72-14,964

PISONI, David Bob, 1945-
ON THE NATURE OF CATEGORICAL PERCEPTION OF SPEECH SOUNDS.

The University of Michigan, Ph.D., 1971
Psychology, experimental

University Microfilms, A XEROX Company, Ann Arbor, Michigan

THIS DISSERTATION HAS BEEN MICROFILMED EXACTLY AS RECEIVED

ON THE NATURE OF CATEGORICAL PERCEPTION OF SPEECH SOUNDS

by
David Bob Pisoni

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Psycholinguistics)
in The University of Michigan
1971

Doctoral Committee:

Professor Irwin Pollack, Chairman
Professor Edwin Martin
Assistant Professor Michael A. H. O'Malley
Associate Professor Michael Studdert-Kennedy, City University of New York
Professor Ronald S. Tikofsky

PLEASE NOTE:

Some pages have indistinct print. Filmed as received.

University Microfilms, A Xerox Education Company

ACKNOWLEDGEMENTS

The experiments reported in this thesis would have been impossible without the use of the unique facilities for speech research at Haskins Laboratories. I should like to express my deepest thanks and appreciation to Dr. Franklin S. Cooper and Professor Alvin M. Liberman for providing me with an opportunity to work at the Laboratories and for their interest, guidance and encouragement. I am particularly indebted to Professor Michael Studdert-Kennedy for his personal interest in my work and the many long hours spent with me during the course of the present investigation.

I would like to thank my advisor and chairman Professor Irwin Pollack for encouraging me to pursue this work and for his never-ending patience and insightful comments during all phases of this dissertation. I am also grateful to the other members of my dissertation committee: Professor Edwin Martin, Professor Michael O'Malley and Professor Ronald Tikofsky for their advice, criticism and suggestions.

Special thanks are also due the following people who have helped in many different ways: Professors Kenneth Stevens and Dennis Klatt of M.I.T. who provided the use of their speech synthesizer for the preparation of vowel stimuli; Professor Kurt Salzinger for making facilities available at Brooklyn Polytechnic Institute; Professor Donald Shankweiler for making facilities available at the University of Connecticut; Professors Arthur Abramson and Leigh Lisker of Haskins for suggestions and advice on various matters dealing with speech synthesis and acoustic phonetics; Professor Louis J. Gerstman who first introduced me to the study of speech perception and who offered helpful suggestions in planning several experiments; and Jim Cutting, Chris Darwin, Ruth Day, Gary Kuhn and Larry Raphael for listening to my ideas, correcting my mistakes and helping me in many ways too numerous to mention.

Finally, I would like to thank Beth Greene for her patience, understanding and assistance in every phase of the present investigation and Lydia Peterson for her careful preparation of the manuscript.

The research reported in this dissertation was supported in part by the following sources: a grant from the National Institute of Child Health and Human Development to Haskins Laboratories; a grant from the

National Science Foundation to Irwin Pollack (Grant GB 14036X); the Phonetics Laboratory of the University of Michigan; a grant from the National Institute of Mental Health to Ronald Tikofsky; a Rackham Prize Fellowship from the Graduate School of the University of Michigan; and a dissertation grant from the Horace H. Rackham School of Graduate Studies which facilitated the use of paid experimental subjects.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	ii
LIST OF TABLES	vi
LIST OF FIGURES	vii
LIST OF APPENDICES	ix
CHAPTER I	1
INTRODUCTION	1
Why Is Speech Special?	1
Auditory and Phonetic Perception of Speech	6
Categorical and Continuous Perception	8
CHAPTER II	12
EXPERIMENT I - Identification and Discrimination of Consonants and Vowels	12
Method	12
Results	19
Summary and Discussion	28
EXPERIMENT II - Effect of Discrimination Training on Categorical Perception	29
Method	31
Results	37
Discussion	46
EXPERIMENT III - Effect of Stimulus Duration on Vowel Perception	49
Method	51
Results	54
Discussion	57
CHAPTER III	62
EXPERIMENT IV - Four-Interval Test of Paired Similarity in Speech Discrimination	62
Method	64
Results	66
Discussion	68
EXPERIMENT V - A Comparison of the 4IAX and ABX Tests with Stop Consonants	72
Method	72
Results	74
Discussion	77

Page

CHAPTER IV..... 78

EXPERIMENT VI - Auditory Short-Term Memory in Speech Discrimination..... 78
Method..... 79
Results..... 82
Discussion..... 87

CHAPTER V..... 93

SUMMARY AND DISCUSSION..... 93
Summary of Experiments..... 93
Explanations of Categorical Perception..... 95

REFERENCES..... 100

APPENDIX A..... 107

APPENDIX B..... 114

APPENDIX C..... 122

APPENDIX D..... 127

LIST OF TABLES

Table		Page
1.	Starting Frequencies of the Second and Third Formant Transitions	14
2.	Formant Frequencies for Vowel Stimuli	16
3.	Response Consistency Values (Redundancy) in Absolute Identification	25
4.	Chi-square Values for Comparison of Predicted and Obtained Discrimination Functions for Stop Consonants and Steady-State Vowels	27
5.	Order of Presentation of the Experimental Tapes	35
6.	Informational Redundancy Values for Identification Before and After Training	41
7.	Summary of Non-Speech Discrimination Tests with Extreme Stimuli	42
8.	Chi-square Values for Comparison of Predicted and Obtained Discrimination Functions for Pre- and Post-Training	45
9.	Formant Frequencies for Short and Long Vowel Stimuli	52
10.	Chi-square Values for Comparison of Predicted and Obtained Discrimination Functions for Short Vowels and Long Vowels	58
11.	Analysis of Variance for Recognition Memory P("D"|D) Scores	86

LIST OF FIGURES

Figure		Page
1.	Schematized Spectrographic Patterns Appropriate for the Syllables /bae/, /dae/ and /gae/.....	15
2.	Average Identification and Discrimination Functions for Stop Consonants and Steady-State Vowels.....	20
3.	Identification and Discrimination Functions for Stop Consonants and Steady-State Vowels for S 3.....	21
4.	Identification and Discrimination Functions for Stop Consonants and Steady-State Vowels for S 7.....	22
5.	Identification and Discrimination Functions for Stop Consonants and Steady-State Vowels for S 8.....	23
6.	Schematized Spectrographic Patterns Representing the Extreme Non-Speech Stimuli used in Training Sessions.....	32
7.	Pre-Training Identification and Discrimination Functions for Individual Ss.....	38
8.	Pre-Training and Post-Training Average Identification and Discrimination Functions for Non-Speech Stimuli.....	39
9.	Post-Training Identification and Discrimination Functions for Individual Ss.....	43
10.	Average Identification and Discrimination Functions for Short Vowels and Long Vowels.....	55
11.	Average Obtained Two-Step Discrimination Functions for Short Vowels and Long Vowels with the New Predicted Functions based on Fujisaki's Formula.....	61
12.	Average Identification and Discrimination Functions for Short Vowels and Long Vowels.....	67
13.	Average Identification and Discrimination Functions for Stop Consonants.....	75
14.	Average Identification Functions for Each of the Four Stimulus Conditions with Discrimination Functions Averaged over All Delay Intervals Superimposed on the Corresponding Identification Functions.....	83

Figure | Page

15. Average Probability of A "Different" Response When the Stimuli Within A Pair Were Different (P "D"|D) as A Function of Delay Interval for Each Stimulus Condition for Within and Between Phoneme Boundary Conditions..... 85

16. Average d' Scores for Within- and Between-Category Comparisons as A Function of Delay Interval for Each Stimulus Condition..... 88

LIST OF APPENDICES

Appendix — Page

A. Description of the Haskins Laboratories Computer Controlled Speech Synthesis System and Pulse Code Modulation (PCM) System..... 107

B. Identification and Discrimination Functions for Stop Consonants and Steady-State Vowels for Individual Ss in Experiment 1..... 114

C. Derivation of Predicted Discrimination Formulae..... 122

D. Identification and Discrimination Functions for Stop Consonants for Individual Ss in Experiment 5..... 127

CHAPTER I

INTRODUCTION

Current theories of speech perception emphasize that the perception of speech sounds may involve processes that are in some way basically different from the processes involved in the perception of other sounds (Liberman, Cooper, Shankweiler and Studdert-Kennedy, 1967; Stevens and House In press). A substantial body of experimental work indicates that when listeners are exposed to certain speech sounds their ability to identify and discriminate between them on an auditory basis is limited to a large degree by their linguistic knowledge. Differences in the perception of certain types of speech sounds have led investigators to propose a "special" speech perception mode to characterize the way certain phonetic segments are heard (Liberman, 1970). Other findings have suggested that a "special" perceptual mechanism may exist for processing speech sounds (Studdert-Kennedy and Shankweiler, 1970) and that this mechanism may make reference to an articulatory or motor component during the perceptual process (Studdert-Kennedy, Liberman, Harris and Cooper, 1970). Evidence supporting these hypotheses has come primarily from acoustic analyses of real speech and perceptual experiments with synthetic speech sounds and will be reviewed briefly in the sections below.

Why Is Speech Special?

Acoustic cues in speech perception. Search for the acoustic cues underlying the perception of speech sounds began with the development of two instruments shortly after World War II. The first instrument, the sound spectrograph, was introduced at Bell Telephone Laboratories in 1946 (Koenig, Dunn and Lacy, 1946). This device provided a means for analyzing the spectral characteristics of the speech signal in terms of a frequency by time by intensity display. Initial work indicated that utterances that sounded alike usually appeared in spectrograms to be similar and utterances that sounded different were usually different in spectrograms (Potter, Kopp and Green, 1947).

The second instrument, the Pattern Playback (Cooper, 1950; Cooper, Liberman and Borst, 1951) was a device for reconverting spectrographic

patterns back into sound. This provided a means for manipulating the acoustic signal so that the perceptual consequence of various acoustic cues could be tested experimentally. The use of speech analysis and synthesis techniques at Haskins Laboratories and elsewhere in the early 1950's was aimed primarily at isolating the important acoustic cues that distinguished different classes of speech sounds (Cooper, Delattre, Liberman, Borst and Gerstman, 1952; Ingemann, 1957; Delattre, 1958).

One of the first observations that emerged about speech sounds was the general lack of one-to-one relation between the acoustic signal and perception. Examination of spectrographic patterns of real speech revealed that there were, in general, no segments in the acoustic signal that corresponded uniquely to segments of the perceived message. Experiments with synthesized speech sounds showed that a single segment of the acoustic signal carried information about successive segments; phonemes were not concatenated successively but were represented by merged and overlapping parts of the acoustic signal. Other experiments showed that, in many cases, there was a lack of invariance between the acoustic signal and perceived message (Liberman, Delattre and Cooper, 1952). Different acoustic signals were perceived as identical and identical acoustic signals were perceived as different. These initial findings led investigators to suspect that there was something peculiar about speech stimuli; there was an intricate and complex restructuring of the message in terms of the acoustic signal.

Research on two applied problems reinforced these initial observations. First, attempts to develop automatic speech recognition devices were unsuccessful because of the lack of invariance and segmentation problems. The variability among different talkers and the hierarchical structure of the message prevented any serious attempts at the development of devices which could automatically convert the continuously varying acoustic signal into a discrete sequence of linguistic units (Peterson, 1963). Second, the development of non-speech auditory codes for reading machines revealed that speech was a highly efficient vehicle for the transmission of linguistic information. Listeners could learn to process non-speech signals at only about one-tenth the rate that they

could process speech signals (Studdert-Kennedy and Liberman, 1963; Studdert-Kennedy and Cooper, 1966).

Perception in the speech mode. A series of experiments on the identification and discrimination of speech sounds revealed that certain classes of speech sounds such as stop consonants were perceived in a special mode, unlike other auditory stimuli. Liberman, Harris, Hoffman and Griffith (1957) found that synthetic speech stimuli which varied in acoustically equal steps through the range sufficient to produce the stop consonants /b/, /d/ and /g/ were perceived as members of discrete categories. When listeners were required to discriminate pairs of these stimuli by an ABX procedure, they were able to discriminate stimuli drawn from different phonetic categories but could not discriminate stimuli drawn from the same phonetic category. The obtained discrimination functions were not monotonic with changes in the physical scale but showed marked discontinuities at points along the continuum that were correlated with changes in identification. The phenomenon of "categorical perception" was considered to be quite unusual when compared with the results typically obtained in most psychophysical studies. Stimuli that lie on a single continuum are ordinarily perceived continuously and the discrimination functions are typically monotonic with the physical scale. Listeners can discriminate many more stimuli than they can identify in absolute terms (Pollack, 1952; 1953).[1]

These findings led to a similar experiment with vowel stimuli that varied in acoustically equal steps through the range /I/, /ɛ/ and /ae/. Fry, Abramson, Eimas and Liberman (1962) reported that these stimuli were perceived continuously, much like non-speech stimuli. The discrimination functions did not yield discontinuities along the continuum that were related to changes in identification but were relatively constant across the whole continuum. Moreover, it was observed that

---

[1]This is true for unidimensional stimuli; however, Pollack and Ficks (1954) have shown that information transmission (in absolute judgements) may be increased by employing multidimensional auditory stimuli. In this connection, Lane (1962) has pointed out that with complex multidimensional stimuli such as speech sounds, the perception of a single attribute may not be invariant under changes in other attributes.

vowels were more discriminable than stop consonants indicating that listeners could perceive many more intra-phonemic differences. Similar differences in the perception between consonants and steady-state vowels have been reported more recently by Stevens, Liberman, Studdert-Kennedy and Ohman (1969). The results of these experiments have led investigators to propose two different modes for the perception of speech sounds: a categorical or phonetic mode and a continuous or auditory mode.

The differences between the perception of consonants and vowels may also reflect in part the differences in the nature of the acoustic cues underlying these speech sounds. For the stop consonants there appears to be a relatively complex relation between the phoneme and its representation as sound; the essential acoustic cue is a rapidly changing spectrum (second and third formant transitions) which is relatively short in duration (50 msec.) and transient in nature. On the other hand, the major cue for the synthetic vowels used in perceptual experiments is the steady-state frequencies of the first three formants, which has a relatively long duration (300 msec.) and remains uniform over the entire length of the stimulus.[2] The acoustic cues for consonants are considered to be complex because they show a lack of one-to-one relation and a lack of invariance with the acoustic signal, while the acoustic cues for vowels show a relatively simpler relation with the acoustic signal. The Haskins investigators have used the term "highly encoded" to characterize the former class of sounds and "unencoded" to describe the latter.

Special perceptual mechanism for speech. Evidence for the existence of a specialized perceptual mechanism for speech has been shown in several recent dichotic listening experiments. Overloading the perceptual mechanisms by putting competing signals simultaneously into the two ears indicated that speech stimuli presented to the right ear are more accurately perceived than speech stimuli presented to the left ear (Broadbent and Gregory, 1964; Kimura, 1961a; 1961b; Shankweiler and Studdert-Kennedy, 1967; Studdert-Kennedy and Shankweiler, 1970). The opposite finding has been reported when non-speech stimuli are presented dichotically (Kimura, 1964; Shankweiler, 1966; Darwin, 1969). Based on clinical and

[2]However, see Lindblom and Studdert-Kennedy (1967) and Stevens (1968) for research dealing with vowels in dynamic context.

experimental work, there is some reason to believe that the left hemisphere is the dominant hemisphere for linguistic processing (Darwin, 1969; Bever, 1970; Shankweiler, 1970). Under dichotic presentation the contralateral (right ear to left brain) representation is more efficient for speech stimuli than non-speech stimuli. Of importance is the observation that a right ear advantage occurs for stop consonants which show a complex relation between sound and message, i.e., highly encoded sounds, while no advantage has been reported at either ear for the simple attributes of steady-state sustained vowel sounds, i.e., unencoded sounds (see Darwin, 1971 and Haggard, 1971). Studdert-Kennedy and Shankweiler (1970) have suggested that while auditory processing may be accomplished by the general auditory system common to both hemispheres, phonetic processing is accomplished primarily by specialized mechanisms in the language-dominant left hemisphere. As a result of these findings, it has been suggested that a "special speech decoder" is required to process highly encoded speech sounds such as stop consonants.

Articulation and speech perception. A strong articulatory-motor involvement has always been present in discussion and theoretical accounts of speech perception. There seem to be two major reasons for this. First, the area of acoustic phonetics and most of the research in speech perception was developed on an articulatory model developed by phoneticians. Thus, the acoustic cues for different classes of speech sounds were related to the articulatory dimensions that distinguished these sounds. Spectrographic analysis and perceptual experiments revealed that the sounds of speech were not arrayed along a single complex dimension, but could be specified in terms of a few simple and independent dimensions (Gerstman, 1957; Liberman, 1957). Acoustic dimensions were found in the early experiments with synthetic speech which provided the same distinctions in perception that the articulatory dimensions provided in speech production and suggested that the perceptual and articulatory dimensions of speech may be intimately linked.

The second reason for implicating articulation and production mechanisms in speech perception involves the unique characteristics of speech sounds. Speech sounds are distinct from all other sounds in the

environment because they are produced by a sound source (Fant, 1960; Flanagan, 1965) which has well defined acoustic constraints for a listener (Fourcin, 1968; Gerstman, 1968). Thus, although the acoustic signal as represented in a sound spectrogram may appear arbitrary and incomprehensible to the eye, the distribution of acoustic cues is naturally determined by the characteristics of the human vocal tract (Fant, 1960). More recent work indicates that there appear to be intimate articulatory-perceptual matches between the changes in vocal tract shape and the acoustic signal (Lieberman, 1970). Stevens (1969) has found that in certain parts of the vocal tract, small changes in the position of the tongue constriction resulted in large variations in the formant frequencies; whereas in other regions which corresponded to the "natural" articulatory positions, large variations in the position of the major tongue constriction had only a small effect on the formant frequencies. Stevens contends that the articulatory positions of phonemes, especially consonants, have been chosen at locations where large deviations in articulation will produce only minimal changes in the acoustic signal. Knowledge of the anatomical properties of a given vocal tract (i.e., length and shape) may be used to predict the acoustic output, but it has not generally been the case that the acoustic output alone can be used to predict the characteristics of the vocal tract or the phonetic message. However, Reddy (1967a; 1967b) has recently reported some limited success with procedures that operate directly on the speech wave without any preprocessing hardware (i.e., filtering, vocoding or spectral analysis).

Auditory and Phonetic Perception of Speech

The differences between categorical and continuous perception referred to above represent a fundamental issue in the study of speech perception. The focal point of this issue which has recently been described by Studdert-Kennedy (In press) concerns the distinction between auditory and phonetic processes involved in the perception of speech. According to Studdert-Kennedy we may think of the auditory level as the first stage in the analysis between the acoustic signal and perceived message. At this level the continuously varying acoustic wave form that impinges upon the ear of a listener is transformed into some "time-varying" neurological pattern of events in the auditory system. A spectrogram has

been considered the closest representation of the information in the auditory pattern (Licklider, 1952). At this stage of analysis acoustic information such as spectral structure, fundamental frequency, intensity and duration is extracted by the auditory system. All subsequent stages of analysis beyond the auditory level are based on these auditory features. The phonetic level,[3] the second stage of analysis, is closely related to the first stage. At this stage, segments and features necessary for phonetic classification are derived from the auditory representations of the acoustic signal. At the output of this stage, the continuously varying auditory stimulus has become transformed into a sequence of discrete phoneme-like units.

Categorical perception may be thought of as reflecting a phonetic mode of perception where each acoustic pattern is always, and only, perceived as a token of a particular phonetic type. Listeners can discriminate between two different acoustic patterns if the stimuli have been categorized into different phonetic categories, but they cannot discriminate two different acoustic patterns that have been categorized into the same phonetic category. Information about the acoustic properties of these stimuli may be unavailable for purely auditory judgements as a consequence of phonetic classification, except for gross changes in frequency, intensity, etc. Although phonetic perception can be considered to reflect a more primitive form of stimulus control, it may represent, on the other hand, a highly efficient mode of response for absolute and rapid decisions such as those required in connected speech.

Continuous perception can be thought of as reflecting an auditory mode of perception where discrimination is independent of category assignment. Although listeners assign acoustically different stimuli

---

[3]It is possible to focus on additional levels of analysis between auditory and phonetic. Fourcin (1971) has suggested that four intermediate stages may lie between the acoustic signal and the phonetic level of analysis: 1) acoustic inputs in the auditory canals are mechanically transformed, 2) mechanical inputs to the inner ear are peripherally analyzed, 3) neural outputs are directed to processing appropriate for speech inputs and 4) normalization of neural inputs takes place so that acoustic outputs of features of different physical sources can be evaluated equivalently.

to the same category, they may still discriminate between tokens selected from the same category. An auditory non-phonetic basis for discrimination is available to the listener.

Categorical and Continuous Perception

The degree to which performance is categorical or continuous has primarily been determined by an examination of performance on two kinds of tests: identification and ABX discrimination. In the identification test, stimuli varying in equal steps along a chosen continuum are presented singly, one at a time. The listener is required to label the different stimuli into a set of categories defined by the experimenter. In the ABX discrimination test, three stimuli are presented and the listener is required to determine whether the third stimulus of each triad was identical to the first or second.

A comparison of the identification and discrimination functions is made by application of the predictive test (Liberman et al., 1957; Studdert-Kennedy et al., 1970) which has two assumptions: 1) the stimuli are perceived categorically (i.e., stimuli are identified absolutely and independent of context); and 2) that subjects use the same categories in discrimination as they use in identification. A discrimination function may be predicted from the identification function on the assumption that the probability of correctly discriminating two stimuli is equal to the probability that the two stimuli are identified as different.

The differences between categorical and continuous modes of perception have been reflected in the differences between the perception of stop consonants and sustained steady-state vowels. For stop consonants, discrimination is little better than can be predicted on the basis of absolute identification. Discrimination is quite good between phonetic categories but close to chance within-categories. On the other hand, vowel discrimination is generally good overall indicating that listeners can discriminate many more stimuli than they can identify absolutely. Three explanations have been advanced for the differences between categorical and continuous perception and they will be considered below.

Articulatory reference and categorical perception. The original explanation of the differences for categorical perception of consonants and continuous perception of vowels was offered by the Haskins investigators (Liberman, Cooper, Harris and MacNeilage, 1963; Lisker, Cooper and Liberman, 1962), in terms of the articulatory differences in the production of these sounds. The motor theory of speech perception[4] proposes that reference to articulation is a mediating stage between the incoming acoustic signal and its perception.

Stop consonants are articulated in a discontinuous way by a constriction at a particular place in the vocal tract, while vowels are produced by continuous changes in the overall shape of the vocal tract. According to the motor theory, even though the appropriate acoustic cues for consonants may be described by a physical continuum, they are perceived discontinuously (i.e., categorically) because the articulations underlying the production of these sounds are essentially discontinuous. In contrast, vowels are perceived continuously because of the continuous changes in the articulators from one position to another.

The Haskins investigators were led to the articulatory reference hypothesis as a result of their work in searching for the acoustic cues which underlie the perception of speech. They have found that the relation between perception and articulation is more nearly one-to-one than is the relation between perception and the acoustic signal. Their theory proposed to account for the overriding problem in speech perception: the general lack of invariance between the perception of certain classes of speech sounds (especially consonants) and their corresponding acoustic cues.

Although the original accounts of the motor theory (Liberman, 1957; Liberman et al., 1963; Lisker et al., 1962) proposed that articulatory movements and their sensory effects mediate perception, more recent

---

[4]The reader should not be led to conclude that the only evidence for a motor theory of speech perception rests on the phenomenon of categorical perception. See Liberman et al. (1967), Haggard (1970) and especially Studdert-Kennedy (In press) for discussions of the motor theory of speech perception. However, categorical perception continues to be implicated as evidence for a motor theory of speech perception (Mattingly, Liberman, Syrdal and Halwes, 1971).

versions of the motor theory (Liberman et al., 1967) have placed the reference at the level of the neuromotor commands that activate the articulators. Nevertheless, a strong articulatory involvement has always been part of the Haskins orientation to speech perception.

Discrimination training. A second explanation of categorical perception has been proposed by Lane (1965), who suggests that categorical perception results from a history of reinforcement with respect to labelling behavior. Lane claims that the postulation of a "special" perceptual mechanism for speech is unwarranted since categorical perception is not unique to speech sounds, but may be obtained with non-speech continua as well. Studdert-Kennedy et al. (1970) contend that Lane's account of categorical perception is deficient in several respects. Moreover, they claim that discrimination training alone is not sufficient to produce categorical perception. According to the Haskins view, categorical perception "reflects some structurally determined process adapted to the complex code that links the sounds of speech to the phonetic message they convey" (Studdert-Kennedy et al., 1970, p. 247).

Short-term memory. Recently Fujisaki and Kawashima (1969; 1970) have proposed a model of the perceptual processes involved in speech discrimination that considers the separate contributions of phonetic and auditory short-term memory. These authors report that the continuous-categorical perception of vowels was related to stimulus duration. Vowels of very short duration tend to be perceived more categorically than longer vowels. Their work suggests that categorical perception may be related to the role of short-term memory in speech discrimination.

Although the differences between categorical and continuous perception have stimulated a great deal of controversy, the differences between these two modes of perception are not well understood. The experiments reported in this thesis were undertaken to examine the differences between categorical and continuous perception. Experiment 1 constitutes a replication of the original Haskins experiments on the perception of stop consonants and steady-state vowels. This experiment was conducted to determine whether the results described by the Haskins investigators could be replicated, and if so, what differences might

appear between these findings and those previously reported. Experiment 2 considers the explanation of categorical perception proposed by Lane in terms of discrimination training. Experiment 3 examines the explanation of categorical perception provided by Fujisaki and Kawashima which is based on auditory and phonetic short-term memory in speech discrimination. Experiments 4 and 5 explore several methodological problems in past speech perception experiments. Finally, Experiment 6 considers the role of auditory short-term recognition memory in speech discrimination and its relation to categorical perception.

CHAPTER II

EXPERIMENT I

Identification and Discrimination of Consonants and Vowels

The purpose of this experiment was to examine the identification and discrimination of synthetic stop consonants and steady-state vowels. The major aim of the study was to attempt a replication of earlier Haskins experiments which compared the perception of consonants and vowels (Liberman et al., 1957; Fry et al., 1962; Stevens et al., 1969), as a baseline for additional studies. These earlier studies have shown that the perception of stop consonants differs in interesting and important ways from the perception of steady-state vowels. The perception of stop consonants has been described by these investigators as "nearly categorical" because listeners can discriminate only slightly better than they can identify absolutely. Performance on an absolute identification task can be used to predict discrimination under the extreme assumption that listeners can discriminate two sounds only to the extent that they can identify them as different. The perception of vowels on the other hand has been described as "more nearly continuous" because listeners can discriminate many more stimuli than they can identify absolutely.

The differences among the outcomes of all these experiments have been described primarily in terms of the degree to which the discrimination functions can be successfully predicted from the corresponding absolute identification tests. Since these basic studies have stimulated much discussion about the mechanisms for the perception of speech sounds (Lane, 1965; Studdert-Kennedy et al., 1970), it was felt that a replication was necessary before additional work could be undertaken. Because there appear to be considerable individual differences among listeners in most speech perception experiments, the same group of subjects was used in both the consonant and vowel tests in this experiment. This procedure has not typically been followed in the original experiments on the differences between consonant and vowel perception.

Method

Description of Stimuli

Stop consonants. A set of voiced stop consonant-vowel syllables was synthesized on the Haskins Laboratories parallel resonant synthesizer.

The set of stimuli consisted of 13 three formant syllables which were 300 msec. in duration. The final 220 msec. of each stimulus was a steady-state vowel appropriate for an American English /ae/, with the first three formants fixed at 743, 1620 and 2862 Hz respectively. During the initial 40 msec. a period of closure voicing was simulated on the synthesizer by a low amplitude F_1 at 150 Hz. This period of prevoicing appropriate for the voiced stops was immediately followed by a 40 msec. transitional period during which the first three formants moved toward the steady-state frequencies of the vowel. Throughout the whole stimulus F_1 and F_2 were of equal amplitude; F_3 was approximately -5 dB lower than F_1 and F_2. The F_1 transition started at 150 Hz in all cases and moved along a non-linear contour toward its steady-state frequency. The starting frequencies of the second and third formant transitions were varied systematically and are shown in Table 1 below. Stimulus 1 had second and third formant frequencies beginning at 1232 and 2180 Hz. For successive stimuli in the series, the F_2 starting frequency increased in approximately equal steps from 1232 Hz to 2156 Hz. The F_3 transition began at higher frequencies for stimuli 1 through 7, and then decreased systematically in equal steps for stimuli 7 through 13. The change in <u>both</u> F_2 and F_3 transitions from stimuli 1 to 13 has been shown to be the major acoustic cue for distinguishing place of production between the syllables /bae/, /dae/ and /gae/. The fundamental frequency was set at 120 Hz for the entire duration of the syllable. Schematized spectrographic patterns of three of the stimuli (1, 7 and 13) are shown in Figure 1. The stimuli were recorded on magnetic tape (7 1/2 ips) and then digitized. The wave forms for all 13 stimuli were stored on the Haskins PCM system (see Appendix A) and the experimental tapes were produced from these stimuli.

<u>Steady-state vowels</u>. A set of 13 steady-state vowel stimuli was synthesized on the vocal tract analogue synthesizer at the Research Laboratory of Electronics, Massachusetts Institute of Technology. Table 2 provides the frequencies of the first three formants for these vowels. The fourth and fifth formants were fixed at 3500 Hz and 4500 Hz respectively. The formant frequencies employed here are identical to those provided by Stevens et al. (1969) in their cross language study of vowel perception.

TABLE 1

Starting Frequencies of the Second and Third Formant Transitions of Stop Consonants

Stimulus Number	Starting Frequencies in Hz	
	F_2	F_3
1	1232	2180
2	1312	2348
3	1386	2525
4	1465	2694
5	1541	2862
6	1620	3026
7	1695	3195
8	1772	3026
9	1845	2862
10	1920	2694
11	1996	2525
12	2078	2348
13	2156	2180

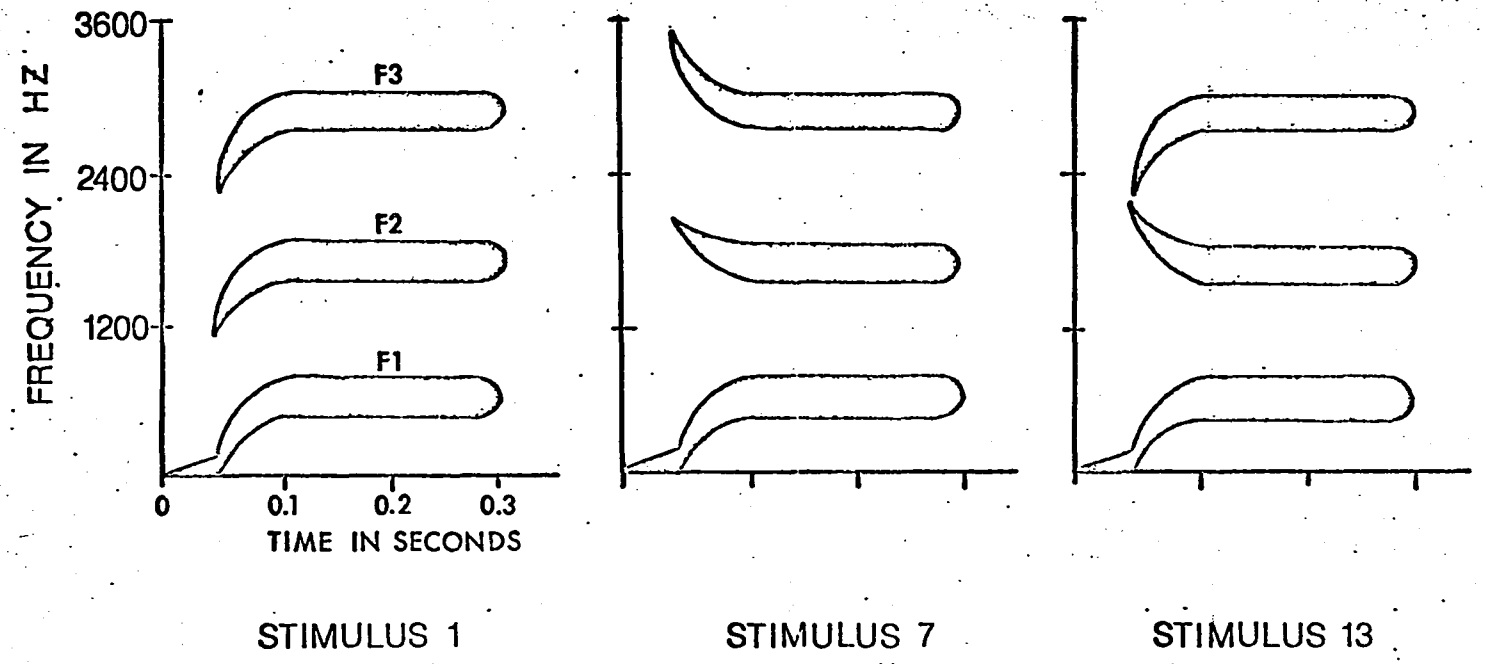


Figure 1. Schematized spectrographic patterns appropriate for the syllables /bae/, /dae/ and /gae/.

TABLE 2

Formant Frequencies for Vowel Stimuli

Stimulus Number	Formant Frequency (Hz)		
	F_1	F_2	F_3
1	270	2300	3019
2	285	2262	2960
3	298	2226	2902
4	315	2180	2836
5	336	2144	2776
6	353	2103	2719
7	374	2070	2666
8	397	2032	2628
9	420	1999	2581
10	444	1966	2556
11	472	1926	2544
12	501	1898	2518
13	530	1858	2492

The thirteen vowels were arranged so that the first three formants varied in approximately equal logarithmic steps through the English vowels /i/, /I/ and /ɛ/.

The bandwidths of the first three formants were fixed at 50 Hz, 80 Hz and 110 Hz. All had a steady-state duration of 300 msec. with a rise and decay time of 50 msec. and had a falling fundamental frequency. F_0 fell linearly from 125 to 80 Hz. These stimuli were originally recorded on magnetic tape at M.I.T. They were then digitized on the Haskins PCM system and the wave forms stored on disk for tape preparation.

Experimental Tapes

All of the experimental tapes were produced on the Haskins PCM system. Magnetic tape recordings were made on an Ampex AG-500 tape deck which was connected directly to the output of the PCM system. A 1000 Hz calibration tone was always placed at the beginning of each tape to insure that the record and playback levels would be uniform throughout the recording and testing sessions.

Identification tapes. Three different 130-item identification test sequences were prepared. Each sequence contained ten different randomizations of an entire series of 13 stimuli. The stimuli were recorded singly with a 5-second interval between presentations and an eight second interval after every ten presentations. Three consonant identification tapes and three vowel identification tapes were prepared from these test sequences.

ABX discrimination tapes. Six different 92-item test sequences were constructed. Members of each AB comparison were made by pairing each stimulus with the next stimulus one or two steps apart along the continuum. There were 12 possible one-step comparisons and 11 two-step comparisons which produced 23 AB pairs. For each AB comparison there are four possible ABX arrangements (ABA, ABB, BAA and BAB) where X is identical to either A or B. Each of the six ABX discrimination tests consisted of a different random arrangement of the 92 triads (i.e., 23 AB comparisons X 4 ABX arrangements). The stimuli within each triad were separated by one second while successive triads were separated by five seconds. After every ten trials there was an additional ten second pause. Six different consonant discrimination test tapes and six different vowel discrimination test tapes were prepared from these original test sequence orders.

Subjects

Ten female undergraduate students at the University of Connecticut served as subjects (Ss) in the present experiment. The Ss were obtained from the Psychology Department subject pool. All Ss were right handed native speakers of English and reported no past history of a hearing disorder or speech impediment. The Ss were run in two groups of approximately equal size. Ss were either paid for their services at a rate of $1.50 per hour or at their request received the equivalent in credit as part of a course requirement. None of the Ss had ever heard any synthetic speech before this experiment.

Procedure

All experimental tapes were reproduced on a General Radio Model 1525 tape recorder and were presented binaurally through Telephonics (TDH-39) matched headphones. The gain of the tape-recorder playback for the 1000 Hz calibration tone was adjusted to give a voltage across the earphones equivalent to 75 dB SPL re 0.0002 dyn/cm^2. Ss were tested in a small experimental classroom which is used for speech perception experiments. The experimenter (E) read a set of instructions aloud to the Ss which explained the nature of the task. In addition, Ss also had a printed copy of these instructions before them and were encouraged to follow along with the E as he read them aloud. Ss were told that this was an experiment concerned with the way people listen to speech sounds.

For the consonant identification tests, Ss were required to identify each syllable as either a /b/, /d/ or /g/ and were told to guess if they were not sure. Similarly, for the vowel identification tests, Ss were required to identify each vowel sound as either a /i/, /I/ or /ɛ/ Since naive Ss usually have difficulty with phonetic symbols, they were instructed to write down the words "beet", "bit" and "bet" on their response sheets and to use the following notation to record their vowel judgements: "EE" as in beet, "I" as in bit and "EH" as in bet.

In the ABX discrimination task Ss were told that the stimuli would be arranged in groups of three and that their task was to decide whether the third stimulus of each triad was identical to either the first or second. Ss were encouraged to guess if they were not sure of a judgement and that they should respond on every trial. All judgements were recorded in a response booklet for later analyses.

Ss were run for an hour a day on five consecutive days. Each of the first four sessions began with either a consonant or a vowel identification test which was then followed by the corresponding ABX discrimination tests. Two of the six discrimination tapes were used each day. On the fifth day all Ss received both consonant and vowel identification and discrimination tapes. When the data from all sessions were combined, each S provided 30 identification responses to each of the thirteen consonant and each of the thirteen vowel stimuli. In addition, each S provided 24 judgements for each of the AB discrimination comparisons for both consonants and vowels.

Results

Absolute Identification Functions for Consonants and Vowels

The average identification functions for consonants and vowels for the ten Ss are shown in Figure 2. Both of the identification functions indicate that Ss partitioned the stimulus continua into three relatively distinct phoneme categories. Inspection of the average functions reveals that the shift from one response category to another appears to be somewhat more abrupt and clean cut for the consonant continuum than for the vowel continuum. These initial findings are in very close agreement with those reported in the previous Haskins experiments on the identification of consonants and vowels (see in particular Stevens et al., 1969).

However, since the phoneme boundaries vary from one subject to the next in their position along the continuum, the average identification functions are, in general, less consistent and perhaps misleading to some degree. For this reason the individual consonant and vowel data for the three best Ss are reproduced in Figures 3, 4 and 5. The criterion for selection was based on the response consistency index described below. These functions show quite clearly that these three listeners are more consistent in identifying the consonants than the vowels. The individual data for all ten Ss are presented in Appendix B.

In most of the original Haskins experiments, it was necessary to select subjects on the basis of how well they categorized the members of a set of synthetic stimuli. Ss who were not consistent on the identification task were eliminated. This was usually done by an informal inspection of the identification functions. However, since the basis for

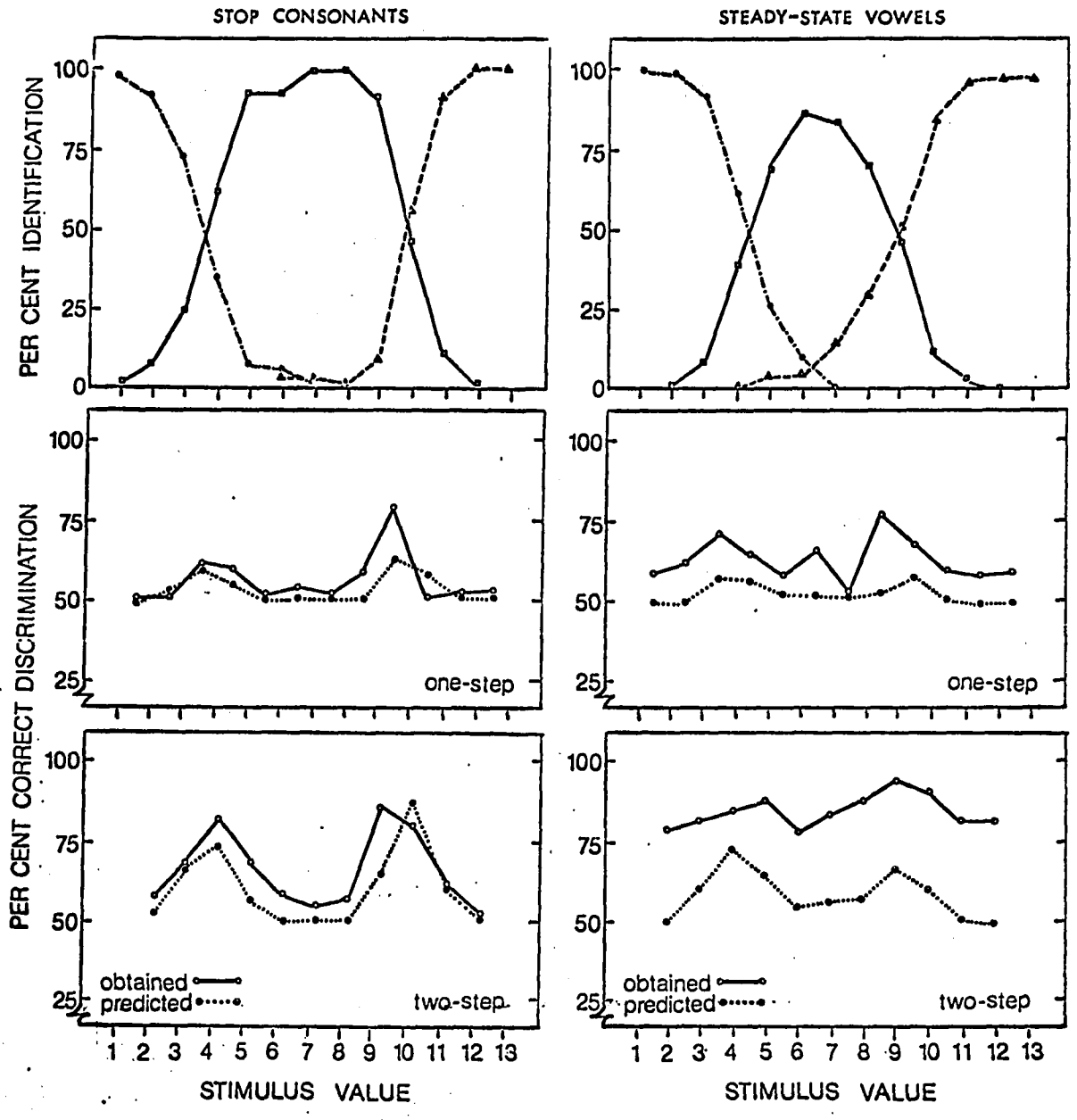


Figure 2. Average identification and discrimination functions for stop consonants and steady-state vowels.

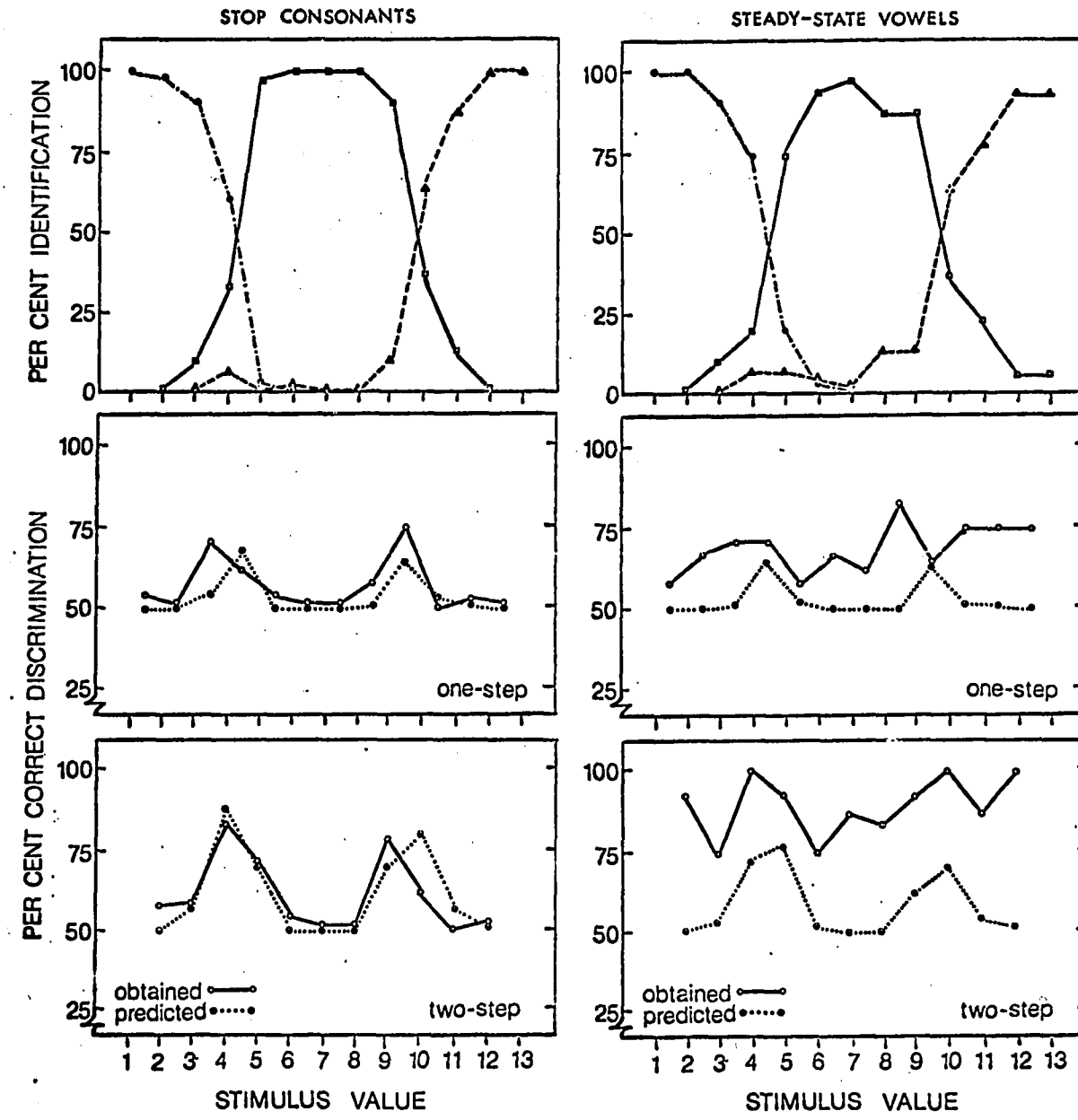


Figure 3. Identification and discrimination functions for stop consonants and steady-state vowels for S 3.

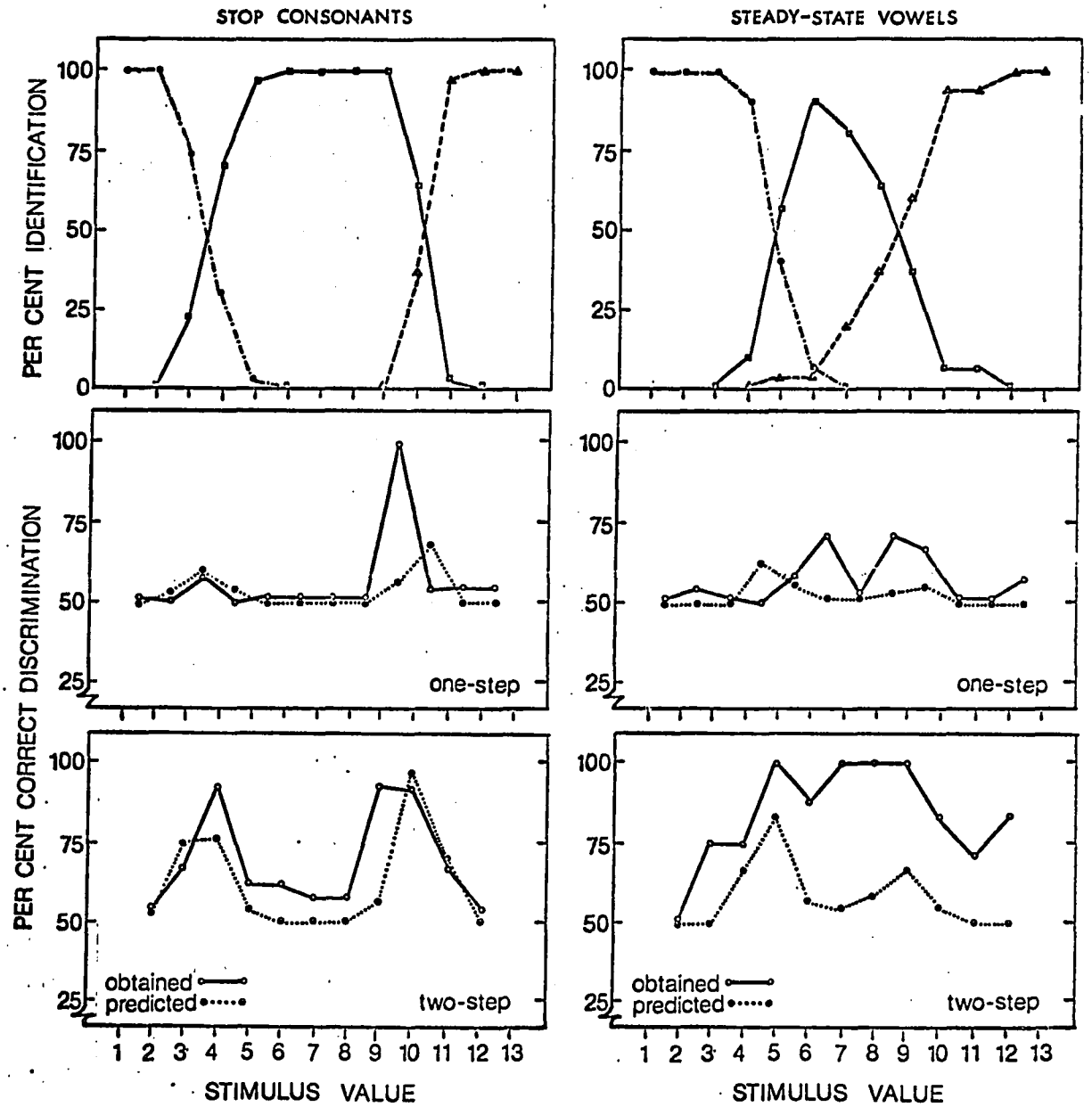


Figure 4. Identification and discrimination functions for stop consonants and steady-state vowels for S 7.

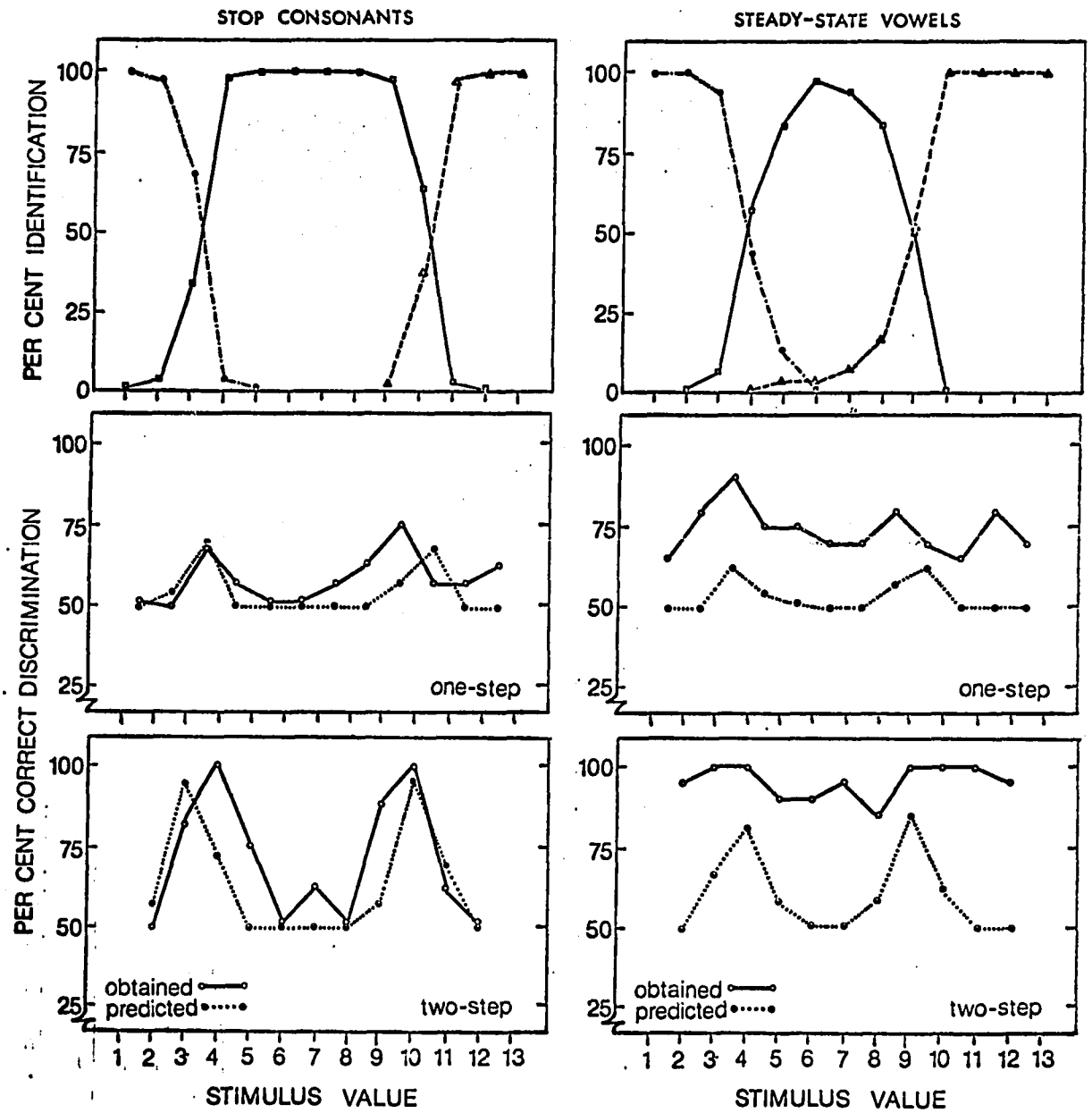


Figure 5. Identification and discrimination functions for stop consonants and steady-state vowels for S 8.

subject selection has never been made explicit, it was felt that some objective measure should be used to describe how subjects categorized the two kinds of stimuli employed in the present experiment.

To obtain a measure of response consistency in absolute identification, the amount of informational uncertainty (H) was calculated for each stimulus number along the continuum. These values were then converted to informational redundancy values so that complete response diversity (i.e., three equiprobable responses) would register as 0 and complete response consistency would equal 1.0. The relative H values were summed across all stimuli for each S. The mean redundancy values are presented separately for each S for the consonant and vowel identification task in Table 3. Response consistency as indexed by the redundancy measure was higher for the consonants than for the vowels in seven of the ten Ss. A Wilcoxon matched-pairs signed-ranks test showed that the slight overall difference in favor of the consonants was not significant ($T = 15.5$, $p > .05$). The relationship between consonant and vowel response consistency was assessed by means of a Spearman rank order correlation. The results of this analysis indicated that there was little relation between performance on the consonant identification test and the vowel identification test ($r_s = + .15$, $p > .05$).

ABX Discrimination Functions

The average one- and two-step ABX discrimination functions for the consonants and vowels are shown below the corresponding identification functions in Figure 2. The solid lines in both figures represent the obtained ABX discrimination functions. The dotted lines show the corresponding average predicted discrimination function which is based on the extreme assumption that stimuli can be discriminated only to the extent that they can be identified as belonging to different categories. The predicted functions were calculated from the identification functions for each S according to the formula provided by Liberman et al. (1957).[5] A description of the formula and its derivation is provided in Appendix C.

[5]There are two different ways to calculate a group predicted discrimination function. The first procedure employs the averaged identification probabilities in the predicted formula while the second method averages the individual predicted scores for all the Ss. The first procedure will provide lower overall predicted scores than the second procedure if there

TABLE 3

Response Consistency Values (Redundancy)
in Absolute Identification

Subject	Stop Consonants		Steady-State Vowels
1*	.5672	<	.7429
2	.7689	>	.7234
3	.8136	>	.6713
4*	.7867	<	.8216
5	.7255	>	.5961
6*	.7575	<	.8236
7	.8433	>	.7234
8	.8690	>	.7901
9	.7879	>	.7701
10	.7246	>	.6756
	Mean= .7644		Mean= .7338

*These three Ss were more consistent in identifying the vowels than the consonants.

An examination of the obtained functions indicates two important differences between the discrimination of vowels and consonants which have been reported in the previous Haskins experiments. First, vowel discrimination is better overall than consonant discrimination. This is especially true for the two-step discrimination functions. Second, the consonant discrimination functions show peaks at points along the continuum which correspond to phoneme boundaries and troughs within phoneme categories, where discrimination is little better than chance. Similar, but less extreme peaks and troughs appear in the vowel data (cf. Stevens et al., 1969).

The differences between the obtained and predicted discrimination functions were always greater for the vowels than for the consonants. In every case the obtained discrimination functions for the vowels were well above the predicted ones. The predicted and obtained functions for the consonants revealed a closer correspondence. The match between identification and discrimination was assessed by comparing the obtained discrimination functions to those predicted from the identification data with a Chi-square goodness of fit test.[6] The results of the separate comparisons for vowels and consonants for each S are presented in Table 4. The Chi-squares for the consonant comparisons show that in general the obtained discrimination functions do not differ significantly from the predicted functions. As expected, the comparisons of the two-step vowel data indicate that the obtained functions differ significantly from the predicted ones. This was true for all of the 10 Ss.

---

is any variability among listeners. The second procedure was followed in all the present experiments whenever group predicted functions are presented. I am grateful to Dr. Pollack for pointing out the discrepancy between these two procedures.

[6]The Pearson χ^2 statistic for goodness of fit which represents an approximation to the Chi-square distribution was employed in the present analyses. The Chi-square values were obtained by summing the individual Chi-squares for each comparison (i.e., predicted vs. obtained) along the continuum. With only two categories, the Pearson χ^2 statistic has the same value as the square of the standardized variable used in testing for a single proportion with the normal approximation to the binomial distribution. The following formula was used in these calculations:

$$\chi^2 = N \frac{(p_o - p_e)^2}{p_e \cdot q_e}$$

TABLE 4

Chi-square Values

Comparison of Predicted and Obtained Discrimination Functions

for Stop Consonants and Steady-State Vowels

	Stop Consonants		Steady-State Vowels	
Subject	One-step	Two-step	One-step	Two-step
1	18.81	40.78*	60.15*	171.80*
2	15.26	15.14	34.07	74.52*
3	5.47	7.00	39.83*	74.52*
4	8.18	18.14	18.44	50.71*
5	13.59	15.68	31.16	82.68*
6	16.71	28.75	24.73	103.50*
7	17.65	21.54	9.26	92.70*
8	9.07	35.72*	58.03*	149.53*
9	5.59	56.53*	18.07	48.95*
10	21.69	14.00	17.06	38.38*

*$p < .001$

df = 12 (one-step)

df = 11 (two step)

Summary and Discussion

The following three general observations concerning the differences in perception between stop consonants and steady-state vowels have been replicated in the present experiment using a single group of listeners.

1. The absolute identification functions for stop consonants tend to be sharper, and for most Ss, more consistent than for the corresponding functions for steady-state vowels. However, the differences in consistency as determined by an information measure were not statistically significant across the subject group.

2. The overall level of discrimination is higher for steady-state vowels than for stop consonants.

3. The perception of stop consonants is "nearly categorical," in the sense that listeners tend to discriminate pairs of stimuli only to the extent that they identify them as different. The perception of steady-state vowels is "more nearly continuous," in the sense that listeners discriminate many more intraphonemic differences than they identify absolutely. The results of the absolute identification tests predicted the obtained consonant discrimination functions more accurately than they predicted the obtained vowel discrimination functions.

The overall results of this experiment are essentially the same as those previously reported by the Haskins investigators. The results found with the vowels are remarkably similar to the findings reported by Stevens et al. (1969) using the same formant frequency values. It can be concluded that there are differences in the perception of vowels and consonants and that the results reported by the Haskins investigators have been verified.

The next experiment considers an alternative account of categorical perception in terms of discrimination training and the principles of auditory psychophysics.

The rationale for using this procedure is described in Hays (1963), pp. 578-614 and McNemar (1969), pp. 245-275. I am grateful to Professor J. E. Keith Smith for his assistance with these analyses, and to Professor Edwin Martin for pointing out several problems connected with the use of the Chi-square statistic.

EXPERIMENT II

Effect of Discrimination Training on Categorical Perception

Lane (1965) has published a critical review of the motor theory of speech perception and devotes a large part of the review to important and insightful criticisms of the Haskins experiments that were concerned with the differences between the perception of speech and non-speech sounds (Liberman, Harris, Eimas, Lisker and Bastian, 1961a; Liberman, Harris, Kinney and Lane, 1961b). In recent papers, Lane (Lane, 1965; 1967) has sought to demonstrate similar categorical perception effects with non-speech stimuli. Using a variety of visual and auditory stimuli, he tried to show that if two stimuli sampled from the same continuum evoke different identification responses, they will be more readily discriminated than two stimuli (separated by equal physical units) that evoke the same identification responses. Lane maintains that the observed relations between identification and discrimination (i.e., categorical perception) which have been used as evidence for a "special" speech perception mode are not peculiar to the speech sounds, but rather are the result of general procedures for discrimination training and testing with discrete response repertoires.

Support for this alternative explanation of categorical perception comes from a series of experiments reviewed in Lane (1965). The primary experiment cited by Lane was an otherwise unpublished study by Lane and Schneider (1963). In this experiment, subjects were trained to identify the extreme members of a non-speech auditory continuum, the inverted spectrograms of the /do/-/to/ control continuum (Liberman et al., 1961b) by a disjunctive conditioning procedure (see Cross and Lane, 1962). Lane and Schneider trained Ss to identify the extreme members of this non-speech continuum as /do/ or /to/. Following this training, the responses generalized to the intermediate stimuli such that the stimulus continuum was partitioned into two equivalent response classes. Before initiating the training procedure with the extreme stimuli, the ABX discrimination functions for the non-speech continuum showed chance performance. However, after training, the discrimination functions showed peaks at the boundary between the two response categories.

If Lane's contention that categorical perception can be obtained "quickly and easily in the laboratory after a few minutes of conditioning" (Lane, 1967, p. 86) is correct, there would be little justification for considering the perception of speech as special, involving perceptual operations and mechanisms basically different from those used in the processing of other auditory stimuli.

Lane's explanation of categorical perception was recently attacked by the Haskins group (Studdert-Kennedy et al., 1970). Studdert-Kennedy et al. focused on a comprehensive examination of the Lane and Schneider experiment in which Lane purported to show categorical perception with non-speech sounds. After reconstructing the data for the Ss which Lane presented, they questioned the effectiveness of his discrimination training procedures in producing categorical perception with non-speech stimuli. Their major criticism of the Lane and Schneider experiment dealt with the ABX discrimination data for the 3 Ss presented by Lane. Studdert-Kennedy et al. contended that two of the conditions for categorical perception were not reliably established in this experiment. First, the discrimination functions were in every case less consistent than the data shown by Lane for the one "representative" S. Second, the discrimination functions after training could not be accurately predicted from the corresponding identification functions for any of the three Ss.

The Haskins group undertook to replicate the Lane and Schneider experiment. Their attempt to produce categorical perception by discrimination training was summarized in one paragraph in their reply to Lane. Their results indicated that they were no more successful than he in establishing categorical perception by discrimination training procedures. Of the five Ss in their replication study, only one achieved better than 75% accuracy in identifying the two extreme stimuli. The effect of discrimination training for this one S was to raise his overall discrimination performance rather than producing the expected selective discrimination function (cf. Parks, Wall and Bastian, 1969).

The conflicting results of the Lane and Studdert-Kennedy experiments with non-speech sounds indicated a real need for an additional attempt at establishing categorical perception by discrimination training. The

present experiment was designed to replicate the original Lane and Schneider training procedures with a stimulus variable shown to be perceived categorically in speech and continuously in non-speech (Mattingly, Liberman, Syrdal and Halwes, 1971). The experimental variable under consideration was the direction and extent of an isolated second formant transition.

The major aim of this experiment was to determine what effects training procedures have on the absolute identification and discrimination of non-speech auditory stimuli.

Method

Materials

Stimuli. A set of seven single formant patterns were designed as non-speech control stimuli for use in discrimination training. These patterns consisted of a steady-state second formant at 1386 Hz and seven initial second formant transitions. Figure 6 shows schematized spectrograms of stimuli 1 and 7. The F_2 transition was approximately 35 msec. in duration while the steady-state portion was 165 msec. The starting frequencies of the transitions were varied systematically from 921 Hz to 1845 Hz in approximately equal steps and moved along a non-linear contour toward the steady-state second formant frequency. The transitions for stimuli 1, 2 and 3 had positive slopes, beginning below the steady-state frequency and rising while stimuli 5, 6 and 7 had negative slopes, beginning above the steady-state frequency and falling. Stimulus 4 had no transition since it began at the same frequency as the second steady-state formant. The fundamental frequency was held steady at 114 Hz for the first 100 msec. and then dropped linearly to 70 Hz over the remainder of the stimulus. The seven stimuli were produced on the Haskins parallel-resonance synthesizer and were recorded on magnetic tape. The stimuli were then digitized and the wave forms stored on the Haskins PCM system. All experimental tapes were prepared from these digitized wave forms. A 100 msec. 1000 Hz code pulse was also digitized and stored on the PCM system. The code pulses were used to trigger the relays and counting equipment used in the training sessions.

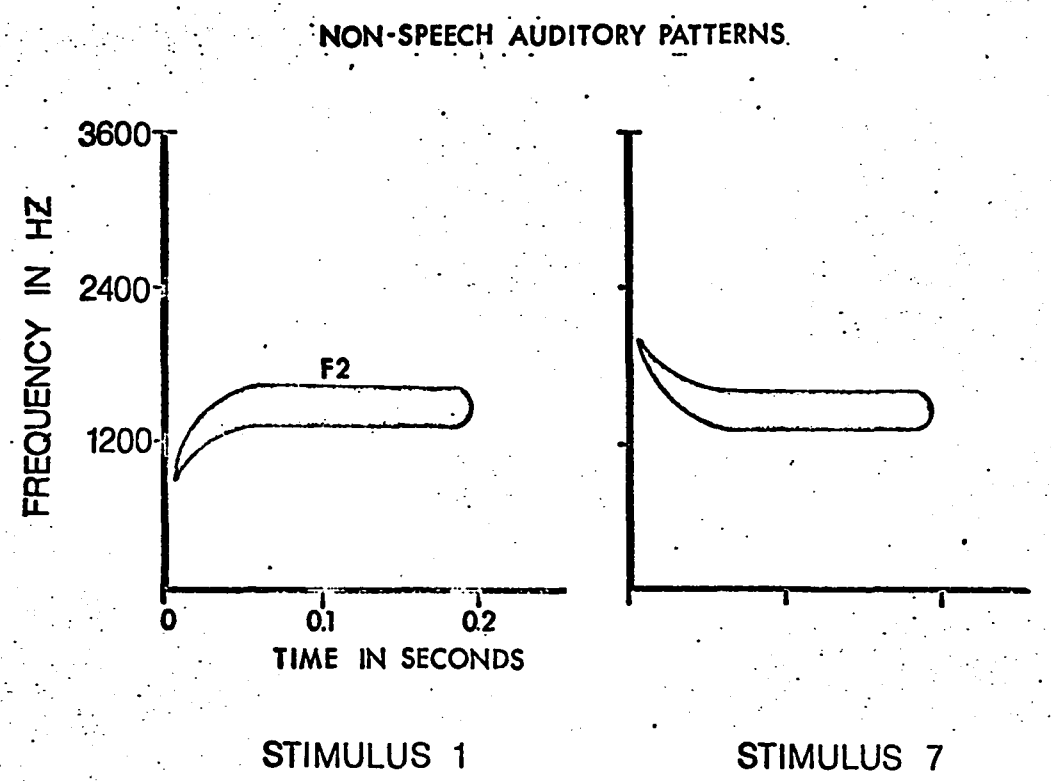


Figure 6. Schematized spectrographic patterns representing the extreme non-speech stimuli used in training sessions.

Experimental Tapes

Four different types of experimental tapes were prepared on the Haskins PCM system. The tapes were recorded on a four track Tandberg stereo tape recorder.

Initial shaping tape. Following the same protocol employed by Lane and Schneider (1963), a shaping tape was prepared with the two extreme stimuli (1 and 7) from the continuum. This tape consisted of 160 trials arranged in the following order. For the first sixty trials, the extreme stimuli were presented in an alternating pattern: ten presentations of stimulus 1 followed by ten presentations of stimulus 7. For the next sixty trials the extreme members were arranged in blocks of five. And for the last twenty trials the stimuli alternated in blocks of two. The trials were separated by 4 seconds. There was an 8-second pause after every block of ten trials. The 1000 Hz code pulses were placed 500 msec. before each stimulus on the second channel of the tape.

Discrimination training tape. A 240-trial discrimination training tape was prepared according to the same format specified by Lane and Schneider. The stimuli were arranged so that each of the two extreme stimuli occurred equally often within each half of the tape. Code pulses were also placed 500 msec. before each stimulus on the second channel of the tape.

Identification tapes. Three different identification tapes were prepared with all seven stimuli from the series appearing in an irregular order. Each tape contained 70 trials with the restriction that each stimulus occurred equally often within each half of the tape. Trials were separated by 4 seconds, with an 8-second pause after every ten trials.

ABX discrimination tapes. Three different ABX discrimination tapes were prepared with all possible one- and two-step comparisons of the seven stimuli along the continuum appearing twice within each tape which contained 88 ABX triads. There are 6 one-step comparisons and 5 two-step comparisons which were arranged in 4 different ways (ABA, ABB, BAB and BAA) for a total of 44 trials. Each tape consisted of two different randomizations of these 44 trials. There was a one second interval

between members of a triad; successive trials were separated by 4 seconds. There was an 8-second pause after every ten trials.

Subjects

Six male undergraduate students at Polytechnic Institute of Brooklyn served as Ss in the present experiment. The Ss were obtained from an introductory psychology course. All Ss were right handed native speakers of English and reported no past history of a hearing disorder or speech impediment. The Ss were paid at a base rate of $1.85 per hour. The opportunity to earn additional money was determined by their performance during the training sessions. None of the Ss had ever heard any synthetic speech before this experiment.

Procedure

All Ss were tested individually in a small experimental room in the Experimental Psychology Laboratory of Polytechnic Institute. Each S served for an hour a day on five consecutive days. The S was seated comfortably in front of a counter, signal lights and a Tandberg Model 12-41 tape recorder. The stimulus tapes were presented binaurally through Sharp HA-10 headphones at a comfortable listening level, approximately 70 dB SPL re 0.0002 dyn/cm^2.

An electronic switch (Grason-Stadler Model 829E) was interposed between the tape recorder output and the headphones. Code pulses which were recorded on the second channel of the training tapes triggered an interval timer. After 450 msec. the electronic switch closed and the stimulus reached the headphones. These code pulses also triggered a sensitive relay (Alton Electronics) that activated a system of electromechanical relays, timers and counters which was used in the training sessions. All the control apparatus was located in a room adjoining the experimental room.

The procedure employed in this experiment involved an examination of the identification and discrimination functions for the non-speech stimuli both before and after discrimination training. Table 5 shows the order of presentation of the different experimental tapes over the five days.

Pre-training tests. On the first day each S was given the ABX discrimination tapes in order to establish a baseline of discrimination

TABLE 5

Order of Presentation of the Experimental Tapes

Day	Type of Session	Tape Description	Number of Trials
1	Pre-training	ABX Discrimination Tapes	264
2	Pre-training	Identification Tape	70
	Training	Initial Shaping Tape	160
	Training	Discrimination Training Tape	240
3		Identification Tapes	140
	Training	Discrimination Training Tape	240
4		Identification Tapes	140
	Training	Discrimination Training Tape	240
5	Post-training	Identification Tapes	210
	Post-training	ABX Discrimination Tapes	264

before training commenced. Ss were told that this was an experiment dealing with the discrimination of complex auditory patterns and that on each trial they would hear a group of three sounds. They were instructed to determine whether the third sound was most like the first sound or most like the second sound. Judgements were recorded in a response booklet for later analyses. After completion of the three ABX discrimination tests, the Ss filled in a brief questionnaire which concerned their impressions of the sounds they had just heard.

On the second day Ss were told that the sounds they had heard previously were really the speech sounds /bae/ and /dae/ which had been modified by a computer to make them appear somewhat different from their original form. Ss were then presented with the identification tapes. They were told that the sounds would now be presented individually and that their task was to identify each sound as a /b/ or a /d/.

Training. After completing the pre-training identification tests the S was introduced to the training procedures with the following set of instructions:

> "In this part of the experiment we are going to train you to identify the original speech sounds. In front of you are two buttons. The one on the left is marked with the syllable /bae/ and the one on the right is marked with the syllable /dae/. You can earn money by simply responding with a /bae/ or a /dae/ at appropriate times. We can't tell you now when or how these responses should be used. That is obviously for you to learn on your own. All you have to do is wear this headset and watch the display unit in front of you. You will hear two sounds. One of them sounds like a /bae/ and the other sounds like a /dae/. Each time you respond appropriately the green light above each response button will flash and 1 point will be added to your score on the counter in front of you. If you do not respond correctly you will not receive a point. Everytime you hear a sound say /bae/ or say /dae/ and then immediately press the appropriate response button. Do not respond until after the end of each sound. The beginning of a new trial will be signaled with a red warning light in the middle of the panel. A half second later the sound will come through your earphones. Say the syllable out loud and then press the appropriate button. Watch to see how you did and then get ready for the next warning signal. You will obviously want to earn as high a score on the counter as possible because the amount we will pay you at the completion of the experiment will be determined by your final score on

> the counter. In other words, the amount of money you can earn will depend on the number of correct responses you make during this session. Be sure to respond to every sound presented even if you have to guess."

The S was then presented with the initial 160-trial shaping tape. After a short break he was given the 240-trial discrimination training tape. A continuous reinforcement schedule was used during this first training phase. Ss received a penny for each correct response.

The procedure for days 3 and 4 was almost identical except that the reinforcement schedule was modified. On day 3, the S received 5 cents for every ten correct responses in succession. Finally, on day 4, the S received 10 cents for every twenty correct responses in succession. Each session began with two identification tapes followed by the 240-trial discrimination training tape.

Post-training tests. In order to assess the effects of the training procedures, on the fifth and final day, the S was given three identification tapes followed immediately by three ABX discrimination tapes. The same tapes that were used for the pre-training tests were presented again for the final tests, although they were arranged in a different random order.

Results

Pre-Training Tests

Identification. The pre-training identification functions for each of the six Ss are presented in the upper portion of Figure 7. Each function shows the percentage of /bae/ and /dae/ responses for each of the seven non-speech stimuli. The average data for the six Ss are shown in Figure 8. The functions are exact complements of one another because of the forced-choice identification task. Examination of these figures reveals a fairly wide variability across Ss. There is some tendency for Ss to partition the stimulus continuum into two categories as the direction (i.e., rising or falling) and extent of the F_2 transition changed systematically from stimulus 1 to stimulus 7. This is quite evident in the data for Ss MB and LK. The identification functions for the remaining four Ss are unstable indicating that these stimuli cannot be reliably identified. An application of the informational uncertainty analysis described in Experiment 1 to the identification data

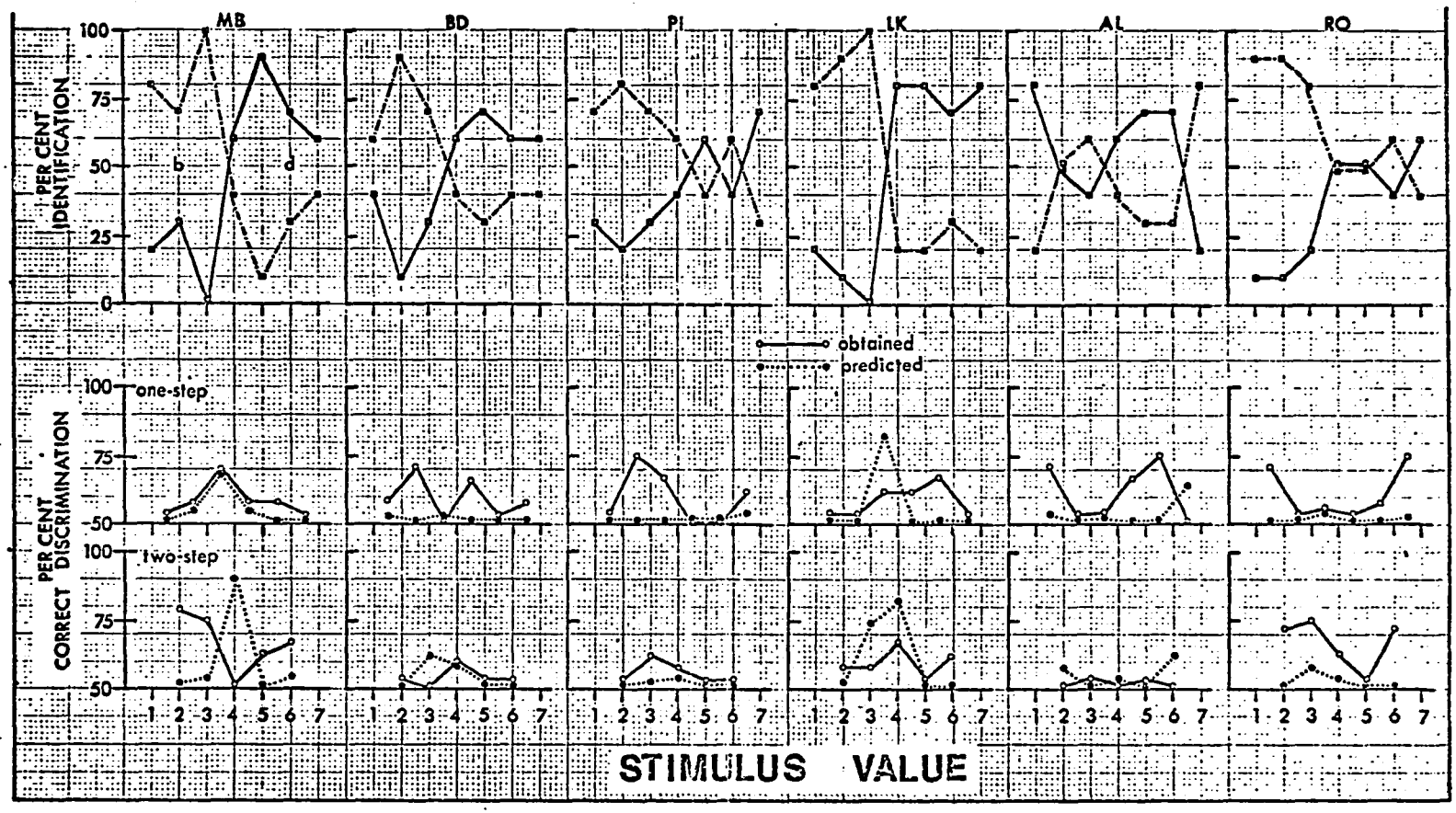


Figure 7. Pre-training identification and discrimination functions for individual Ss.

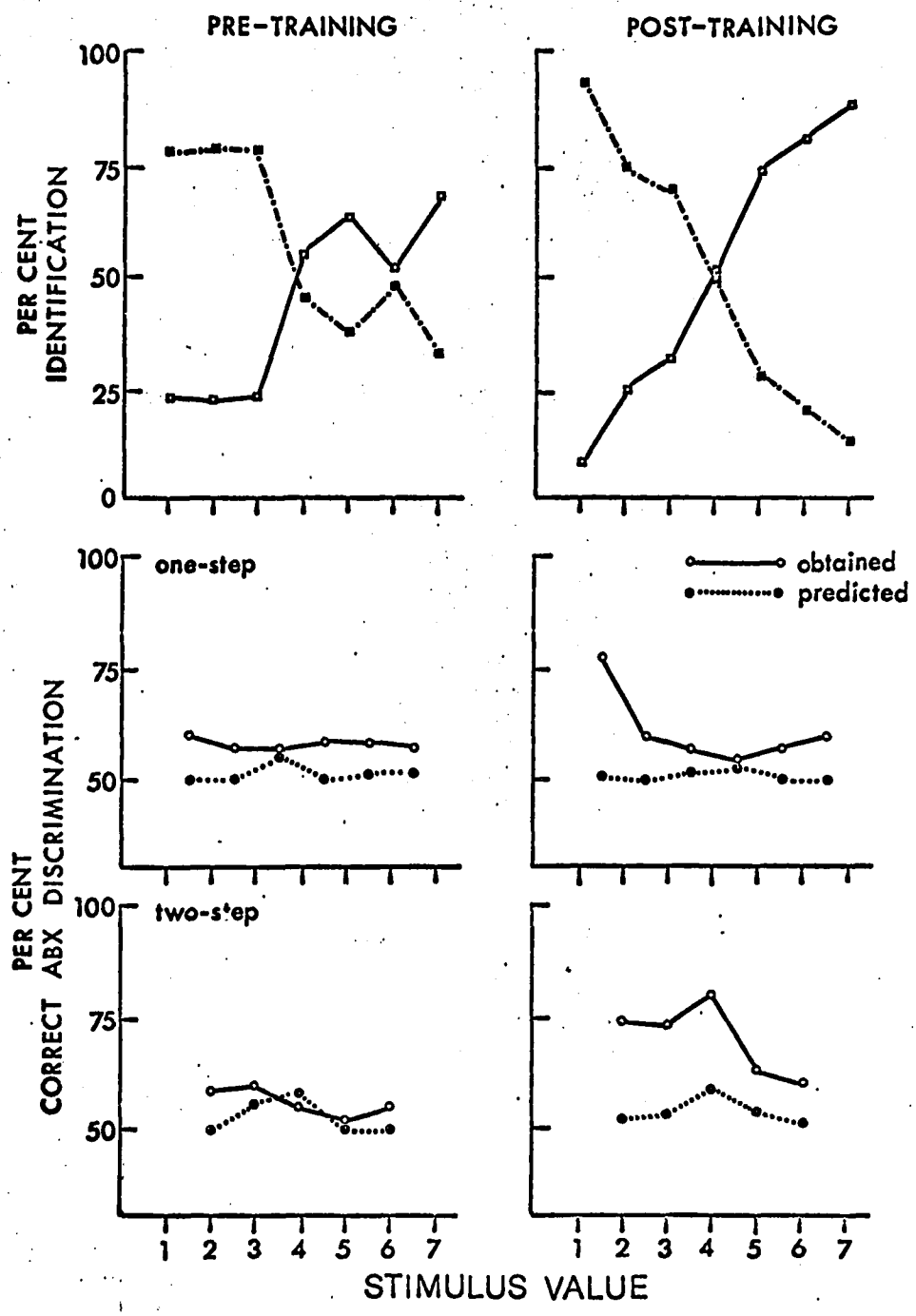


Figure 8. Pre-training and post-training average identification and discrimination functions for non-speech stimuli.

confirms these initial observations. Table 6 shows the mean redundancy values for each S. It can be seen that Ss MB and LK have substantially higher redundancy scores than the other four Ss.

Discrimination. The one- and two-step predicted and obtained ABX discrimination functions are shown in Figure 7 below the corresponding identification functions for each S. Figure 8 shows the group data. The solid lines represent the obtained values while the dotted lines provide the predicted values. The predicted functions were calculated from the identification data according to the procedure followed in Experiment 1.

The discrimination functions for most of the Ss show performance only slightly above chance. Ss MB and LK may show a slight increase in their obtained function at the boundary between /bae/ and /dae/ (MB: one-step; LK: two-step), as might be predicted from their identification data.

On the basis of the pre-training identification and discrimination tests, therefore, two relatively distinct groups of Ss have emerged from the initial six Ss. No arbitrary selection criterion was imposed and all six Ss were given the same amount of discrimination training during the second part of the experiment.

Discrimination Training

Table 7 presents a summary of the discrimination training sessions with a breakdown of the percent correct responses for each S over three days. The accuracy in identifying the extreme stimuli during the training session was quite good, about 84% overall, although it remained relatively stable for each S over the three sessions. The differences between training sessions for the six Ss were not significant by a Friedman two-way analysis of variance on the total percent correct responses ($X_r^2 = 3.62$, $p > .05$). Apparently some Ss did not benefit from increased discrimination training with the extreme members of the stimulus set. In addition, Ss MB and LK began training at high levels of accuracy and a ceiling might have prevented further gains.

Post-Training Tests

Identification. The post-training identification functions for each of the Ss are shown in Figure 9. The average data for the six Ss

TABLE 6

Informational Redundancy Values for Identification Before and After Training

Subject	Pre-training	Post-training
1 (MB)	.2992	.4425
2 (BD)	.1264	.4188
3 (PI)	.1031	.1383
4 (LK)	.3946	.5998
5 (AL)	.1217	.5065
6 (RO)	.1998	.2993

TABLE 7

Summary of Non-Speech Discrimination Tests with Extreme Stimuli

Subject	Percent Correct Responses											
	Day 2			Day 3			Day 4			Days 2-4		
	b	d	$\bar{X}$	b	d	$\bar{X}$	b	d	$\bar{X}$	b	d	$\bar{X}$
1 (MB)	94	84	89	96	84	90	100	86	93	96.6	84.6	90.6
2 (BD)	72	82	77	74	68	71	82	70	76	76.0	69.2	72.6
3 (PI)	76	76	76	80	70	75	86	80	83	80.6	75.4	78.0
4 (LK)	98	94	96	98	98	98	100	96	98	98.6	96.0	97.3
5 (AL)	64	68	66	96	92	94	94	92	93	84.6	84.0	84.3
6 (RO)	90	64	77	86	74	80	92	82	87	89.2	73.4	81.3
Means	83.0	77.0	80.0	88.2	81.8	85.0	92.0	84.0	88.0	87.6	80.8	84.3

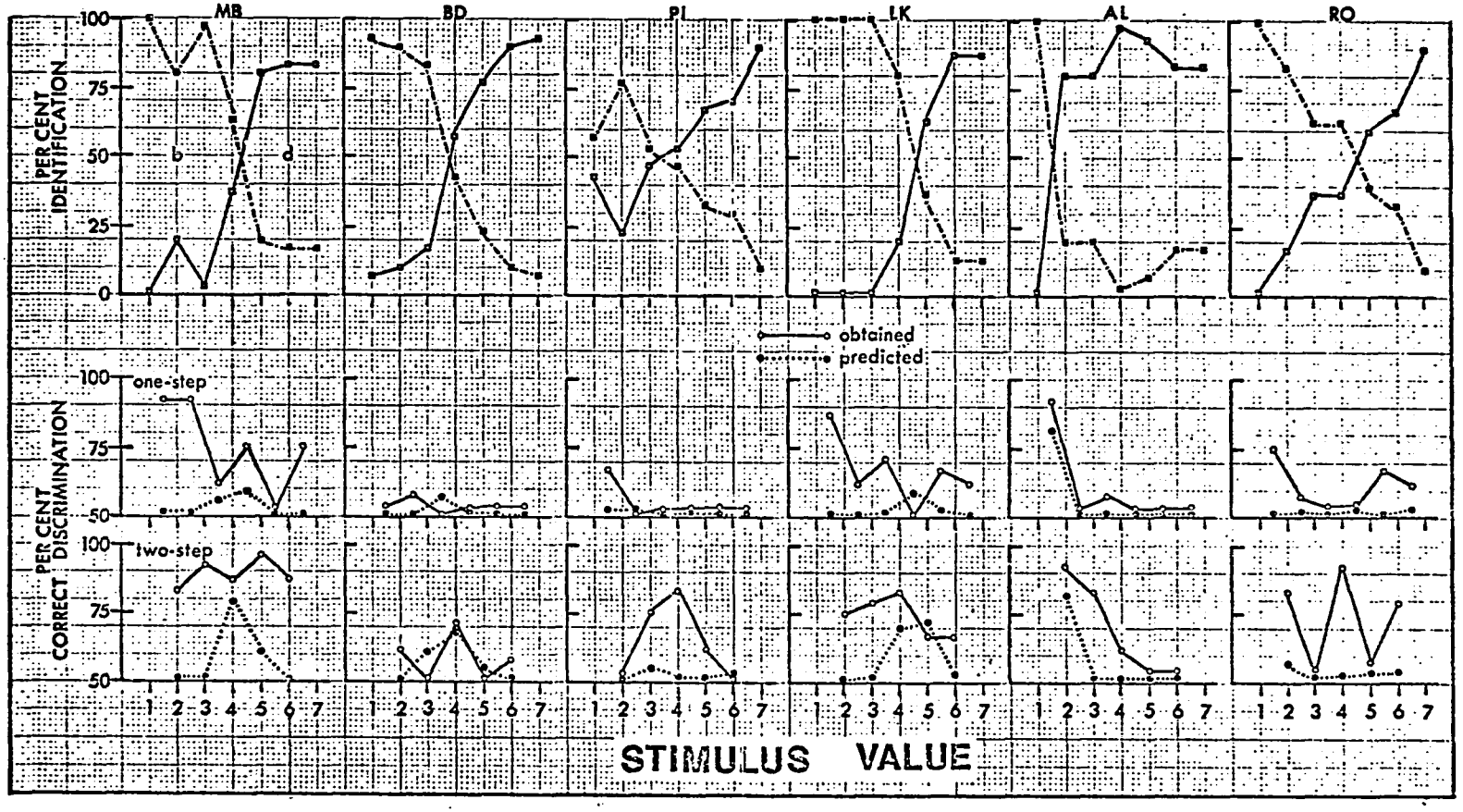


Figure 9. Post-training identification and discrimination functions for individual Ss.

are presented in Figure 8. A comparison of Figure 7 with Figure 9 shows the effects of the discrimination training procedure on absolute identification. The mean redundancy values for these identification functions are presented in Table 6 next to the corresponding pre-training data. Examination of Table 6 indicates that the response consistency as indexed by the redundancy measure increased for every S after training. The differences between the pre- and post-test redundancy values in absolute identification were significant by a Wilcoxon matched-pairs signed-ranks test ($T = 0$, $p < .05$).

Discrimination. The post-training discrimination functions are presented below the identification functions in Figure 9. The average one- and two-step functions for the group are shown in Figure 8. In order to assess the effects of the training procedure on discrimination performance, a separate three-factor analysis of variance was carried out with the one-step and with the two-step discrimination data. The three variables under consideration were training (pre-test, post-test), discrimination (predicted, obtained) and stimulus comparison.

The results of the analysis of variance on the one-step data revealed that only the main effect for discrimination was significant when evaluated against the corresponding mean square term including Ss, $F_{(1,5)} = 18.73$, $p < .01$. The obtained one-step scores were greater than the predicted scores.

On the other hand, the analysis of variance for the two-step data showed that both the training procedure and stimulus comparison main effects were significant beyond the .025 level, $F_{(1,5)} = 30.26$ and $F_{(4,20)} = 4.33$ respectively. Interestingly, the main effect of discrimination was not significant in the two-step analysis, although the training by discrimination two-way interaction was significant at the .025 level, $F_{(1,5)} = 12.32$. The remaining two-way interactions and the three-way interaction were not significant.

The match between predicted and obtained discrimination scores was assessed with the Chi-square goodness of fit test. Table 8 shows the one- and two-step Chi-square values for each S before and after training. It can be seen that there was only one significant Chi-square before training while four significant Chi-squares appeared after training. In addition, for the two step functions, there is an increase in the Chi-square values for every S except MB, whose Chi-squares were significant both before and after training.

TABLE 8

Chi-square Values

Comparison of Predicted and Obtained Discrimination Functions for Pre- and Post-training

	Pre-training		Post-training	
Subject	One-step	Two-step	One-step	Two-step
1 (MB)	.82	84.05*	40.67*	53.13*
2 (BD)	7.32	1.92	3.88	5.08
3 (PI)	9.93	2.80	2.45	18.70
4 (LK)	10.99	9.98	24.15*	17.14
5 (AL)	14.38	5.26	2.87	14.13
6 (RO)	10.71	12.09	9.61	27.52*

*$p < .001$

df = 6 (one-step)

df = 5 (two-step)

Discussion

The results of the present experiment indicate that the Lane and Schneider training procedure improves identification as measured by the response consistency index and enhances overall discrimination, at least for the two-step obtained discrimination functions. The improvement for the obtained discrimination functions is relatively greater than the improvement for identification in terms of the equivalent predicted discrimination functions, which were derived from the identification functions. However, there was a lack of any consistent effect of the training procedure across all Ss and large individual differences among Ss. The results suggest that the previous findings reported may be due as much to subject selection as to the experimental procedures employed.

Since the one-step data did not reveal any differences due to the training procedures, the following discussion will be concerned primarily with the two-step discrimination data. It is interesting to note here that both Lane and Studdert-Kennedy were also unsuccessful in obtaining any systematic differences in the one-step obtained discrimination functions (see in particular Lane, 1963; 1965). The one-step comparisons probably do not represent large enough physical differences to be affected by these training procedures.

Three major findings from the present experiment deserve special attention. First, although the main effect of stimulus comparison was significant, its lack of significant interaction with either training or discrimination indicates that the shape of the function was not different across the predicted and obtained functions or across pre- and post-training sessions. From the latter result, it may be concluded that the training procedure had little effect on the shape of the discrimination function.

Second, the significant main effect of training plus its interaction with discrimination indicates that the training procedure affected only the obtained discrimination function and not the predicted discrimination function. The effect of the training procedure was to raise the overall level of the obtained discrimination above that predicted from the identification data. The fact that training did not significantly affect the analysis of variance of the predicted

functions (which were in fact derived from the identification functions) is interesting in view of the significant Wilcoxon test on the information measures between pre- and post-training identification scores. Although the differences for the information measures between pre- and post-training were significant, the overall level of response consistency was substantially lower with the non-speech stimuli than with the speech stimuli used in Experiment 1. The discrepancy is probably due to the relative insensitivity of the prediction formula to small changes in identification response probabilities. For example, in order to obtain a 10% change in the predicted function near chance performance, a difference of 45% is needed between the identification probabilities of the two stimuli to be discriminated.

Third, although there is no consistent overall sharpening of discrimination at category boundaries, such sharpening is shown in isolated cases. It is possible that by inspection the selection of individuals as "representative" can be made to support either the Lane position or the Haskins position with respect to the effects of training procedures on categorical perception. For example, if we arbitrarily used the subject selection criterion employed by Lane which was "30 correct identifications in succession" within the first 400 training trials, we would have only two Ss remaining from the original six (i.e., Ss MB and LK). It will be recalled that these were the two Ss who partitioned the stimulus continuum into two categories even before training had commenced (see Figure 7). Similarly, we could have chosen Ss BD and AL, since their post-training predicted and obtained discrimination functions match closely (see Figure 9) in the expected direction. Finally, to support the Haskins position we could have selected Ss PI, RO or MB who all showed larger gains in obtained two-step discrimination than in the predicted functions.

However, the introduction of the Chi-square technique permits a more quantitative assessment of the outcome of the training procedures. If the training procedure suggested by Lane produced categorical perception we would expect the Chi-square values between predicted and obtained functions to decrease after training. However, as shown in Table 8 the Chi-squares on the two-step data increase for all but one

S indicating that the effect of training is to increase the level of discrimination above the predicted functions.

Lane's contention that categorical perception can be obtained "quickly and easily in the laboratory after a few minutes of conditioning" (Lane, 1967, p. 86) must be qualified in light of the present findings. Even though each S received 880 training trials with the Lane procedure over a three-day period, the effect was not consistent, as we have seen, across all Ss. Although learning undoubtedly plays some role in speech perception, it is clear from the present experiment that this learning may be very different from the kind of learning that results from the training procedures propounded by Lane.

The results of this experiment are not conclusive with respect to resolving the controversy over the effects of discrimination training on categorical perception. However, little evidence has been found to support the Lane position. The outcome of this experiment does not permit us to exclude the possibility that fuller and more adequate training procedures may improve the fit between predicted and obtained discrimination functions.

EXPERIMENT III

Effect of Stimulus Duration on Vowel Perception

Recent developments in information processing have suggested that different levels of cognitive activity are not autonomous but rather can be thought of as representing a continuum of successive and interdependent stages of analysis between stimulus and response. Examination of the relations between stimulus and response characteristics at various stages after stimulation may provide an understanding of the flow and content of information during analysis. Since the observer typically reports his judgement after the stimulus has ended, some type of perceptual memory is required in most experiments dealing with detection, recognition or identification. Successive discrimination tasks have primarily been employed in speech perception studies, thus confounding the effects of perceptual memory with those of immediate perception. Since speech stimuli can be rapidly coded into a more durable phonetic code, it may be quite difficult with existing experimental techniques to isolate the perceptual processes involved in speech perception from those dependent on memory. Nevertheless, it is important to acknowledge the relatively large memory component inherent in traditional speech perception experiments and to take this into account in the discussion of experimental findings and the formulation of theoretical accounts.

Recently, Fujisaki and Kawashima (1969) have offered an elaboration upon categorical perception in terms of a model of the auditory and phonetic processes involved in speech perception. According to their model, when a listener is required to discriminate two <u>different</u> phonetic types he makes a decision based on the derived phonetic properties and upon features of the auditory stimulus as represented in "phonetic short-term memory." In this case, he determines whether the two stimuli have been identified as belonging to different phonetic categories with a binary decision (i.e., "same phoneme" or "different phoneme"). On the other hand, if the listener is required to discriminate two <u>identical</u> phonetic types (allophones) he must then rely primarily on a stored auditory representation of the acoustic parameters of the two stimuli in "auditory short-term memory" in order to make this decision. In this case, he must make a comparative judgement attending to the

specific acoustic properties of the stimuli. The Fujisaki model states that categorical perception is related to the degree to which these two types of memory stores, and in turn these two types of judgements, are employed in the decision process during an ABX discrimination task.

An auditory short-term store, analogous to Sperling's (1960) "iconic store", will be dependent on at least the following two factors: 1) the nature of the acoustic properties of the stimulus and 2) the length of time that the material is required to remain in store as the auditory transform of the acoustic signal.

It may be that the differences in perception between consonants and vowels reported in Experiment 1 are related to properties of the auditory short-term memory store rather than to articulatory differences in the production of these speech sounds. Vowel discrimination is better than consonant discrimination perhaps because auditory short-term memory for the sustained steady-state formant patterns (300 msec.) that distinguish isolated vowels is more dependable than auditory short-term memory for the brief transient acoustic cues (50 msec. transitions) that distinguish stop consonants. Thus, the differences in discrimination between the various classes of speech sounds (i.e., vowels, semi-vowels, stop consonants, fricatives, etc.) may reflect differences in auditory short-term memory for the corresponding "acoustic cues" which separate one class of speech sounds from another class.

The auditory processing model deserves rather close attention in view of the experimental findings that have been reported. Fujisaki and Kawashima (1968; 1969; 1970) have shown that non-categorical perception of vowels is primarily due to acoustic characteristics of the stimuli and the specific set of experimental conditions employed in studying them. Their results suggest that steady-state isolated vowels tend to be perceived more categorically if the stimulus duration is systematically reduced or if the vowel appears in a fixed context. Experiments by Stevens (1968) and Sachs (1969) provide additional support for the latter proposal. Stevens has shown that the vowels /i/, /I/ and /ɛ/ in the medial context /bV̲l/ are perceived more categorically than the same vowels in isolation. Sachs has reported similar results for the vowels /a/ - /ae/ in isolation and in the word context /badəl/ and /baedəl/.

The purpose of the present experiment was to determine the effect of stimulus duration upon both the identification and discrimination of synthetic steady-state vowels. A subsidiary goal of this study was to replicate and extend, with an additional experimental condition (long duration vowels), the Fujisaki findings which showed that the perception of vowels was more nearly categorical at very short durations.

Method

Materials

Stimuli. Two sets of seven steady-state vowels were synthesized on the vocal tract analogue synthesizer at the Research Laboratory of Electronics, Massachusetts Institute of Technology. Table 9 provides the frequencies of the first three formants for both sets of vowels. The fourth and fifth formants were fixed at 3500 Hz and 4500 Hz respectively. The formant frequencies were identical to the first seven vowel stimuli used in Experiment 1. The seven stimuli were arranged so that the first three formants varied in equal logarithmic steps from the English vowels /i/ through /I/.

One set of stimuli (long vowels) had a steady-state duration of 300 msec. (the equivalent of approximately 30 glottal pulses) with a rise and decay time of 50 msec. The second set of stimuli (short vowels) had a steady-state duration of 50 msec. (the equivalent of five glottal pulses) with a rise and decay time of 10 msec. Both sets of stimuli had identical formant frequency values and had a falling fundamental frequency. F_0 fell linearly from 125 Hz to 80 Hz for the long vowels and from 125 Hz to 100 Hz for the short vowels. The stimuli were originally recorded on magnetic tape at MIT. They were then digitized on the Haskins PCM system and the wave forms stored on disk for tape preparation.

Experimental Tapes

All the experimental tapes were produced under computer control using the Haskins PCM system. A 1000 Hz calibration tone was placed at the beginning of each tape to insure that the playback levels would be uniform throughout the testing sessions.

Identification tapes. Four different 70-item identification test sequences were prepared. Each sequence contained ten different randomizations of an entire set of 7 stimuli. The stimuli were recorded

TABLE 9

Formant Frequencies for Short and Long Vowel Stimuli

	Formant Frequency (Hz)		
Stimulus Number	F_1	F_2	F_3
1	270	2300	3019
2	285	2262	2960
3	298	2226	2902
4	315	2180	2836
5	336	2144	2776
6	353	2103	2719
7	374	2070	2666

singly with a 4-second interval between presentations and an 8-second interval after every ten trials. Four long vowel and four short vowel identification tapes were prepared from these test sequences.

ABX Discrimination tapes. Four different 88-item ABX discrimination tapes were constructed for each stimulus set. All possible one- and two-step comparisons of the seven stimuli appeared twice within each tape. The tapes were balanced with the restriction that each ABX triad occur equally often within each half of the test. Stimuli within each triad were separated by 1 second while successive triads were separated by 4 seconds. There was an 8-second pause after every ten trials.

Subjects

Five female undergraduate students at the University of Connecticut served as Ss in the present experiment. The Ss were obtained from the Psychology Department subject pool. All Ss were right handed native speakers of English with no past history of a hearing disorder or speech impediment. The Ss were paid for their services at a rate of $1.50 per hour. None of the Ss had ever heard any synthetic speech before the present experiment.

Procedure

The experimental tapes were reproduced on a General Radio Model 1525 tape recorder and were presented binaurally through Telephonics (TDH-39) matched headphones. The gain of the tape recorder playback for the 1000 Hz calibration tone was adjusted to give a voltage across the earphones equivalent to 75 dB SPL re 0.0002 dyn/cm^2. In order to compensate for the differences in loudness between long and short vowels due to stimulus duration the gain for the calibration tone on the short vowel tapes was adjusted to be +8 dB above the long vowel tapes. Measurements were made on a Hewlett Packard VTVM (Model 400) before presentation of each experimental tape. Ss were run in a group and were tested in a small experimental classroom.

The E read a set of instructions aloud to the Ss which explained the nature of the experiment. Ss also had a set of printed instructions before them. Ss were told that this was an experiment dealing with

speech perception. For the identification tests, Ss were required to identify each stimulus as either an /i/ or an /I/. They were instructed to use the following notation in recording their judgements: "EE" as in the word beet and "I" as in the word bit. In the ABX discrimination task Ss were told that the stimuli would be arranged in groups of three and that their task was to decide whether the third sound was identical to either the first sound or the second sound. Ss were told to guess if they were not sure but to respond on every trial. Judgements were recorded in a response booklet for later analyses.

Ss were run for an hour a day on five consecutive days. On each day Ss received both the long vowel and short vowel stimulus conditions in a counterbalanced arrangement. An identification test for a given stimulus condition was always followed immediately by the corresponding ABX discrimination test. When the data are combined over the five sessions, each S provided 50 identification judgements to each of the 7 stimuli in both the long and short vowel conditions. Each S provided 40 judgements for each of the AB comparisons in each stimulus condition.

Results

The average identification and discrimination functions for the six Ss are shown in Figure 10. The identification functions for both stimulus conditions are shown at the top of the figure. The corresponding one- and two-step obtained and predicted discrimination functions are shown below the appropriate identification functions.

Identification

The identification functions for both stimulus conditions are quite sharp and consistent. The mean informational redundancy values for absolute identification are .6816 for the short vowel condition and .6896 for the long vowel condition. An examination of these two functions reveals that they are very nearly exact complements of each other. A one step shift in the phoneme boundary between /i/ and /I/ can be observed. When stimulus duration is reduced from 300 msec. to 50 msec., the boundary shifts in favor of the short lax vowel /I/.

Discrimination

Inspection of the one- and two-step discrimination functions shows consistent differences in the shape and absolute level between the two

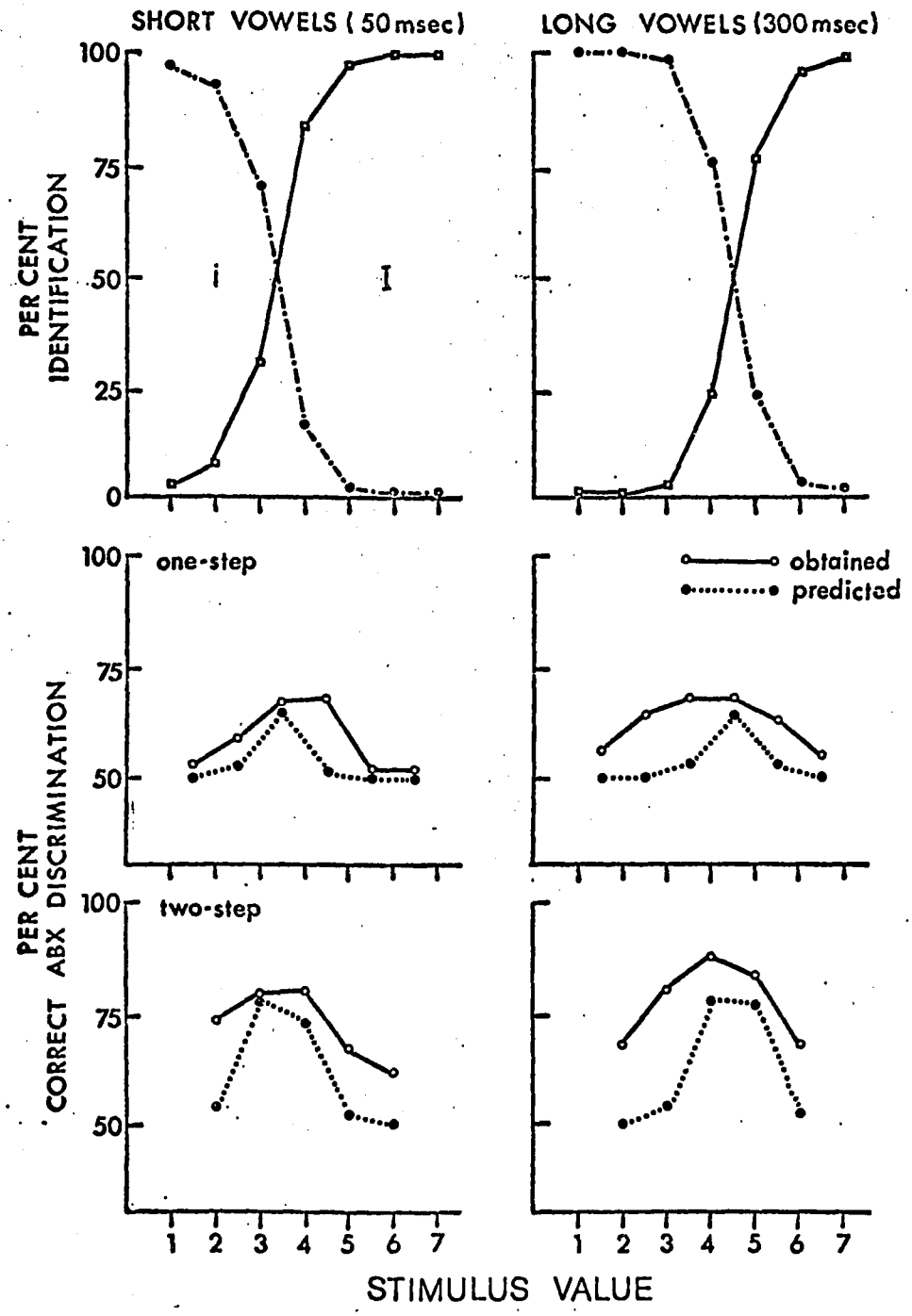


Figure 10. Average identification and discrimination functions for short vowels and long vowels.

stimulus conditions. To assess the effects of stimulus duration on discrimination, separate three-factor analyses of variance were carried out with the one- and two-step discrimination data. The variables under consideration were: (V) stimulus conditions, long vowels vs. short vowels; (D) discrimination, predicted vs. obtained; and (C) stimulus comparisons along the continuum.

For both the one- and two-step data the main effect of stimulus condition was significant, $F_{(1,4)} = 10.17$, $p < .05$ and $F_{(1,4)} = 15.66$, $p < .025$ respectively. Long vowels were discriminated more accurately than short vowels. The main effect for discrimination was also significant for the one- and two-step data, $F_{(1,4)} = 8.55$, $p < .05$ and $F_{(1,4)} = 8.33$, $p < .05$ respectively. The obtained discrimination scores were always greater than the predicted discrimination scores. This result appears primarily to be due to the larger discrepancy between the obtained and predicted scores in the long vowel condition. A significant two-way interaction between (V) stimulus condition and (D) discrimination, $F_{(1,4)} = 6.57$, $p < .05$, for the two-step data supports this observation. The main effect for stimulus comparison (C) summed over the other two variables was also significant for both the one- and two-step analyses, $F_{(5,20)} = 7.65$, $p < .001$ and $F_{(4,16)} = 18.23$, $p < .001$ respectively.

None of the one-step interactions was significant. A significant interaction between (V) stimulus condition and (C) stimulus comparison was obtained in the two-step analysis, $F_{(4,16)} = 13.09$, $p < .001$. Discrimination was better for the short vowels at comparison 2-4 and better for the long vowels at comparison 4-6. This interaction is due entirely to the differences in the location of the phoneme boundary. The differential discriminability at specific comparisons along the continuum is correlated with the shift in the boundary under the two stimulus conditions. Both of the two-way interactions combined to produce significant three-way interaction V X D X C, $F_{(4,16)} = 3.93$, $p < .025$.

Predicted and Obtained Discrimination Functions

To assess the effects of vowel duration on categorical perception, the predicted discrimination scores derived from the identification functions were compared with the obtained discrimination scores with a

Chi-square goodness of fit test for each S. The results of separate analyses for the one- and two-step data for both stimulus conditions are shown in Table 10. These analyses indicate that there are substantial individual differences between the predicted and obtained discrimination scores for long and short vowels.

Discussion

General Findings

The results of this experiment are in substantial agreement with the findings reported by Fujisaki. The following results have emerged concerning the effects of stimulus duration on the identification and discrimination of synthetic steady-state vowels.

1. In an absolute identification task with two response alternatives, vowel duration had little effect on response consistency as measured by the informational redundancy index. Identification functions for long vowels and short vowels were clean-cut and highly consistent. However, a shift in the phoneme boundary was observed favoring the phonologically short lax vowel /I/ as a consequence of reducing stimulus duration from 300 msec. to 50 msec.

2. The overall level of discriminability was related to changes in vowel duration. Long vowels were discriminated more accurately than the short vowels.

3. The difference between the predicted and obtained discrimination scores was greater overall for the long vowel condition than for the short vowel condition, as revealed by the analysis of variance. However, the separate Chi-square analyses on the predicted and obtained discrimination functions showed that there were large individual differences between Ss.

4. The results of this experiment indicate that stimulus duration is an important stimulus parameter related to categorical perception. Short vowels are perceived more categorically than long vowels.

Auditory Short-term Memory in Discrimination

The overall differences in discrimination performance for long and short vowels could be due to the differential reliability of auditory short-term memory in the ABX discrimination task. To test this hypothesis

TABLE 10

Chi-square Values

Comparison of Predicted and Obtained Discrimination Functions for Short Vowels and Long Vowels

	Short Vowels		Long Vowels	
Subject	One-step	Two-step	One-step	Two-step
1	2.53	10.50	5.49	14.02
2	9.00	33.92*	26.34*	29.10*
3	8.76	17.67	22.28	54.54*
4	27.31*	71.68*	14.64	69.42*
5	5.46	19.02	12.28	6.91

*$p < .001$

df = 6 (one-step)

df = 5 (two-step)

the contribution of auditory short-term memory to discrimination in both stimulus conditions was evaluated according to the procedures developed by Fujisaki. The following two assumptions are implicit in his model: 1) discrimination is based on identification if the first two members of an ABX triad (e.g., A and B) are judged by the listener to be <u>different</u> phonemes; 2) discrimination is based on auditory short-term memory if the first two members of an ABX triad have been judged to be the <u>same</u> phoneme.

Following Fujisaki, the predicted correct ABX discrimination score can be expressed by the following two components:

$$C_{ABX} = C_{A \neq B} + C_{A = B}$$

$$= C_{A \neq B} + M_{A = B} \cdot P_{A = B}$$

where $C_{A \neq B}$ = the probability that a correct discrimination occurs on the basis of identification.

$C_{A=B}$ = the probability that a correct discrimination occurs on the basis of auditory short-term memory.

$P_{A=B}$ = the probability that stimuli A and B are identified as the same phoneme.

$M_{A=B}$ = the conditional probability that a correct discrimination takes place when A and B are identified as the same phonemes. This quantity indicates the degree to which judgements are based on auditory short-term memory and is equal to the asymptotic value of C_{ABX} at the extremes of the stimulus range.

These components are related according to the following equations:

$$C_{A \neq B} = 1/2 \left[(P_1 - P_2) + P_1(1 - P_2) + P_2(1 - P_1)\right]$$

$$C_{A=B} = \left[P_1P_2 + (1 - P_1)\ (1 - P_2)\right] \cdot M_{A=B}$$

P_1 and P_2 represent the probabilities that stimuli A and B in the triad are identified as the same phoneme.[7]

---

[7] It should be noted here that Fujisaki obtained the identification probabilities with the "context-in" procedure. <u>Ss</u> were required to identify the stimuli as they appeared in the ABX triads rather than individually. The probabilities used in the present analysis were obtained with the context-out procedure where <u>Ss</u> identified the stimuli singly. Studdert-Kennedy et al. (1970) have pointed out the differences between the use of these two procedures in determining identification probabilities.

A new set of predicted ABX discrimination scores were calculated based on the Fujisaki procedure outlined above. Figure 11 shows the obtained two-step discrimination functions along with the new predicted functions for both stimulus conditions. The new predicted discrimination functions match the obtained functions much more closely than the standard Haskins predictions for both long vowels and short vowels. Better agreement is to be expected since one point on the obtained discrimination function has been used to calculate a new predicted discrimination function.

The results of this experiment suggest that the differences between categorical perception and continuous perception are dependent to some extent on the acoustic properties of the stimuli. The present experiment confirms the basic predictions derived from Fujisaki's model of the auditory and phonetic processes entailed in speech perception. The results further suggest that auditory short-term memory may be involved in the decision processes in the ABX discrimination task.

One other finding merits attention. In Fujisaki's experiments, he employed a two-response category continuum which was comparable to the present short vowel condition. However, it can be observed that even the long vowel discrimination functions appear somewhat categorical. Moreover, the individual Chi-square analyses indicate that the differences between obtained and predicted functions are not always greater for the long vowel condition as compared with the short vowel condition across all Ss. It is possible that the two-response category situation may constrain judgements differently than the three-category continuum used in Experiment 1. Vowel duration may not be the only determinant of the present results since the two-response category situation may have contributed to the categorical perception observed in some Ss for the long vowel condition.

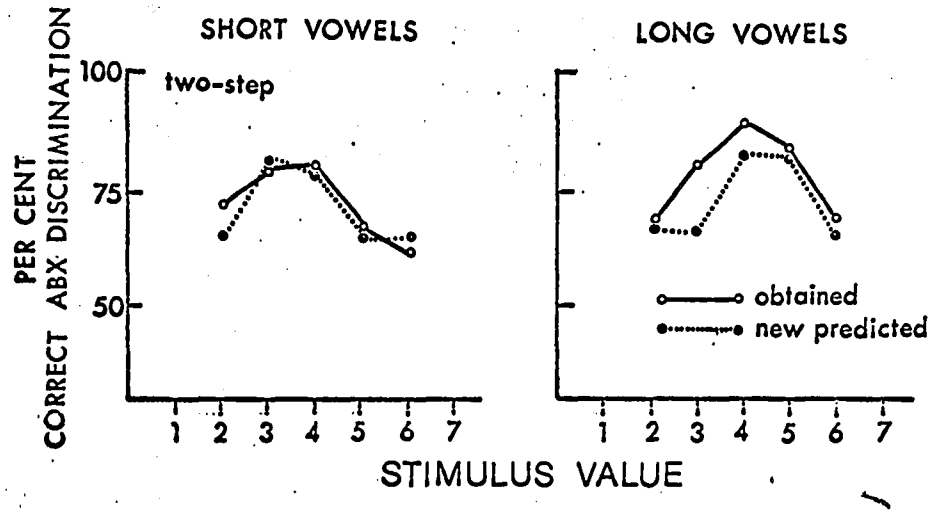


Figure 11. Average obtained two-step discrimination functions for short vowels and long vowels with the new predicted functions based on Fujisaki's formula.

CHAPTER III

EXPERIMENT IV

Four-Interval Test of Paired Similarity in Speech Discrimination

The procedure used to assess discrimination in all of the early speech perception experiments was limited exclusively to the ABX test (Liberman et al., 1957; 1961a; 1961b). In the ABX test, the S is presented with three stimuli in succession. The first two stimuli are always different; the third stimulus is identical to either the first or second stimulus in the triad. The S is required to indicate whether the third stimulus is more like the first or more like the second stimulus. Although more recent speech perception experiments have used the oddity technique (Mattingly et al., 1971), the ABX technique accounts for the overwhelming mass of data in speech perception studies.

The ABX procedure has one important advantage over the traditional procedures such as the method of constant stimulus differences, the AX test. The listener need not concern himself with the nature of the differences between stimuli. To this great advantage, we must balance one serious disadvantage. The ABX procedure appears to place an unusually heavy load on memory. Harris has perceptively described the complexities of the ABX procedure in 1952:

> "Although the instructions to the subject in ABX can be very simple, it is clear that a really very involved judgement process is called for. The subject is not making a single judgement, as good psychophysical practice usually requires, but a compound one. He first has the opportunity to compare B with A and probably always does make this comparison; then he compares X with B; then he has to reach back in time (italics mine DBP), so to speak, and compare X with A; and finally he has to make his judgement of whether the XB or XA comparisons is minimal. I do not see how data from this chain of comparisons can be interpreted as clearly as data from a technique using a single judgement." p. 417.

With pure tones, for example, where both the AX and ABX procedures may be used, difference thresholds for frequency are at least twice as great when measured by ABX as when measured by AX procedures (Rosenblith and Stevens, 1953). We may assume that in the pure tone situation the ABX test is less sensitive to purely auditory or acoustic differences

among stimuli than the AX test, since pure-tone discrimination judgements are presumably relatively free of categorization.

The development of sensitive and efficient procedures for speech discrimination testing has unfortunately not paralleled developments in auditory psychophysics (Swets, 1961; Green, 1961). Moreover, although auditory short-term memory appears to provide an important limitation on psychoacoustic performance (Massaro, 1970; 1971), short-term memory has received very little attention in speech perception experiments. By failing to reveal differences in discrimination due to acoustic factors, the sole use of the ABX procedure for the description of speech discrimination may have unduly underestimated listener's ability to detect intraphonemic differences among certain classes of speech sounds.

The purpose of the present experiment was to examine the use of a different discrimination procedure in speech perception: the four-interval forced-choice test of pair-similarity (4IAX). In the 4IAX test, which is a variant of the "method of cartwheels" (Coombs, 1964), a sequence of four sounds is presented to a listener (i.e., AA AB or AB AA). The S is told that on each trial one of the pairs of sounds will always be the same and one of the pairs of sounds will always be different. The S is required to indicate which pair of sounds was more alike (i.e., first or second). This technique shares the excellent property with the ABX test that the basis for discrimination need not be specified for the listener. It has the additional advantage over the ABX test in that the correct response can be achieved on a pair-by-pair comparison. It is not necessary to compare the first and third stimuli as is usually done in the ABX test. More importantly, this procedure requires the S to respond only to magnitudes of difference between stimuli rather than to the absolute magnitude of individual stimuli. The 4IAX test was selected over the more traditional AX two-interval same-different test, where the criterion for "same" or "different" is not under the control of the experimenter.

An additional aim of the present experiment was to determine whether the acoustic characteristics of the stimuli interact with the particular experimental procedures employed to assess discrimination. The two discrimination techniques (ABX and 4IAX) were compared in the present experiment under two conditions of vowel duration.

Method

Description of Stimuli

Two sets of 13 steady-state vowel stimuli were synthesized on the vocal tract analogue synthesizer at the Research Laboratory of Electronics, Massachusetts Institute of Technology. The frequencies of the first three formants are identical to those used in Experiment 1. The thirteen vowels were arranged so that the first three formants varied in approximately equal logarithmic steps through the three English vowels /i/, /I/ and /ɛ/.

The bandwidths of the first three formants were fixed at 50 Hz, 80 Hz and 110 Hz respectively. One set of stimuli (long vowel condition) had a steady-state duration of 300 msec. with a rise and decay time of 50 msec. The second set of stimuli (short vowel condition) had a steady-state duration of 50 msec. with a rise and decay time of 10 msec. Both sets of vowels had identical formant frequencies and had a falling fundamental frequency. F_0 fell linearly from 125 to 80 Hz for the long vowel set and from 125 to 100 Hz for the short vowel set.

These stimuli were originally recorded on magnetic tape at M.I.T. They were then digitized on the Haskins PCM system and the wave forms stored on disk for the preparation of all experimental tapes.

Experimental Tapes

All experimental tapes were produced under computer control on the Haskins Laboratories PCM system. The recording procedures were identical to those outlined in Experiment 1.

Identification tapes. Four different 130-item identification sequences were prepared for each stimulus condition. Each sequence contained ten different randomizations of an entire series of 13 stimuli. The stimuli were recorded singly with a 5-second interval between presentations and an 8-second pause after every ten presentations.

ABX discrimination tapes. Three different 92-item ABX tests were constructed for each stimulus condition. Each of the three ABX tests consisted of a different random order of the 92 possible ABX triads for stimuli one and two steps apart along the continuum. The test sequences were the same ones used in Experiment 1. The stimuli within each triad were separated by one second; successive trials were separated by five seconds. There was a ten second pause after every ten trials.

4IAX discrimination tapes. Three different 92-item 4IAX tests were also prepared for each stimulus condition. The 12 possible one-step comparisons and the 11 possible two-step comparisons for the 13 stimulus continuum were arranged in each of the following 4IAX sequences: AA AB, AA BA, AB AA, BA AA. Each of the three tests consisted of a different random arrangement of the 92 trials (i.e., 23 one- and two-step comparisons X the four 4IAX sequences). The stimuli within each pair were separated by 150 msec., while stimulus pairs were separated by 1 second. Successive trials were separated by five seconds. After every ten trials there was a 10-second pause.

Subjects

Six undergraduate students at Yale University served as Ss in the present experiment. The Ss were obtained from the volunteer subject pool at the Department of Psychology. All Ss were right handed native speakers of English and reported no past history of a hearing disorder or speech impediment. The Ss were run in three groups of equal size. Ss were paid $1.75 per hour for their services. None of the Ss had ever heard any synthetic speech before the present experiment. All Ss served over all experimental conditions.

Procedure

The experiment was conducted in an acoustically treated room at Haskins Laboratories. All experimental tapes were reproduced on an Ampex AG-500 tape recorder and were presented binaurally through Telephonics (TDH-39) matched and calibrated headphones. The gain of the tape recorder playback was adjusted to give a voltage across the earphones equivalent to 75 dB SPL re 0.0002 dyn/cm^2 for a 1000 Hz calibration tone. To compensate for the differences in loudness between long and short vowels due to the stimulus duration, the gain for the calibration tone on the short vowel tapes was adjusted to be +8 dB above the long vowel tapes. All measurements were made on a Hewlett Packard VTVM (Model 400) before the presentation of each experimental tape.

The instructions for the identification tests and the ABX discrimination tests were identical to those used in Experiment 1. In the 4IAX discrimination task, Ss were told that they would hear two pairs of

stimuli on each trial and that one of the pairs would always be the same while the other pair would always be different. They were told that their task was to determine which pair was the same (i.e., first pair or second pair). Ss were told to guess if they were not sure, but to respond on every trial. Ss recorded their judgements for both discrimination tasks in a prepared booklet containing IBM test sheets.

Ss were run for two hours a day for three consecutive days. An identification test for each of the two stimulus conditions was presented on the first day. Either a long vowel or a short vowel identification test was presented on the second day. The remaining identification test was presented on the third day. Discrimination tests were presented after the identification tests.

Each of the three groups received a different random arrangement of all of the discrimination tests every day. However, none of the discrimination tests was presented more than twice to any group on a given day.

Results

Identification Functions

The average identification functions for both vowel conditions for the six Ss are shown in Figure 12. Each point is based on 120 judgements summed over the six Ss. Both figures reveal that Ss partitioned the continua into three distinct phoneme categories. The proportion of /I/ judgements is somewhat lower for the long vowel condition as compared with the short vowel condition. The identification function for the long vowels is quite similar to that shown in Experiment 1 for the same stimuli.

In order to assess response consistency, informational redundancy measures were obtained for both sets of identification judgements for each S. Although the short vowel judgements were more consistent overall ($R = .7585$) than the long vowel judgements ($R = .7025$), the difference was not significant across subjects by a Wilcoxon matched-pairs signed-ranks test ($T = 10$).

ABX and 4IAX Discrimination Functions

The average one- and two-step ABX and 4IAX discrimination functions for both vowel conditions are shown below the identification functions in

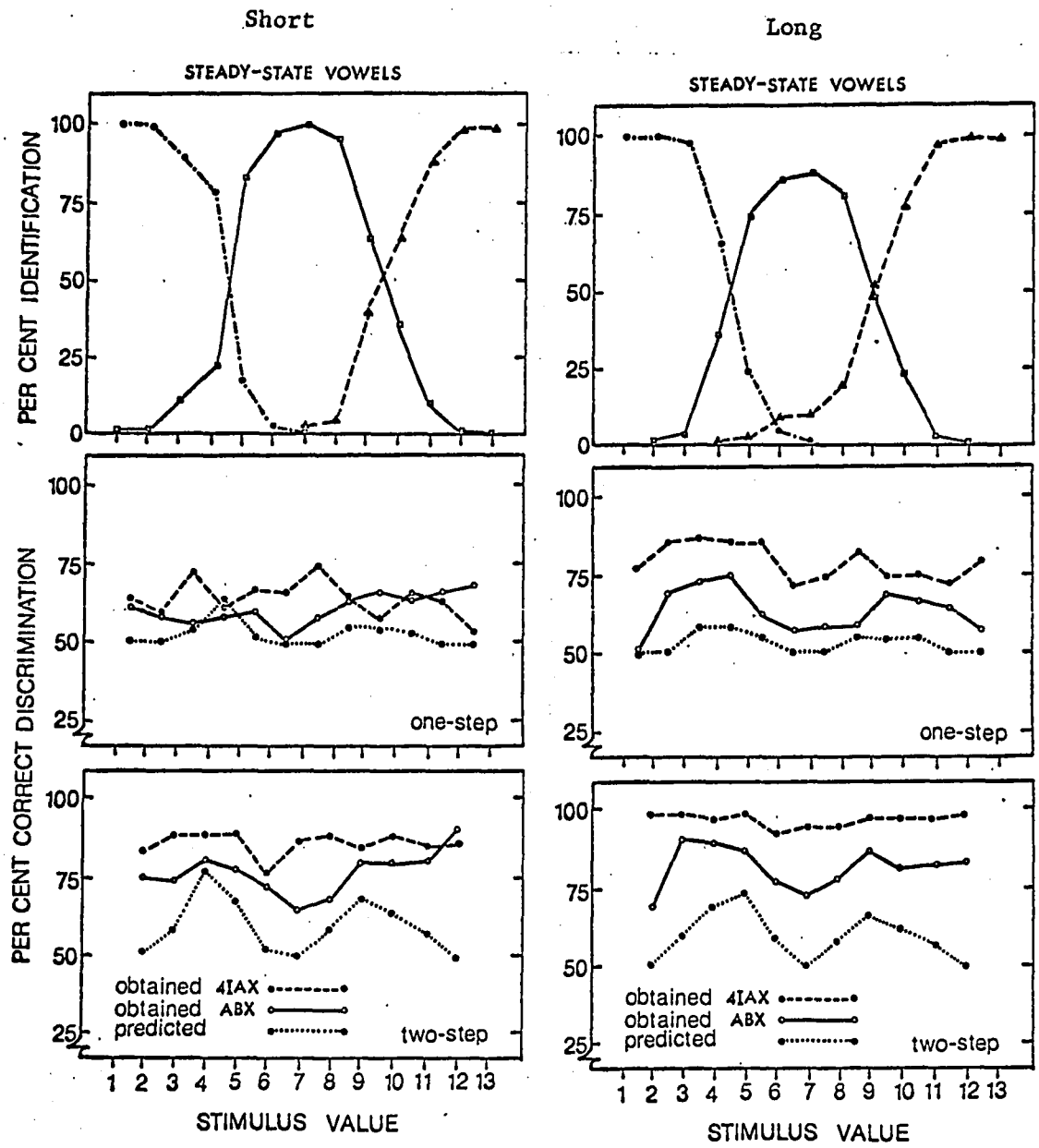


Figure 12. Average identification and discrimination functions for short vowels and long vowels.

Figure 12. Each comparison is based on 144 judgements across the 6 Ss. The solid lines represent the ABX discrimination function while the dashed lines show the 4IAX discrimination function. The predicted discrimination function based on the Haskins procedure is presented for comparison and is represented by the dotted lines. The predicted functions are based on the extreme assumption that stimuli can be discriminated only to the extent that they are identified differently.

An inspection of the discrimination functions indicates that there are substantial and consistent differences between the ABX and 4IAX procedures for vowel discrimination. The differences are quite apparent for both the one- and two-step long vowel functions. Overall discrimination for accuracy improves from 63% to 79% for the one-step data and from 82% to 96% for the two-step data. The differences between the ABX and 4IAX tests are not as obvious for the one-step short vowel data, although the differences are revealed in the two-step functions. Overall accuracy improves from 61% to 64% for the one-step data and from 77% to 86% for the two-step data.

Separate three-factor analyses of variance for the one- and two-step discrimination scores established that the differences between the two discrimination techniques were highly significant, $F_{(1,5)} = 20.17$, $p < .01$ for the one-step functions and $F_{(1,5)} = 36.13$, $p < .005$ for the two-step functions. Discrimination accuracy was greater for the 4IAX as compared with the ABX test. The difference between the long and short vowels was highly significant, $F_{(1,5)} = 25.54$, $p < .005$ for the one-step data and $F_{(1,5)} = 39.66$, $p < .005$ for the two-step data. Long vowels were discriminated better than short vowels even though the identification functions were quite similar with regard to the response consistency index.

The main effect for stimulus comparison along the continuum was not significant for either the one-step or two-step data. All of the first and second order interactions did not reach significance.

Discussion

The results of the present experiment indicate that there are substantial differences in overall scores between the standard ABX discrimination test and the new 4IAX test of paired similarity. One possible

explanation of the differences between the two discrimination procedures may be found in an examination of the short-term memory requirements for each task. For example, in the ABX test, the S compares X with B and then X with A before arriving at a final decision. The procedure requires the S to retain information about the acoustic properties of the stimuli over a relatively long period of time. On the other hand, in the 4IAX test, the S can base his final decision on a pair-wise comparison and respond to the magnitude of the difference between pairs of stimuli. The information that is retained over time for a final discrimination judgement will most probably be information about a decision or an evaluation such as degree of similarity on the first pair of stimuli rather than specific information about the characteristics of the stimuli (i.e., a judgement about the absolute magnitude of individual stimuli). It would appear from these results that retention of information about an initial decision (i.e., "same" or "different") is more reliable for the final discrimination judgement than retention of information about the specific properties of the acoustic stimuli.

Another explanation of the differences between the two discrimination tests may be related to the difficulty of the task. The ABX test may require a more difficult judgement for the S than the 4IAX test. However, this seems unlikely since almost all of the Ss reported that the 4IAX test was more difficult and required them to pay closer attention to the stimuli than the ABX test.

A comparison of the ABX and 4IAX tests under the two stimulus conditions showed that the differences in discrimination performance were independent of vowel duration. The stimulus condition by discrimination technique interaction did not reach significance indicating that both the acoustic properties of the stimuli (i.e., vowel duration) and the procedures used to assess discrimination make separate contributions to speech discrimination. Even with very short vowels of 50 msec. duration, discrimination improved with the 4IAX test.

The results of the present experiment differ in several respects from the results obtained in Experiment 3 with the two-category vowel continua. The discrimination functions for the short vowels in the present experiment do not show the peaks and troughs which were present

in the discrimination functions for Experiment 3. This may be due in part to the increased number of stimuli in this experiment and the extension to a third vowel category in contrast with the two of Experiment 3 or the averaging over Ss. The range of the stimuli employed and the number of phoneme categories encompassed may be an important factor in speech discrimination. An increase in the number of stimuli and phonetic classes may cause greater acoustic contrast, thus increasing the possible overall context effects obtained in the experiment. It should be noted here that the Haskins experiments on vowel perception also used three categories. Although large context effects can be observed in the three-category vowel situation, there is little evidence of context effects in Experiment 3 where only two phoneme categories were used. Context effects have been determined by an inspection of the identification functions. The degree to which identification response categories are clearly delineated and whether there are abrupt shifts from one category to another have been cited as evidence for context effects (Eimas, 1962).

Another difference worth noting concerns the number of identification judgements which took place before discrimination. Only 20 identification judgements were obtained for each stimulus in this experiment, as compared with 50 judgements per stimulus in Experiment 3. From the shape of the functions it seems apparent that listeners did not identify the stimuli into categories during the discrimination task and thus failed to show the characteristic peaks and troughs.

Cross and Lane (1964) have shown that instructions before discrimination testing produce substantial differences in the shape of the obtained discrimination function. In their experiment, the /do/-/to/ stimuli (Liberman et al., 1961b) were presented to two groups of Ss. One group of Ss was told that these stimuli were the speech sounds /do/ or /to/ while the other group was told that these stimuli were complex auditory tones. The discrimination functions obtained with the first group showed peaks at the phonetic boundaries and troughs within-categories; the discrimination functions for the other group did not show discontinuities correlated with changes in identification. Popper (1967) has reported that practice in labelling the stimuli during the identification task raises the Ss criterion for different responses to within-category

comparisons which are presumably based on acoustic differences among stimuli.

Although the differences between the two discrimination procedures in the present experiment were quite large, they were nevertheless obtained with vowel stimuli which are typically more discriminable than consonants. It remains to be seen whether the differences reported here between the ABX and 4IAX task also can be obtained with stop consonant stimuli.

In summary, in discrimination tasks where substantial improvement can result from noting acoustic differences among vowels, lower scores are obtained with the ABX discrimination technique than with other procedures. We suspect that the heavy memory-related task requirement associated with ABX is concealing differences among stimuli which are due to acoustic factors. Moreover, the categorical perception effects that have been inferred with the ABX procedure may be related in part to auditory short-term memory for the acoustic properties of different classes of speech sounds. The next experiment was designed to test this hypothesis with a stop consonant continuum.

EXPERIMENT V

A Comparison of the 4IAX and ABX Tests with Stop Consonants

The results of the previous experiment indicated that the 4IAX test was more efficient in revealing the differences between stimuli than the traditional ABX procedure. It was suggested that the differences between the procedures may be found in the use of acoustic information in the 4IAX procedure and the severe memory requirements associated with the ABX procedure. Conversely, conclusions about the role of categorical perception may be related to short-term auditory memory. The exclusive use of this one discrimination procedure in past speech perception experiments may have led investigators to underestimate S's ability to employ acoustic information in speech discrimination.

However, the differences between the two discrimination procedures reported in the previous experiment were obtained with vowel stimuli, which are both more discriminable than consonants and have been shown in previous work to display non-categorical perception. Even the ABX discrimination scores were significantly higher in both Experiment 4 and Experiment 1 than the predicted scores indicating that listeners can discriminate vowels better than would be predicted from absolute identification alone.

The purpose of the present experiment was to determine whether the differences between the 4IAX and ABX procedures, which were found in Experiment 4, can also be obtained with a consonant continuum which is more likely to be perceived categorically. If differences between the two discrimination procedures still appear with consonants, and if we assume that differences associated with the procedures are related to their differential memory requirements, we will have reason to further examine the role of short-term memory in categorical perception.

Method

Description of Stimuli

A set of voiced stop consonant-vowel syllables was synthesized on the Haskins parallel resonant synthesizer. The set of stimuli consisted of 13 three formant syllables, 300 msec. in duration. These stimuli were identical to those employed in Experiment 1. The starting frequencies of the F_2 and F_3 transitions were varied systematically to

produce a series of stimuli ranging between the syllables /bae/, /dae/ and /gae/.

The stimuli were originally recorded on magnetic tape. They were then digitized on the Haskins PCM system and the wave forms stored on disk for the preparation of all experimental tapes.

Experimental Tapes

All experimental tapes were produced under computer control on the PCM system. The recording procedures were identical to those outlined in Experiment 1.

Identification tapes. Four different 130-item identification sequences were prepared. Each sequence contained ten different randomizations of an entire series of 13 stimuli. The stimuli were recorded singly with a 5-second interval between presentations and an 8-second pause after every ten presentations.

ABX discrimination tapes. Three different 92-item ABX tests were constructed. Each of the three ABX tests consisted of a different random order of the 92 possible ABX triads for one and two steps apart along the continuum. The test sequences were the same ones used in Experiment 1.

4IAX discrimination tapes. Three different 92-item 4IAX tests were also prepared. Each of the three 4IAX tests consisted of a different random order of the 92 trials. These test sequences were identical to those used in Experiment 4. Stimuli within each pair were separated by 150 msec., while stimulus pairs were separated by one second. Trials were separated by five seconds. There was an additional 10-second pause after every ten trials.

Subjects

Five undergraduate students at the University of Michigan served as Ss. The Ss were obtained from the paid subject pool at the Mental Health Research Institute. They were all right handed, native speakers of English and reported no past history of a hearing disorder or speech impediment. The Ss were paid at the rate of $1.75 per hour for their services. None of the Ss had ever heard any synthetic speech before the present experiment.

Procedure

The experiment was conducted in an anechoic chamber located in the Phonetics Laboratory at the University of Michigan. The experimental tapes were reproduced on an Ampex 351-2 tape recorder and were presented binaurally through Telephonics (TDH-39) matched and calibrated headphones. The gain of the tape recorder was adjusted to give a voltage across the earphones equivalent to 75 dB SPL re 0.0002 dyn/cm^2 for a 1000 Hz calibration tone. Measurements were made on a Ballantine VTVM (Model 300) before the presentation of each experimental tape.

The instructions for the identification test and the ABX discrimination test were identical to those used in Experiment 1. For the identification task, Ss were required to identify each stimulus as a /b/, /d/ and /g/. For the ABX task they were told that the stimuli would be arranged in groups of three and that their task was to decide whether the third sound was identical to the first sound or the second sound. The instructions for the 4IAX test were identical to those used in Experiment 4. Ss were told that they would hear two pairs of stimuli on each trial and that one pair was always the same and one pair was always different. Ss were instructed to determine which pair was the same. Judgements were recorded in prepared booklets containing IBM test sheets for later analyses.

Ss were run for an hour a day on five consecutive days. Each of the first four sessions began with an identification tape which was followed by an ABX tape and a 4IAX tape. On the fifth day each discrimination test was presented twice. The order of presentation of the discrimination tests was counterbalanced over the five days.

Results

Identification Functions

The average identification for all five Ss is shown in Figure 13. Each point is based on 200 judgements summed over the five Ss. The group function is less consistent than that obtained in Experiment 1 for the same stimuli. The average informational redundancy values for the identification functions confirm this observation, R = .7644 in Experiment 1 and R = .6264 in the present experiment. It appears that these Ss are less consistent in their identification judgements than the Ss employed in Experiment 1. Because of the lack of agreement

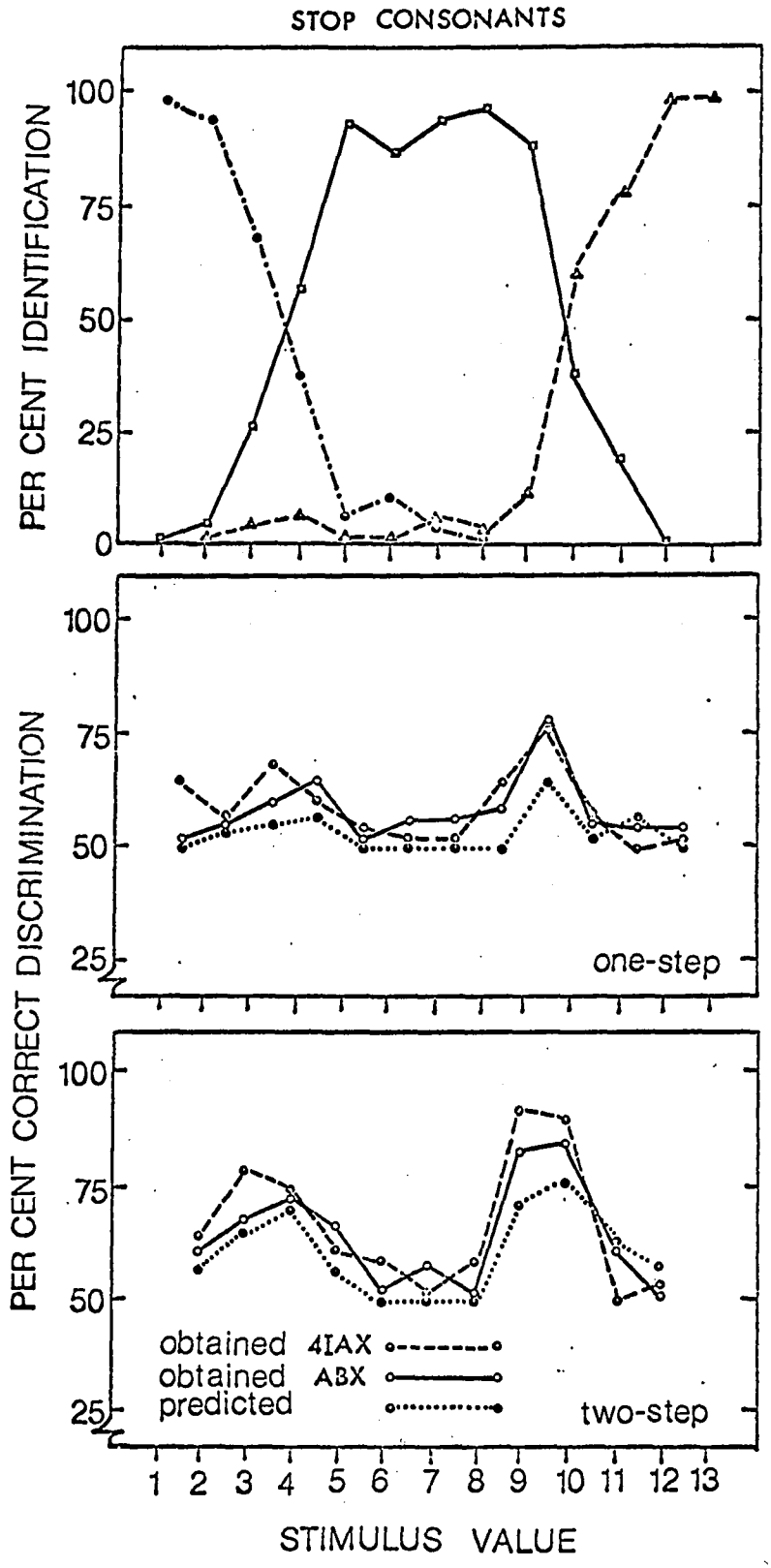


Figure 13. Average identification and discrimination functions for stop consonants.

among Ss, the individual data for each Ss is presented in Appendix D for examination.

ABX and 4IAX Discrimination

The average one- and two-step ABX and 4IAX discrimination functions are shown below the identification function in Figure 13. Each comparison is based on 140 judgements averaged across the five Ss. The individual discrimination functions for each S are shown below the corresponding identification functions in Appendix D. The solid lines show the ABX function; the dashed lines show the 4IAX functions. The Haskins predicted function is represented by the dotted lines.

Inspection of the discrimination functions indicates that the differences between the ABX and 4IAX procedures with the consonant stimuli are less obvious than those found for the vowels in Experiment 4. However, separate two-factor analyses of variance on the discrimination scores showed that the main effect of discrimination task was significant for both the one- and two-step data, $F_{(2,8)} = 10.95$ $p < .01$ and $F_{(2,8)} = 30.62$, $p < .001$ respectively. For the one-step data a Newman-Keuls multiple comparison of means test (Winer, 1962, pp. 80-85) indicated that the 4IAX scores differed significantly from the predicted scores but, not from the ABX scores. A similar analysis on the two-step data revealed that the means for all three discrimination scores differed significantly from each other. The discrimination scores for the two-step data were higher with the 4IAX test as compared with the ABX test. In this context, it is particularly puzzling that the significant advantage of 4IAX discrimination is more marked at the peaks (where categorization should be contributing most heavily to discrimination) than in the troughs (where acoustic memory might be expected to be working against ABX discrimination).

The main effect for stimulus comparison along the continuum was highly significant for both one- and two-step discrimination scores, $F_{(11,44)} = 7.35$, $p < .001$ and $F_{(10,40)} = 10.54$, $p < .001$ respectively. Discrimination by comparison interaction was not significant in either the one- or two-step analyses.

Discussion

The results of the present experiment suggest that the 4IAX discrimination technique is more sensitive to the acoustic differences between consonants than the traditional ABX procedure. However, the differences between the two discrimination procedures are not nearly as great for the consonants as they were for the vowels. The small but consistent overall differences between the two discrimination techniques resulted from the sensitive within Ss design in the analysis of variance. From an examination of the average discrimination functions, it becomes clear that while the differences are statistically significant, they are not well reflected in the group data.

Two additional points should be noted. First, the significant main effect for stimulus comparison indicates that there were large and consistent peaks and troughs in both the ABX and 4IAX consonant discrimination functions. This contrasts with the vowel functions yielded by these techniques in Experiment 4. Vowels are perceived less categorically than the consonants under both discrimination procedures. Second, the absence of a significant interaction between discrimination technique and stimulus comparison indicates that the shape of the discrimination function was basically the same for both types of discrimination task. The peaks and troughs, characteristic of the predicted discrimination functions for categorically perceived speech stimuli, still remain with the 4IAX procedure. This last finding is important because it reaffirms the consistent differences in perception between consonants and vowels that have been reported by Haskins investigators. Discrimination of consonants is only slightly better than could be expected on the basis of absolute identification. Discrimination across phoneme boundaries is far superior to discrimination within phoneme boundaries even with the 4IAX discrimination technique.

CHAPTER IV

EXPERIMENT VI

Auditory Short-Term Memory in Speech Discrimination

The results of the previous experiments have suggested that the differences between consonant and vowel perception may be related in some way to their differential fate in auditory short-term memory. This explanation was based on an examination of two experimental findings. First, in Experiment 3, stimulus duration was shown to be related to the degree of categorical perception. Vowels of very short duration tend to be perceived more categorically than identical vowels of longer duration. Secondly, in Experiments 4 and 5, the ABX test was shown to be less sensitive to the acoustic differences between stimuli than the 4IAX procedure. Differences between the two discrimination procedures, it was argued, were due to the differential memory requirements in the two tasks. But the argument is, at best, indirect. And, since there was little information available in the literature about the role of auditory short-term memory in speech discrimination tasks, an experiment was designed to explore this problem.

The purpose of the present experiment was to test the hypothesis that consonants and vowels differ in the degree to which auditory short-term memory is employed in their discrimination. If there are differences in auditory short-term memory for vowels and consonants, it should be possible to examine these differences with several procedures. Among the possibilities available is a delayed-comparison recognition memory task. Several recent studies have examined the relationship between duration of the delay interval and pitch discrimination (Wickelgren, 1966; Massaro, 1970). This procedure, however, has been used sparingly to study recognition memory in speech discrimination. Popper (1967) has reported some evidence that speech discrimination is better at shorter inter-stimulus intervals (i.e., 500 msec.) than longer inter-stimulus intervals (i.e., 5 sec.) but his study was confined exclusively to consonant stimuli and there were several artifacts in his procedures.

It is important to distinguish two types of memory in speech discrimination: short-term phonetic memory and short-term auditory non-

phonetic memory. Short-term auditory memory refers to the retention of the auditory transform of the acoustic properties of the stimuli. Short-term phonetic memory refers to the retention of abstract derived phonetic features from the auditory signal. Following Fujisaki's model, we will assume that a listener's decision in a speech discrimination task is dependent on both of these memory components, and that the listener relies on each of these components differentially. For example, if a listener has identified two sounds as phonetically different, we assume that he can make a discrimination judgement based solely on phonetic memory. He has only to retrieve information about the phonetic identification of the two stimuli. On the other hand, if a listener is able to discriminate two sounds that he has identified the same phonetically, we assume that he relies on auditory short-term memory in discrimination. The listener bases his discrimination judgement on the auditory transform of specific acoustic properties of the stimuli, since both stimuli belong to the same phonetic category. We can produce a relative preponderance of the two types of judgements by selecting pairs of stimuli of equal physical differences from either between phoneme boundaries or within phoneme boundaries. The present experiment was designed to assess the temporal course of the accuracy of recognition memory for both vowel and consonant stimuli both within and between phoneme boundaries.

Method

Experimental Design

The design of the present experiment involved the manipulation of three independent variables: stimulus conditions (4 levels, two classes of vowels and two classes of consonants); stimulus comparisons (2 levels, within phoneme boundary comparisons and between phoneme boundary comparisons); and delay interval (0.0, .25, .50, 1.0 and 2.0 seconds). Ss were assigned to one of four different stimulus conditions, with stimulus distributed over all of the levels of the other two independent variables.

Subjects

Sixteen undergraduate students at the University of Michigan served as Ss in the present experiment. The Ss were obtained from the paid subject pool at the Mental Health Research Institute. All Ss were right

handed native speakers of English and reported no history of a hearing disorder or speech impediment. Ss were paid for their services at the rate of $2.00 per hour. None of the Ss had ever heard any synthetic speech before the experiment.

Description of Stimuli

The following four sets of synthetic speech stimuli were prepared and correspond to the different stimulus conditions in the experiment. All of the stimuli were digitized and their wave forms were stored on the Haskins PCM system. All stimuli were initiated and terminated at the temporal baseline so that transients were avoided, even with 0 sec delay.

Voiced stop consonants (/bae-dae/). The first seven voiced stop consonant-vowel syllables were selected from the set of stimuli employed in Experiment 1. The seven stimuli varied in approximately equal steps from the syllable /bae/ to /dae/.

Bilabial stop consonants (/ba-pa/). A set of seven three formant bilabial stop consonant-vowel syllables was synthesized on the Haskins Laboratories parallel resonant synthesizer. The stimuli were 300 msec. in duration. The seven stimuli varied in 10 msec. steps along the voice onset time (VOT) continuum from 0 msec. through + 60 msec. This range of VOT values has been shown by Lisker and Abramson (1967) to be sufficient to characterize the voicing and aspiration distinctions between the initial stop consonants in the syllables /ba/ and /pa/.

Long steady-state vowels (/i - I/). The seven 300-msec. steady-state vowels used in the long vowel condition of Experiment 3 were employed here. The seven stimuli varied in equal steps from the vowel /i/ through /I/.

Short steady-state vowels (/i - I/). The seven 50-msec. steady-state vowels used in the short vowel condition of Experiment 3 were also employed here. The seven stimuli varied in equal steps from the vowel /i/ through /I/.

Experimental Tapes

All of the experimental tapes were produced under computer control on the Haskins PCM system. The recording procedures were identical to

those described in Experiment 1. Two types of tests were prepared for each stimulus condition: an absolute identification test and a delayed-comparison recognition memory test.

Absolute identification tests. Two different 70-item identification tests were prepared for each of the four stimulus conditions. Each identification test contained ten different randomizations of an entire series of the seven stimuli in the set. The stimuli were recorded singly with a four second interval between presentations and an eight second interval after every ten presentations.

Delayed-comparison recognition memory tests. Four different delayed-comparison recognition memory tests were constructed for each of the four stimulus conditions. Each test contained two separate replications of 50 trials. A set of fifty trials consisted of ten basic test pairs at each of five delay intervals. The ten test pairs were constructed in the following manner. Stimulus numbers 1, 3, 5, and 7 were selected from each of the four original sets of seven stimuli. Each of these four stimuli was paired with the next stimulus one-step apart along the continuum and then arranged in two AB permutations. This operation produced 6 "different" AB pairs at a given delay interval. Each of the four stimuli was also paired with itself once resulting in 4 "same" AA pairs at each delay interval. Each set of 50 trials appeared in a different random arrangement in each test.

The stimuli were recorded in pairs, with a five second interval between successive trials. The delay intervals (0, .25, .50, 1.0, 2.0 sec.) were arranged under computer control. A 100 msec. 1000 Hz tone was recorded 750 msec. prior to the onset of the first stimulus in each pair as a ready signal.

Procedure

The experiment was conducted in an anechoic chamber located in the Phonetics Laboratory at the University of Michigan. The experimental tapes were reproduced on an Ampex 351-2 tape recorder and were presented binaurally through Telephonics (TDH-39) matched and calibrated headphones. The gain of the tape recorder was adjusted to give a voltage across the earphones equivalent to 75 dB SPL re 0.0002 dyn/cm^2 for a

1000 Hz calibration tone. Measurements were made on a Ballantine VTVM (Model 300) before the presentation of each experimental tape.

The instructions for the identification task were identical to those used in the previous experiments. The Ss were told that the stimuli would be presented individually and that they were required to identify each stimulus into one of two categories depending on the particular stimulus condition employed.

For the delayed-comparison recognition memory task, Ss were told that they would hear two stimuli separated by a varying interval on each trial and their task was to decide whether the two stimuli were the "same" or "different". They were told that approximately half of all of the pairs were the same and half of the pairs were different. They were encouraged to guess if they were not sure of a judgement. Judgements were recorded in prepared booklets containing IBM test sheets for later analyses.

Ss were run for an hour a day on two consecutive days. Each session began with a 70-item identification test which was followed by four 100-trial delayed-comparison recognition memory tests. The order of presentation for the delayed-comparison tests was reversed on the second day.

Results

Identification Functions

The average identification function for each of the four stimulus conditions is shown in Figure 14. Each point is based on 80 judgements summed over the four Ss in each condition. In general, it can be observed that Ss partitioned each of the stimulus continua into two relatively distinctive phoneme categories. The discrimination functions averaged across all delay intervals are shown by the dotted lines in Figure 14.

Delayed-Comparison Discrimination

The within-boundary discrimination scores were obtained by averaging the judgements for stimulus comparisons 1-3 and 5-7 in each condition. The between-boundary discrimination scores were represented by the judgements for the stimulus comparison 3-5 for each stimulus condition.

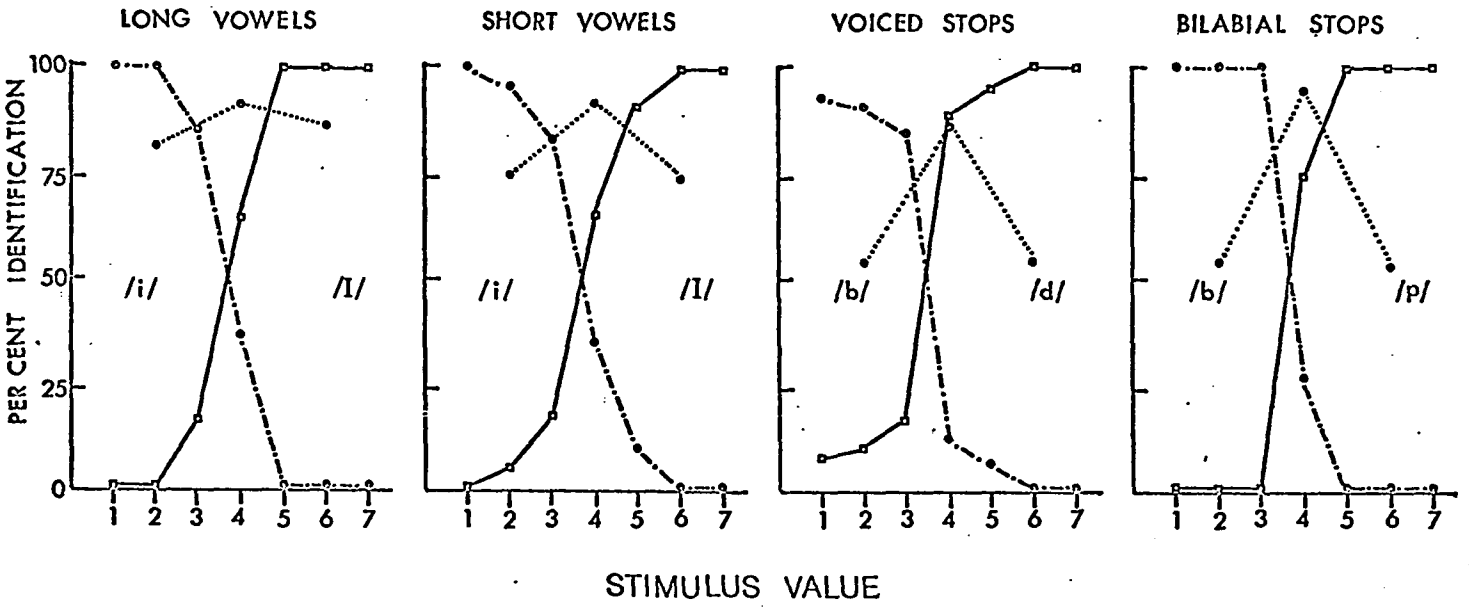


Figure 14. Average identification functions for each of the four stimulus conditions with discrimination functions averaged over all delay intervals superimposed on the corresponding identification functions.

P("D"|D) discrimination scores. In the present experiment, discrimination probabilities are obtained from conditions in which the pairs of stimuli are different. Figure 15 shows the average probability of a "different" response when the stimuli were different (i.e., P("D"|D)) as a function of delay interval for each stimulus condition. The filled circles represent between phoneme boundary comparisons while the open circles represent within phoneme boundary comparisons.

The P("D"|D) scores were analyzed by means of a three-factor analysis of variance for mixed designs. The results of the analysis of variance are shown in Table 11. All F-ratios were evaluated against the corresponding mean square term including Ss. An inspection of the analysis of variance table reveals that each of the three main effects, their two-way interactions and their three way interaction was significant.

The overall P("D"|D) scores for between-category comparisons are quite high and relatively stable across the delay intervals for three of the four stimulus groups. The one exception appears to be the long vowel condition where there is a decline in the between-category discrimination scores as the delay interval increases. However, Newman-Keuls tests on the differences among between-category means for each group indicated that they did not differ significantly from each other. Inspection of the within-category scores, however, shows very clearly the differences between the discrimination of vowels and consonants. The within-category consonant scores are very low and do not appear to be related in any way to changes in the delay interval. Newman-Keuls tests showed that there were no significant differences for the consonant means across delay intervals.

On the other hand, the within-category vowel scores are much higher and are systematically related to changes in the delay interval. For both vowel conditions, within-category accuracy is maximum at .25 sec. and then decreases systematically with increasing delay intervals. Newman-Keuls tests established that the .25 interval was significantly different from the other delay intervals for both vowel conditions.

d'discrimination scores. In order to separate the effects of recognition memory and response bias on the observed judgements, a signal detection analysis was performed. False alarm rates were obtained

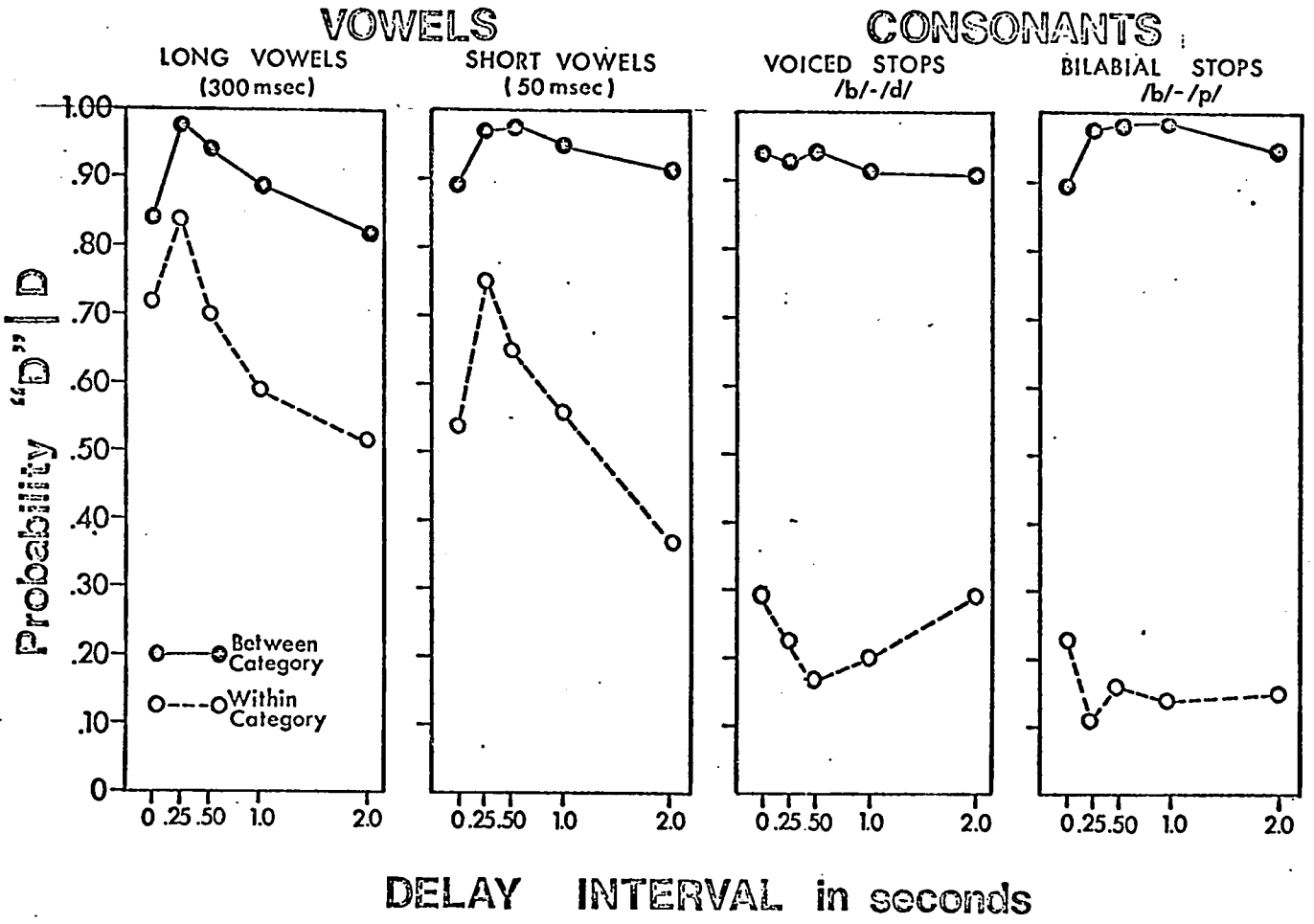


Figure 15. Average probability of a "different" response when the stimuli within a pair were different (P "D"|D) as a function of delay interval for each stimulus condition for within and between phoneme boundary conditions.

TABLE 11

Analysis of Variance for Recognition Memory

P("D"|D) Scores

Source of Variation	Sum of Squares	df	Mean Square	F	p
Mean	719446.51	1	719446.51		
G(Stimulus Condition)	16063.37	3	5354.46	16.47	<.001
I(Delay Interval)	2077.46	4	519.37	5.58	<.005
C(Stimulus Comparison)	108212.01	1	108212.01	238.35	<.001
S(G)	3895.02	12	324.58		
GI	3105.54	12	258.79	2.77	<.01
GC	22303.37	3	7434.46	16.37	<.001
IC	905.84	4	226.46	3.65	<.025
SI(G)	4481.60	48	93.37		
SC(G)	5444.32	12	453.69		
GIC	1643.16	12	136.93	2.20	<.05
SIC(G)	2976.80	48	62.02		

from trials on which Ss responded "different" when the pairs of stimuli were the same, i.e., P("D"|S). Since these false alarm rates were available for each stimulus at each delay interval, it was possible to obtain a d' score and use this as a reliable measure of detection which is independent of response bias. The d' value has been used as a measure of trace strength in recognition memory (Wickelgren, 1966; Massaro, 1970). Figure 16 shows the average d' scores for within- and between-category comparisons as a function of delay interval for each stimulus condition. Better discrimination accuracy is shown by higher d' levels.

The results shown in Figure 16 indicate that discrimination accuracy decreases for both within- and between-category comparisons as the delay interval increases beyond .25 sec. The effect also appears to be greater for the two vowel conditions than for the consonants. A three-factor analysis of variance was applied to the d' scores. The analysis revealed essentially the same results as the analysis of variance performed on the P("D"|D) scores except that one important difference appeared between the two analyses. The stimulus condition by delay interval interaction which was significant in the first analysis with the P("D"|D) scores did not reach significance with the d' scores. The triple interaction also did not reach significance.

It is clear from the results of both analyses of variance that increases in the delay interval affect vowel discrimination accuracy much more than consonant discrimination. Moreover, the differences appear to be most pronounced for the within-category vowel comparisons.

Discussion

The outcome of the present experiment strongly supports the claim that the differences between consonant and vowel discrimination are related in some way to the differential use of auditory short-term memory for these two classes of speech sounds. Furthermore, the results provide some insight into the temporal course of recognition memory in speech discrimination.

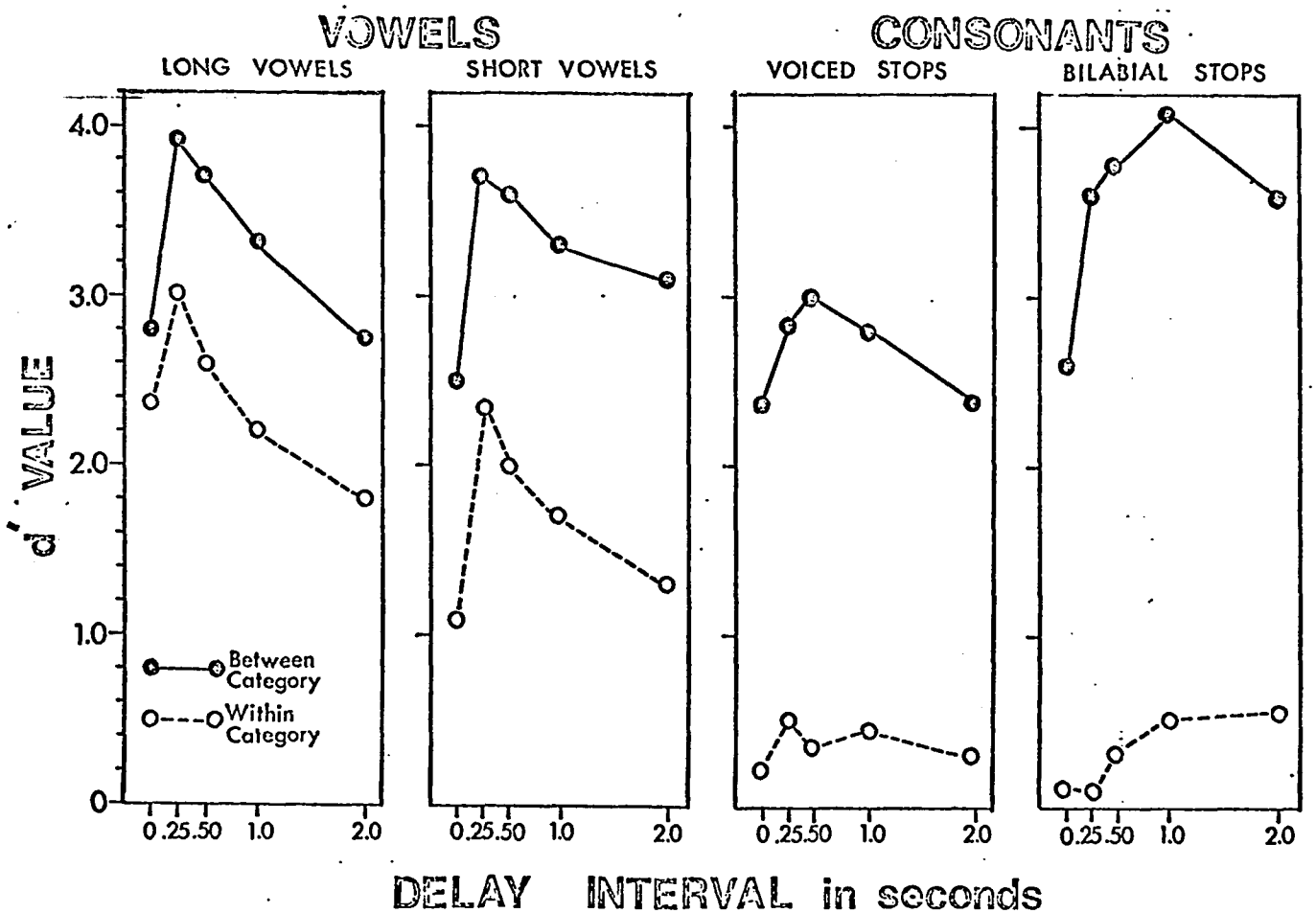


Figure 16. Average d' scores for within- and between-category comparisons as a function of delay interval for each stimulus condition.

Auditory Short-term Memory for Consonants and Vowels

The major finding of this experiment can be illustrated by a consideration of the two types of discrimination trials employed: within phoneme boundary comparisons and between phoneme boundary comparisons. It was argued at the outset that discrimination judgements for these comparisons may be considered to represent two types of memory components; short-term auditory non-phonetic memory and short-term phonetic memory, respectively. The results presented in Figure 15 based on the P("D"|D) scores indicate that only within-category vowel discrimination was related to changes in the comparison delay interval. The within-category consonant discrimination was not only quite poor overall, but it also showed no relationship to changes in the delay interval. These differences indicate that while auditory short-term memory facilitates vowel discrimination within categories, it contributes little to within-category consonant discrimination. On the other hand, the P("D"|D) scores for between-category discrimination for both vowels and consonants were quite high and appeared to be unaffected by changes in the delay interval. Phonetic memory appears to be quite good for both vowels and consonants.

The d' analysis which took account of the response bias inherent in the same-different task revealed several additional findings which were obscured by the P("D"|D) scoring analysis. First, there appears to be some tendency even for the between-category d' scores to decrease with increases in the delay interval. This was most noticeable for the long vowel condition. Secondly, the absence of a significant stimulus condition by delay interval interaction suggests that consonant discrimination is also affected to some extent by changes in the delay interval. The analysis of variance for the P("D"|D) scores showed this interaction to be significant. An inspection of the data revealed that the interaction was due entirely to the differences in the within-category vowel discrimination scores. The conclusion that the course of short-term memory is different for vowels than for consonants is, therefore, a premature over-simplification. The differences that obtain are entirely due to within-category pairs. For these pairs, large and consistent differences do obtain in both the P("D"|D) and d' analyses.

While it can be concluded from these results that auditory short-term memory for consonants is different from auditory short-term memory for vowels, some explanation for the differences is warranted. It is apparent that the acoustic information needed to discriminate two physically different but phonetically identical consonants is somehow not available for use in discrimination, even at very short delay intervals. However, it is not clear why this acoustic information is unavailable for use in discrimination.

One possible explanation was suggested by Fujisaki and Kawashima. Their explanation can be called the cue duration hypothesis. According to this hypothesis, the major factor responsible for the inferior auditory short-term memory with consonants is the duration of the critical information in the signal. The acoustic information in consonants is relatively short in duration and presumably cannot be stored well in auditory memory. The acoustic information in vowels extends the entire length of the stimulus. The results of Experiment 3 suggested that this explanation was sufficient to characterize the differences in ABX discrimination between 50 msec. short vowels and 300 msec. long vowels. The short vowels were perceived more categorically than the long vowels. However, if this hypothesis were correct, we would expect to find in the present experiment that the short vowels are more similar to the stop consonants than to the long steady-state vowels. The recognition memory data presented in this study argue against this prediction. The short vowels behave almost identically to the long vowels. Although discrimination is somewhat lower overall in the short vowel condition, the effect of the delay interval is still present.

As defined by the relationship between identification and discrimination functions, short vowels exhibit categorical perception. Nonetheless, this type of categorical perception may in fact be qualitatively different from that exhibited by the stop consonants. As pointed out previously, the use of the ABX procedure prevents a direct comparison between successive stimuli. The categorical perception observed with the short vowels in Experiment 3 may be attributed to the use of the ABX discrimination procedure. This procedure was also employed by Fujisaki in all of his experiments with short vowels.

Temporal Course of Recognition Memory in Speech Discrimination

Two additional findings deserve some comment. First, there is a noticeable decrease in discrimination performance at the 0.0 delay interval for both the P("D"|D) and d' scores. It is possible that this decrease represents an interruption of perceptual processing at an early stage of analysis. Using a backward masking paradigm, Massaro (1970) has shown that perceptual processing of a brief tone is terminated if a masking tone follows the test tone at very short delay intervals. Massaro suggests that the masking tone interrupted a readout of information from a preperceptual sensory store, which holds the image of a stimulus until the features needed for identification can be extracted. It is interesting to note that the interruption at 0.0 delay interval is more apparent for the within-category vowel comparisons than within-category consonant comparisons. Moreover, the interruption is less marked overall for the between-category comparisons than for the within-category comparisons. The direction is that anticipated if the subsequent signal terminated the processing of auditory features, assuming that categorical processing was more nearly completed.

Secondly, discrimination accuracy for within-category vowel comparisons reaches a maximum at the .25 second delay interval. Successive increases in the delay interval beyond .25 sec. produces a steady decline in the discrimination performance. This value may represent the processing time necessary for auditory recognition. If perceptual processing is completed within .25 seconds, then the acoustic information is still relatively salient for use in discrimination.

The most important difference between categorically and continuously perceived stimuli appears to rest on the level of within-category discrimination performance. Although the discrimination functions for vowels show a peak at the category boundary, the level of within-category discrimination is still well above chance. This suggests that categorization is not absolute and that purely auditory information is still available for use in discrimination judgements. Since prosodic information is carried almost entirely in vowels, it seems reasonable that auditory information would be available for use in discrimination.

However, the situation with respect to the consonants is peculiar. Although the discrimination functions for the consonants show comparable peaks at phonetic boundaries, the level of within-category discrimination is very close to chance indicating that categorization is absolute and binding. The extraction of relevant features from the signal for consonant recognition may preclude the further use of auditory information for non-phonetic judgements.

These results also bear on an important misunderstanding made by Lane (1965) in his attack on the Haskins work. By focusing most of his attention on the peaks in the discrimination functions, Lane neglected what appear to be the more important troughs (Studdert-Kennedy et al., 1970). The important question is why is discrimination so poor within categories for certain classes of speech sounds rather than why is discrimination so good across phonetic boundaries for all speech sounds.

In summary, the results of the present experiment suggest that the differences between consonant and vowel discrimination are primarily due to the differential availability of auditory short-term memory for the acoustic cues which distinguish these two classes of speech sounds. Vowel discrimination within-categories was considerably better than consonant discrimination within-categories. As a result, we can conclude that auditory short-term memory for the acoustic properties of vowels is better than auditory short-term memory for the acoustic properties of consonants.

CHAPTER V

SUMMARY AND DISCUSSION

In this chapter we shall begin with a review of the major findings of the six experiments dealing with categorical perception. The results of each of these experiments will then be discussed in reference to the three explanations offered for categorical perception in the introduction.

Summary of Experiments

The first experiment replicated the original Haskins experiments on the identification and discrimination of stop consonants and steady-state vowels. It was found that the description provided by the Haskins investigators was accurate in describing the differences in perception between consonants and vowels. The perception of stop consonants was found to be "nearly categorical" in the sense that listeners tend to discriminate pairs of stimuli only to the extent that they identify them as different. On the other hand, the perception of steady-state vowels was found to be "more nearly continuous" in the sense that listeners discriminate many more intraphonemic differences than they identify absolutely. The predicted discrimination functions derived from the absolute identification task matched the obtained consonant discrimination functions more closely than the obtained vowel discrimination functions. It was also observed that with a single group of Ss there were no significant differences in response consistency for identification of consonants and vowels.

The second experiment was an attempt to replicate the findings reported by Lane on the effects of discrimination training in producing categorical perception with non-speech stimuli. It was found that the general effect of the training procedure was to raise the overall level of obtained two-step discrimination above that predicted from the identification data. There was no consistent overall sharpening of discrimination at category boundaries, although such sharpening was observed in isolated cases. The results of this experiment were equivocal with respect to determining whether the procedures described by Lane are sufficient to produce categorical perception with non-speech stimuli.

The third experiment replicated the results reported by Fujisaki and Kawashima. It was found that steady-state isolated vowels tend to be perceived more categorically at brief stimulus durations. Vowel duration had little effect on absolute identification with two response alternatives; identification functions for both long vowels (300 msec.) and short vowels (50 msec.) were highly consistent and showed abrupt transitions between response categories. The contribution of auditory short-term memory to discrimination was evaluated according to the procedures developed by Fujisaki. The results of this experiment confirmed the predictions derived from Fujisaki's model which suggested that auditory short-term memory may be involved in the decision process in speech discrimination, but this explanation was at best only indirect.

The fourth experiment established that the ABX discrimination technique was less sensitive to the acoustic differences between vowel stimuli than at least one other discrimination testing procedure. It was argued that differences in discrimination performance between the ABX and the new 4IAX test of paired similarity were in part due to auditory short-term memory for the acoustic properties of the stimuli. Furthermore, it was suggested that categorical perception may be related to both the acoustic properties of the stimuli (vowel duration) and the experimental procedures that were used to assess discrimination in past speech perception experiments.

The fifth experiment compared the ABX and 4IAX discrimination procedures with a stop consonant continuum. It was found that the 4IAX technique was more sensitive to the acoustic differences between consonants than the traditional ABX procedure. However, the peaks and troughs characteristic of categorically perceived stimuli were still present in the 4IAX discrimination functions reaffirming the consistent differences between consonant and vowel perception.

The sixth experiment tested the hypothesis that consonants and vowels differ in the degree to which auditory short-term memory is employed in their discrimination. Using a delayed-comparison recognition memory task, it was found that accuracy of discrimination for vowels both within and between phoneme boundaries was related to the magnitude of the comparison interval. In contrast, discrimination of stop

consonants remained relatively stable both within and between boundaries. It was concluded that the differences in discrimination between consonants and vowels were primarily due to the differential availability of auditory short-term memory for the acoustic cues distinguishing these two classes of speech sounds. The results suggested that the major difference between categorically and continuously perceived stimuli appears to rest on the level of within-category discrimination rather than between-category discrimination.

Explanations of Categorical Perception

In the introductory chapter, three explanations of categorical perception were described. These were: 1) the Haskins position that the differences between categorical perception of consonants and the continuous perception of vowels were related to the differences in the production of these classes of speech sounds; 2) the Lane position that categorical perception results from discrimination training with verbal labels; and 3) the Fujisaki position that categorical perception is related to the degree to which auditory short-term memory is employed in the discrimination task.

The Haskins explanation. All of the experiments reported in the present investigation have examined the relationship between absolute identification and discrimination. Throughout these experiments, the procedures developed by the Haskins investigators were employed to assess the degree of categorical perception. The results of the "predictive test" which compared the obtained discrimination function to the predicted function based on absolute identification was found to be sufficient to describe the differences in perception between consonants and vowels. In general, no evidence has been found in the present investigation to suggest that the description of the differences in perception between vowels and consonants provided by the Haskins investigators was inadequate. The assumption that listeners can discriminate certain speech sounds only to the extent that they can identify them differently served as a strong basis for evaluating the effects of different experimental procedures.

There was one finding, however, which did not agree with the Haskins account of categorical perception and should be mentioned briefly. In

several of the Haskins experiments, it has been reported that the identification functions for vowels are less consistent than the identification functions for consonants. A clearly delineated identification function with sharp transitions between categories has been considered to be a necessary but not sufficient condition for a categorically perceived continuum. The results of the present investigation support the view that examination of identification functions alone is not sufficient to determine whether a continuum is perceived categorically or continuously. In Experiment 1, where the same group of Ss served in both vowel and consonant conditions, there was no significant difference in response consistency for vowels or consonants, even though the consonant data were slightly more consistent than the vowel data. This finding provides more direct support for a point recently raised by the Haskins investigators in their reply to Lane's claim about determining categorical perception:

> "...the degree to which perception is categorical or continuous can only be unequivocally determined from an examination of both identification and discrimination data." (Studdert-Kennedy et al., 1970, p. 256.)

The outcome of Experiment 1 is not at variance with the original explanation of categorical perception in terms of the articulatory differences in the production of consonants and vowels. However, the results of Experiment 3 suggest that the "strong motor theory" account of the differences in perception between consonants and vowels may no longer be tenable. Three findings may be offered to support this claim. First, vowels may be perceived categorically if stimulus duration is reduced. Second, the ABX procedure may be in part responsible for the observed categoricalness of vowel perception. Third, the range, spacing and number of categories may also be an important factor contributing to the categorical perception observed with vowels. It is important to stress that the categorical perception of vowels may be qualitatively different from that obtained with the stop consonants. None of the equivalent experimental procedures seems to have produced any convincing differences in the discrimination of consonants. This seems to be particularly obvious in the results of Experiment 6 where the delay interval had a strong effect on both classes of vowels but little on the consonants.

Phoneme boundary effects and categorical perception. Several investigators (i.e., Lane, 1965; Darwin, 1969; Haggard, 1969; 1970) have equated the phoneme boundary effect with categorical perception. The phoneme boundary effect refers to the relative differences in discrimination for within phoneme category and across phoneme category comparisons. Discrimination between two sounds which are identified as different phonemes is better than discrimination between two sounds which are individually identified as the same phoneme, even though the acoustic distance between the two sounds of each pair is comparable. The strongest phoneme boundary effects occur as we have seen, with a categorically perceived continuum such as stop consonants. In this case, discrimination of within-category differences is very close to chance and the obtained discrimination functions can be predicted accurately from absolute identification. However, the phoneme boundary effect also occurs for vowels, although discrimination of within-category differences is well above the level predicted from the identification functions. The results of the present investigation indicate that the important difference between categorical and continuous perception rests on the magnitude of the phoneme boundary effect, especially the level of within-category discrimination performance. For highly encoded speech sounds such as stop consonants, within-category discrimination is so poor as to suggest that information other than a binding phonetic categorization is unavailable to the listener for use in discrimination. In fact, several recent Haskins papers have suggested that a "special speech decoder" may be required to operate on only the encoded speech sounds. Thus, it is possible that the unavailability of acoustic information about the stop consonants may be a consequence of processing by this specialized device.

The Lane explanation. As noted in Experiment 2, the explanation of categorical perception offered by Lane in terms of discrimination training has been the subject of some recent controversy. Since the results of this experiment were ambiguous with respect to either position, neither conclusion about the effects of discrimination training on categorical perception finds strong support here. This appears to be a fair assessment of the controversy since the judicious selection

of Ss in the present experiment can be made to support either the Lane position or the Haskins position. Haggard (1970) has suggested that one of the difficulties in resolving the Lane-Haskins disagreement stems from a mutual failure "to enumerate levels or aspects of the perceptual process and make separate statements about them" (p. 6). The overall results suggest that if training does produce categorical perception, it is certainly not as rapid or easy to obtain as Lane has claimed. Only future research can provide additional information for the resolution of this problem.

The Fujisaki explanation. Experiment 3 considered an explanation of categorical perception in terms of a model of the auditory and phonetic processes involved in speech perception. This model was proposed by Fujisaki and Kawashima who suggested that the discrepancy in discrimination between the obtained ABX scores and those predicted from the standard Haskins labelling formula were due to a component they called auditory memory for timbre information. To compensate for the discrepancy in discrimination between obtained and predicted scores, they developed a new prediction formula which included both the contribution due to labelling and a contribution due to auditory memory. The memory component in the new predicted formula was estimated from the asymptotic value of the obtained discrimination function and then employed in the predicted discrimination formula. The procedure introduced by Fujisaki and Kawashima substantially improved the fit of the obtained and predicted discrimination functions and suggested that auditory short-term memory may be involved in some way in the decision processes in speech discrimination tasks. The outcome of Experiment 3 indicated that the procedures developed by Fujisaki and Kawashima were basically correct. Prediction of the obtained discrimination function can be improved by accounting for some component in the prediction formula. These workers chose to identify this component as "auditory memory," but, in absence of additional tests, this must remain a weak inference. Moreover, in replicating the results of the Fujisaki experiments, it was found that vowels of short duration are perceived more categorically than vowels of longer duration. This finding suggested that categorical perception may be related to the characteristics of the stimuli and the degree to which auditory short-term memory is employed in discrimination.

A stronger basis for the inference with respect to the role of auditory short-term memory in speech discrimination was obtained in the final experiment where experimental operations previously related to memory were employed. Vowel discrimination was found to be better than consonant discrimination because auditory short-term memory for the acoustic properties of vowels was more readily available for use in discrimination while auditory short-term memory for the acoustic properties of consonants was less readily available for use in discrimination.

In final summary, the perception of consonants and vowels is related to the short-term memory requirements associated with the procedure used to assess discrimination. The basis for between-category discrimination, which is based on phonetic classification, appears to be equivalent for the two classes of speech sounds. Only for within-category discrimination, which is based on auditory memory for the acoustic characteristics of the stimuli, are the two classes of speech sounds differentiated. Acoustic information about stop consonants is unavailable for use in discrimination of within-category comparisons. On the other hand, acoustic information about the vowels may be available but this is related to task requirements. It may be concluded that the differences between categorical phonetic perception and continuous auditory perception are related to the degree to which auditory short-term memory is employed in speech discrimination.

REFERENCES

Bever, T. G. The nature of cerebral dominance in speech behavior of the child and adult. In R. Huxley and E. Ingram (Eds.), Mechanisms of language development. London: Academic Press, 1970. In press.

Broadbent, D. E. and Gregory, M. Accuracy of recognition for speech presented to the right and left ears. Quarterly Journal of Experimental Psychology, 1964, 16, 359-360.

Coombs, C. H. A theory of data. New York: John Wiley, 1964.

Cooper, F. S. Spectrum analysis. Journal of the Acoustical Society of America, 1950, 22, 761-762.

Cooper, F. S., Delattre, P. C., Liberman, A. M., Borst, J. M. and Gerstman, L. J. Some experiments on the perception of synthetic speech sounds. Journal of the Acoustical Society of America, 1952, 24, 597-606.

Cooper, F. S., Liberman, A. M., and Borst, J. M. The interconversion of audible and visible patterns as a basis for research in the perception of speech. Proceedings of the National Academy of Sciences, 1951, 37, 318-325.

Cooper, F. S. and Mattingly, I. G. Computer-controlled PCM system for investigation of dichotic speech perception. Status Report on Speech Research (SR-17/18), Haskins Laboratories, New York, 1969, 17-21.

Cross, D. V. and Lane, H. L. On the discriminative control of concurrent responses: The relations among response frequency, latency and topography in auditory generalization. Journal of the Experimental Analysis of Behavior, 1962, 5, 487-496.

Cross, D. V. and Lane, H. L. An analysis of the relations between identification and discrimination functions for speech and non-speech continua. Progress Report No. 6, Behavior Analysis Laboratory, University of Michigan, Ann Arbor, 1964.

Darwin, C. J. Auditory perception and cerebral dominance. Unpublished doctoral dissertation, University of Cambridge, 1969.

Darwin, C. J. Ear differences in the recall of fricatives and vowels. Quarterly Journal of Experimental Psychology, 1971, 23, 46-62.

Delattre, P. C. Les indices acoustique de la parole: Premier rapport. Phonetica, 1958, 2, 108-118, 210-251.

Eimas, P. D. A study of the relation between absolute identification and discrimination along selected sensory continua. Unpublished doctoral dissertation, University of Connecticut, 1962.

Fant, G. Acoustic theory of speech production. The Hague: Mouton, 1960.

Flanagan, J. L. Speech analysis, synthesis and perception. New York: Academic Press, 1965.

Fourcin, A. J. Speech source inference. IEEE Transactions on Audio and Electroacoustics, 1968, 16, 65-67.

Fourcin, A. J. Perceptual mechanisms at the first level of speech processing. Proceedings of the 7th International Congress of Phonetic Sciences, Montreal, August 1971.

Fry, D. B., Abramson, A. S., Eimas, P. D. and Liberman, A. M. The identification and discrimination of synthetic vowels. Language and Speech, 1962, 5, 4, 171-189.

Fujisaki, H. and Kawashima, T. The influence of various factors on the identification and discrimination of synthetic speech sounds. Reports of the 6th International Congress on Acoustics, Tokyo, August 1968.

Fujisaki, H. and Kawashima, T. On the modes and mechanisms of speech perception. Annual Report of the Engineering Research Institute. Vol. 28, Faculty of Engineering, University of Tokyo, Tokyo, 1969, 67-73.

Fujisaki, H. and Kawashima, T. Some experiments on speech perception and a model for the perceptual mechanism. Annual Report of the Engineering Research Institute, Vol. 29, Faculty of Engineering, University of Tokyo, Tokyo, 1970, 207-214.

Gerstman, L. J. Perceptual dimensions for the friction portions of certain speech sounds. Unpublished doctoral dissertation, New York University, 1957.

Gerstman, L. J. Classification of self-normalized vowels. IEEE Transactions on Audio and Electroacoustics, 1968, 16, 1, 78-80.

Green, D. M. Psychoacoustics and detection theory. Journal of the Acoustical Society of America, 1960, 32, 1189-1203.

Haggard, M. P. Perception of semivowels and laterals. Journal of the Acoustical Society of America, 1969, 46, 1, 115.

Haggard, M. P. Theoretical issues in speech perception. Speech Synthesis and Perception, Report No. 4, Psychological Laboratory, Cambridge University, Cambridge, England, 1970.

Haggard, M. P. Encoding and the REA for speech signals. Quarterly Journal of Experimental Psychology, 1971, 23, 34-45.

Harris, J. D. Remarks on the determination of a differential threshold by the so-called ABX technique. Journal of the Acoustical Society of America, 1952, 24, 417.

Hays, W. L. Statistics for psychologists. New York: Holt, Rinehart and Winston, 1963.

Ingemann, F. Rules for synthesizing American English on the pattern playback. Unpublished paper, Haskins Laboratories, New York, September 1957.

Kimura, D. Some effects of temporal lobe damage on auditory perception. Canadian Journal of Psychology, 1961, 15, 156-165. (a)

Kimura, D. Cerebral dominance and the perception of verbal stimuli. Canadian Journal of Psychology, 1961, 15, 166-171. (b)

Kimura, D. Left-right differences in the perception of melodies. Quarterly Journal of Experimental Psychology, 1964, 16, 355-358.

Koenig, W., Dunn, H. K. and Lacy, L. Y. The sound spectrograph. Journal of the Acoustical Society of America, 1946, 17, 19-49.

Lane, H. L. Psychophysical parameters of vowel perception. Psychological Monographs, 1962, 76(44, Whole No. 563).

Lane, H. L. The motor theory of speech perception: A critical review. Progress Report No. 4, Behavior Analysis Laboratory, University of Michigan, Ann Arbor, 1963.

Lane, H. L. The motor theory of speech perception: A critical review. Psychological Review, 1965, 72, 275-309.

Lane, H. L. A behavioral basis for the polarity principle in linguistics. In K. Salzinger and S. Salzinger (Eds.), Research in verbal behavior and some neurophysiological implications. New York: Academic Press, 1967. Pp. 79-96.

Lane, H. L. and Schneider, B. A. Discriminative control of concurrent responses by the intensity, duration and relative onset time of auditory stimuli. Unpublished report, Behavior Analysis Laboratory, University of Michigan, Ann Arbor, 1963.

Liberman, A. M. Some results of research on speech perception. Journal of the Acoustical Society of America, 1957, 29, 117-123.

Liberman, A. M. Some characteristics of perception in the speech mode. In D. A. Hamburg (Ed.), Perception and Its Disorders, Proceedings of A. R. N. M. D. Baltimore: Williams and Wilkins Co., 1970. Pp. 238-254.

Liberman, A. M., Cooper, F. S., Harris, K. S. and MacNeilage, P. F. A motor theory of speech perception. In C. G. M. Fant (Ed.), Proceedings of the Speech Communication Seminar, Stockholm, 1962. Speech Transmission Laboratory, Royal Institute of Technology, Stockholm, 1963.

Liberman, A. M., Cooper, F. S., Shankweiler, D. and Studdert-Kennedy, M. Perception of the speech code. Psychological Review, 1967, 74, 431-461.

Liberman, A. M., Delattre, P. C. and Cooper, F. S. The role of selected stimulus variables in the perception of the unvoiced stop consonants. American Journal of Psychology, 1952, 65, 497-516.

Liberman, A. M., Harris, K. S., Eimas, P., Lisker, L., and Bastian, J. An effect of learning on speech perception: The discrimination of durations of silence with and without phonemic significance. Language and Speech, 1961, 4, 175-195. (a)

Liberman, A. M., Harris, K. S., Hoffman, H. S., and Griffith, B. C. The discrimination of speech sounds within and across phoneme boundaries. Journal of Experimental Psychology, 1957, 54, 358-368.

Liberman, A. M., Harris, K. S., Kinney, J. A., and Lane, H. L. The discrimination of relative onset time of the components of certain speech and non-speech patterns. Journal of Experimental Psychology, 1961, 61, 379-388. (b)

Licklider, J. C. R. On the process of speech perception. Journal of the Acoustical Society of America, 1952, 24, 6, 590-594.

Lieberman, P. Towards a unified phonetic theory. Linguistic Inquiry, 1970, 1, 3, 307-322.

Lindblom, B. E. F. and Studdert-Kennedy, M. On the role of formant transitions in vowel recognition. Journal of the Acoustical Society of America, 1967, 42, 4, 830-843.

Lisker, L. and Abramson, A. S. Some experiments in comparative phonetics. Proceedings of the 6th International Congress of Phonetic Sciences, Prague, September 1967.

Lisker, L., Cooper, F. S. and Liberman, A. M. The uses of experiment in language description. Word, 1962, 18, 82-106.

Mattingly, I. G. Synthesis by rule of general American English. Supplement to: Status Report on Speech Research, Haskins Laboratories, New York, 1968.

Mattingly, I. G., Liberman, A. M., Syrdal, A. K. and Halwes, T. Discrimination in speech and non-speech modes. Cognitive Psychology, 1971, 2, 2, 131-157.

Massaro, D. W. Retroactive interference in short-term recognition memory for pitch. Journal of Experimental Psychology, 1970, 83, 1, 32-39.

Massaro, D. W. The role of preperceptual images, perceptual units, and processing time in auditory recognition. Studies in Human Information Processing, Report 71-2, Wisconsin Mathematical Psychology Program, University of Wisconsin, Madison, 1971.

McNemar, Q. Psychological statistics. (4th ed.) New York: John Wiley and Sons, 1969.

Parks, T., Wall, C. and Bastian, J. Intercategory and intracategory discrimination for one visual continuum. Journal of Experimental Psychology, 1969, 81, 241-245.

Peterson, G. E. Basic problems in automatic speech recognition. In C. G. M. Fant (Ed.), Proceedings of the Speech Communication Seminar, Stockholm, 1962. Speech Transmission Laboratory, Royal Institute of Technology, Stockholm, 1963.

Pollack, I. The information in elementary auditory displays. Journal of the Acoustical Society of America, 1952, 24, 745-749.

Pollack, I. The information in elementary auditory displays II. Journal of the Acoustical Society of America, 1953, 25, 765-769.

Pollack, I. and Ficks, L. Information of elementary multidimensional auditory displays. Journal of the Acoustical Society of America, 1954, 26, 155-158.

Pollack, I. and Pisoni, D. B. On the comparison between identification and discrimination tests in speech perception. Psychonomic Science, 1971. In press.

Popper, R. D. Linguistic determinism and the perception of synthetic voiced stops. (Doctoral dissertation, University of California, Los Angeles) Ann Arbor, Michigan: University Microfilms, 1967. No. 68-231.

Potter, R. K., Kopp, G. A. and Green, H. C. Visible speech. New York: van Nostrand, 1947.

Reddy, D. R. Computer recognition of connected speech. Journal of the Acoustical Society of America, 1967, 42, 2, 329-347. (a)

Reddy, D. R. Phoneme grouping for speech recognition. Journal of the Acoustical Society of America, 1967, 41, 5, 1295-1300. (b)

Rosenblith, W. A. and Stevens, K. N. On the DL for frequency. Journal of the Acoustical Society of America, 1953, 25, 5, 980-985.

Sachs, R. M. Vowel identification and discrimination in isolation vs. word context. Quarterly Progress Report No. 93, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts, 1969, 220-229.

Shankweiler, D. P. Effects of temporal-lobe damage on perception of dichotically presented melodies. Journal of Comparative and Physiological Psychology, 1966, 62, 115-119.

Shankweiler, D. P. An analysis of laterality effects in speech perception. In D. L. Horton and J. J. Jenkins (Eds.), The perception of language. Columbus, Ohio: Charles E. Merrill, 1970.

Shankweiler, D. P. and Studdert-Kennedy, M. Identification of consonants and vowels presented to left and right ears. Quarterly Journal of Experimental Psychology, 1967, 19, 59-63.

Sperling, G. The information available in brief visual displays. Psychological Monographs, 1960, 74 (11, Whole No. 498).

Stevens, K. N. On the relations between speech movements and speech perception. Zeitschrift fur Phonetik, Sprachwissenschaft und Kommunikationsforschung, 1968, 21, 102-106.

Stevens, K. N. The quantal nature of speech: Evidence from articulatory-acoustic data. In E. E. David, Jr. and P. B. Denes (Eds.), Human communication: A unified view. New York: McGraw-Hill, 1969. In press.

Stevens, K. N. and House, A. S. Speech perception. In J. Tobias (Ed.), Foundations of modern auditory theory. New York: Academic Press. In press.

Stevens, K. N., Liberman, A. M., Studdert-Kennedy, M. and Ohman, S. E. G. Cross-language study of vowel perception. Language and Speech, 1969, 12, 1-23.

Studdert-Kennedy, M. The perception of speech. In T. A. Sebeok (Ed.), Current trends in linguistics, Vol. XII. The Hague: Mouton. In press.

Studdert-Kennedy, M. and Cooper, F. S. High-performance reading machines for the blind; psychological problems, technological problems, and status. Paper presented at the meeting of St. Dunstan's International Conference on Sensory Devices for the Blind, London, June 1966.

Studdert-Kennedy, M. and Liberman, A. M. Psychological considerations in the design of auditory displays for reading machines. Proceedings of the International Congress on Technology and Blindness, 1963. Pp. 289-304.

Studdert-Kennedy, M., Liberman, A. M., Harris, K. and Cooper, F. S. The motor theory of speech perception: A reply to Lane's critical review. Psychological Review, 1970, 77, 3, 234-249.

Studdert-Kennedy, M. and Shankweiler, D. P. Hemispheric specialization for speech perception. Journal of the Acoustical Society of America, 1970, 48, 2, 579-594.

Swets, J. A. Is there a sensory threshold? Science, 1961, 134, 168-177.

Wickelgren, W. A. Consolidation and retroactive interference in short-term recognition memory for pitch. Journal of Experimental Psychology, 1966, 72, 250-259.

Winer, B. J. Statistical principles in experimental design. New York: McGraw-Hill, 1962.

APPENDIX A

Description of the Haskins Laboratories Computer Controlled Speech Synthesis System and Pulse Code Modulation (PCM) System

APPENDIX A

Description of the Haskins Laboratories Computer Controlled Speech Synthesis System and Pulse Code Modulation (PCM) System

Speech Synthesis System

The speech synthesis system currently in use at Haskins Laboratories consists of a parallel resonance synthesizer and a Honeywell DDP-224 digital computer. Various peripheral input, storage and output devices are linked to the computer for increased speed and flexibility in producing experimental tapes. A program for controlling the synthesizer was written and developed by Ignatius Mattingly (1968). The executive program serves as the major software tool for the production and manipulation of synthetic utterances. The two sections below based on Mattingly (1968) briefly describe the operation of the synthesizer and the executive program. The interested reader is advised to consult Mattingly (1968) for additional details on the development and operation of the Haskins speech synthesis system.

The synthesizer. The synthesizer was designed by Franklin S. Cooper. The resonant circuits have been designed in parallel. Figure A1 presents a block diagram of the Haskins speech synthesizer. The first four formants of the vocal tract are represented by the resonant circuits 1, 2, 3 and 4. The frequency of resonance 4 is fixed at 3600 Hz. Each of the first three resonances can be varied through a limited range (F_1 = 100 - 894 Hz, F_2 = 543 - 2910 Hz and F_3 = 1190 - 3698 Hz). The band widths of the four resonances are 60 Hz, 90 Hz, 120 Hz and 300 Hz and were designed to remain constant across the whole range of each formant. In addition to the resonant circuits, there are three high pass filters which shape high frequency noise at 4000 Hz, 2400 Hz and 1000 Hz appropriate for fricative sounds and stop consonant bursts.

Vocal cord excitation is simulated by a buzz generator which can be varied from 0 Hz to 1000 Hz. Turbulent excitation is simulated by a hiss (random noise) generator. If the buzz switch is closed, a pulse generator excites resonances 1, 2, 3 and 4. If hiss switch 1 is closed, low-level noise, intended to increase the naturalness of the synthetic speech, excites resonances 2, 3 and 4. If hiss switch 2 is closed,

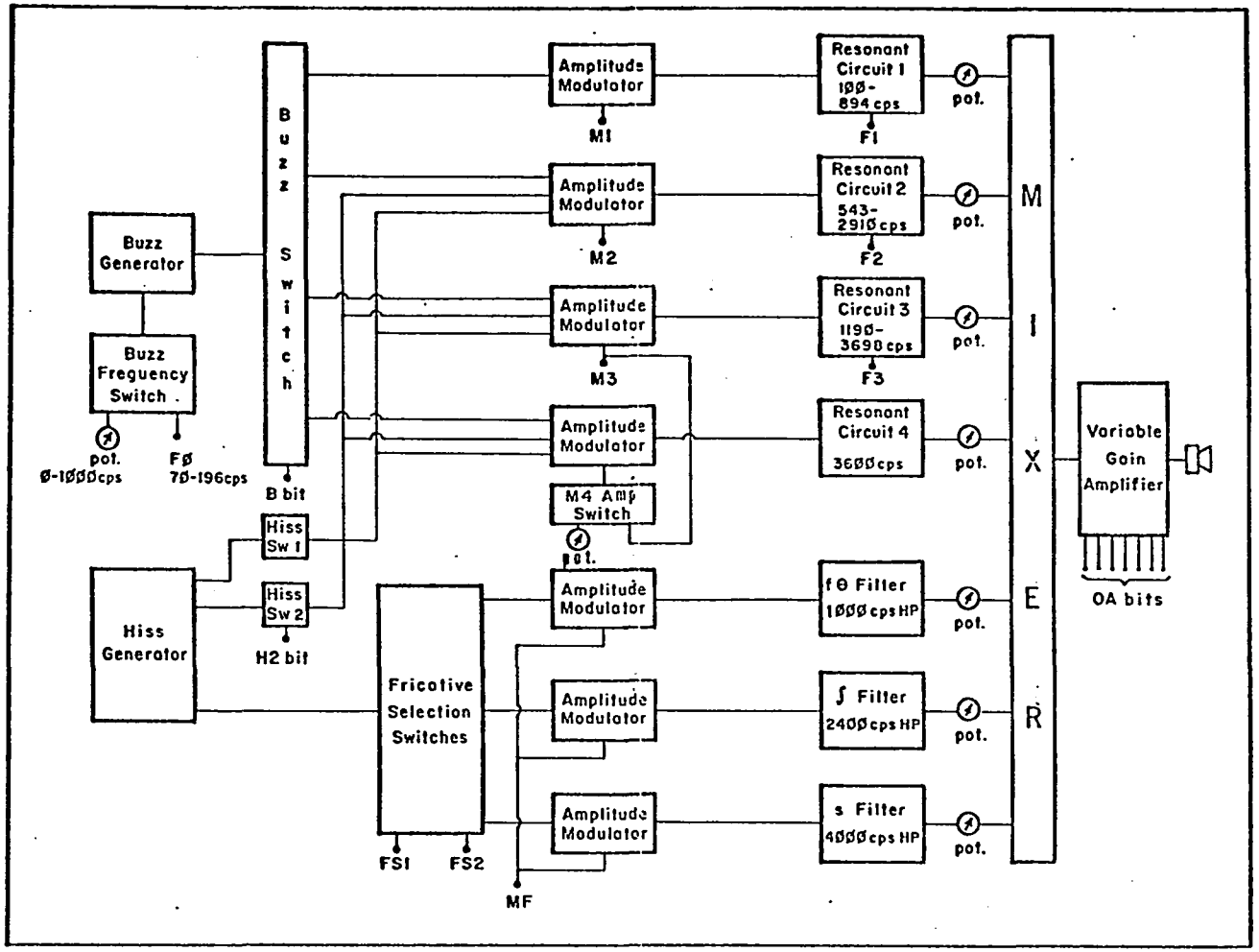


Figure A1. Block diagram of the Haskins Resonance Synthesizer (from Mattingly, 1968).

higher level noise excites the same resonances. When any of these switches is open the corresponding excitation source is silent. By setting the buzz and hiss selectors appropriately, it is possible to excite the higher resonances with any combination of buzz or hiss. The hiss generator can also excite one of the three high-frequency filters, although this depends on the particular settings of the fricative selection switch.

The amplitude of the excitation of each of the resonances and filters is regulated by a modulator. The maximum amplitude of the output of each of these circuits is in turn regulated by a potentiometer. All outputs are added together in a mixer. The overall amplitude of the resulting signal is further regulated by a variable gain amplifier before being converted into sound.

The synthesizer is controlled by means of two digital control registers in the DDP-224 which contain digital values for the various parameters. Each of the digital control registers contains a 24-bit computer control word which contains coded values for each of the parameters. The synthesizer is linked to the computer by a 24-bit buffer storage unit. The synthesis program transfers control words to the synthesizer.

The synthesis program assumes that the utterance to be synthesized is represented in memory by a series of pairs of 24-bit control words. The first and then the second word in each pair is transferred from memory to a buffer storage unit and then to the synthesizer. After one pair of control words has been transmitted to the synthesizer, the computer waits several milliseconds before transmitting another pair. During the waiting period, the synthesizer is controlled by the parameter value set contained in the recent pair of control words. After the entire series of control words has been transmitted, the computer pauses and then begins synthesis again. An utterance is synthesized repeatedly until the experimenter intervenes by executing a new control command at the computer console.

The executive program. The executive program which was written by Mattingly (1968) consists of a series of input, output and editing routines which are used to provide control information to the synthesizer.

The experimenter specifies a given routine by typing one of a set of mnemonic commands on a Sanders CRT unit. When a command has been executed the CRT unit displays "NEXT" which indicates that the computer is ready for another command. Several of these commands will be described below.

The command "SAVE" takes the parameter control data currently in core and stores it on a removable magnetic disk pack. The control data are transferred to a specific utterance location on this disk. The command "PULL" retrieves parameter information previously stored in a location on disk and places it back into core.

An extended series of utterances may be synthesized by means of two commands. Utterance storage locations and appropriate inter-stimulus intervals are specified with a "TYSQ" command. The command "SEQU" executes a previous TYSQ specification by retrieving the control parameter data in the disk locations and synthesizing the corresponding utterances with the appropriate time intervals.

An additional set of routines are available for editing the synthesis control parameters which have been stored in core. The command "TERP" provides either a linear or non-linear interpolation between two specified values of a given parameter. Each interpolated value is calculated as a binary number and is inserted in the proper control word segment.

Pulse Code Modulation System (PCM)

The purpose of the PCM system at Haskins Laboratories is to enable an experimenter to digitize a short utterance, edit the digitized utterance and store the wave form on a magnetic disk for tape preparation. The PCM system consists of the Honeywell DDP-224 computer connected with a Control Data Company disc file, an analog-to-digital (A-D) converter and two digital-to-analog (D-A) converters, a storage oscilloscope and an Ampex AG-500 tape recorder.

The sections below briefly describe the operation of the PCM system. This system was designed by Franklin S. Cooper and Ignatius Mattingly and the interested reader is advised to consult a recent paper by these authors (Cooper and Mattingly, 1969) for additional details on the operation of this sytem at Haskins Laboratories.

Input. The input to the PCM system consists of an utterance which has been recorded on audio tape. The maximum length that can be read in is 3.6 seconds although only half of this (i.e., 1.8 seconds) may be stored in digital form on the disk file. Utterances are recorded on one channel of an audio tape and a 10 KHz tone is recorded on the other channel of the tape beginning slightly before the utterance. When the audio tape with the 10 KHz tone on channel 2 passes under the tape recorder heads, the PCM unit begins digitizing at a rate of 8000 samples per second and continues until 14,592 samples have been stored in buffer (i.e., 1.824 seconds).

Editing. The editing routines are employed to adjust the overall amplitude and duration of the utterance. The wave form is displayed on a storage scope and the experimenter can edit the beginning (head edit) or end (tail edit) of the utterance to desired specifications. The overall amplitude of the utterance may be adjusted in 1 dB steps over a 7 dB range in order to insure that utterances may be equated in terms of intensity. Four letter commands are used for automatic editing under computer control. After the wave form of an utterance has been edited, it may be stored in a disk file for later retrieval in preparing test tapes.

Output. The output of the PCM system consists of two D-A channels which are connected to the input of the tape recorder. The experimenter can record either channel independently or can record both channels simultaneously. The latter procedure has been used for the preparation of tapes for dichotic listening experiments.

A test order is declared by specifying appropriate inter-stimulus intervals and stimulus location numbers. A whole test sequence may be specified and then stored on the disk file. After the test order specification has been entered, an experimental tape may be produced by retrieving the test order and executing its contents. All timing intervals are placed automatically under control of the computer's internal clock. The experimenter places a fresh magnetic tape on the recorder, retrieves the desired test sequence order from disk file, starts the tape recorder and then executes the particular sequence.

When the sequence has been properly executed, the computer provides a message on the display scope that indicates that the sequence has been executed properly.

Digitized materials (stimuli and test orders) that are on disk may also be transferred to magnetic tape for storage and then later placed back on disk for additional test order preparations.

APPENDIX B

Identification and Discrimination Functions for Stop Consonants and Steady-state Vowels for Individual Ss in Experiment 1

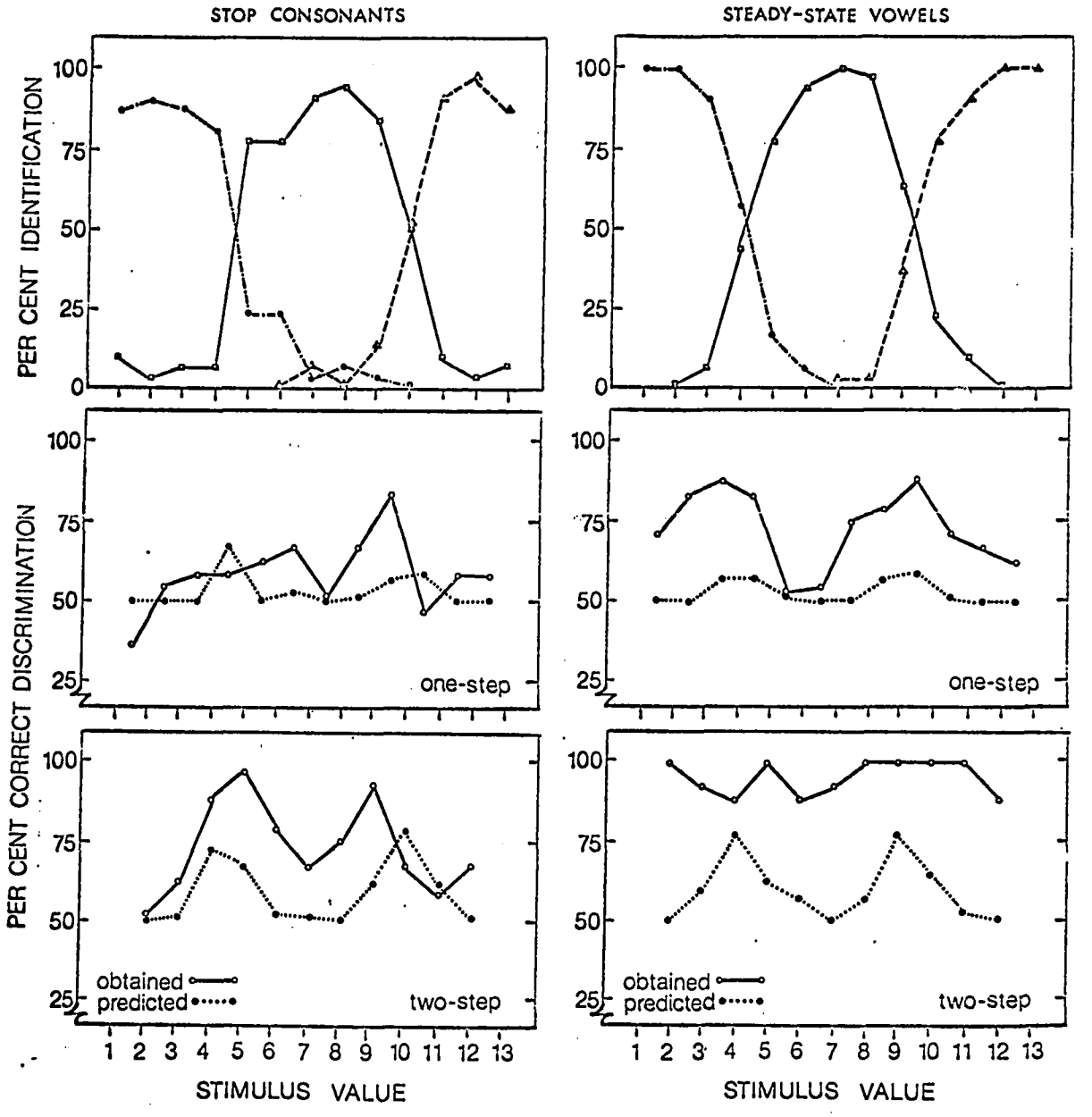


Figure B1. Identification and discrimination functions for stop consonants and steady-state vowels for S 1.

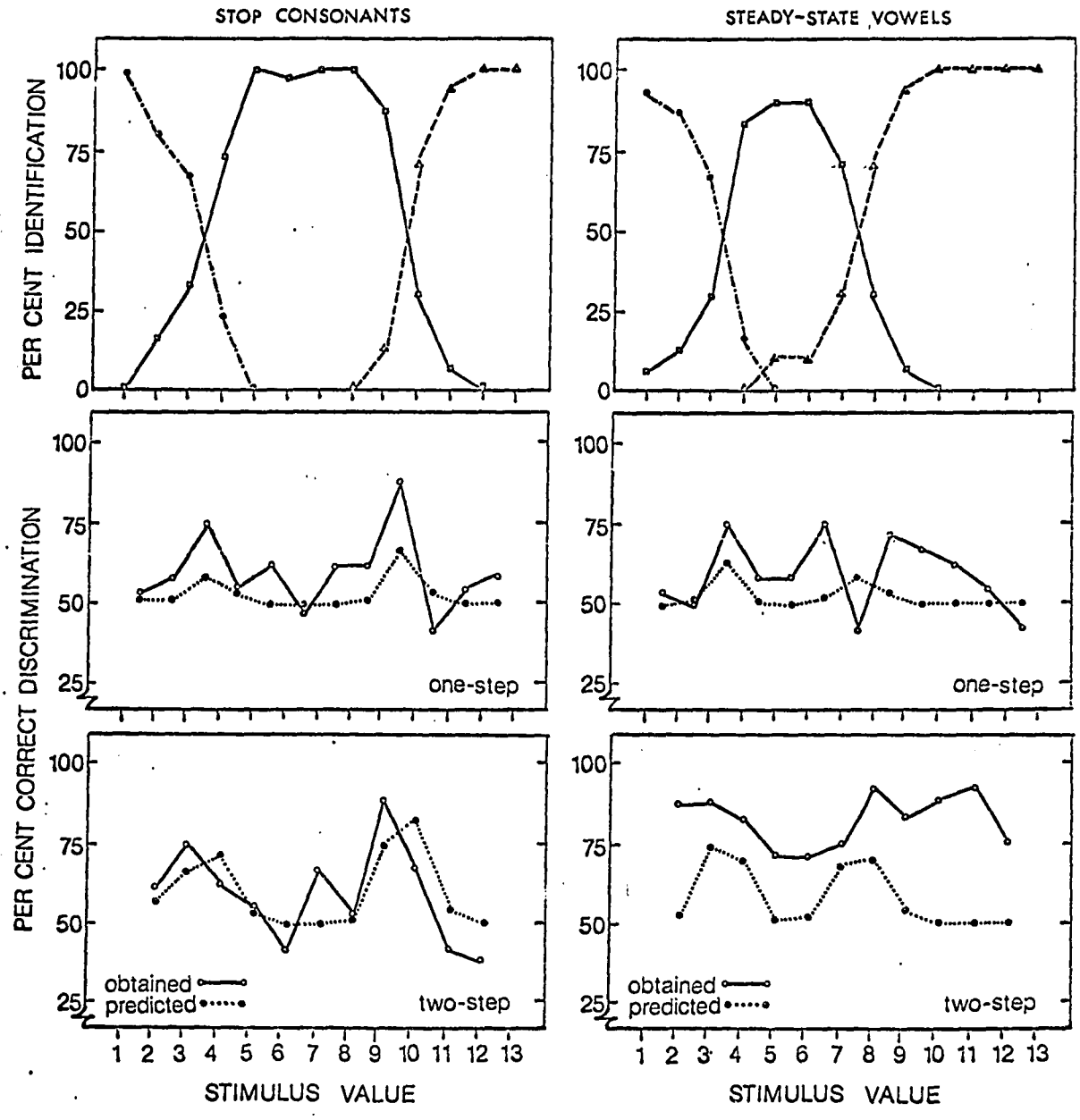


Figure B2. Identification and discrimination functions for stop consonants and steady-state vowels for S 2.

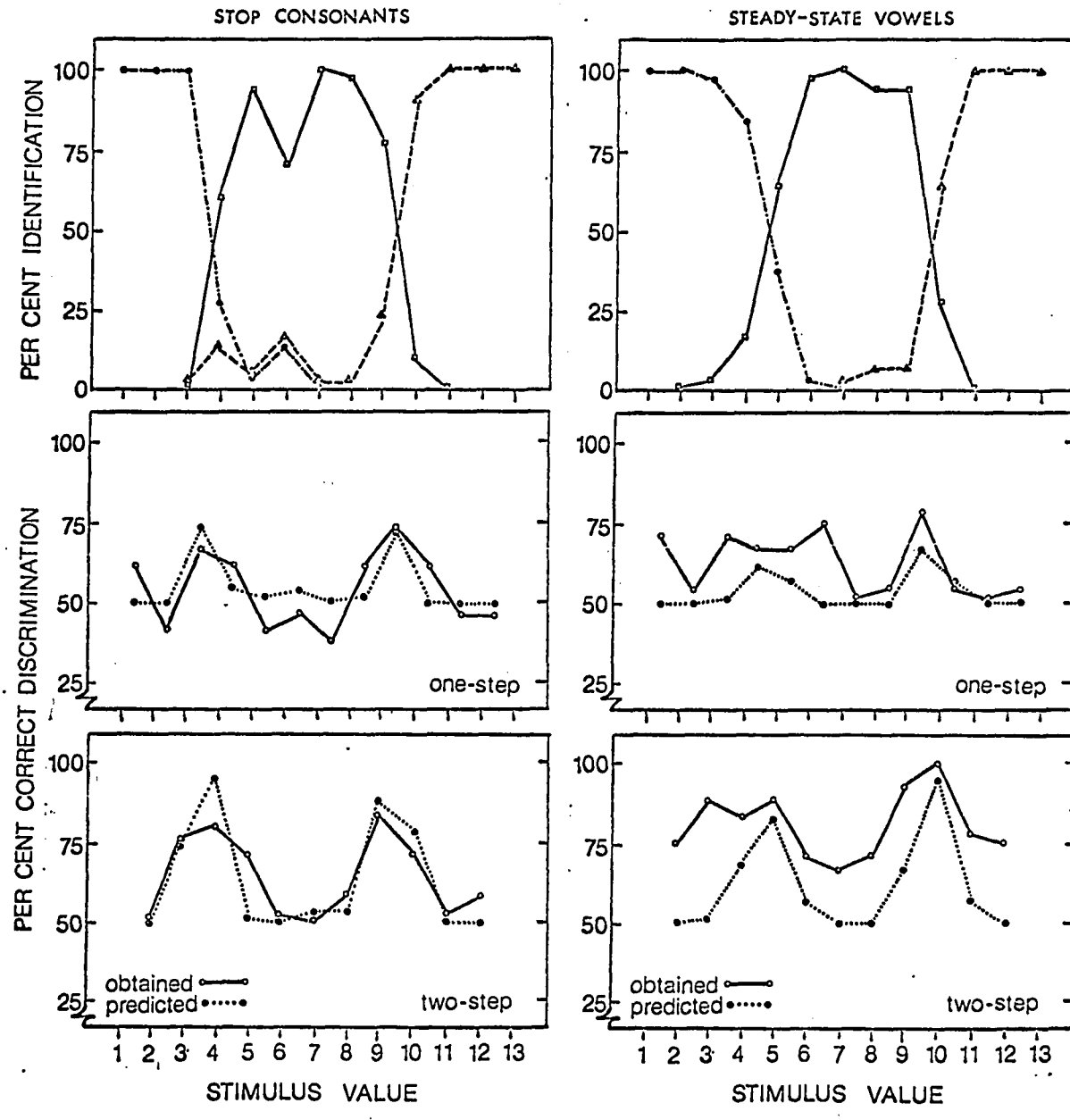


Figure B3. Identification and discrimination functions for stop consonants and steady-state vowels for S 4.

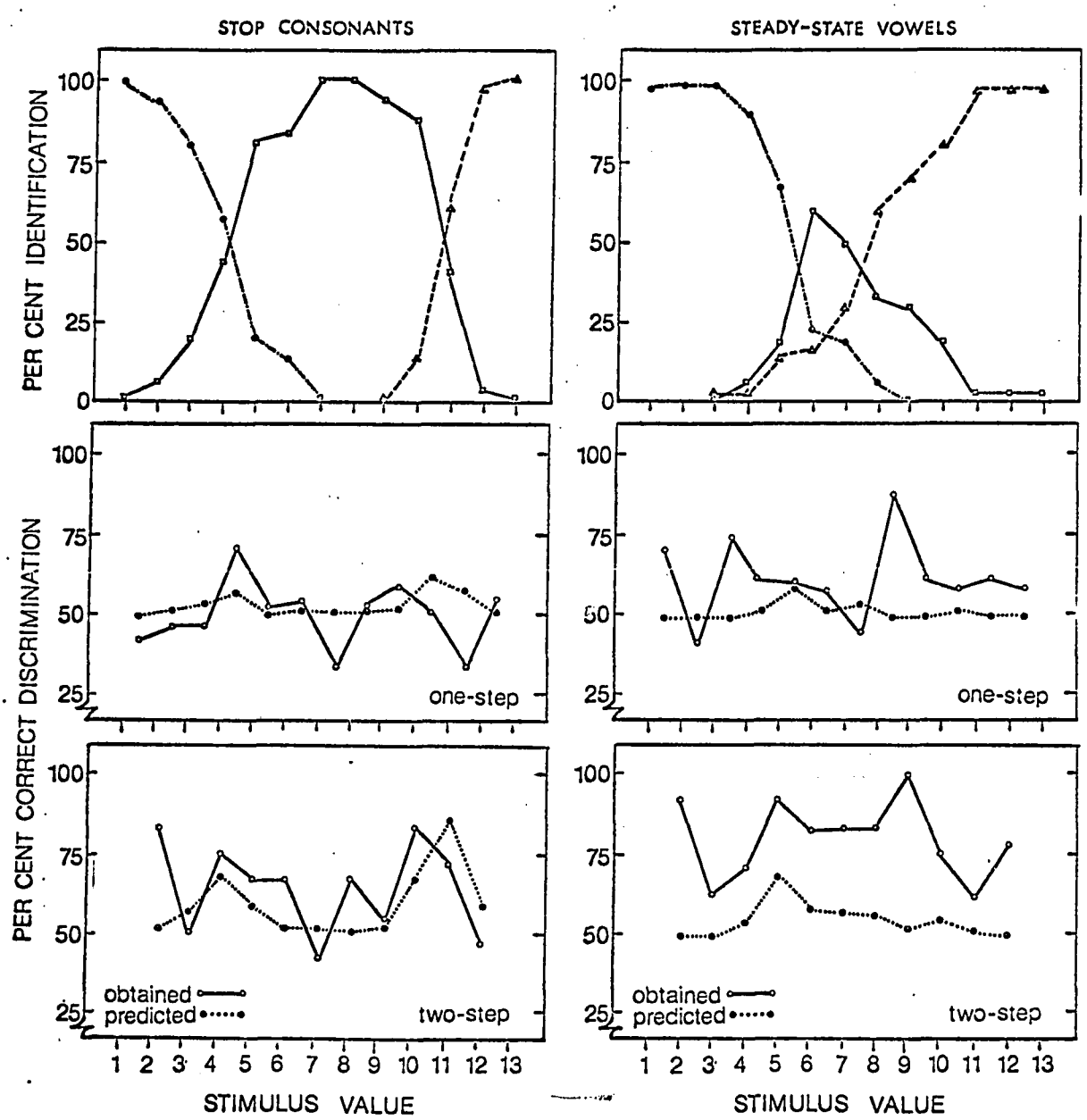


Figure B4. Identification and discrimination functions for stop consonants and steady-state vowels for S 5.

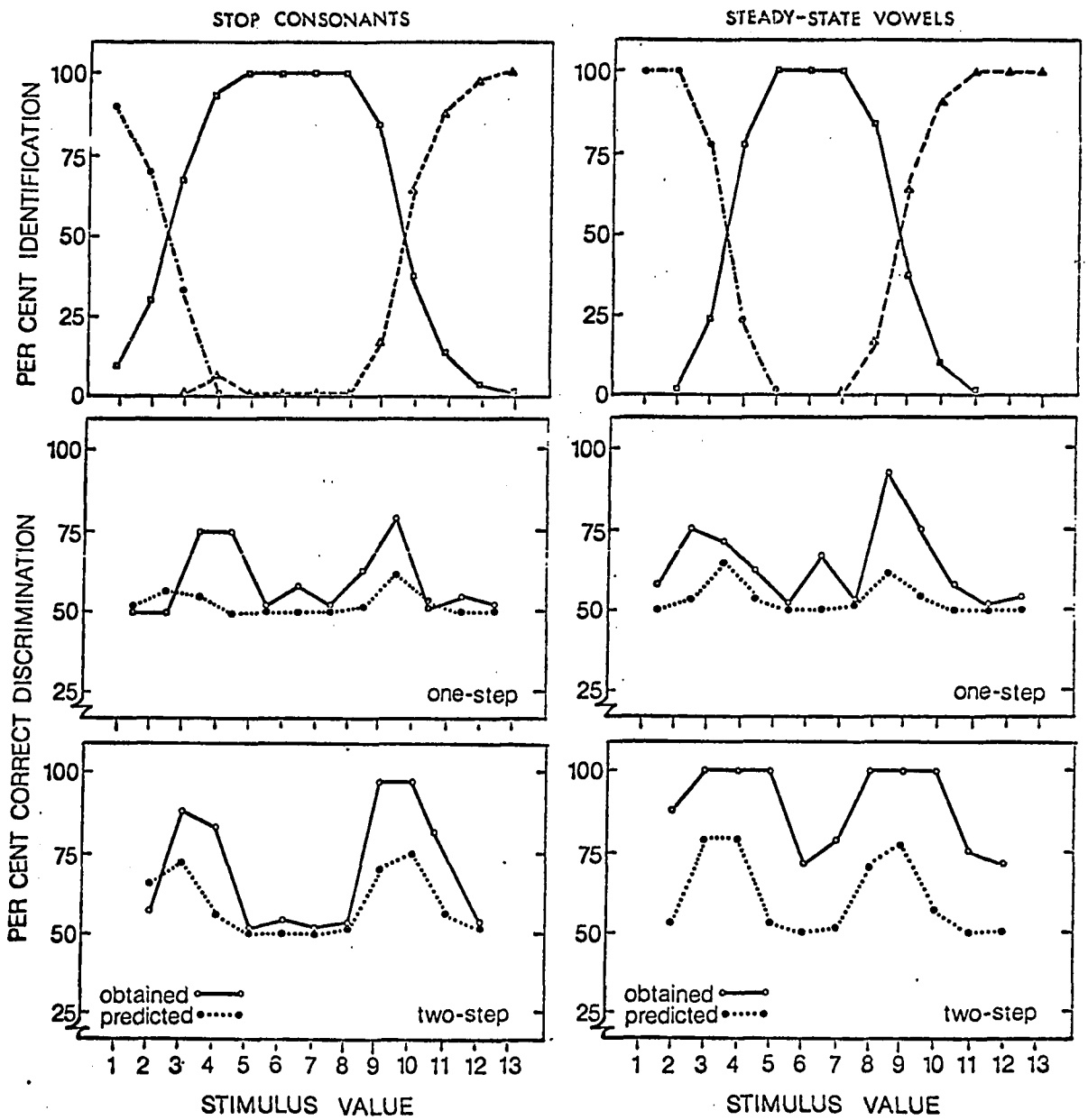


Figure B5. Identification and discrimination functions for stop consonants and steady-state vowels for S 6.

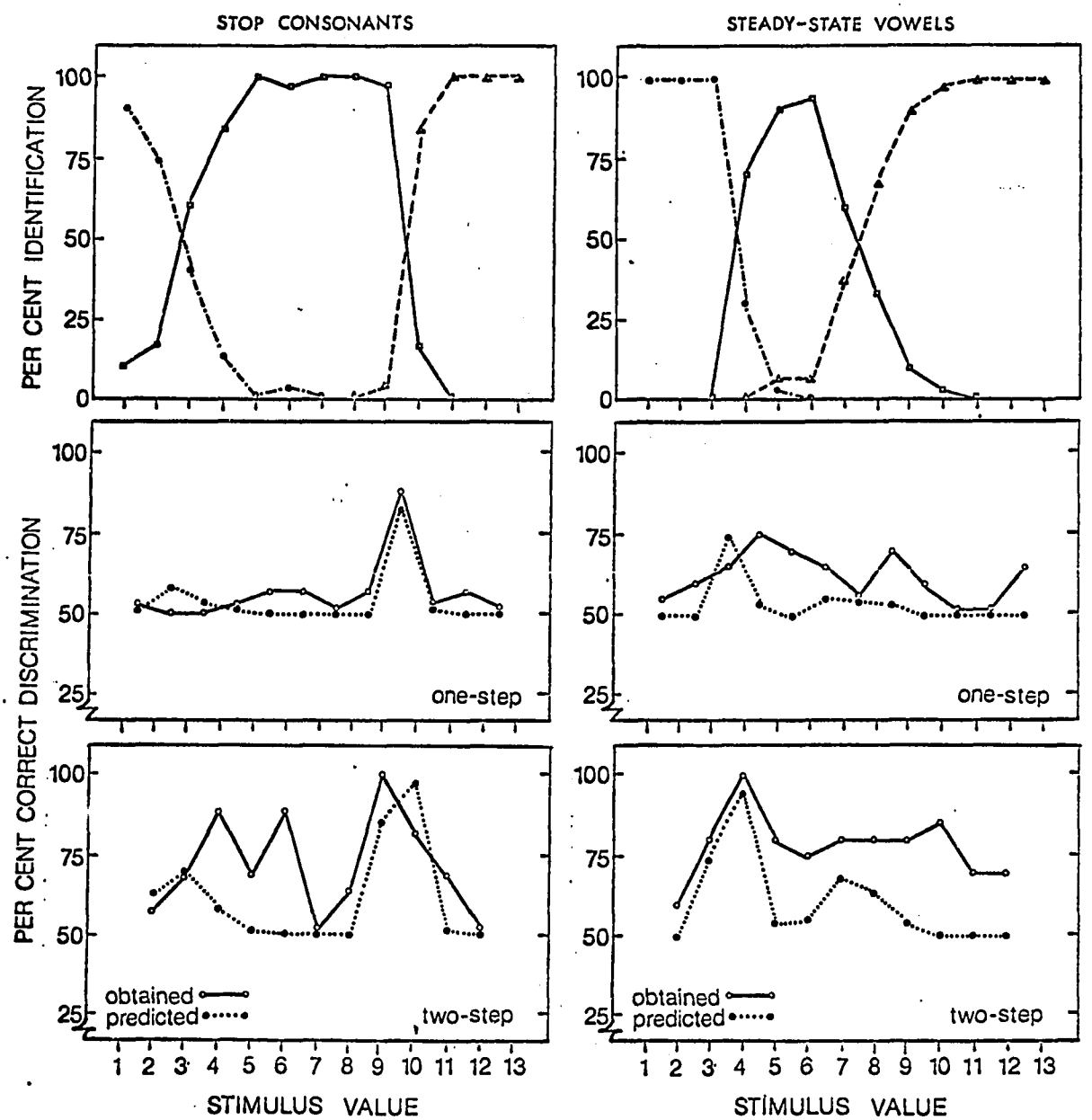


Figure B6. Identification and discrimination functions for stop consonants and steady-state vowels for S 9.

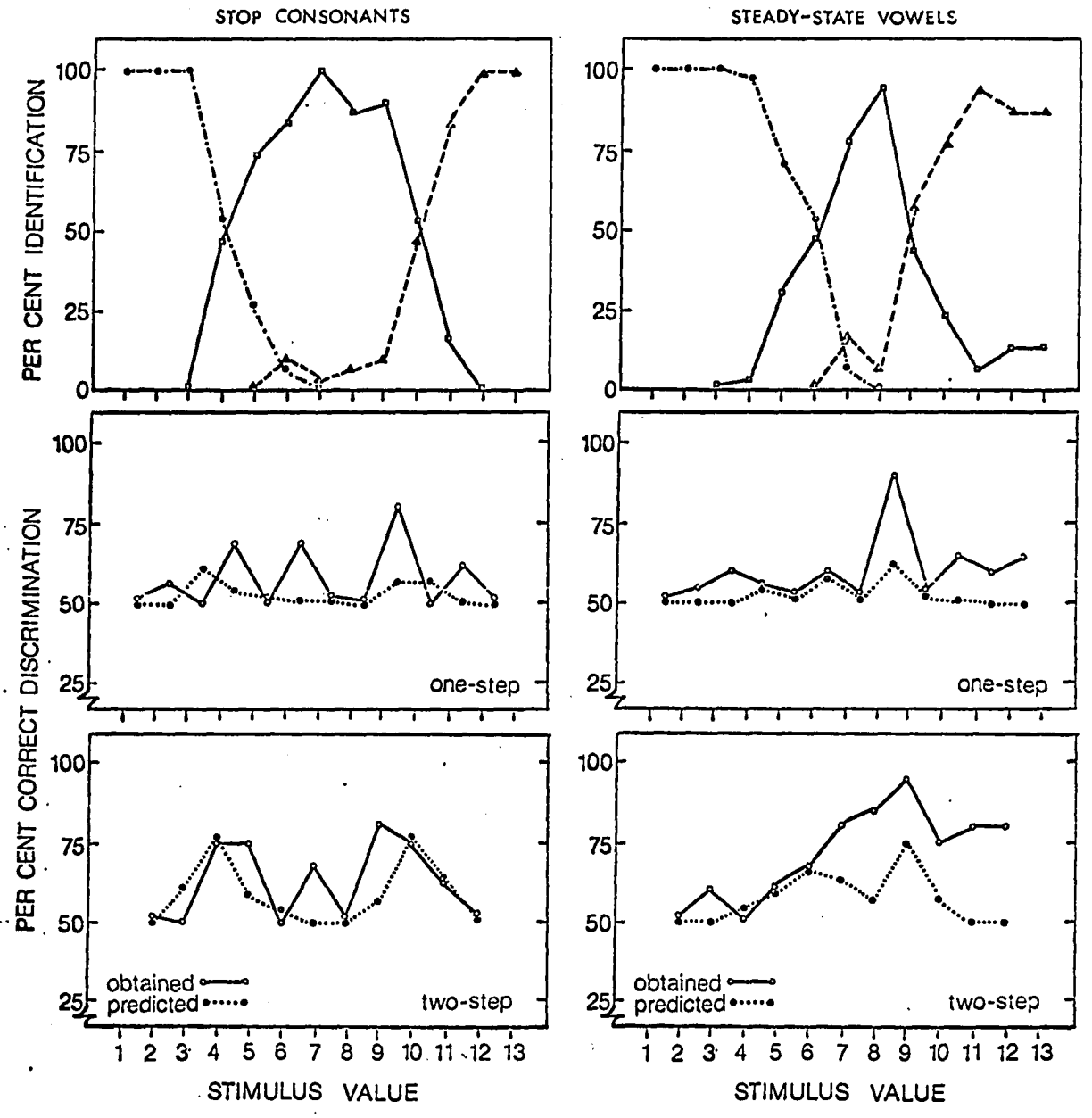


Figure B7. Identification and discrimination functions for stop consonants and steady-state vowels for S 10.

APPENDIX C

Derivation of Predicted Discrimination Formulae

APPENDIX C

Derivation of Predicted Discrimination Formulae from Absolute Identification

The procedure used to obtain the predicted discrimination scores is derived and examined in this appendix. The basic assumption of this procedure is that two sounds will be non-discriminable whenever they are assigned the same label in an identification task and will be discriminated only if assigned different labels. The original labelling hypothesis and prediction formula were devised by Liberman et al. (1957). Extensions to the two-interval same-different test (2IAX) and the four-interval test of paired similarity (4IAX) are also examined here and are available in a recent paper by Pollack and Pisoni (1971).

In the following derivations we will assume that labelling probabilities have been obtained from an absolute identification test for three responses. In this task the listener is presented with the stimuli one at a time and is required to label the sounds into a set of categories defined by the experimenter, e.g., "bae", "dae" and "gae". The probabilities obtained in the identification test may be used to predict a discrimination function under the extreme assumption that a listener can discriminate two stimuli only to the extent that he can identify them as different.

Let us consider the 2IAX test in which two stimuli are presented on each trial and the listener is required to determine whether the stimuli are the 'same' or 'different'. If each of the two sounds of a pair of stimuli is assigned the same label in the identification test, the assumption is that the listener will respond 'same' in the discrimination tests; if each pair of sounds is assigned a different label in the identification test, we will assume the listener will respond 'different' in the discrimination tests.

Table C1 illustrates the derivation. The proportion of b, d and g responses in the identification test to the A stimulus in an AX pair is represented as a b, d or g, abbreviated from $P(b|A)$, $P(d|A)$ or $P(g|A)$, respectively. Similarly, the proportion of b, d or g responses in the identification test to the B stimulus in an AX pair is represented

TABLE C1

Derivation of Predicted Discrimination Formulae

A. Notation:

$b = P(b|A) \qquad b' = P(b|B)$

$Q = bb + dd + gg$

$R = bb' + dd' + gg'$

$S = b'b' + d'd' + g'g'$

$1-Q = bd + bg + db + dg + gb + gd$

$1-R = bd' + bg' + db' + dg' + gb' + gd'$

$1-S = b'd' + b'g' + d'b' + d'g' + g'b' + g'd'$

B. Derivation:

Test	Type	Correct Example	Correct Prob.	Guess Example	Guess Prob.
2IAX	AA	bb	Q		
	AB	bd'	(1-R)		
	BA	d'b	(1-R)		
	BB	b'b'	S		
4IAX	AA AB	bb bd'	Q(1-R)	bb bb' bd bd'	QR (1-Q)(1-R)
	BB AB	b'b' bd'	S(1-R)	b'b' dd' b'd' bd'	SR (1-S)(1-R)
ABX	ABA	bd'b		bb'b bb'd bd'g	
	ABB	bd'd'		bb'b' bb'd' bd'g'	

C. Formulas:

2IAX $\quad 4P(C) = Q + S + 2(1-R)$

4IAX $\quad 4P(C) = 2Q(1-R) + 2S(1-R) + QR + (1-Q)(1-R) + SR + (1-S)(1-R)$

$= Q + S + 2(1-R)$

ABX $\quad 4P(C) = Q + (1-R) + S + (1-R)$

$= Q + S + 2(1-R)$

$2 + (i-i')^2$ Short Form

as b', d' or g', abbreviated from P(b|B), P(d|B) or P(g|B), respectively. The next set of terms under "Notation" in Table C1 provides a pair-wise classification in terms that will be useful in the derivation.

There are four types of trials in the 2IAX test, as shown, each with a probability of 0.25. Examples of AA trials are: bb, dd and gg. The contribution to a correct 'same' response, given by an AA trial, is (bb + dd + gg) or Q. The contribution to a correct 'different' response, given by an AB trial, is (bd' + bg' + db' + dg' + gb' + gd') or (1-R). The contribution to a correct 'different' response, given by a BA trial, is, assuming no time order effects, (1-R). And, the contribution to a correct 'same' response, given by a BB trial, is (b'b' + d'd' + g'g') or S. The final proportion of correct trials, P(C), is then:

$$4P(C) = Q + S + 2(1-R)$$

The same basic procedure is employed for evaluating the 4IAX and ABX procedures. In the 4IAX test, a sequence of four sounds is presented: AA AB or AB AA. The task of the listener is to indicate which pair was the 'same' (i.e., first or second). Upon AA AB trials, for example, a correct discrimination response in the 4IAX test is registered if the label assigned to each A stimulus of the first pair is in agreement, and if the labels assigned to each member of the second pair are in disagreement. [The following notation is employed in the examples; if a specific mismatch is demanded, the typical example employs a 'g'; if any mismatch can be used, the typical example employs a 'd'; if any response can be used, the typical example employs a 'b'.] When possible quadruplets are enumerated, the contribution to correct responses is [bb(1-R) + dd(1-R)] or Q(1-R). Similarly, upon BB AB trials, a correct selection is made if the label assigned to each B stimulus of the first pair is in agreement, and the labels assigned to the stimulus members of the second pair are in disagreement. There are two ways to arrive at a correct response by guessing; when the labels agree both within the first pair and within the second pair; and when the labels disagree both within the first pair and within the second pair. If guessing is random, the first way contributes (QR)/2 or (SR)/2; whereas the second way contributes (1-Q)(1-R)/2 or (1-S)(1-R)/2. Again, the same formula is obtained, as in the 2IAX tests, for P(C).

In the ABX discrimination test, a sequence of three sounds is presented: e.g., ABA or ABB, where A and B are stimuli employed in the identification tests. The first two stimuli, A and B, differ; the third is equally-likely to be either A or B. The task of the listener is to indicate whether the third sound was more like the first or the second sound.

The ABX procedure has been examined previously in detail (Liberman, et al., 1957). A correct response in an ABA trial is obtained when the labels assigned to the first and third A stimuli are in agreement, and, in turn, differ from that assigned to the B stimulus. Now there are three ways to arrive at a correct response through guessing: when all three stimuli give rise to the same label; when the first two stimuli yield the same label, but differ from the third; and when the labels differ to each of the three stimuli. The collation of trials into the present notation follows: the weighting to correct responses of the correct examples is twice that of the guess examples. We, therefore, duplicate the correct examples: one set combines with the first guess example to yield (bd'b + bg'b + bb'b) or bb(b' + d' + g') or bb, which leads to the Q term for the ABA trials and to the S term for the ABB trials; one set combines with the second and third guess examples to yield (bd'b + g'b + b'd + b'g + d'g + g'd) or b(1-S), which leads to the (1-S) term for ABA tests and b'(1-S), which leads to the (1-S) term for ABB tests.

The three tests predict the same conversion formula. An alternative short form is also presented in Table C1, paralleling that of Liberman, et al., 1957, where i represents each of the three alternatives, in turn; j represents either of the remaining alternatives; and k represents the remaining alternative. Thus $(i-i')^2$ implies $(b-b')^2 + (d-d')^2 + (g-g')^2$.

Finally, it should be noted that differences in obtained discrimination scores with these procedures do not arise from differential prediction by the labelling hypothesis. Rather, differences in the obtained discrimination scores across the procedures are presumably to be sought in the different task requirements embodied within the individual tests.

APPENDIX D

Identification and Discrimination Functions for Stop Consonants for Individual Ss in Experiment 5

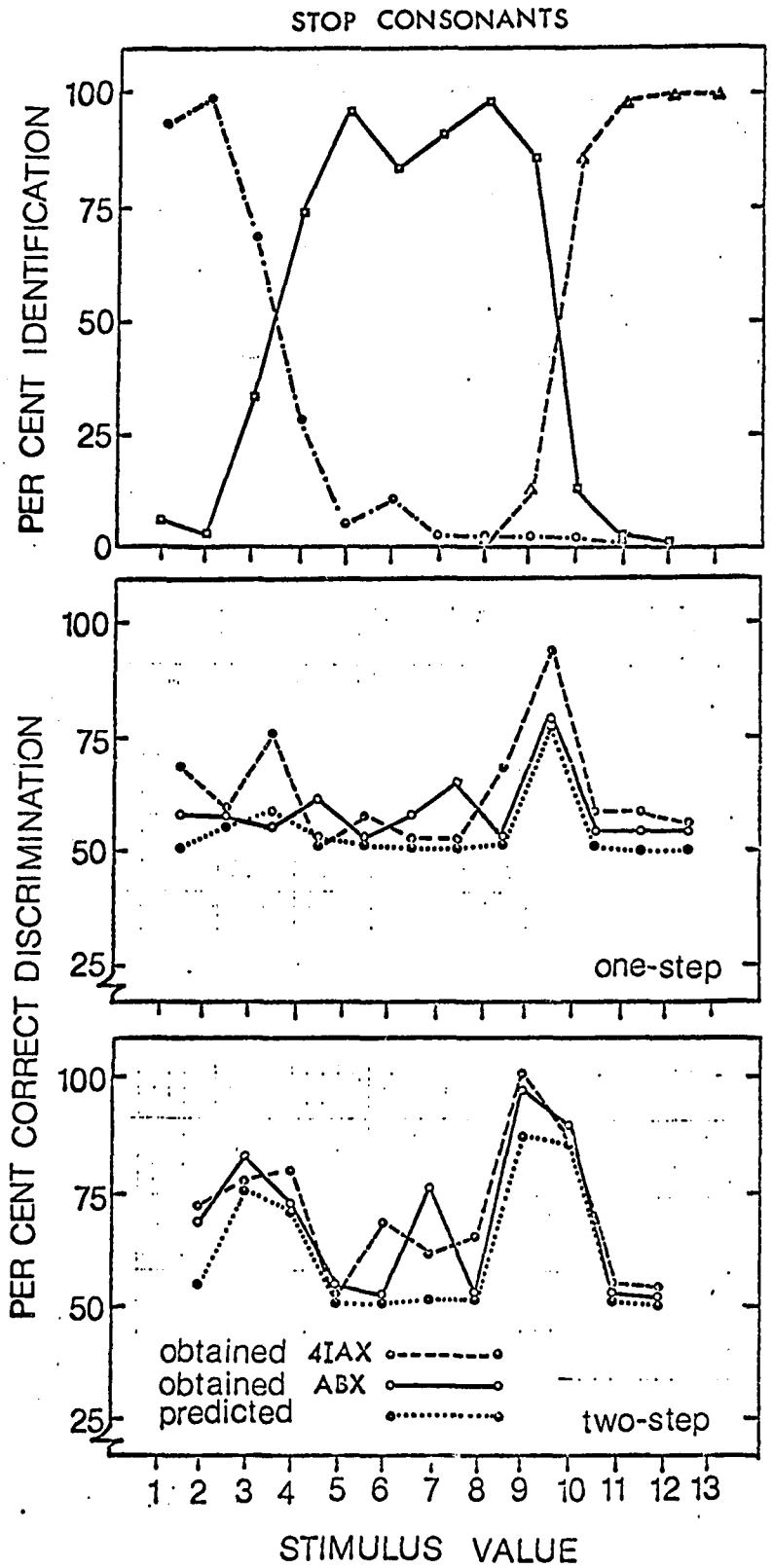


Figure D1. Identification and discrimination functions for stop consonants for S 1.

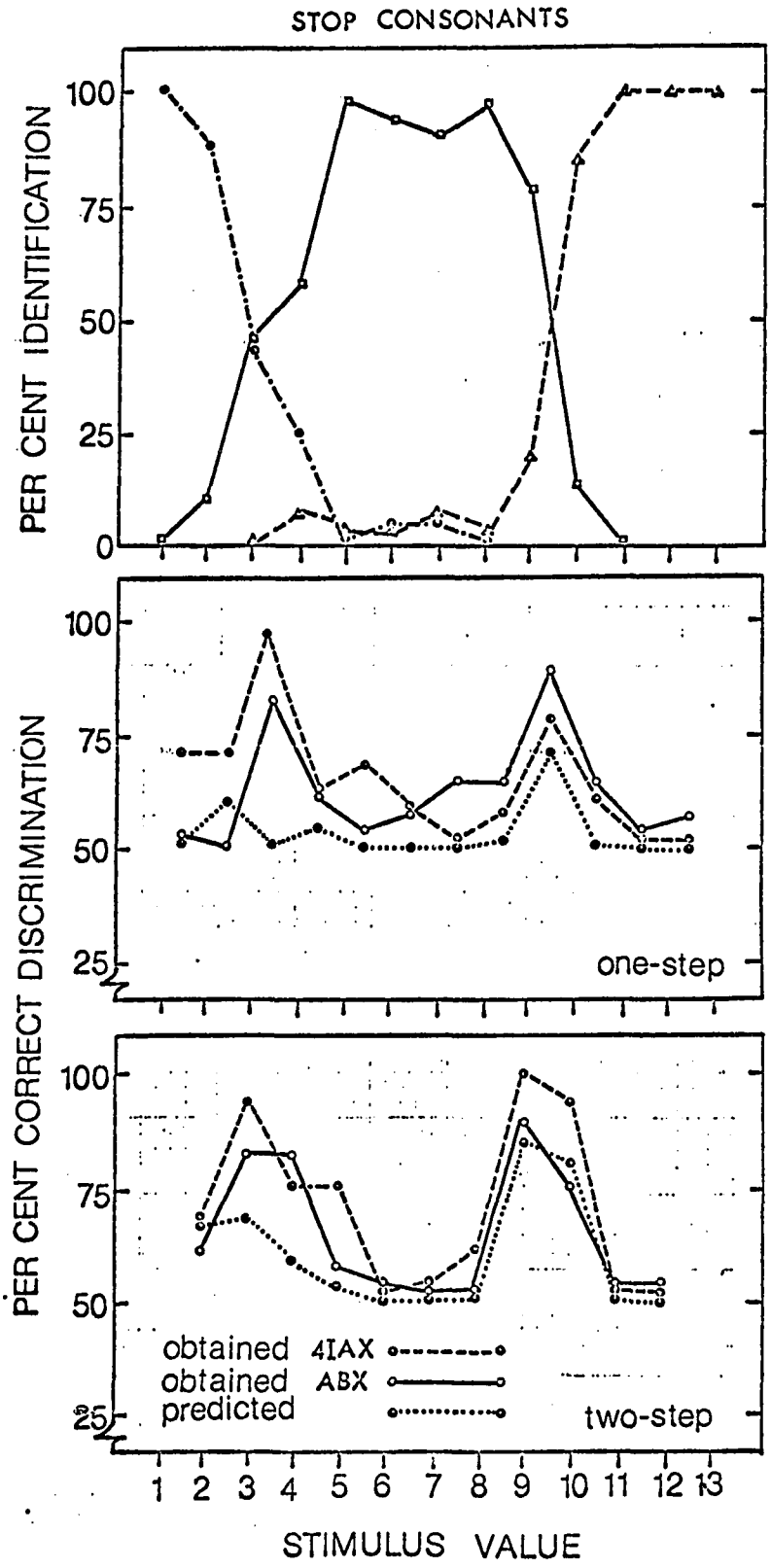


Figure D2. Identification and discrimination functions for stop consonants for S 2.

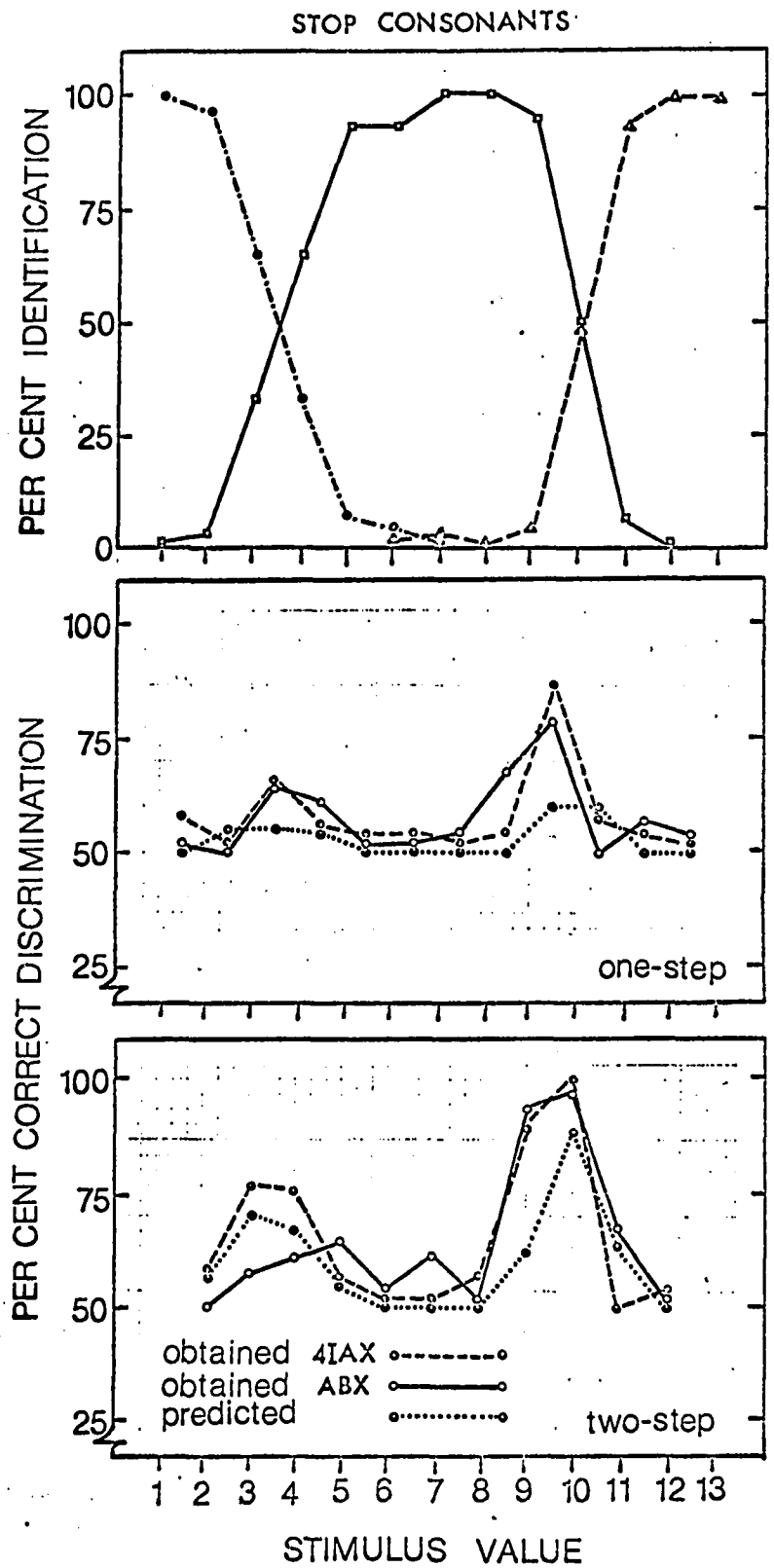


Figure D3. Identification and discrimination functions for stop consonants for S 3.

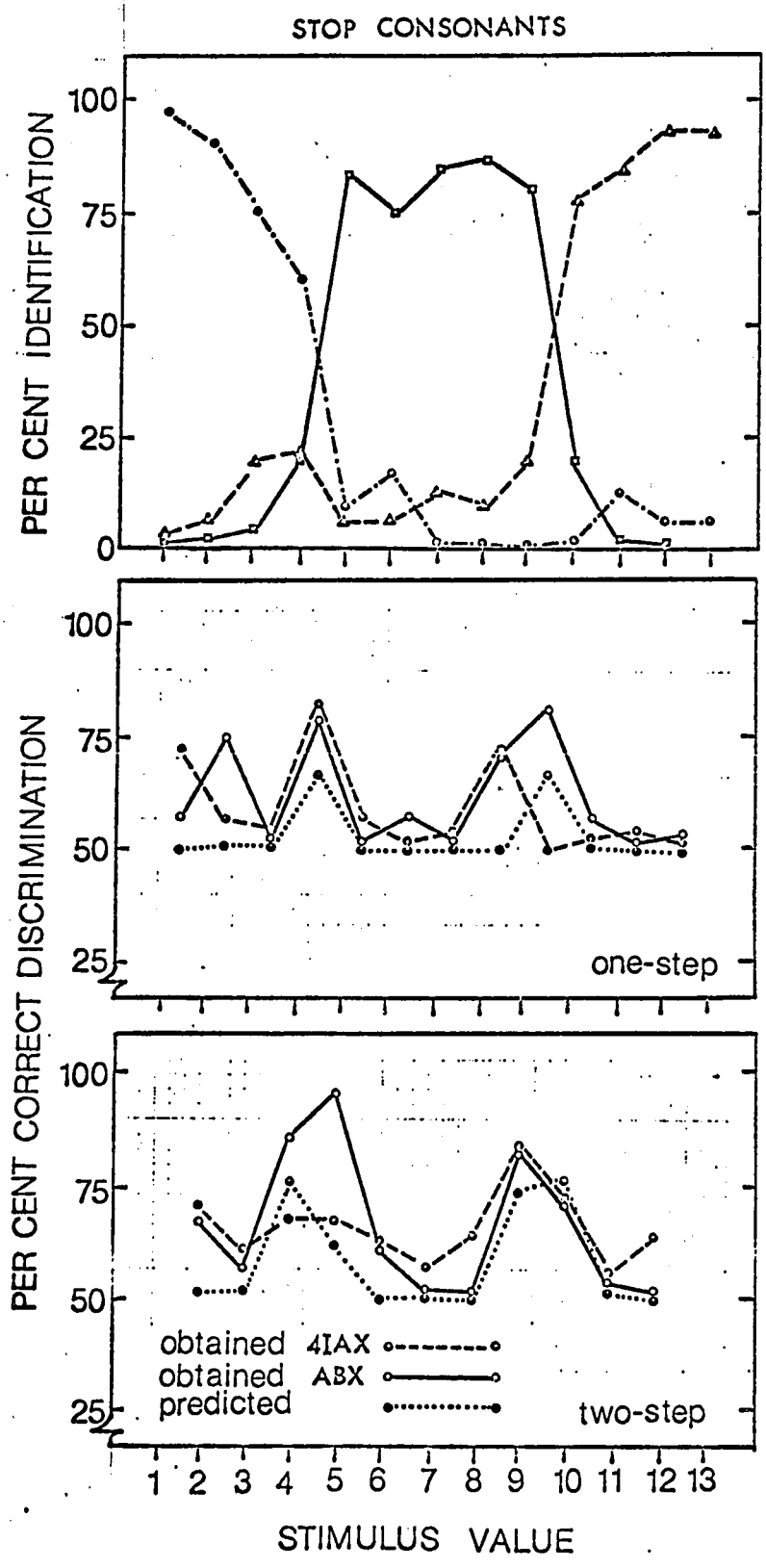


Figure D4. Identification and discrimination functions for stop consonants for S 4.

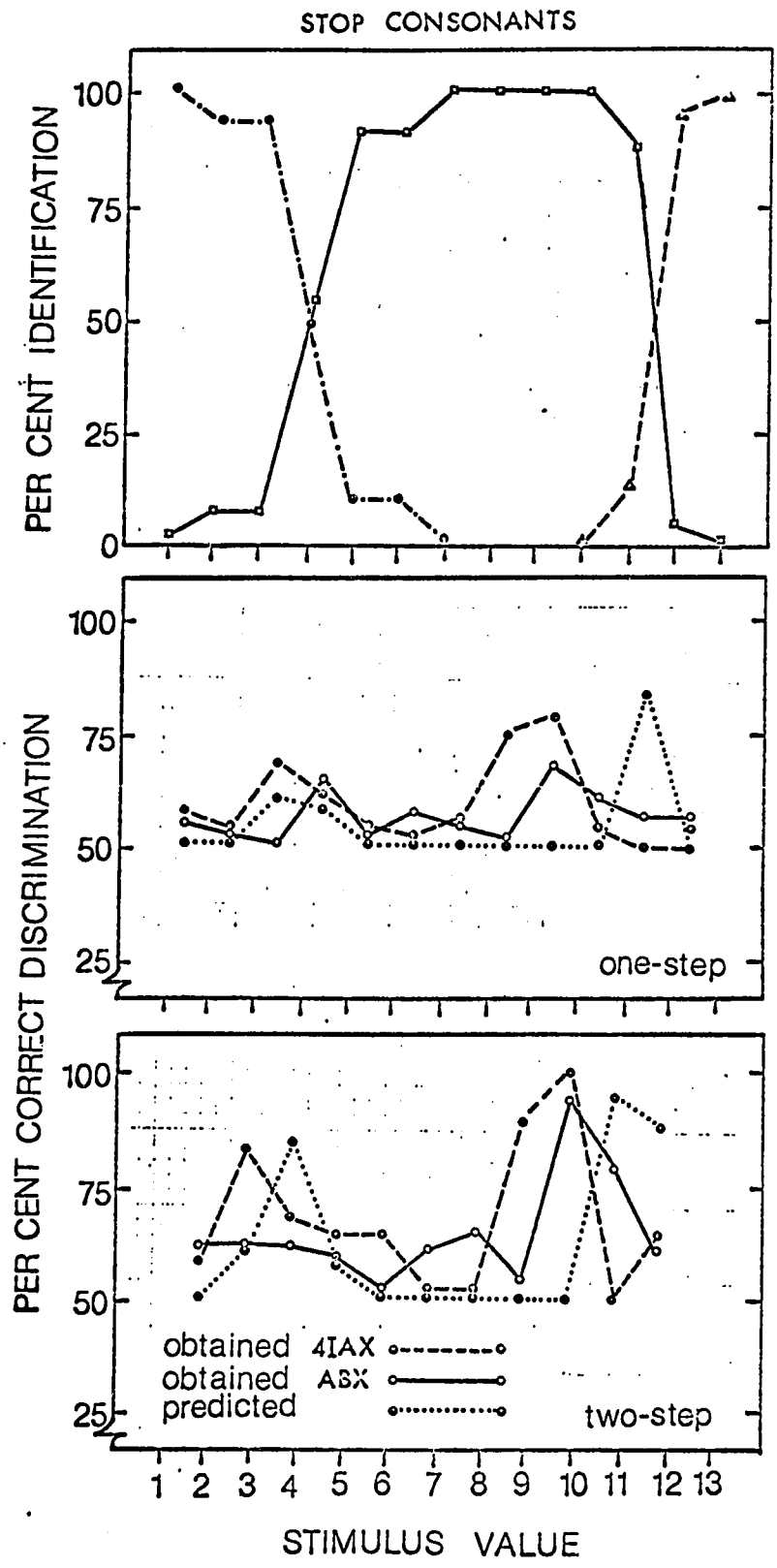


Figure D5. Identification and discrimination functions for stop consonants for S 5.