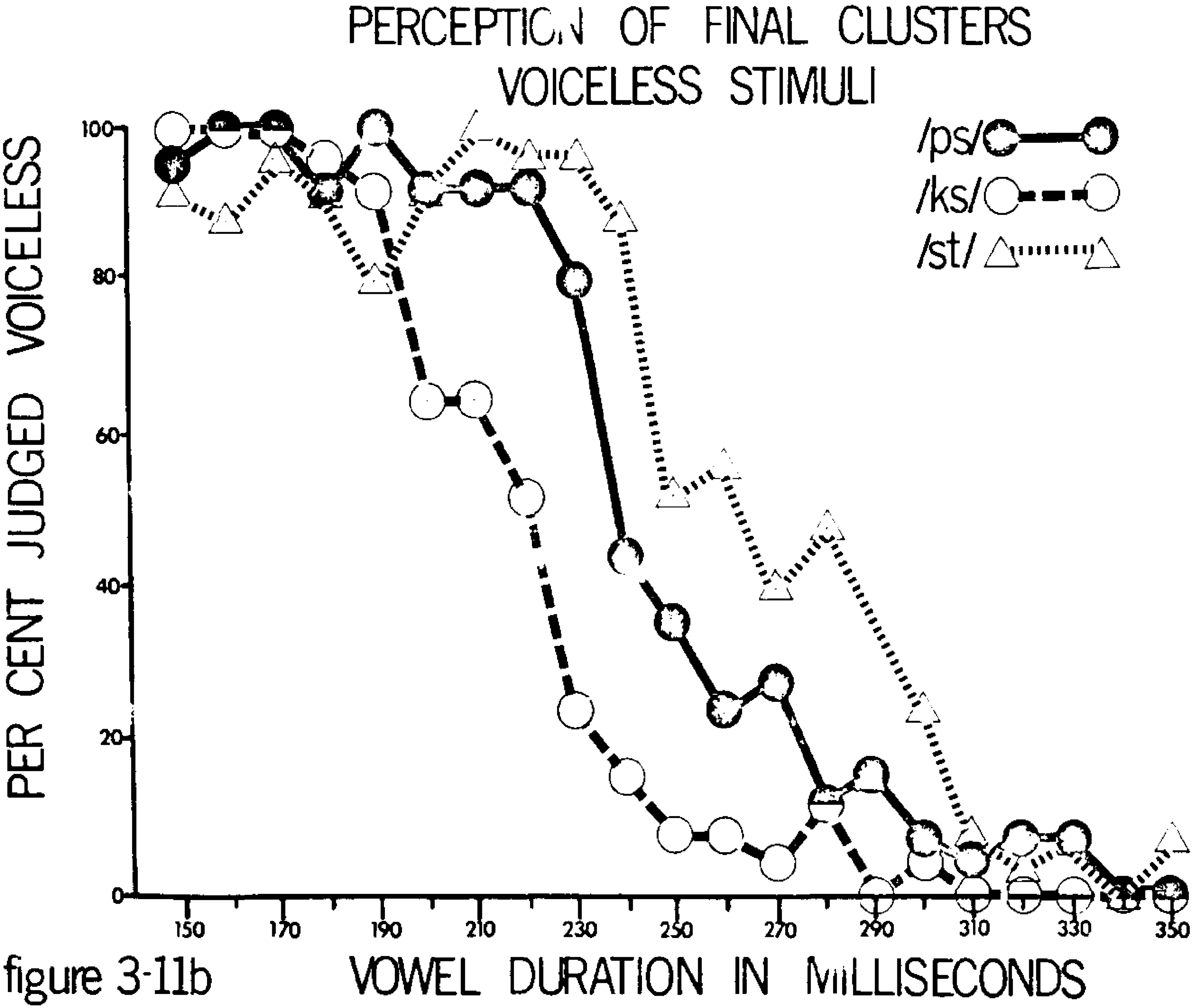

PERCEPTION OF FINAL CLUSTERS
VOICELESS STIMULI
/ps/
/ks/
/st/
PER CENT JUDGED VOICELESS
100
80
60
40
20
0
150
170
190
210
230
250
270
290
310
330
350
VOWEL DURATION IN MILLISECONDS

figure 3-11b

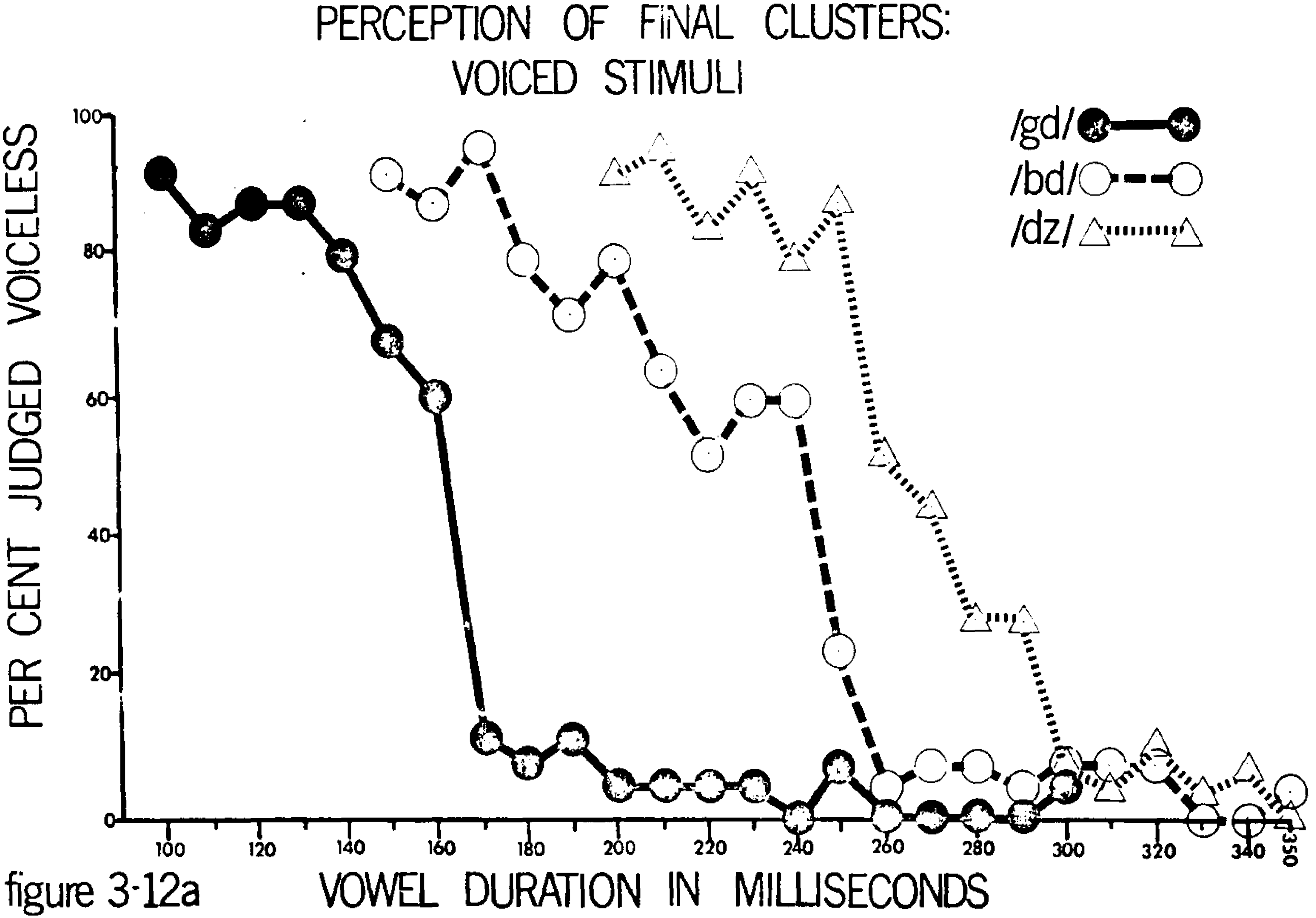

PERCEPTION OF FINAL CLUSTERS:
VOICED STIMULI
PER CENT JUDGED VOICELESS
100
80
60
40
20
0
100
120
140
160
180
200
220
240
260
280
300
320
340
350
VOWEL DURATION IN MILLISECONDS
/gd/
/bd/
/dz/

figure 3-12a

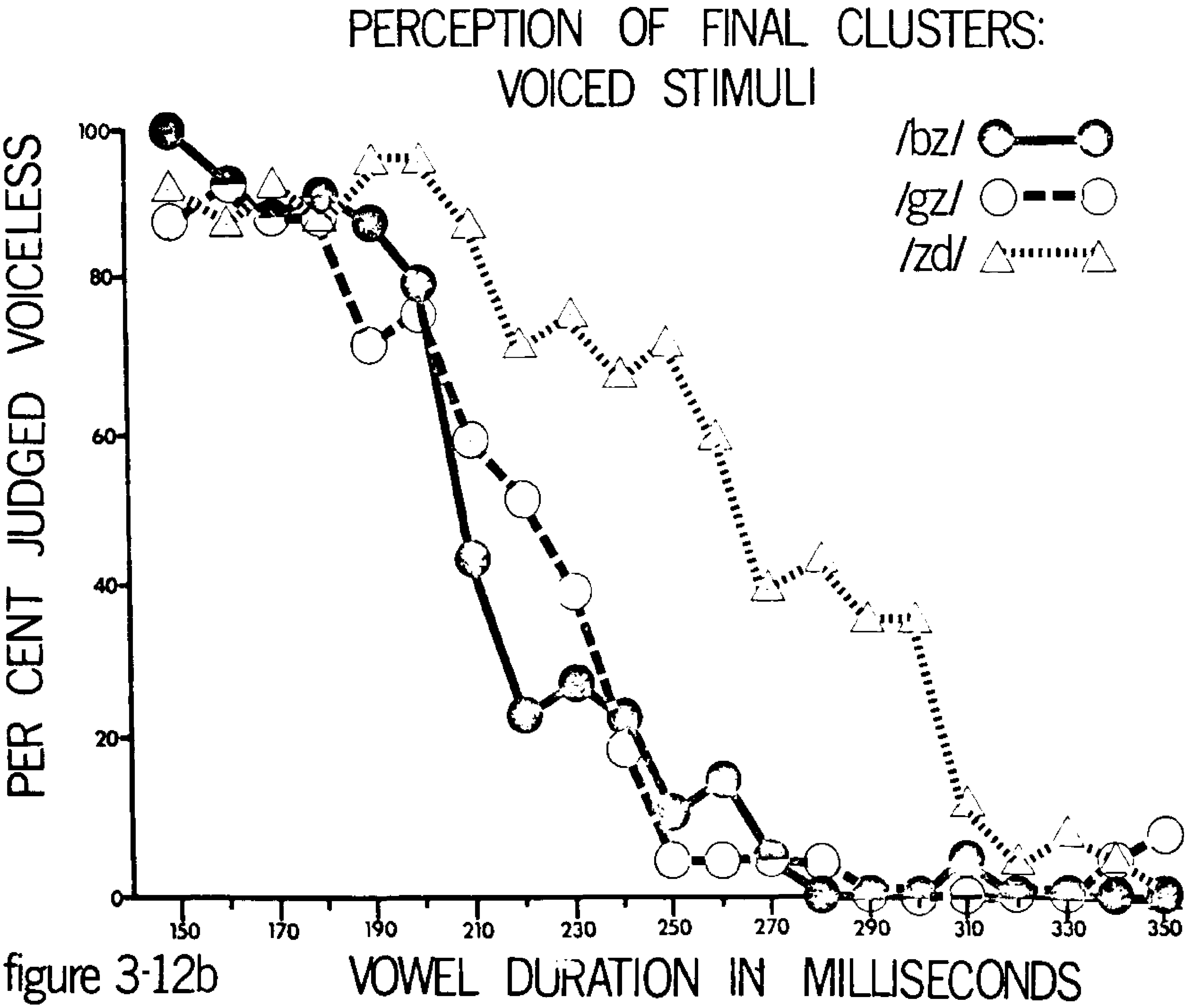

PERCEPTION OF FINAL CLUSTERS:
VOICED STIMULI
/bz/
/gz/
/zd/
PER CENT JUDGED VOICELESS
100
80
60
40
20
0
150
170
190
210
230
250
270
290
310
330
350
VOWEL DURATION IN MILLISECONDS

figure 3-12b

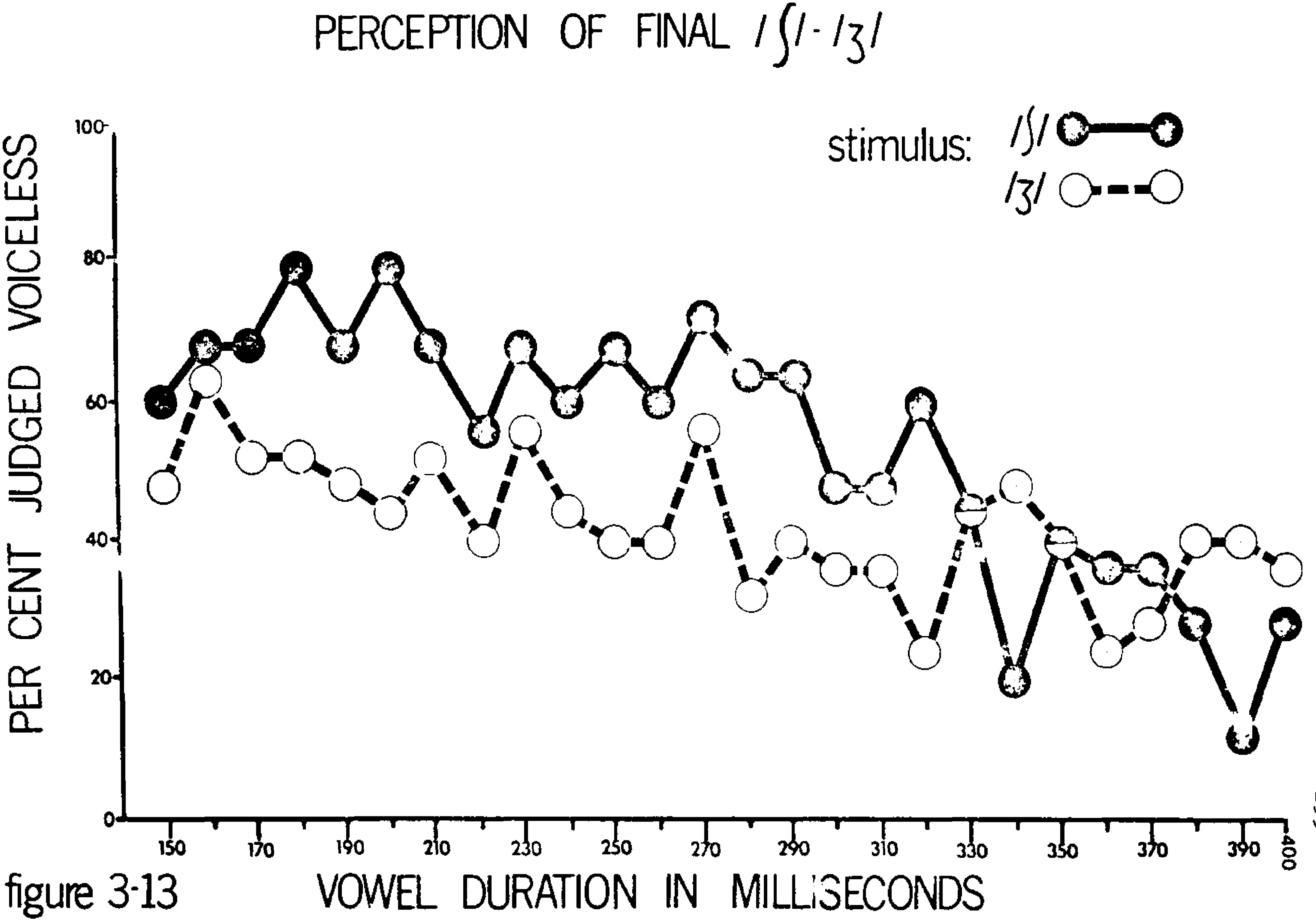

PERCEPTION OF FINAL /ʃ/-/ʒ/
stimulus: /ʃ/
/ʒ/
PER CENT JUDGED VOICELESS
100
80
60
40
20
0
150
170
190
210
230
250
270
290
310
330
350
370
390
400
VOWEL DURATION IN MILLISECONDS

figure 3-13

stimulus belonged to one class rather than to another.

Three explanations are suggested here for the failure of the /ʃ-ʒ/ opposition to behave as did the other final consonants. First, these two sounds are not well synthesized by the Pattern Playback. The high noise level of the machine itself tends to make the synthesis of any voiceless fricative difficult. Thus, the cue of short vowel duration, rather than offsetting the cues of presence or absence of voicing during the friction, tends merely to make the stimuli ambiguous.

Second, the /ʃ/-/ʒ/ opposition is one of extremely low yield in English. It is thus possible that the Ss were not used to listening for it and so lacked the experience necessary for making rapid decisions as to the class of the final consonant of the stimuli.

Finally, the minimal pair used, as mentioned above, was somewhat artificial: It contained what might have been an unusual or even unknown word for some of the listeners, and lacked the difference in vowel articulation usually used by speakers of New York City English in distinguishing the members of the pair.[8]

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

Hypothesis II

The results of the experiment generally confirm the second hypothesis: The changes in perception on the part of the listeners were continuous rather than categorical.

The question of categorical vs. continuous perception is of particular interest in the experiment described above. Previous findings have indicated that synthetic stop consonants (in particular) are perceived categorically with regard to changes in place and voicing,[9] and that changes in vowel perception are continuous with regard to quality and duration.[10] In the experiments described here, 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 hypothesized.

The verification of this hypothesis consisted of an experiment in which the stimuli previously labelled by the Ss were presented in a test to determine how well listeners could discriminate one stimulus from another. 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 labelling curves and that obtained from the discrimination test itself.[11]

In those cases where perception was continuous across a range of varying stimuli, the curves did not fit, the obtained discrimination function lying at a considerably higher level than the predicted function.[12] Further, if perception were continuous and Ss discriminated more accurately than they labelled, the greatest distance between the curves was at the points where stimuli within a phoneme boundary were being compared, rather than at points where stimuli across the phoneme boundary were presented for comparison.[13]

In the discrimination experiment reported here, one member of each class of 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/. Further, the sounds selected represented either the shortest or the second shortest critical vowel duration[14] of their classes. 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 prepared for the test in an "odd-ball" format, in which the Ss heard a triad of stimuli, one of which was different from the other two presented. Ss were asked to pick the odd member of each triad. The stimuli used were taken primarily from the critical vowel duration. Each stimulus was matched in a triad with the stimulus which differed from it by one, two, three, and seven stimulus steps. Further, each stimulus was heard once in the triad in all three positions, so that three triads were prepared for each oddball stimulus for each of the stimulus steps tested. An example of the result for a stimulus in the middle of the range of stimuli being tested was as follows:

> For a stimulus with a vowel duration of 220 msec, six triads were prepared in the one-step test:
>
> | | |
> |---|---|
> | 220 210 210 | 220 230 230 |
> | 210 220 210 | 230 220 230 |
> | 210 210 220 | 230 230 220 |
>
> Six more triads were prepared for the two-step test:
>
> | | |
> |---|---|
> | 220 200 200 | 220 240 240 |
> | 200 220 200 | 240 220 240 |
> | 200 200 220 | 240 240 220 |
>
> And a third set of six triads were prepared for the three-step test:
>
> | | |
> |---|---|
> | 220 190 190 | 220 250 250 |
> | 190 220 190 | 250 220 250 |
> | 190 190 220 | 250 220 220 |

For stimuli at the end of the test range, of course, fewer triads were prepared because of the absence of stimuli to the left (at the short end of the range), or to the right (at the long end of the range).

In order to compare discrimination across phoneme boundaries with that within phoneme boundaries, a seven-step test was prepared. Seven steps were necessary because of the extended critical vowel duration which separated the phonemic labels applied by the Ss (see Figures 3-8, 10, 11b). Because of this relatively large number of steps and because perceptual changes occurred at the short end of the stimulus range, there were only three or four stimuli in the seven-step test. In all cases, one stimulus was from the voiceless range on the labelling test, and two were from the voiced range. This allowed for one comparison across and one within the phoneme boundary. In the discrimination test for <u>dues</u> /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 within any voiceless phoneme category.

The discrimination test tapes were prepared and played to twenty-five undergraduate phonetics students at Lehman College of the City University of New York. The students had first been given the labelling test so that the range of stimuli for the discrimination test could be determined. It should be noted that the range selected coincided almost exactly with the range that would have been selected from the labelling curves obtained from the Ss in the experiment reported in the discussion of Hypothesis I.

The results of the one-, two-, and three-step discrimination tests are shown in Figures 3-14 to 3-16.[15] Because the results of these tests contain no intra-phonemic comparisons, they are not as revealing as the seven-step test results. They do, however, indicate a generally

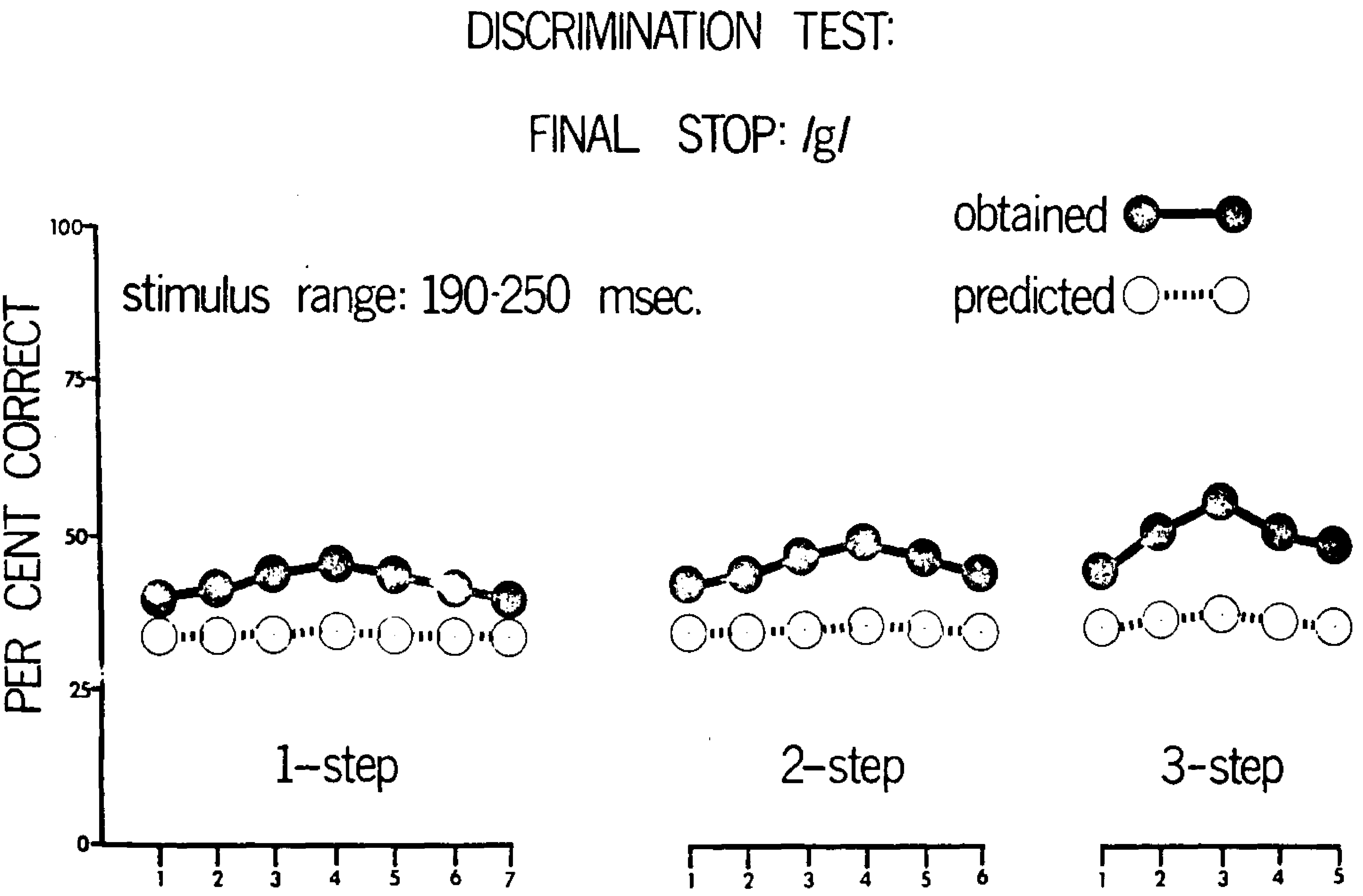

DISCRIMINATION TEST:
FINAL STOP: /g/
stimulus range: 190-250 msec.
obtained
predicted
PER CENT CORRECT
100
75
50
25
0
1-step
2-step
3-step
1
2
3
4
5
6
7
1
2
3
4
5
6
1
2
3
4
5
STIMULUS NUMBER

figure 3-14

DISCRIMINATION TEST:

FINAL FRICATIVE: /z/

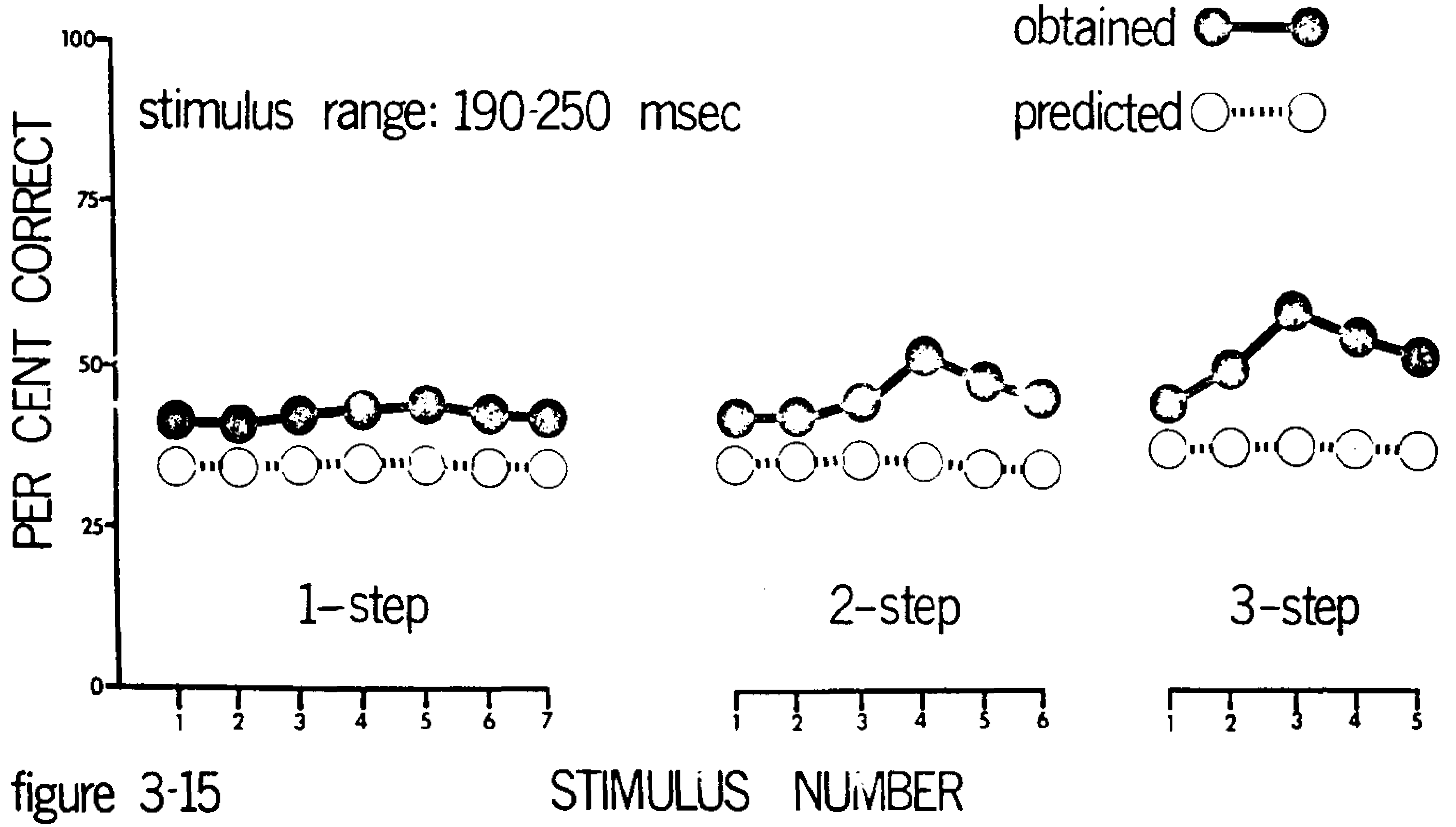


figure 3-15

DISCRIMINATION TEST:

FINAL CLUSTER: /ks/

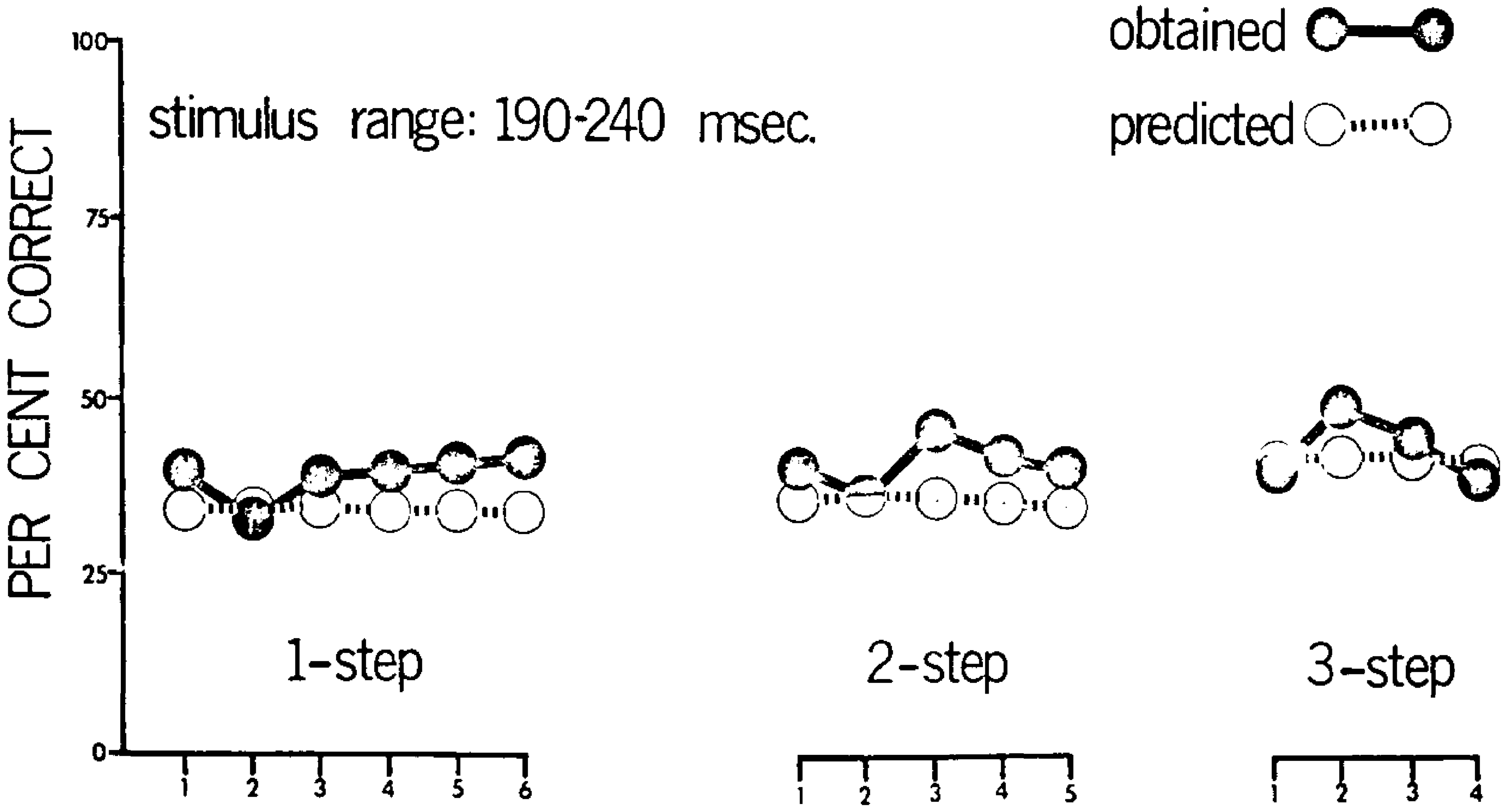


figure 3-16

higher level of discrimination than would be expected from the predicted curve. This is particularly true in the case of the final stop and fricative tested, especially as the step-function increases from one to three. The final cluster results for <u>picks</u> /pɪks/ are ambiguous, with little difference between the obtained and predicted functions, and occasional points where the obtained function actually dips below the predicted function.[16] Thus, although the indications from the one-, two-, and three-step tests are that perception is continuous, these are by no means conclusive findings.

More convincing results may be seen in the seven-step test (Figure 3-17), which allowed for inter- and intra-phonemic stimuli comparisons. In all three categories tested, the obtained discrimination function lies at extremely high values, both for inter- and intra-phonemic comparisons. On the other hand, the predicted levels, while high in the inter-phonemic comparisons, still lie considerably below the obtained functions. Further, the predicted levels for the intra-phonemic comparisons are virtually at the chance level, while the obtained functions for the same comparisons are at or near the ninety per cent level.

From these results it appears that Ss can discriminate far better than they can label, both across and, especially, within phoneme boundaries, and that when the voiced-voiceless opposition is cued by the duration of the

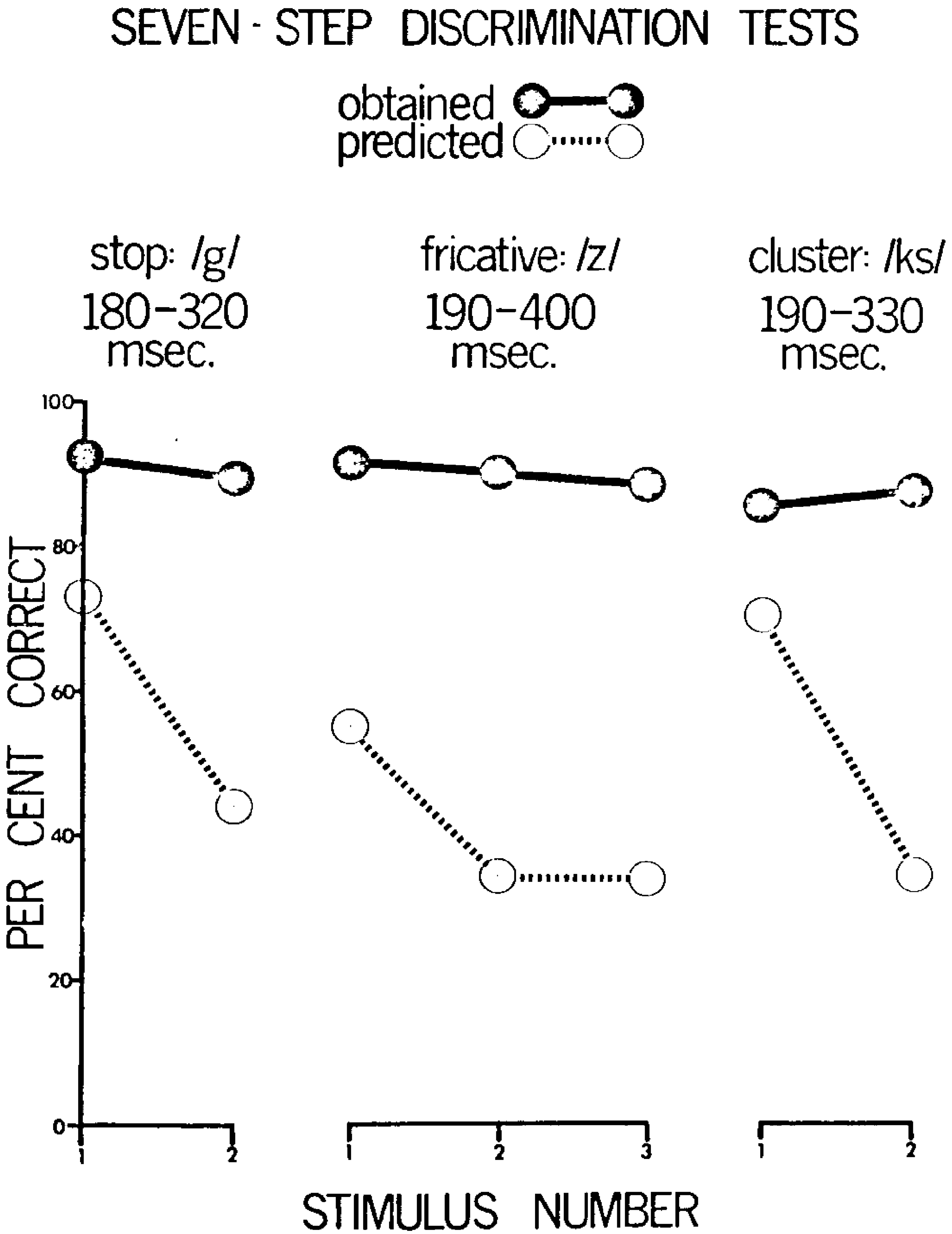

SEVEN - STEP DISCRIMINATION TESTS
obtained
predicted
stop: /g/
180–320
msec.
fricative: /z/
190–400
msec.
cluster: /ks/
190–330
msec.
100
80
60
40
20
0
PER CENT CORRECT
1
2
1
2
3
1
2
STIMULUS NUMBER

figure 3 - 17

preceding vowel, that perception is continuous, rather than categorical.

Hypothesis III

The third hypothesis predicted that the minor cue of voicing during final consonant closure would cause a difference between the durations of vowels needed to effect equivalent perceptions of sounds synthesized as cognates. This hypothesis is confirmed by the fact that the curves for the voiced stimulus series generally descend earlier than those of their voiceless cognate stimulus series (see Figures 3-18 to 3-21). There are some differences, however, between the classes of sounds tested.

Three points on the curves may be taken as indicative of the differences between the effects of vowel duration on the perception of the stimuli:

1. The beginning of the critical vowel duration. The critical vowel duration is taken to be the range of stimuli over which perceptual judgments change from eighty per cent to twenty per cent agreement of voicelessness.
2. The first fifty per cent crossover point.
3. The end of the critical vowel duration.

Table III-8 displays each of these points for each of the contrasts tested.

It was assumed that the particular vowel used in the synthesis of any series of stimuli did not materially affect any of the measures used to confirm Hypothesis III, or

COMPARISON OF PERCEPTUAL CHANGE:
VOICED vs. VOICELESS STOPS

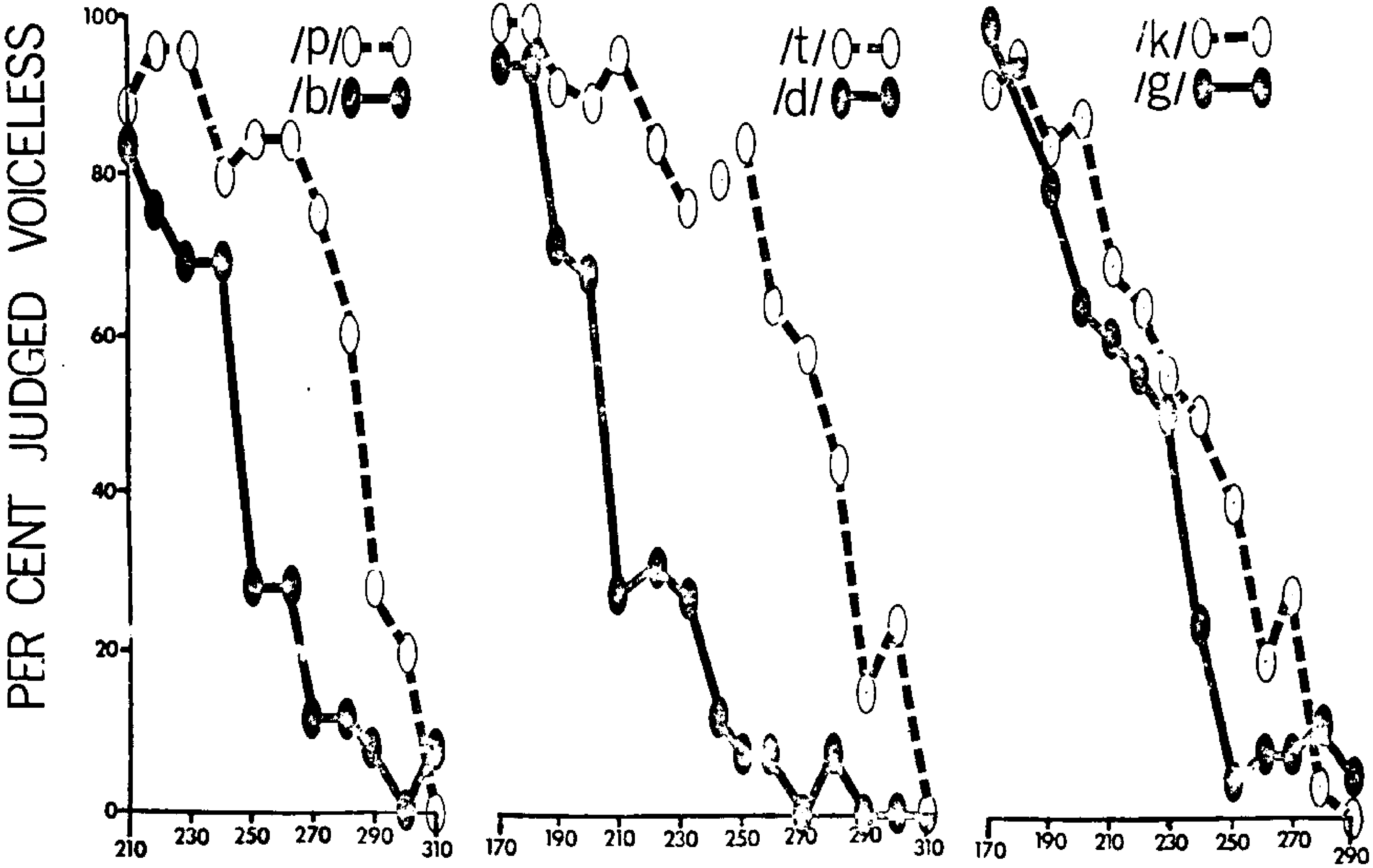
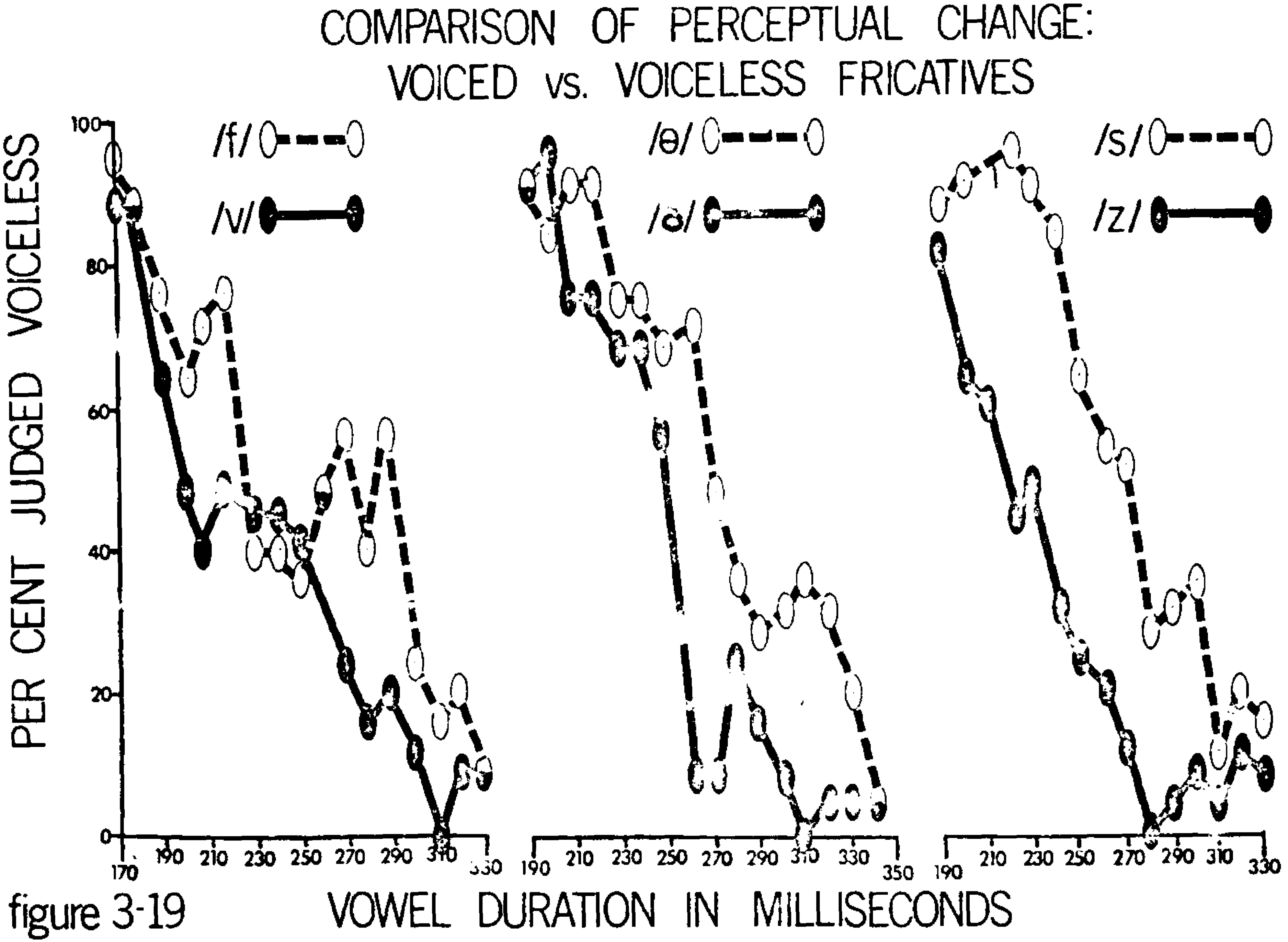


figure 3-18 VOWEL DURATION IN MILLISECONDS



COMPARISON OF PERCEPTUAL CHANGE:
VOICED vs. VOICELESS FRICATIVES

figure 3-19

COMPARISON OF PERCEPTUAL CHANGE:
VOICED vs. VOICELESS CLUSTERS

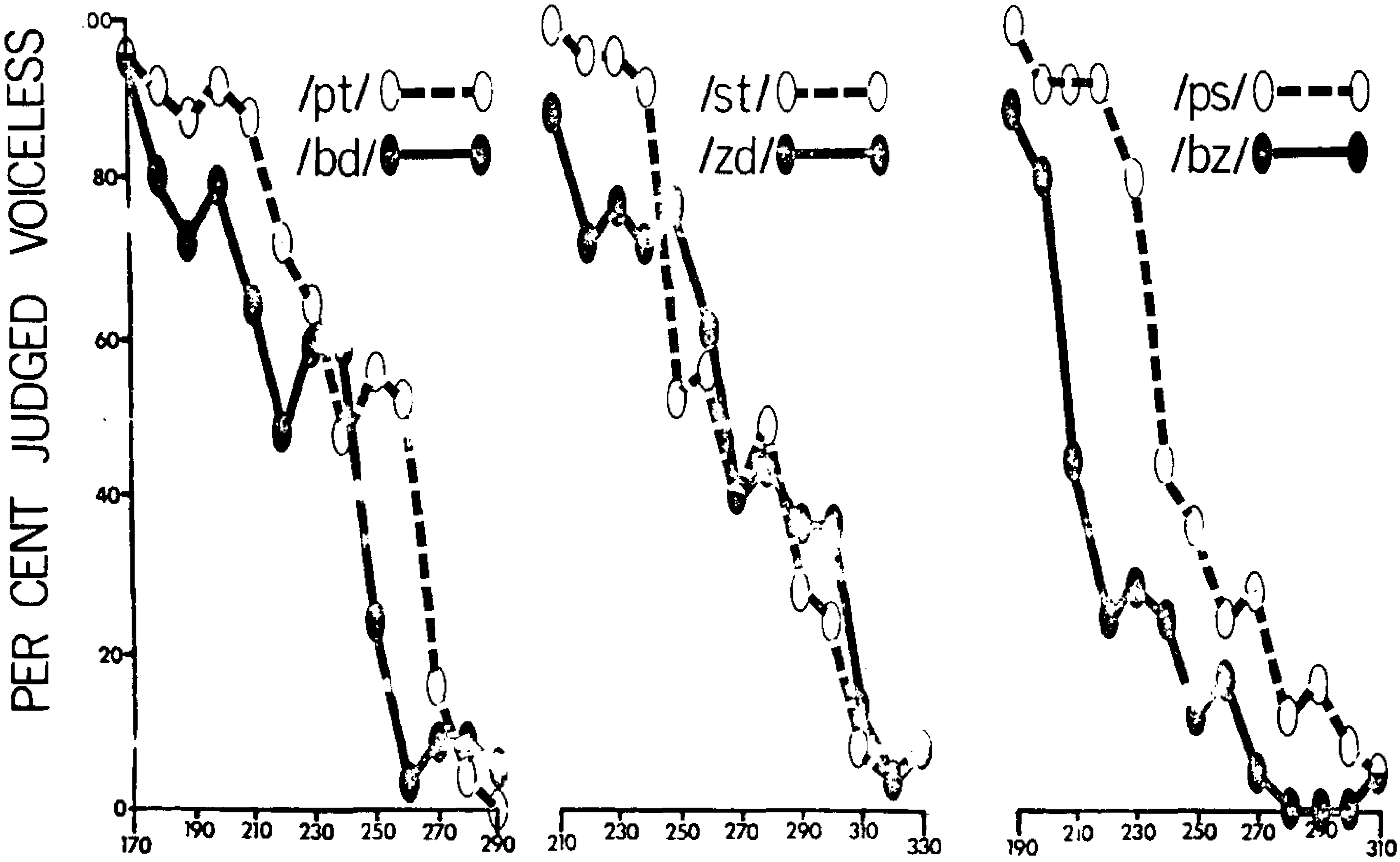


figure 3-20

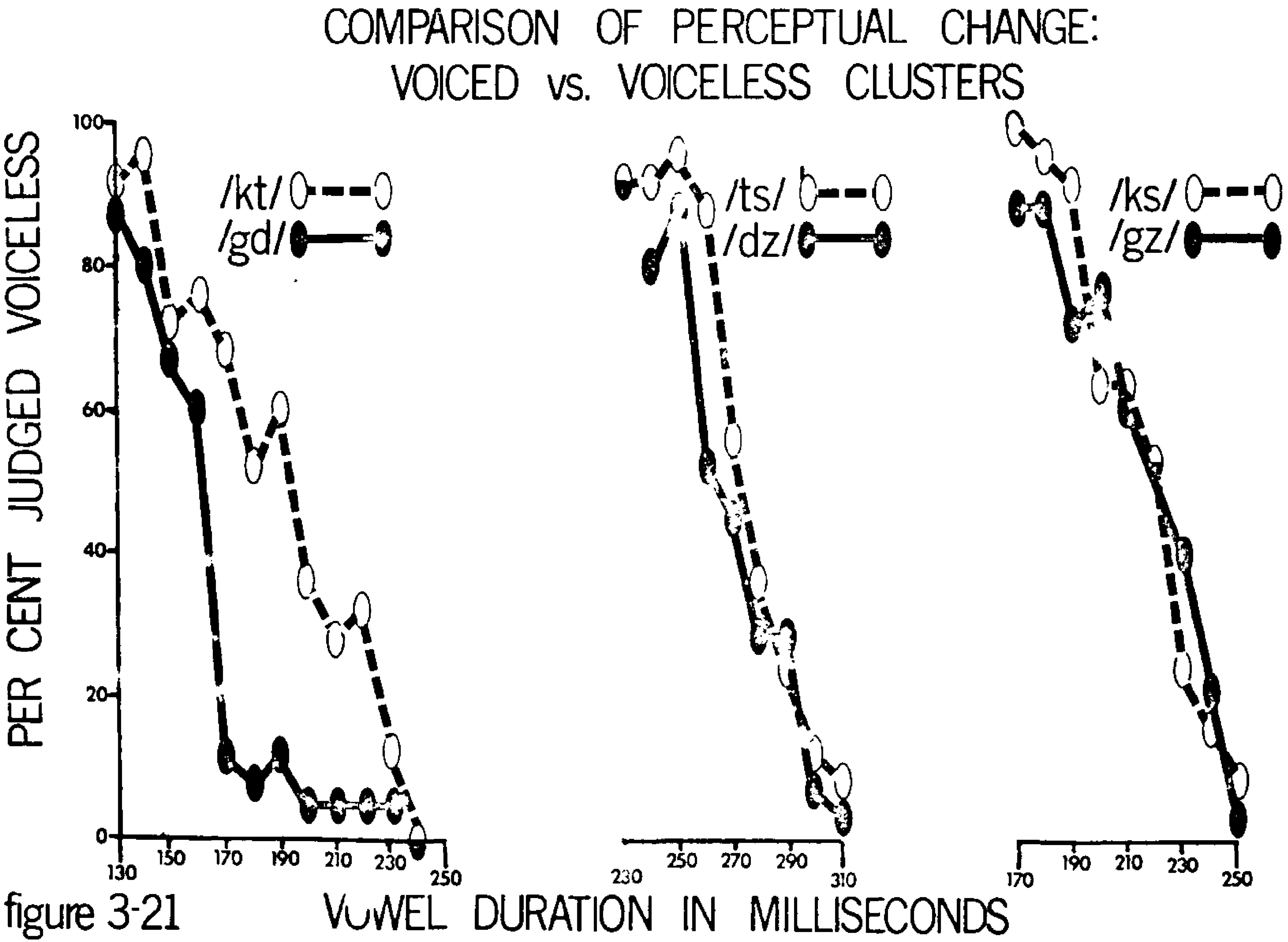

COMPARISON OF PERCEPTUAL CHANGE:
VOICED vs. VOICELESS CLUSTERS
PER CENT JUDGED VOICELESS
100
80
60
40
20
0
/kt/
/gd/
/ts/
/dz/
/ks/
/gz/
130
150
170
190
210
230
250
230
250
270
290
310
170
190
210
230
250
VOWEL DURATION IN MILLISECONDS

figure 3-21

TABLE III-8

CRITICAL VOWEL DURATIONS
POINTS OF MEASUREMENT (MILLISECONDS)

	Beginning Point		1st 50% Crossover		End Point	
Opposition Tested	Voiceless Stimulus	Voiced Stimulus	Voiceless Stimulus	Voiced Stimulus	Voiceless Stimulus	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/	186.6	183	228	198	305	275
/θ-ð/	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	

Hypothesis IV, in which the measures were compared between stimulus classes. This assumption seems reasonable, whatever view one takes of inherent vowel duration in English. Obviously, if one assumes that there exist no inherently long or inherently short vowels, then the particular vowel used is of no immediate significance in determining critical vowel duration, or the location of the first fifty per cent crossover point. On the other hand, if one assumes that there are inherent durations, and that vowels may be divided into long and short groups, then it is possible to point out that vowels from both groups were used before all the classes of final consonants tested, and that this did not seem to make a relative difference in the critical vowel durations or in the locations of the first fifty per cent crossover points.

The stops conform most consistently to the hypothesis. In every case there is at least a 12 millisecond difference between any pair of points for each contrast tested. The average difference between the three points measured for each class of sounds is greater than 35 milliseconds in the predicted direction.

The fricatives are about as consistent as the stops. With the exception of the /f-v/ opposition, where the values for the beginning of the critical vowel duration are only 3.6 milliseconds apart, every pair of points is separated by at least 18 milliseconds, and the average difference between the three points measured for each class of sounds is greater than 25 milliseconds in the predicted direction.

The findings are somewhat different in the case of the clusters. Although the average difference between any of the three pairs of points is in the predicted direction, it is never more than 17.7 milliseconds, and as low as 13.5 milliseconds. Further, there are two cases where the values for a pair of points are identical, and still three others where the difference between the values, though small, is not in the predicted direction. It would seem that in the case of the final clusters, there were fewer differences between the perceptual changes in the voiced and the voiceless stimuli than between the voiced and voiceless stimuli of the final stops and fricatives.

The overall tendency for a voiced stimulus to be heard as voiced after vowels of shorter duration than a voiceless (cognate) stimulus heard as voiced, supports the conclusion of Chapter II that the presence or absence of voicing during the closure for the final consonant does contribute to perception in some minor way.

Hypothesis IV

The fourth hypothesis held that changes in perception caused by varying vowel duration would be more readily effected before stops than before fricatives. Further, it was predicted that the clusters would resemble the stops more than the fricatives in the way in which vowel duration affected their perception.

Two interrelated measures confirm this hypothesis. One of these is the critical vowel duration. If the voicing during consonant closure is a more important cue in the case of fricatives than in the case of stops (see Chapter II, p. 123), one would expect that the critical vowel duration for fricatives would be greater than that for stops. That is, the range of durations over which the vowel must be varied to overcome the effect of voicing during closure and thus cause a change in perception will be greater before fricatives (which are relatively long in duration) than before stops (which are relatively short in duration). The data in Table III-9 confirms this assumption.

The slopes of the perception curves during the critical vowel duration (which are, of course, a function of that duration) present the same evidence in another form. (See Table III-10.) Although there is one case of overlap between the slopes for the two categories, the averages are clearly separated. The figures indicate a much steeper slope and thus a much more rapid change in perception per unit of vowel duration in the case of stops than in the case of fricatives.

The clusters, as predicted, do resemble the stops more than the fricatives in the way in which their perception is affected by preceding vowel duration. Although the range of variation within the cluster category is greater than the range within the other two categories, the average critical vowel duration for the clusters (51.2 msec) is much closer

TABLE III-9

CRITICAL VOWEL DURATIONS (MILLISECONDS)

Opposition Tested	Voiceless Stimulus Series		Voiced Stimulus Series
/p-b/	35		50
/t-d/	37		48
/k-g/	58		52
Average:			
Each series:	43.3		50.0
Average:			
Both series		46.6	
/f-v/	118.4		92
/θ-ð/	102		50
/s-z/	65		95
Average:			
Each series	95.1		79.0
Average:			
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
Average:			
Each series	51.8		50.5
Average:			
Both series		51.2	

TABLE III-10

SLOPE OF CRITICAL VOWEL DURATION

Voiceless Stimulus			Voiced Stimulus
	Stops		
gape	1.715	1.200	gabe
bet	1.740	1.240	bed
burke	1.070	1.150	berg
Average: Voiceless Stops	1.508	1.197	Average: Voiced Stops
	Fricatives		
duff	.507	.645	dove
teeth	.585	1.200	teethe
deuce	.945	.858	dues
Average: Voiceless Fricatives	.679	.901	Average: Voiced Fricatives
	Clusters		
bopped	1.110	1.150	bobbed
cost	1.330	.655	caused
cops	1.385	1.330	cobs
tucked	.765	2.070	tugged
tights	1.915	1.460	tides
picks	1.460	1.090	pigs
Average: Voiceless Clusters	1.328	1.293	Average: Voiced Clusters

Average: All Stops	1.353
Average: All Fricatives	.790
Average: All Clusters	1.310

to that for the stops (46.6 msec) than to that for the fricatives (87.1 msec). Similarly, the average slope of the critical vowel duration for the clusters (1.310) is much closer to that for the stops (1.353) than to that for the fricatives (0.790).

One can infer from these results that it is the presence of the stop in each of the final clusters that conditions both the vowel length and the rapidity of the perceptual change as manifested by the critical vowel duration. This is true even for those clusters containing a fricative and independently of the ordering of the stop and fricative in the cluster. Table III-11 presents data revealing that, in real speech, vowel duration before final clusters of all three types tested more closely resembles that of final stops than that of final fricatives.[17]

It can be concluded, then, that identical variations in preceding vowel duration will cause a greater and more rapid change in the perception of stops and clusters than in the perception of fricatives. It should be noted, however, that this difference, manifested in the generally greater critical vowel duration of fricative syllables, is not the only one distinguishing the classes of sounds. The manner in which the perception changes for the fricatives is often markedly different from that of the stops and clusters. The latter, during their relatively brief critical vowel durations, produce curves that are generally smooth and which do not often reverse direction. The few reversals

TABLE III-11

AVERAGE VOWEL DURATIONS

Vowel Precedes:	Msec
<u>Stop + Stop</u>	
voiceless	163
voiced	242
<u>Stop + Fricative</u>	
voiceless	150
voiced	219
<u>Fricative + Stop</u>	
voiceless	157
voiced	240
<u>All Clusters</u>	
voiceless	156
voiced	230
<u>All Stops</u>	
voiceless	164
voiced	316
<u>All Fricatives</u>	
voiceless	210
voiced	362.2

that do occur are most usually at the extremes of the critical vowel duration. That is, once the listeners' perceptions begin to change, they continue to change until the cognate of the sound originally perceived is heard.

This is not the case for the fricatives, where there is often a zone of confusion during the critical vowel duration. Within the central portion of this zone the percentage of judgments tends to change in small steps, and changes in the direction of the curve are not uncommon until the perception changes rapidly to that of the cognate first perceived. These characteristics of the various curves are evident in Figures 3-8 to 3-10, as well as in the functions derived from the tape splicing experiment in Chapter II (Figures 2-17 to 2-23).

In Figure 3-22 the critical vowel durations of all the fricative and stop stimuli have been set equal to one hundred per cent, and the average values for all the curves for both classes of stimuli have been plotted together. The resultant curves are much smoother and more unidirectional than any of the curves for a specific series of stimuli, but even here the differences between the classes of sounds can be noted: The center of the critical range (from 60 per cent to 40 per cent) of total judgments covers a span of 26 per cent of the total critical vowel duration for the stops, and 48 per cent of the total critical vowel duration for the fricatives.

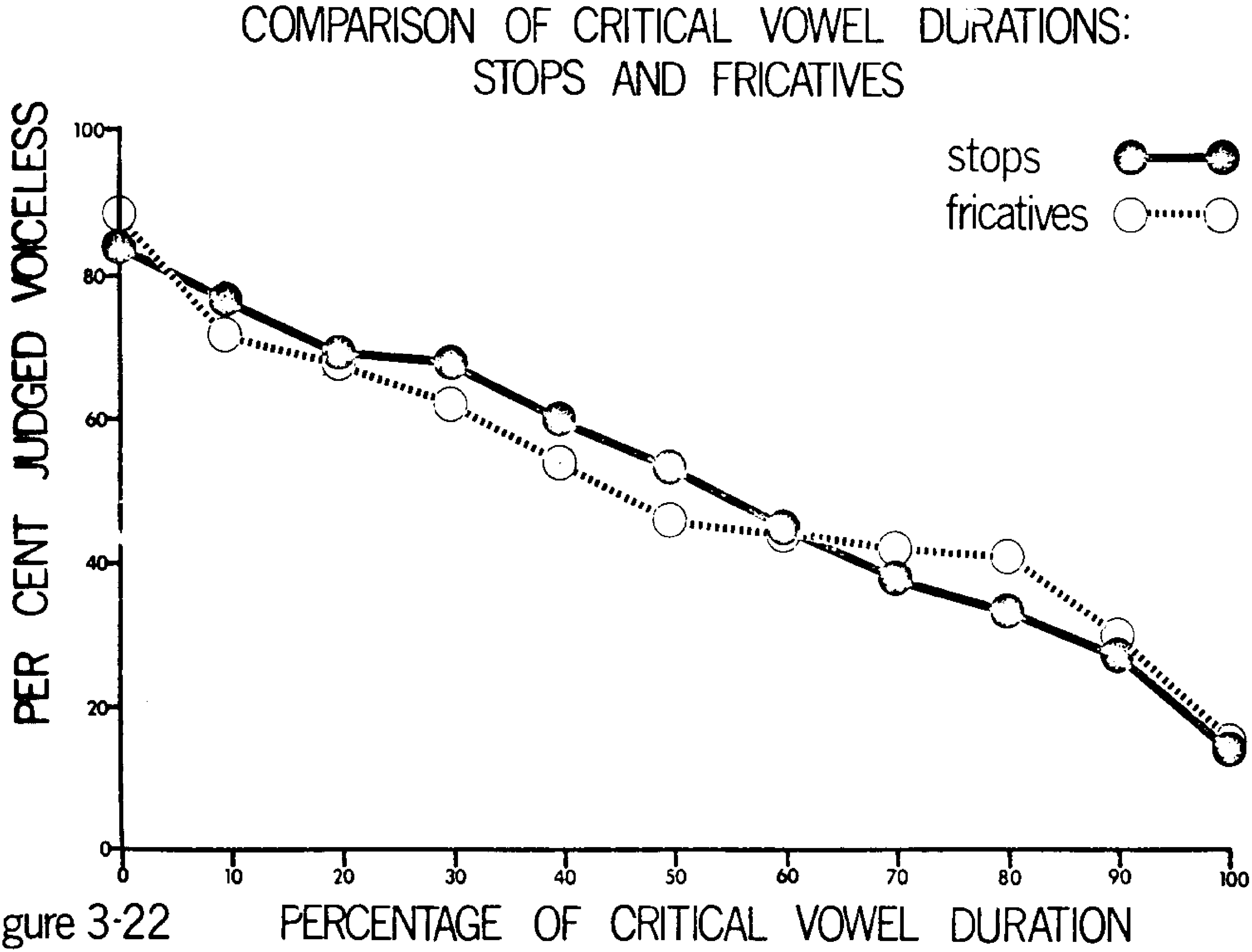

COMPARISON OF CRITICAL VOWEL DURATIONS:
STOPS AND FRICATIVES
stops
fricatives
PER CENT JUDGED VOICELESS
100
80
60
40
20
0
0
10
20
30
40
50
60
70
80
90
100
PERCENTAGE OF CRITICAL VOWEL DURATION

figure 3-22

Footnotes - Chapter III

[1] D. B. Fry, A. S. Abramson, P. D. Eimas, and A. M. Liberman, "The Identification and Discrimination of Synthetic Vowels," Language and Speech, 5 (October-December, 1962), 171-189.

[2] J. Bastian and A. S. Abramson, "Identification and Discrimination of Phonemic Vowel Duration," Journal of the Acoustical Society of America, 34 (1962), 743 (Abstract). See also A. M. Liberman, F. S. Cooper, D. P. Shankweiler, and M. Studdert-Kennedy, "Perception of the Speech Code," Psychological Review, 74 (1967), 443.

[3] See below, p. 183, for the data comparing vowel duration before stops, fricatives and clusters.

[4] Howard S. Hoffman, "Study of Some Cues in the Perception of the Voiced Stop Consonants," Journal of the Acoustical Society of America, 30 (1958), 1035. For a more detailed description of the Pattern Playback, see F. S. Cooper, A. M. Liberman, and J. M. Borst, "The Interconversion of Audible and Visible Patterns as a Basis for Research in the Perception of Speech," Proceedings of the National Academy of Sciences, 37 (1951), 318-325.

[5] All initial consonants in these experiments were synthesized as stops in order to facilitate the accurate determination of the duration of the vowel in each stimulus.

[6] Preliminary experiments showed no significant

difference in listener judgments between stops synthesized with releases and those synthesized without releases.

[7] There is some disagreement in the literature as to whether [æ] is a tense or a lax vowel. Bronstein, for example, notes that [æ] is "commonly considered a lax vowel," but adds that "a clearly tense variety . . . exists in all parts of the country." *The Pronunciation of American English* (New York: Appleton-Century-Crofts, 1960), p. 154. The vowel has been classified here as lax, partially because it is, like most other lax vowels, found almost exclusively in checked position. Since there is no objective evidence confirming the existence of a tense-lax opposition of vowels, the linguistic criterion seems no more arbitrary than any other that might be chosen.

[8] The difference in vowel quality alluded to here is the use of a lower, perhaps laxer [æ] in *cas*(ual), and a raised, perhaps tenser [æ˔], [ɛ˕], or even [ɛ˕ə] in *cash*. Even with this difference omitted in the synthetic stimuli, the results suggest that if the range of vowel durations had been extended in both directions, perceptions might have approached unanimity. Such an extension, however, would have contained vowels both much longer and much shorter than could be heard in real speech.

[9] Fry et al., "Identification and Discrimination," pp. 171-189.

[10] Bastian and Abramson, "Identification and Discrimination," p. 743.

[11] That is, on the extreme assumption that Ss can discriminate no better than they can label. Thus, if perception is categorical, the discrimination curve which is predicted from the labelling curve should not differ in any significant way from the discrimination curve obtained from the discrimination test. See A. M. Liberman, K. S. Harris, H. S. Hoffman, and B. C. Griffith, "The Discrimination of Speech Sounds Within and Across Phoneme Boundaries, Journal of Experimental Psychology, 54 (November, 1957), 358-368; A. M. Liberman, K. S. Harris, J. A. Kinney, and H. Lane, "The Discrimination of Relative Onset-Time of the Components of Certain Speech and Nonspeech Patterns, Journal of Experimental Psychology, 61 (May, 1961), 379-388; A. M. Liberman, K. S. Harris, P. Eimas, L. Lisker, and J. Bastian, "An Effect of Learning on Speech Perception: the Discrimination of Durations of Silence With and Without Phonemic Significance," Language and Speech, 4 (October-December, 1961), 175-195.

[12] Fry et al., "Identification and Discrimination," pp. 180-182.

[13] Ibid.

[14] "Critical vowel duration" is used here to refer to the range of stimuli over which perception changed most rapidly for a given stimulus series. The boundaries of the critical vowel duration were determined by the points where the perception function intersected the ordinate values of 80 per cent and 20 per cent.

[15] The equation used to predict the discrimination function from the labelling results was $(p_1-p_2)^2(.66) + .33$. For the derivation of this equation, see Appendix **5**.

[16] Although spectrographic evidence and inspection of the painted patterns used in the synthesis do not reveal any error, there are clear indications that at least one stimulus from the long end of the range was faulty in some way. This is suggested by the stimuli comparisons 6, 5, and 4 in the 1, 2, and 3-step tests, respectively. Instead of the expected increase of the obtained function there is a decrease from the 1- to the 2-step test, and a rise from the 2- to the 3- step test. Further, the final obtained value in the 3-step test for the 4 comparison is, in fact, lower than that for the 6 comparison in the 1-step test.

[17] Samples are taken from the same speaker sources as those in Chapter II, Tables II-1 and II-2.

CHAPTER IV

SUMMARY AND IMPLICATIONS

On the basis of the results of the tape-cutting and synthetic speech experiments, it is possible to reach the following conclusions:

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

2. The perception cued by the preceding vowel duration is continuous, rather than categorical.

3. The presence of voicing during closure does have some cue value, although it is minor compared to that of vowel duration.

4. The cue of preceding vowel duration is more effective before clusters (of the type tested here) and stops, than before fricatives.

That preceding vowel duration should be a powerful cue is not surprising, especially if one contrasts it with the nature of the other potential cues to the voicing characteristics of final consonants: Variation in vowel duration is the most consistently present of all the cues. Examples of voiced consonants with little or 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.[1] The cue of stop release, it will be recalled, is often absent in American English in final position,[2] and fricative duration differences, although consistently present,[3] may be accompanied by contradictory vocal pulsing cues, depending on the context of the utterance.[4] Since the reliability of these cues is not great, it might be expected that the more consistently present cue of preceding vowel duration would attain a 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 destructive effects of context becoming the most significant cue of all.

That vowel duration is a necessary cue to voicing is indicated by the fact that when all of the potential cues mentioned above were synthesized to be appropriate for one of the categories of voicing, a preceding vowel of inappropriate duration caused the listeners to perceive the final consonant as belonging to the other voicing category.[5]

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. For example, perceptual tests of synthetic stimuli involving F1 cutback and VOT for initial position stops, and for real speech tests on intervocalic stop closure duration, reveal a perceptual pattern in which listeners can discriminate stimuli only about as well as they can label them.[6]

There is, then, something more here than just another case in which different cues to the same distinctions 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 sound or class of sounds from another are almost universally assumed to be present during, and as a direct result of, the articulation of the sound(s) in question. That is, for example, the phonetic feature that differentiates a voiced stop or fricative from a voiceless one is generally expected to be coincidental in time with the closure for that sound. 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."[7] One could further hold with Lisker that "each acoustic segment might be said to supply cues to a single inguistic segment, while any 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."[8] 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."

It appears, then, that any system of distinctive feature analysis which purports to account for the production and perception of English sounds, or, more generally, for all sounds in all natural languages, should take durational factors into consideration, at least in the sense in which they have been investigated here.

Footnotes - Chapter IV

[1] L. Lisker and A. S. Abramson, "Transillumination of the Larynx in Running Speech," Haskins Laboratories Status Reports on Speech Research, 5-6 (1966), 4.4.

[2] Harry Rositzke, "The Articulation of Final Stops in General American Speech," American Speech, 18 (1943), 39. See also, A. Malécot, "The Role of Releases in the Identification of Released Final Stops," Language, 34 (1958), 370.

[3] P. C. Delattre, "Acoustic Cues in Speech: First Report," translated by the author from the article appearing in Phonetica, 2 (1958), 108-118, 226-251. Translated version distributed in mimeographed form by Haskins Laboratories, pp. 29-30.

[4] Lisker and Abramson, "Transillumination," p. 4.4.

[5] The claim that vowel duration is a necessary cue is made only for the synthetic stimuli used in this study. It may well be that the addition of other cues to the synthetic speech, or the presence of other cues in real speech could render vowel duration less significant as a cue.

[6] A. M. Liberman, F. S. Cooper, D. P. Shankweiler, and M. Studdert-Kennedy, "Perception of the Speech Code," Psychological Review, 74 (1967), 442-443.

[7] L. Lisker, "Linguistic Segments, Acoustic Segments, and Synthetic Speech," Language, 33 (July-September, 1957), 372.

[8] Ibid.

Appendix 1

ITEMIZED CORPUS OF WORDS USED FOR SPECTROGRAPHIC MEASUREMENTS OF POTENTIAL CUES

Preceding Vowel Sound	/p-b/	/t-d/	/k-g/
/i/	heap-Hebe	pleat-plead	leak-league
/I/	nip-nib	bit-bid	wick-wig
/e/	gape-Gabe	rate-raid	*fake-vague
/ɛ/	rep-reb	bet-bed	peck-peg
/æ/	*lap-lab	hat-had	*sack-sag
/ɑ/	mop-mob	pot-pod	clock-clog
/ɔ/		sought-sawed	*hawk-dog
/o/	lope-lobe	moat-mowed	broke-brogue
/ʊ/		*put-good	
/u/	loop-lube	suit-sued	*duke-Krug
/ʌ/	sup-sub	but-bud	tuck-tug
/ɝ/	*slurp-Herb	hurt-heard	Burke-Berg
/aɪ/	tr'pe-tribe	height-hide	
/ɔɪ/		*Hoyt-Boyd	
/aʊ/		clout-cloud	
/ə/		pallet-pallid	

---

*sub-minimal pair

Dummy Corpus (incorporated in randomized list with test corpus above)

Preceding Vowel Sound			
/i/	heave	please	leash
/ɪ/	niche	bin	wish
/ɛ/	wretch	Bess	pen
/e/	gave	rain	**raise
/æ/	**lax	hang	**sash
/ɑ/	mom	palm	clotch
/ɔ/		song	**dawn
/o/	loaf	moan	broach
/ʊ/		**puss	
/u/	lose	soothe	dune
/ʌ/	sun	bus	touch
/ɝ/	**hearse	Hersh	birch
/aI/	tries	wine	
/ɔɪ/		**boys	
/aʊ/		clown	
/ə/		palace	

**dummy item for sub-minimal pair

Appendix 2

TEST INSTRUCTIONS: TAPE-SPLICING EXPERIMENTS

The speaker on the tape you are about to hear will be saying one of the two words listed on the top of response sheet A each time he speaks. Your task will be to determine which of the two words the speaker intended to say each time you hear him speak. To indicate your decision, circle the word that you think was intended next to the appropriate item number on the response sheet. Since a particular word was intended in every case, you should respond to every item spoken, even if you are not sure of the speaker's intention, and so have to guess.

There are no "right" or "wrong" answers on this test. We simply want to find out what you hear, individually, and as a group. There is therefore no need to be concerned if you are uncertain of your response, or if your responses are not in agreement with those of your classmates, in whole or in part.

I will now play five practice items for you, for which space has been provided on the top of the sheet. These are the items labelled A, B, C, D, and E. Are you ready? Then here are the practice items.

(The practice items are played)

Are there any questions you would like to ask at this time? (Any questions which are not likely to contaminate the experimental results are answered.)

We are now ready to begin the test proper. We will begin with item number one (1). Remember, you must respond to every item you hear by circling the word that you think was intended, or by guessing if you are not sure.

(The first test tape is played)

Turn now to response sheet B.[1] Your task this time is exactly the same, but with another pair of words, the pair listed on the top of the new response sheet. Here again are five practice items.

(The practice items are played)

Are there any questions you would like to ask at this time? We are now ready to begin the next test. Remember to respond to every item you hear.

(The second test tape is played)

---

[1] This section of the instructions, with appropriate changes in the response sheet letter, is repeated for each succeeding test tape played.

SAMPLE RESPONSE SHEET: TAPE-SPLICING EXPERIMENTS--VOICING OFFSET TIME, VOWEL DURATION

* * *

NAME________________________________

RESPONSE SHEET A

PECK PEG

PRACTICE ITEMS:

A. PECK PEG

B. PECK PEG

C. PECK PEG

D. PECK PEG

E. PECK PEG

TEST ITEMS:

1. PECK PEG

2. PECK PEG

3. PECK PEG

4. PECK PEG

5. PECK PEG

6. PECK PEG

7. PECK PEG

8. PECK PEG

9. PECK PEG

10. PECK PEG

(over)

Appendix 3

TEST INSTRUCTIONS: SYNTHETIC-SPEECH LABELLING EXPERIMENTS

The speech you are about to hear is spoken by a machine. It was produced by a speech synthesizer called the Pattern Playback, which was developed at Haskins Laboratories in New York City. The speech does not sound like human speech, but this is basically due to a lack of melody (the machine speaks on a monotone), and to a difference between the voice quality of a machine and a human being. The actual speech sounds spoken are, however, quite intelligible, and you should have very little difficulty understanding the words and sentences produced by the synthesizer.

In order for you to become familiar with this type of speech, I am going to play some samples of the synthesizer saying the sentence: "I am a speech synthesizer."

(Play: "I am a speech synthesizer." three times.)

Here is another sentence. This time I will not tell you what it is until you have heard it twice.

(Play: "Alexander is an intelligent conversationalist."
two times.)

The sentence you heard was, "Alexander is an intelligent conversationalist."

(Play: "Alexander is an intelligent conversationalist."
once.)

Is there anyone here who could not understand the second sentence after I told you what it was? (If there is, he is instructed to write that fact on the top of the first response sheet. His results will be discounted for use in the study.)[1]

On the tape you are about to hear now, the synthesizer, each time it speaks, will say one of the two words listed on the top of response sheet A. Your task will be to determine which of the two words the synthesizer said, each time you hear it speak. You will indicate your decision by circling the word that you heard, next to the appropriate number on the response sheet. You should respond to every item, even if you are not sure which of the two words you heard. If you are not sure, therefore, you must guess.

There are no "right" or "wrong" answers in this test. We simply want to find out what you hear individually and as a group. There is, therefore, no need to be concerned if you are uncertain of your response, or if your responses are not in agreement with those of your classmates, in whole or in part.

I will now play five practice items for you, for which space has been provided on the top of response sheet A, in the items labelled A, B, C, D, and E. Are you ready? Then here are the practice items.

---

[1] There were no such subjects found in the course of the listening tests.

(Practice items are played)

Are there any questions you would like to ask at this time? (Any questions which are not likely to contaminate the experimental results are answered.)

We are now ready to begin the test proper. Remember, you must respond to every item you hear by circling one of the two choices next to the appropriate number. Circle the word you hear in each case, or guess if you are not sure.

(The first test tape is played)

Turn now to response sheet B.[2] Your task this time is exactly the same, but with another pair of words, the pair listed at the top of the new response sheet. Here again are five practice items:

(The practice items are played)

Are there any questions that you would like to ask at this time?

We are now ready to begin the next test. Remember to respond to every item you hear.

---

[2] This section of the instructions, with appropriate changes in the response sheet letter, is repeated for each succeeding test tape played.

SAMPLE RESPONSE SHEET: SYNTHETIC-SPEECH LABELLING EXPERIMENT

* *

NAME________________________________

RESPONSE SHEET A

<u>PICKS</u> <u>PIGS</u>

PRACTICE ITEMS:

A. PICKS PIGS
B. PICKS PIGS
C. PICKS PIGS
D. PICKS PIGS
E. PICKS PIGS

TEST ITEMS:

1. PICKS PIGS
2. PICKS PIGS
3. PICKS PIGS
4. PICKS PIGS
5. PICKS PIGS
6. PICKS PIGS
7. PICKS PIGS
8. PICKS PIGS
9. PICKS PIGS
10. PICKS PIGS

Appendix 4

TEST INSTRUCTIONS: DISCRIMINATION OF SYNTHETIC STIMULI

The tapes you are about to hear contain synthesized speech of the same sort as that which you heard on the last set of test tapes. Your task this time is somewhat different. On these tapes you will hear three successive synthetic utterances, separated from each other by one-second silences. Two of these three items are identical; the third differs from the other two. Your task is to pick out the odd stimulus from each group of three.

You will indicate your choice by circling the appropriate letter next to the proper utterance number on the list: circle A if you feel that the first item of the three was different from the second two; circle B if you feel that the second item differed from the first and the third; finally, if you feel that the first two items were the same and that the third was the "oddball" in the group, circle the letter C.

You must respond once to every group of three utterances that you hear, even if you have to guess because of uncertainty as to which item is unique among the three.

As with the other test tapes, there are no "right" or "wrong" answers. We just want to see what you hear, both individually and as a group. There is, therefore, no need to be concerned if you are uncertain of your responses, or if your responses are not in agreement with those of your classmates, in whole or in part.

I will now play five practice items for you, for which space has been provided on the top of response sheet A in the items labelled A, B, C, D, and E. Are you ready? Here are the practice items.

(Practice items are played)

Are there any questions you would like to ask at this time? (Any questions which are not likely to contaminate the experimental results are answered.)

We are now ready to begin the test proper. Remember, you must respond once to every group of three items that you hear by circling the letter of the odd item in the group. Guess if you are not sure.

(The first test tape is played)

Turn now to response sheet B.[1] Your task this time is exactly the same, but the utterances will be taken from the words listed on the top of the new response sheet. Here again are a few practice items.

(The practice items are played)

Are there any questions that you would like to ask at this time?

---

[1] This section of the instructions, with appropriate changes in the response sheet letter, is repeated for each succeeding test tape played.

We are now ready to begin the next test. Remember to respond once to every group of three items that you hear, circling the letter of the item which is different from the other two.

(The second test tape is played)

SAMPLE RESPONSE SHEET:

DISCRIMINATION OF SYNTHETIC STIMULI

* *

NAME________________________________

RESPONSE SHEET A

PICKS PIGS

PRACTICE ITEMS:

A. a b c

B. a b a

C. a b c

D. a b c

E. a b c

TEST ITEMS:

1. a b c

2. a b c

3. a b c

4. a b c

5. a b c

6. a b c

7. a b c

8. a b c

9. a b c

10. a b c

Appendix 5

DERIVATION OF THE DISCRIMINATION PREDICTION EQUATION

The subjects in the discrimination test are assumed to be able to discriminate only as well as they have labelled on the perception test. They have two alternatives on the perception test, and the same two possible ways of identifying the stimuli on the discrimination test.

Each stimulus can be discriminated in two unique ways as it appears in a triad: one in which it is the majority member of the triad (occurring twice), while another stimulus is the "odd ball," e.g. A-B-A; and one in which it is the odd member of the triad, e.g. A-B-B.

In the prediction equation (using the /p/-/t/ opposition as an example here) we shall let:

P_t = the probability of an S hearing the A stimulus as /t/.

P_t' = the probability of an S hearing the B stimulus as /t/.

P_d = the probability of an S hearing the A stimulus as /d/.

P_d' = the probability of an S hearing the B stimulus as /d/.

Thus the subject can hear the A stimulus as /t/ or as /d/ and the B stimulus as /t/ or as /d/. This means that there are eight ways in which any given triad can be heard, each of which will cause the subject to make a certain decision or type of decision as to which stimulus in the triad is the "odd ball." Some ways will lead to a correct decision, others to an incorrect decision, others to a guess.

For a triad of the type A-B-A, the possible ways are:

Leading to a correct decision	Leading to an incorrect decision	Leading to a guess
t-d-t	t-t-d	t-t-t
d-t-d	t-d-d	d-d-d
	d-t-t	
	d-d-t	

For a triad of the type A-B-B, the possibly ways are:

Leading to a correct decision	Leading to an incorrect decision	Leading to a guess
t-d-d	d-t-d	t-t-t
d-t-t	t-d-t	d-d-d
	d-d-t	
	t-t-d	

Since both types of triads are possible, but only one can occur at a time, the probability of a correct decision in the case of any triad will be one-half the sum of the probabilities of a correct decision for each of the two basic types:

$$P_{corr} = 1/2(P_{corr}\ (ABA) + P_{corr}\ (ABB))$$

Thus:

$$\begin{aligned} P_{corr}\ (ABA) &= 1 \times P_{corr} + 0 \times P_{incorr} + .33\ P_{guess} \\ &= P_t\ P_d'\ P_t + P_d\ P_t'\ P_d + .33(P_t\ P_t'\ P_t + \\ &\qquad P_d\ P_d'\ P_d) \end{aligned}$$

Since $P_d = 1-P_t$ and $P_d' = 1-P_t'$ we may substitute as follows:

$$= P_t^2(1-P_t') + P_t'(1-P_t)^2 + .33(P_t^2P_t' + ((1-P_t'))((1-P_t))^2)$$

$$= P_t^2-P_t^2P_t' + P_t'(1-2P_t+ P_t^2) + .33(P_t^2P_t' +((1-P_t'))$$
$$((1-2P_t + P_t^2)))$$

$$= P_t^2-P_t^2P_t' +P_t'-2P_tP_t' + .33(P_t^2P_t' + 1-P_t'-2P_t + 2P_tP_t'$$
$$+ P_t^2-P_t^2P_t')$$

$$= P_t^2+P_t'-2P_tP_t' + .33-.33P_t'-.66P_t + .66P_tP_t' + .33P_t^2$$

$$P_{corr}\ (ABA) = 1.33P_t^2 + .66P_t' - 1.33P_tP_t' - .66P_t + .33$$

$$P_{corr}\ (ABB) = 1 \times P_{corr} + 0 \times P_{incorr} + .33P_{guess}$$

$$= P_tP_d'P_d' + P_dP_t'P_t' + .33(P_tP_t'P_t' + P_dP_d'P_d')$$

Since $P_d = 1-P_t$ and $P_d' = 1-P_t'$ we may substitute as follows:

$$= P_t(1-P_t')^2 + (1-P_t)P_t'^2 + .33(P_tP_t'^2 + ((1-P_t))$$
$$((1-P_t'))^2)$$

$$= P_t(1-2P_t' + P_t'^2) + P_t'^2-P_tP_t'^2 + .33(P_tP_t'^2 + ((1-P_t))$$
$$((1-2P_t' + P_t'^2)))$$

$$= P_t -2P_tP_t' +P_tP_t'^2 + P_t'^2 -P_tP_t'^2 + .33(1-P_t -2P_t'$$
$$+2P_tP_t' +P_t'^2)$$

$$= P_t -2P_tP_t' +P_t'^2 + .33 -.33P_t -.66P_t' + .\underline{6}6P_tP_t'$$
$$+.33P_t'^2$$

$$P_{corr}\ (ABB) = .66P_t - 1.33P_tP_t' + 1.33P_t'^2 -.66P_t' + .33$$

Adding P_{corr} (ABA) to P_{corr} (ABB) we get:

$$P_{corr} = 1.33P_t^2 - 2.66P_tP_t' + .66 + 1.33P_t'^2$$

$$= 1.33(P-P')^2 + .66$$

Dividing by two yields:

$$P_{corr} = .66(P-P')^2 + .33$$

BIBLIOGRAPHY

Abercrombie, David. *Elements of General Phonetics.* Chicago: Aldine Publishing Company, 1967.

Abramson, Arthur S.; Lisker, Leigh, and Cooper, Franklin S. "Laryngeal Activity in Stop Consonants." *Haskins Laboratories Status Reports on Speech Research*, 4 (November, 1965), 6.1-6.13.

Abramson, Arthur S., and Lisker, Leigh. "Voice Onset Time in Stop Consonants: Acoustic Analysis and Synthesis." *Haskins Laboratories Status Reports on Speech Research*, 5 (August, 1965), 1.1-1.7.

Bastian, Jarvis; Eimas, Peter D.; and Liberman, Alvin M. "Identification and Discrimination of a Phonemic Contrast Induced by Silent Interval." *Journal of the Acoustical Society of America*, 33 (June, 1961), 842A.

Bastian, Jarvis, and Abramson, Arthur S. "Identification and Discrimination of Phonemic Vowel Duration." *Journal of the Acoustical Society of America*, 34 (May, 1962), 743-44A.

Bastian, Jarvis; Delattre, Pierre C.; and Liberman, Alvin M. "Silent Interval as a Cue for the Distinction Between Stops and Semivowels in Medial Position." *Journal of the Acoustical Society of America*, 31 (November, 1959), 1568A.

Belasco, Simon. "Influence of Force of Articulation of Consonants on Vowel Duration." *Journal of the Acoustical Society of America*, 25 (September, 1953), 1015-16.

Black, John W. "The Pressure Component in the Production of Consonants." *Journal of Speech and Hearing Disorders*, 15 (June, 1960), 207-10.

Bronstein, Arthur J. *The Pronunciation of American English.* New York: Appleton-Century-Crofts, 1960.

Cooper, Franklin S. "Spectrum Analysis." *Journal of the Acoustical Society of America*, 22 (November, 1950), 761-62.

Cooper, Franklin S.; Delattre, Pierre C.; Liberman, Alvin M.; Borst, John M.; and Gerstman, Louis J. "Some Experiments on the Perception of Synthetic Speech Sounds." Journal of the Acoustical Society of America, 24 (November, 1952), 597-606.

Cooper, Franklin S.; Liberman, Alvin M.; and Borst, John M. "The Interconversion of Audible and Visible Patterns as a Basis for Research in the Perception of Speech." Proceedings of the National Academy of Science, 37 (May, 1951), 318-25.

Delattre, Pierre C. "Acoustic Cues in Speech." Phonetica, 2 (1958), 108-18, 226-51.

________. "Some Factors of Vowel Duration and Their Cross-Linguistic Validity." Journal of the Acoustical Society of America, 34 (August, 1962), 1141-43.

________. "The Physiological Interpretation of Sound Spectrograms." PMLA, 66 (September, 1951), 864-76.

Delattre, Pierre C.; Liberman, Alvin M.; and Cooper, Franklin S. "Acoustic Loci and Transitional Cues for Consonants." Journal of the Acoustical Society of America, 27 (July, 1955), 769-73.

________. "Formant Transitions and Loci as Acoustic Correlates of Place of Articulation in American Fricatives." Studia Linguistica, 16 (1964), 104-21.

Denes, Peter. "Effect of Duration on the Perception of Voicing." Journal of the Acoustical Society of America, 27 (July, 1955), 761-64.

Eimas, Peter D. "The Relation Between Identification and Discrimination Along Speech and Non-Speech Continua." Language and Speech, 3 (October-December, 1963), 206-17.

Elert, Claes-Christian. Phonologic Studies of Quantity in Swedish. Uppsala: Almqvist and Wiksells, 1964.

Fischer-Jørgensen, Ely. "Acoustic Analysis of Stop Consonants." Miscellanea Phonetica, 2, Nos. 1-3 (1954), 42-59.

Fromkin, Victoria A. "Some Phonetic Specifications of Linguistic Units: An Electromyographic Investigation." Working Papers in Phonetics, UCLA, Whole No. 3 (1965).

Fry, Dennis B. "Perception and Recognition in Speech." For Roman Jakobson. Edited by Morris Halle. 's-Gravenhage: Mouton Press, 1956, 169-73.

Fry, Dennis B.; Abramson, Arthur S.; Eimas, Peter D.; and Liberman, Alvin M. "The Identification and Discrimination of Synthetic Vowels." Language and Speech, 5 (October-December, 1962), 171-89.

Gerstman, Louis J. "Cues for Distinguishing Among Fricatives, Affricate and Stop Consonants." Unpublished Ph.D. Dissertation, New York University, 1957.

Gimson, A. C. An Introduction to the Pronunciation of English. London: Edward Arnold Ltd., 1962.

Gleason, Henry A., Jr. An Introduction to Descriptive Linguistics. New York: Holt, Rinehart and Winston, 1958.

Halle, Morris. "Phonology in a Generative Grammar." Word, 18 (April-August, 1962), 54-72.

Halle, Morris; Hughes, G. W.; and Radley, J-P. A. "Acoustic Properties of Stop Consonants." Journal of the Acoustical Society of America, 29 (January, 1957), 107-16.

Harris, Katherine S. "Cues for the Discrimination of American English Fricatives." Language and Speech, 1 (January-March, 1958), 1-7.

Harris, Katherine S.; Lysaught, Gloria F.; and Schvey, Malcom M. "Some Aspects of the Production of Oral and Nasal Labial Stops." Language and Speech, 8 (July-September, 1965), 135-47.

Harris, Katherine S.; Hoffman, Howard; Liberman, Alvin M.; Delattre, Pierre C.; and Cooper, Franklin S. "Effect of Third-Formant Transition on the Perception of the Voiced Stop Consonants." Journal of the Acoustical Society of America, 30 (February, 1958), 122-26.

Harris, Katherine S.; Bastian, Jarvis; and Liberman, Alvin M. "Mimicry and the Perception of a Phonemic Contrast Induced by Silent Interval: Electromyographic and Acoustic Measures." Journal of the Acoustical Society of America, 33 (June, 1961), 842A.

Heffner, Roe-Merrill S. "An Adjunct to the Graphic Method." American Speech, 16 (February, 1941), 32-38.

________. General Phonetics. Madison: University of Wisconsin Press, 1964.

________. "Notes on the Length of Vowels." American Speech, 12 (May, 1937), 128-34.

________. "Notes on the Length of Vowels (IV)." American Speech, 16 (October, 1941), 204-07.

________. "Notes on the Length of Vowels (V)." American Speech, 17 (February, 1942), 42-48.

Heffner, Roe-Merrill S., and Locke, W. N. "Notes on the Length of Vowels (II)." American Speech, 15 (February, 1940), 74-79.

Heffner, Roe-Merrill S., and Lehman, W. P. "Notes on the Length of Vowels (III)." American Speech, 15 (December, 1940), 377-80.

________. "Notes on Length of Vowels (VI)." American Speech, 18 (October, 1943), 208-15.

Hockett, Charles F. "A Manual of Phonology." International Journal of American Linguistics, Memoir 11, 1955.

Hoffman, Howard. "Study of Some Cues in the Perception of the Voiced Stop Consonants." Journal of the Acoustical Society of America, 30 (November, 1958), 1035-41.

House, Arthur S. "On Vowel Duration in English." Journal of the Acoustical Society of America, 33 (September, 1961), 1174-78.

House, Arthur S., and Fairbanks, Grant. "The Influence of Consonant Environment upon the Secondary Acoustical Characteristic of Vowels." Journal of the Acoustical Society of America, 25 (January, 1953), 105-13.

Householder, F. W. "Unreleased ptk in American English." For Roman Jakobson. Edited by Morris Halle. 's-Gravenhage: Mouton Press, 1956, pp. 235-44.

Hughes, G. W., and Halle, Morris. "Spectral Properties of Fricative Consonants." Journal of the Acoustical Society of America, 28 (March, 1956), 303-10.

Hultzén, Lee S. "Voiceless Lenis Stops in Prevocalic Clusters." Word, 18 (December, 1962), 307-12.

Jakobson, Roman; Fant, C. Gunnar; and Halle, Morris. Preliminaries to Speech Analysis. Cambridge, Mass.: MIT Press, 1967.

Jakobson, Roman, and Halle, Morris. "Tenseness and Laxness." Supplement to Preliminaries to Speech Analysis. Cambridge, Mass.: MIT Press, 1967, pp. 57-61.

Jones, Daniel. The Phoneme: Its Nature and Use. Cambridge: W. Heffner and Sons Ltd., 1950.

________. The Pronunciation of English. 4th ed. Cambridge: Cambridge University Press, 1966.

Kantner, Claude E., and West, Robert. Phonetics. Revised ed. New York: Harper and Brothers, 1960.

Liberman, Alvin M. "Some Results of Research on Speech Perception." Journal of the Acoustical Society of America, 29 (January, 1957), 117-23.

Liberman, Alvin M.; Delattre, Pierre C.; and Cooper, Franklin S. "Some Cues for the Distinction Between Voiced and Voiceless Stops in Initial Position." Language and Speech, 1 (July-September, 1958), 153-67.

Liberman, Alvin M.; Harris, Katherine S.; Hoffman, Howard S.; and Griffith, Belver C. "The Discrimination of Speech Sounds Within and Across Phoneme Boundaries." Journal of Experimental Psychology, 54 (November, 1957), 358-68.

Liberman, Alvin M.; Harris, Katherine S.; Kinney, Jo Anne; and Lane, Harlan. "The Discrimination of Certain Speech and Nonspeech Patterns." Journal of Experimental Psychology, 61 (May, 1961), 379-88.

Liberman, Alvin M.; Harris, Katherine S.; Eimas, Peter; Lisker, Leigh; and Bastian, Jarvis. "An Effect of Learning on Speech Perception: The Discrimination of Durations of Silence with and without Phonemic Significance." Language and Speech, 4 (October-December, 1961), 175-95.

Liberman, Alvin M.; Ingemann, Frances; Lisker, Leigh, Delattre, Pierre C.; and Cooper, Franklin S. "Minimal Rules for Synthesizing Speech." Journal of the Acoustical Society of America, 31 (November, 1959), 1490-99.

Liberman, Alvin M.; Cooper, Franklin S.; Harris, Katherine S.; MacNeilage, Peter F.; and Studdert-Kennedy, Michael. "Some Observations for the Perception of Speech and Visual Form." Models for the Perception of Speech and Visual Form. Edited by Weiant Wathen-Dunn. Cambridge, Mass.: MIT Press, 1967, pp. 68-87.

Liberman, Alvin M.; Cooper, Franklin S.; Shankweiler, Donald P.; and Studdert-Kennedy, Michael. "Perception of the Speech Code." Psychological Review, 74 (November, 1967), 431-61.

Liberman, Alvin M.; Studdert-Kennedy, Michael; Harris, Katherine S.; and Cooper, Franklin S. "A Reply to 'Identification and Discrimination Functions for a Visual Continuum and Their Relation to the Motor Theory of Speech Perception,' by Cross, Lane, and Sheppard." Haskins Laboratories Status Reports on Speech Research, 3 (August, 1965), 3.1-3.14.

Liberman, Alvin M.; Delattre, Pierre C.; Cooper, Franklin S.; and Gerstman, Louis J. "The Role of Consonant-Vowel Transitions in the Perception of the Stop and Nasal Consonants." Psychological Monographs, 68, No. 8, whole No. 379 (1954).

Liberman, Alvin M.; Delattre, Pierre C.; and Cooper, Franklin S. "The Role of Selected Stimulus Variables in the Perception of Unvoiced Stop Consonants." American Journal of Psychology, 65 (October, 1952), 497-516.

Liberman, Alvin M.; Delattre, Pierre C.; Gerstman, Louis J.; and Cooper, Franklin S. "Tempo of Frequency Change as a Cue for Distinguishing Classes of Speech Sounds." Journal of Experimental Psychology, 52 (August, 1956), 127-137.

Lieberman, Philip. Intonation, Perception and Language. Cambridge, Mass.: MIT Press, 1967.

Lisker, Leigh. "Closure Duration and the Intervocalic Voiced-Voiceless Distinction in English." Language, 33 (January-March, 1957), 42-49.

________. "The English Stops after /s/ at Word Boundary: A Three-Way Contrast." 11th Final Report on Speech Research and Instrumentation, Haskins Laboratories (March, 1965), 1-7.

________. "On Hultzen's 'Voiceless Lenis Stops in Prevocalic Clusters.'" Word, 19 (December, 1963), 376-87.

________. "Linguistic Segments, Acoustic Segments, and Synthetic Speech." Language, 33 (July-September, 1957), 370-74.

________. "Measuring Stop Closure Duration from Intraoral Pressure Records." Haskins Laboratories Status

Reports on Speech Research, 7/8 (July-December, 1966), 5.1-5.6.

________. "Supraglottal Air Pressure in the Production of English Stops." Haskins Laboratories Status Reports on Speech Research, 4 (November, 1965), 3.1-3.15.

________, and Abramson, Arthur S. "A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements." Word, 20 (December, 1964), 384-422.

________. "Some Effects of Context on Voice Onset Time in English Stops." Language and Speech, 10 (January-March, 1967), 1-28.

________. Stop Categorization and Onset Time." Proceedings of the Fifth International Congress of Phonetic Sciences, Münster, 1964. Edited by E. Zwirner and W. Bethge. New York: S. Karger, 1964, pp. 389-391.

Lisker, Leigh; Abramson, Arthur S.; Cooper, Franklin S.; and Schvey, Malcom H. "Transillumination of the Larynx in Running Speech." Haskins Laboratories Status Reports on Speech Research, 5-6 (November-June, 1965-6), 4.1-4.8.

Lorge, Beatrice. "A Study of the Relationship Between Production and Perception of Initial and Intervocalic /t/ and /d/ in Individual English Speaking Adults." Haskins Laboratories Status Reports on Speech Research, 9 (January-March, 1967), 3.1-3.18.

Lotz, John; Abramson, Arthur S.; Gerstman, Louis J.; Ingemann, Frances; and Nemser, William J. "The Perception of English Stops by Speakers of English, Spanish, Hungarian, and Thai: A Tape-Cutting Experiment." Language and Speech, 3 (April-June, 1960), 71-76.

MacNeilage, Peter F. "Electromyographic and Acoustic Study of the Production of Certain Final Clusters." Journal of the Acoustical Society of America, 35 (April, 1963), 461-63.

Malécot, Andre. "The Effectiveness of Intra-Oral Air-Pressure-Pulse Parameters in Distinguishing Between Stop Cognates." Phonetica, 14, No. 2 (1966), 65-81.

________. "An Experimental Study of Force of Articulation." Studia Linguistica, 9, No. 1 (1955), 35-44.

________. "Mechanical Pressure as an Index of 'Force of Articulation.'" *Phonetica*, 14, No. 3 (1966), 169-80.

________. "The Role of Releases in the Identification of Released Final Stops." *Language*, 34 (July-September, 1958), 370-80.

Noll, John Douglas. "The Perceptual Significance of Certain Acoustical Correlates of Consonant Voicing Contrasts." Unpublished Ph.D. Dissertation, University of Iowa, 1960.

Oswald, V. A. "Voiced /t/--A Misnomer." *American Speech*, 18 (February, 1943), 18-25.

Parmenter, C. E., and Treviño, S. N. "The Length of the Sounds of a Middle Westerner." *American Speech*, (May, 1935), 129-33.

Peterson, Gordon E., and Lehiste, Ilse. "Duration of Syllable Nuclei in English." *Journal of the Acoustical Society of America*, 32 (June, 1960), 693-703.

Pilch, H. "The Rise of the American English Vowel Pattern." *Word*, 11 (April, 1955), 57-93.

Reeds, J. A., and Wang, W-S. Y. "The Perception of Stops After *s*." *Phonetica*, 6, Nos. 1-3 (1961), 78-81.

Rositzke, Harry. "The Articulation of Final Stops in General American Speech." *American Speech*, 18 (February, 1943), 39-42.

________. "A Note on Final Stops in General American Speech." *Le Maître Phonétique*, 13 (October-December, 1938), 64-66.

________. "Vowel-Length in General American Speech." *Language*, 15 (April-June, 1939), 99-109.

Roudet, L. *Élements de Phonétique Générale*. Paris: H. Welter, 1910.

Schatz, Carol D. "The Role of Context in the Perception of Stops." *Language*, 30 (January-March, 1954), 47-56.

Sharf, Donald J. "Distinctiveness of 'Voiced t' Words." *American Speech*, 35 (May, 1960), 105-09.

________. "Duration of Post-Stress Intervocalic Stops and Preceding Vowels." *Language and Speech*, 5 (January-March, 1962), 26-30.

Sholes, George N. "Acoustic Cues for Final /s/ and /z/." *Haskins Laboratories Quarterly Progress Reports*, No. 31 (1959), Appendix 3.

________. "Hiss Alone as a Cue for Final /z/ on Voback." *Haskins Laboratories Quarterly Progress Reports*, No. 30 (1959), Appendix 1.

________. "Synthesis of Final /z/ Without Voicing." *Journal of the Acoustical Society of America*, 31 (November, 1959), 1568A.

Stetson, R. H. *Motor Phonetics*. Amsterdam: North Holland Publishing Company, 1951.

"Stop Consonants in Final Position." *Haskins Laboratories Quarterly Progress Reports*, No. 21 (1956), Appendix 5.

Stevens, Joseph C., and Hall, James W. "Brightness and Loudness as Functions of Stimulus Duration." *Perception and Psychophysics*, 1 (September, 1966), 319-27.

"Terminal Stops and Nasals." *Haskins Laboratories Quarterly Progress Reports*, No. 13 (1954), Appendix 2.

Thomas, Charles Kenneth. *An Introduction to the Phonetics of American English*. 2nd ed. New York: Ronald Press, 1958.

Wang, William S-Y. "Transition and Release as Perceptual Cues for Final Plosives." *Journal of Speech and Hearing Research*, 2 (March, 1959), 66-73.

Wise, Claude M. *Introduction to Phonetics*. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1957.

Zimmerman, S. A., and Sapon, S. M. "A Note on Vowel-Duration Seen Cross-Linguistically." *Journal of the Acoustical Society of America*, 30 (February, 1958), 152.

AUTOBIOGRAPHICAL STATEMENT

I was born in New York City in 1938 and educated in the Public Schools there. I entered Queens College of the City University of New York in 1955 and was graduated with a B.A. in general speech in 1959. After a year of graduate work at the University of Illinois I returned to New York City where I taught in the high schools from 1960 to 1965. I was married to the former Miss Carolyn Bauling in 1963. In June, 1965, I received an M.A. from Queens College in the areas of English Linguistics and Speech Science. In 1965 I began to teach at Bronx Community College of the City University of New York, and shortly thereafter entered the City University doctoral program in Speech, taking an interdisciplinary course of study in Speech, English, Anthropology and Psychology. From 1967 to 1969 I taught in the Department of Communication Arts and Sciences at Queens College of the City University of New York. In September, 1969, I took a position at Herbert H. Lehman College of the City University of New York as a lecturer in the Department of Speech and Theatre.