# SOME QUANTITATIVE STUDIES OF RUSSIAN CONSONANT PHONEME DISCRIMINATION

by

Patrick Suppes, Edward Crothers, Ruth Weir,
and Edith Trager

TECHNICAL REPORT NO. 49

September 14, 1962

PSYCHOLOGY SERIES



INSTITUTE FOR MATHEMATICAL STUDIES IN THE SOCIAL SCIENCES
Applied Mathematics and Statistics Laboratories
STANFORD UNIVERSITY
Stanford, California

## TABLE OF CONTENTS


| | | Page |
|---|---|---|
| 1. | INTRODUCTION | 1 |
| 2. | EXPERIMENTAL METHOD | 3 |
| | Experiment I | 3 |
| | Experiment II | 8 |
| 3. | EXPERIMENTAL RESULTS | 10 |
| | Experiment I | 10 |
| | Experiment II | 14 |
| 4. | LINGUISTIC INTERPRETATION OF RESULTS | 24 |
| 5. | QUANTITATIVE APPLICATION OF STIMULUS-RESPONSE THEORY | 30 |
| | The Models | 30 |
| | Tests of the Models | 42 |
| | Tentative Conclusions | 51 |
| 6. | REFERENCES | 53 |
| 7. | APPENDICES A,B,C,D | 55 |

## 1. INTRODUCTION

This Technical Report is the first in a series of detailed quantitative studies attempting to analyze second language learning in terms of stimulus-response theory.* It is the joint work of structural linguists and mathematical psychologists.

Specifically, this first Report is concerned with a quantitative analysis of Russian consonant discriminations in initial position by native speakers of American English. The selection of the materials studied is described in the section on Experimental Method, together with an outline of the procedure utilized. The section on Experimental Results is a detailed account of the quantitative findings which in turn are interpreted from a linguistic point of view in the next section, Linguistic Interpretation of Results. Although primarily based on articulatory phonetics, an acoustic analysis is also utilized in this section whenever it seemed pertinent to a better understanding of the data.

Although the linguistic results present some data of general interest, the most important aim of the study is to apply a mathematical theory of learning to second language acquisition. The results of this analysis have been placed in the section on Quantitative Application of Stimulus-Response Theory, the final section of this Report, in order to make the other aspects of the study more accessible to linguists and other readers who are not primarily interested in this feature of the work.

The Bibliography contains references of works utilized in the study, as well as reference to more general references on terminology in linguistics and mathematical learning theory. For linguistics, we

* The research reported herein was conducted pursuant to contract SAE 9514 between Stanford University and the U.S. Office of Education, and was also partially supported by a grant from the Carnegie Corporation of New York. We wish to acknowledge the assistance of Elise Belenky in preparing and conducting the experiment.

refer the reader to Gleason (1961), and for mathematical learning theory to Bush and Estes (1959), Bush and Mosteller (1955), Estes (1961), and Suppes and Atkinson (1960).

Before turning to the details of the two experiments reported, we want to make some additional remarks about the orientation and purpose of this series of studies. Perhaps the first observation that might be made about the present Report by someone interested in second language acquisition is that if this much detailed analysis is going to be made of each aspect of such acquistion, it seems doubtful that problems of adequate linguistic complexity will be reached. He might also appropriately say that there seems to be only a slight prospect that the mathematical models used in the final section of the Report will have any significant application to other than the learning of the simplest phonetic material. He might indeed ask what hope is there of being able to use such models to explain how American English speakers learn Russian grammar or the semantic categories of Russian.

Our defense is disarmingly simple. We do not in fact yet see how these more complicated applications are to be made. Applications of the kind of models we use to relatively complicated concept formation experiments in the past year or two offer some grounds for hope, but it would take us much too far afield to outline our present ideas and future plans.

At this juncture we are satisfied to limit ourselves to linguistically simple material, which is already relatively complicated in comparison with the stimulus material used in most learning experiments. We also emphasize that we are more concerned to study quantitatively a few fundamental problems than to produce new pedagogical materials or results of immediate classroom interest.

## 2. EXPERIMENTAL METHOD

The descriptions of the two experiments are presented separately.

### Experiment I

Subjects. Twenty Stanford University students from an introductory logic class served as subjects. Each subject's native language was English. Only students who did not speak Russian were eligible for the experiment. Each subject appeared for one-half hour daily for five consecutive days. After the final session, each subject received $6.00 for participating in the experiment.

Before starting the first experimental session, each subject was asked to complete a brief questionnaire in order to determine his or her language background. Due to Stanford University's admission policy, the subjects' foreign language background is considerable. Of the nine subjects who started the study of Latin in high school, none of them continued it in college; however, all except one continued the other foreign languages studied in high school at Stanford University. Six studied French for two to three years, two studied Spanish for one and two years respectively. All three who started French in high school continued for two to three quarters in college. Out of the seven who started the study of Spanish in high school, five continued the language in college. One switched to German, the other to French. The only other language studied in high school was Italian by one subject who lived in Italy for six months. However, he also studied French in high school and college, in addition to one quarter of Spanish. Several of those studying either French or Spanish in high school took up another foreign language in addition. The languages occasionally heard at home were: Polish, Czech, Yiddish, Hebrew and Spanish (one subject each) and German (two subjects). Two subjects participated for two quarters at Stanford in France. One subject stutters.

Materials. A basic list of Russian syllables in phonemic transcription was constructed. There are 32 initial consonant phonemes (the "j" indicates palatalization): /p,pj, t,tj, c,ch, k, b,bj, d,dj, g, f,fj, s,sj, sh, x, v,vj, z,zj, zh, m,mj, n,nj, l,lj, r,rj, y/. The phoneme /y/ was not used but the phonemic sequence /šč/ was used. The list consisted of these consonants and the sequence, followed by the vowel phonemes /a, e, i, o, u/. Excluding the CV syllables (a single consonant followed by a single vowel) which are exceedingly rare, there are 144 such syllables, some of which are spelled in two different ways (e.g., "cho", "chë"). These 144 syllables were grouped into contrasting pairs differing only in the initial consonant phoneme. These contrasts will be denoted by $/C_1V_1 - C_2V_1/$. Then the contrasts were classified into sets which were ordered in terms of expected difficulty of discrimination and production for an American subject with no knowledge of Russian. A fuller description of the sets appears in Experiment II. The sets judged easiest were chosen for Experiment I. The estimates were based on the linguists' judgements of relative difficulty of pronunciation of the pairs, since little *a priori* information concerning difficulty of discrimination was available. They are the following:

| Type | Consonant contrasts |
|---|---|
| voiceless : voiced plain fricatives | /f : v, sh : zh, s : z/ |
| voiceless : voiced plain stops | /p : b, t : d, k : g/ |

The /f : v/ and /k : g/ contrasts included only the vowel phonemes /a, o, u/, since those consonants are most frequently palatalized before /e/ and /i/. All five vowels were used with each of the other consonant contrasts, making a total of 26 CV contrasts. From these 26 contrasts of the form $/C_1V_1 : C_2V_1/$, the stimulus items for Experiment I were constructed in the following manner. For each contrast, the four CV pairs $C_1V_1 : C_2V_1$, $C_2V_1 : C_1V_1$, $C_1V_1 : C_1V_1$, and $C_2V_2 : C_2V_2$ were constructed. (The first two pairs are called minimal pairs, since their members differ by only one phoneme.) Since, when the vowel member is held constant, each contrast involves four pairs (e.g., ba : ba, ba : pa, pa : ba, pa : pa) each contrast is referred to as a 4-concept. Hence, 26 4-concepts, or 104 pairs, were used in Experiment I. The 52 pairs having the second CV member the same as the first will be called same (s) pairs, and the 52 pairs having as the second member a CV syllable different from the first syllable will be called different (d) pairs. The 104 pairs are listed in Appendix A.

There are five vowel phonemes, and, as mentioned, with /f : v/ and /k : g/ three were used in the contrasts. It must be borne in mind, however, that the allophones of these vowels differ according to the preceding consonant. In the case of /a, e, o, u/ the different allophones sound much alike to English speakers. But the allophonic variations of /i/ are particularly marked (see Section 4).

After the pairs had been formed, lists of pairs to be recorded for presentation to the subject were constructed as follows. For the first day, three randomizations of the order of the 104 minimal pairs were prepared. For the second day, three new randomizations were made. For the final three days, the pairs judged (on the basis of pilot evidence, and on linguistic grounds) to be too easy for further experimentation) were eliminated. The 20 /p : b/, 20 /t : d/ and 12 /k : g/ pairs were retained. Six randomizations of these 52 pairs were prepared for each of the last three days' material. One randomization of the pairs will be called a list. Lists 1-6 constituted the material for days 1-2, and Lists 7-24 for days 3-5.

Recordings. High quality tape recordings of the 24 lists of CV pairs were made in a heavily sound-proofed room in the Division of Speech Pathology and Audiology laboratory at the Stanford Medical School.[1/] Recordings were made on Scotch 111 tape at 7.5 inches per second using a boom-mounted Altec 26-M microphone system and an Ampex 351 stereo tape recorder. The microphone was placed at a distance of 4-in. and at an angle of $115^{\circ}$ from the speaker's lips in order to avoid air-blast.

The phonetic peak of all syllables was held above a minimum VU reading. No attempt was made to equate phonetic peaks; instead, we operated with the natural difference in vowel energy. The levels, once established, were not changed during the course of the recording. All recorded items (the CV pairs) were self-approved by the native speaker of Russian and by the monitoring linguist.

The native speaker lived in a Russian-speaking environment from birth (1906) until settlement in the United States in 1928 and has spoken Russian daily throughout her life. Her father was born in Moscow and her mother in Vladivostok. She has lived in both cities, and is from an upper socio-economic background. She received her secondary education in Russia, and

1/ We wish to acknowledge the assistance of the Speech Pathology and Audiology Laboratory staff, and particularly the invaluable help of Professor Dorothy Huntington.

her higher education in France and Belgium.

Apparatus. The recorded material was played back on a Sony recorder, Model 262SL. For playing, the highest possible volume setting was used which still kept the sound free from distortion. The tone was appropriately adjusted. The volume and tone settings were the same for all subjects. The other piece of equipment which consisted of a 75-watt lamp fitted with a 1-in. diameter green reflector and mounted on an 11-in. x 4-in. x 2-in. black metal box permitted the experimenter to deliver a light signal to the subject after each incorrect response. The lamp was illuminated whenever the experimenter pressed a simple doorbell-type button. Another room was equipped with the identical equipment, and two subjects were run concurrently, one in each room.

Procedure. The subject was seated facing the Sony speaker at a distance of six feet. First he completed a written questionnaire pertaining to his background in foreign languages. Then the following instructions, recorded by a native American linguist were played over the speaker.

> You will now hear 104 pairs of syllables, one pair at a time. Each pair will be followed by a short pause. Listen carefully to each pair of syllables. Decide whether the two syllables are the same or different. If they sound the same, say "same". If they sound different, say "different". Answer each time, even when you are not sure. If you are wrong, the green light will flash (the light flash was demonstrated). If you are right, there will be no flash.

Next, the subject was asked if he had any questions about the procedure. Questions were answered by paraphrasing the appropriate portion of the instructions.

Next, the CV pairs were played one at a time over the loudspeaker.

The rate of presentation was 14 pairs per minute, with a 3-sec. pause between pairs. The presentation phase continued without interruption until all pairs had been presented.

## Experiment II

Only the changes from Experiment I will be noted. The method of Experiment II differed from that of Experiment I in the selection of subjects and stimulus material.

Subjects. Twenty Stanford University students, eight from an introductory logic class, and twelve who were secured through the employment bureau, served as subjects. Only students who did not speak Russian were eligible for the experiment. As in Experiment I, each subject appeared for one-half hour daily for five consecutive days and received $6.00 for participating in the experiment. The subjects language background in this experiment did not differ appreciably from those in the previous experiment. Seven subjects had studied Latin in high school for two to three years, and some of them continued the language at the University. In addition all seven had taken a modern language in high school and college. French was the language studied by most (eleven), followed by Spanish (eight) and German (seven). German was started by three subjects for one quarter only at the University; two studied Greek, and one studied Italian in addition to spending six months at Stanford-in-Italy. In regard to language background at home, one subject spoke Spanish frequently and another Hungarian occasionally.

Materials. Again the stimulus items were CV pairs. The contrasts having the highest error rate in Experiment I were retained, that is, the

/p : b/ and /t : d/ combined with all vowel phonemes, and /k : g/ with /a, o, u/.

In order to determine which contrasts were most difficult, and hence of greatest experimental interest, two pilot studies were run. The contrasts presented as stimuli included plain versus palatalized phonemes, and /sh, šč, ts, ch/. Two complete lists of stimuli, one for each pilot study, appear in Appendix B.

An analysis of these studies is given in Appendix C. The obtained order of difficulty was somewhat different from what one would expect on the basis of ease of pronunciation. The more difficult pairs, as indicated by Experiment I and the two pilot studies, were selected for Experiment II.

Of the plain : palatalized contrasts, the voiceless and voiced sibilants were chosen /s : sj, z : zj/ in addition to /d : dj, n : nj/, and the laterals /l : lj/. All of them were combined with /i/ as presenting greatest difficulty of discrimination, and the laterals were also combined with /a/ as an additional vowel.

Fricative : affricate contrasts in initial positions were included, that is, /s : ts/, combined with all five vowel phonemes. One contrast of stop : fricative was also used, namely, /k : x/, thus giving us a list of 25 sets of 4-concepts or 100 pairs. The 52 pairs which pilot data had indicated to be easiest were eliminated after Day 2. The contrasts presented on all five days were the /k : x, z : zj, s : ts, p : b/ pairs. These contrasts included 4, 4, 20 and 20 pairs, respectively.

## 3. EXPERIMENTAL RESULTS

The results for the two experiments are presented separately. Roughly speaking, we first give the analysis of item difficulty and then turn to group and individual learning data.

### Experiment I

Relative difficulty of consonant discriminations. Table 1 indicates the difficulty of the pairs classified according to their vowel and consonant members. For example, the entry .006 at the intersection of the /f : v/ row and the /u/ column is the proportion of errors over all presentations of the four pairs /fu : fu, fu : vu, vu : fu, vu : vu/. The blanks in the table indicate that certain 4-concepts (which are rare or non-existent in Russian) were not presented in the experiment.

---

Insert Table 1 about here

---

Table 1 is arranged so that items are of increasing difficulty, reading from left to right and from top to bottom. The column and row proportions were computed from the overall frequency of errors on the consonant discrimination or vowel indicated. It will be recalled that no /f : v/ or /g : k/ pairs were presented with the vowels /e/ or /i/, and that the "easy" pair types (those listed in the first three rows of the table) were presented only in Lists 1-6. For 4-concepts presented in lists 1-6 only, each entry is based on 480 observations (4 pairs x 20 subjects x 6 lists). Similarly, for 4-concepts presented in all 24 lists, each entry is based on 1,920 observations.

The table shows the following order of discrimination of consonant

TABLE 1

Proportion of Errors in Experiment I on Each Set of Four Pairs which Present the Same Consonant Contrast and Contain the Same Vowel[1]

| Consonant Contrast | Vowel a | o | u | e | i | Mean, Lists 1-6 | 7-24 | 1-24 |
|---|---|---|---|---|---|---|---|---|
| /f : v/ | .006 | .017 | .006 | - | - | .010 | - | - |
| /s : z/ | .010 | .013 | .015 | .010 | .029 | .015 | - | - |
| /sh : zh/ | .030 | .028 | .017 | .021 | .021 | .023 | - | - |
| /k : g/ | .024 | .040 | .037 | - | - | .067 | .022 | .034 |
| /t : d/ | .039 | .050 | .038 | .038 | .054 | .077 | .033 | .044 |
| /p : b/ | .052 | .055 | .071 | .066 | .070 | .139 | .037 | .063 |
| Mean, Lists | | | | | | | | |
| 1-6 | .042 | .053 | .059 | .063 | .081 | | | |
| 7-24 | .028 | .035 | .030 | .032 | .038 | | | |
| 1-24 | .034 | .042 | .041 | .045 | .055 | | | |

[1] The proportions are based on Lists 1-6 data for pairs presented only in Lists 1-6. and on Lists 7-24 data for pairs presented on all 24 lists.

pairs, here listed in ascending order of difficulty: /f : v, s : z, sh : zh, k : g, t : d, p : b/. As for the vowels, /a/ is the easiest, followed by /e, o, u/ which are of about equal difficulty, and /i/ which is the most difficult. Two analyses of variance were performed to determine (a) whether the consonants differed significantly from one another in difficulty, and (b) whether the vowels differed significantly from one another in difficulty.

The first analysis of variance involved the data from Lists 1-6 for the /a, o, u/ vowels and all consonants. Hence it was a 6 consonants x 3 vowels x 20 subjects design. The consonant x vowel x subject mean square was taken as the error term in the computations of *F*. As Table 2 shows, all the main effects and two-way interactions are significant, indicating reliable inter-consonant and inter-vowel differences in difficulty.

The second analysis of variance used the data of Lists 1-24 from the /t : d/ and /p : b/ pairs with the vowels /e/ and /i/. The results, given in Table 3, indicate significant difference in difficulty between /t : d/ and /p : b/. From Table 1, it may be seen that /t : d/ was easier than /p : b/. The variance attributable to vowels (/e/ and /i/ in Table 3) was not significant; hence, they appear to be of about equal difficulty in the present case.

---

Insert Tables 2 and 3 about here

---

*Relative difficulty of s and d pairs*. We now ask whether a pair is more difficult when the correct judgment is "different" (*d*) than when the correct judgment is "same" (s). Table 4 affirms that detection of the difference when the two members of the pair contain different consonants is the more difficult task. For the s pairs, the error rates in Lists 1-6 and 7-24 were .041 and .030, respectively, while the corresponding figures for the d pairs were .075 and 0.34. A sign test in which the total number of errors on s pairs by a given subject was paired with his total on the d pairs was significant at the .01 level on Lists 1-6, but was not significant on Lists 7-24 data, On /k : g/ pairs, 63.4% of the errors were "s" responses to d pairs. The corresponding figures for /t : d/ and /p : b/ were 63.2% and 62.2%, respectively.

---

Insert Table 4 about here

---

The errors on the d pairs were classified according to whether the voiced phonemes /g, d, b/ appeared in the first or second syllable of the pair. Cases where the voiced phonemes appeared in the first syllable comprised 65.0%, 66.3% and 55.2% of the /k : g, /t : d/ and /p : b/ errors on d pairs, respectively (N the number of observations was 123, 264, and 375, respectively).

Learning. The proportion of errors over all subjects and pairs decreased from .11 in the first list to .02 in the last list of the experiment. These proportions were computed for sets of three successive lists and appear in Table 5. The divisions between the daily sessions occurred after Lists 3, 6, 12 and 18. Since no abrupt increase in errors

TABLE 2

Vowels x Consonants x Subjects Analysis of

Variance in Total Errors on Lists 1-6 in Experiment I

| Source of Variance | d.f. | Mean Square | F | p |
|---|---|---|---|---|
| Vowels | 2 | 4.80 | 5.16 | <.01 |
| Consonants | 5 | 62.93 | 67.64 | <.001 |
| Subjects | 19 | 15.72 | 16.90 | <.001 |
| Vowels x Consonants | 10 | 4.32 | 4.64 | <.001 |
| Vowels x Subjects | 38 | 1.41 | 1.52 | <.05 |
| Consonants x Subjects | 95 | 2.57 | 2.76 | <.01 |
| Vowels x Consonants x Subjects | 190 | .93 | - | - |

TABLE 3

Vowels x Consonants x Subjects Analysis of
Variance in Total Errors on Lists 1-24 in Experiment I

| Source of Variance | d.f. | Mean Square | F | p |
|---|---|---|---|---|
| Vowels | 1 | 17.11 | 3.17 | n.s. |
| Consonants | 1 | 159.61 | 29.60 | <.01 |
| Subjects | 19 | 34.20 | 1.92 | n.s. |
| Vowels x Consonants | 1 | 2.81 | 6.34 | <.01 |
| Vowels x Subjects | 19 | 6.53 | 1.21 | n.s. |
| Consonants x Subjects | 19 | 3.72 | 1.45 | n.s. |
| Vowels x Consonants x Subjects | 19 | 5.39 | - | - |

TABLE 4

Proportions of Errors for Pairs Consisting of Two d Syllables and for Pairs Consisting of Two s Syllables in Experiment I

| | Lists | | |
|---|---|---|---|
| Pair Type | 1-6 | 7-24 | 1-24 |
| s | .041 | .030 | .034 |
| d | .075 | .034 | .051 |

occurred when the "easy" pairs were eliminated (after List 6) it seems

---

Insert Table 5 about here

---

that the learning of the difficult discriminations progressed steadily, albeit slowly. From Table 4, it is clear that most of the learning occurred on the d pairs. A sign test in which the proportion of errors on d pairs in Lists 1-6 for a given subject was paired with his proportion in Lists 7-24, was significant at the .01 level, indicating that learning had occurred. A similarly computed sign test on the s pairs was also significant at the .01 level, indicating that learning of the s pairs was also taking place, even though the initial error rate was quite small.

Because of the low initial error rate, the learning data were not subjected to further analysis. For the same reason, no attempt was made to apply mathematical models to the data, since a sensitive discrimination among models cannot be made in the absence of sufficient errors.

Error rates on the pairs presented in pilot studies. Appendix C lists the proportion of errors for each type of pair presented in the two pilot experiments intervening between Experiments I and II. For the first pilot study, the proportions are based on a total of 1872 observations from six subjects, while the number of observations from each of nine subjects in the second pilot study was 72 per 4-concept. The proportion of errors was highest (.28) for the 4-concept consisting of the /so : tso/ pairs, and varied between .17 and .00 for the other 4-concepts.

Experiment II

Relative difficulty of consonant discriminations. Table 6 classifies the pairs of syllables according to their vowel member for each set of pairs that present the same consonant contrast and indicates the proportion of errors for each class. The rows and columns of Table 6 are ordered in terms of increasing difficulty of pairs in Lists 1-6, reading from

---

Insert Table 6 about here

---

top to bottom and from left to right. The Lists 1-6 column of Table 6 shows that the order of difficulty, here listed in ascending order, is the following: (1) the fricative-stop contrast, /k : x/; (2) the plain-palatalized contrasts, /d : dj/, /l : lj/, /n : nj/, /z : zj/, /s : sj/; (3) the stop contrast /k : g/; (4) the fricative-affricate contrast /s : ts/; (5) the dental and labial stop contrasts /t : d/ and /p : b/.

The /p : b/, /t : d/ and /k : g/ pairs exhibit the same order of relative difficulty as obtained in Experiment I. In fact, it is instructive to compare the error rates in Lists 1-6 for the items that appeared in both Experiments I and II. For the /t : d/ items the error rate of .077 in Experiment I contrasts with the .220 value obtained in Experiment II. Likewise, the proportions of errors on /p : b/ items are .139 and .222 in Experiments I and II, respectively. For the /k : g/ pairs, the corresponding figures are .067 and .117. The proportions of errors on the /k : x/ and /z : zj/ pairs are relatively low compared to what one would expect from the pilot data, although for the /z : zj/ pairs the error

TABLE 5

Proportions of Errors on Sets of Three Successive Lists
in Experiment I

| Lists | p(error) | Lists | p(error) |
| --- | --- | --- | --- |
| 1-3 | .077 | 13-15 | .029 |
| 4-6 | .039 | 16-18 | .040 |
| 7-9 | .036 | 19-21 | .022 |
| 10-12 | .039 | 22-24 | .024 |

TABLE 6

Proportion of Errors on Each Set of Four Pairs which Present the Same Consonant Contrast and Contain the Same Vowel[1] - Experiment II

| Consonant Contrast | Vowel /a/ | /i/ | /u/ | /e/ | /o/ | Mean, Lists 1-6 | 7-24 | 1-24 |
|---|---|---|---|---|---|---|---|---|
| /d : dj/ | - | .019 | - | - | - | .019 | - | - |
| /k : x/ | - | .018 | - | - | - | .033 | .013 | .018 |
| /l : lj/ | .050 | .048 | - | - | - | .049 | - | - |
| /n : nj/ | - | .056 | - | - | - | .056 | - | - |
| /z : zj/ | - | .085 | - | - | - | .067 | .092 | .085 |
| /s : sj/ | - | .104 | - | - | - | .104 | - | - |
| /k : g/ | .060 | - | .142 | - | .148 | .117 | - | - |
| /s :ts/ | .142 | .204 | .104 | .166 | .146 | .168 | .147 | .152 |
| /t : d/ | .131 | .283 | .165 | .238 | .281 | .220 | - | - |
| /p : b/ | .119 | .219 | .151 | .141 | .179 | .222 | .140 | .162 |
| Mean, Lists | | | | | | | | |
| 1-6 | .108 | .139 | .151 | .192 | .210 | | | |
| 7-24 | .124 | .115 | .118 | .148 | .149 | | | |
| 1-24 | .117 | .125 | .131 | .163 | .173 | | | |

[1] The proportions are based on Lists 1-6 data for pairs presented only in Lists 1-6, and on Lists 7-24 data for pairs presented in all 24 lists.

rate increased from Lists 1-6 to Lists 7-24.

An items x subjects analysis of variance was performed on the Lists 1-6 plain-palatalized items. Both the concepts and subjects sources of variance were significant at beyond the .001 level (Table 7) indicating significant inter-concept and inter-subject differences.

If we judge the relative difficulty of the vowels on the basis of all the pairs presented (Lists 1-24 data) the vowels ranked in order of increasing difficulty are: /a, i, u, e, o/. Note that, as in Experiment I, the stops preceding /i/ are difficult. Table 8 presents the results of the Lists 1-6 analysis of variance for each of the consonants /s : ts/, /t : d/, /p : b/ with each of the vowels /a, e, i, o, u/. All the main effects and two-way interactions are highly significant.

---

Insert Tables 7 and 8 about here

---

As in Experiment I, the number of errors on voiceless : voiced and on voiced : voiceless d pairs were compared. For /p : b, /t : d/ and /k : g/ 54.7%, 60.3% and 66.7% respectively of the total errors on d pairs occurred on voiced : voiceless pairs. The table also shows that voiceless s pairs are harder than voiced s pairs. Combining this with the previous finding, and without offering an interpretation, we may say that pairs whose second syllable is voiceless are harder than pairs whose second syllable is voiced. Also, 65.9% of the errors on /s : ts/ d pairs occurred when /s/ was first. A regularity which undoubtedly is related to this order effect is that s pairs involving either /s/ or /b/ yielded consistently more errors than s pairs involving either /ts/ or /p/ respectively.

We turn now to the effects of the vowels on consonant discrimination. The relevant data here are the column entries for a given row of Table 6. The order of difficulty generally agrees with that found in Experiment I, since pairs containing /i/ are most difficult, and those containing /a/ are easiest. However, while the pairs containing /o, e, u/ were of equal difficulty in Experiment I, the /u/ pairs in the /p : b/ and /t : d/ 4-concepts appear to be relatively easier this time.

Proportion of errors computed over all subjects and pairs for each list. Figure 1 shows the mean learning curve. For the first six lists, each data point represents two thousand observations; for the last 18 lists, each point represents 1,040 observations. We note that nearly all of the reduction in errors occurred between Lists 1-6 and between Lists 15-24. Also, it is interesting to note that the curve appears to

---

Insert Fig. 1 about here

---

be approaching an asymptotic proportion of errors which is definitely greater than zero (about .10). In the analyses immediately following, the learning curves are considered separately for each of the various categories of pairs.

Relative difficulty of s and d pairs. When the data for those pairs whose correct response is "same" are tallied separately from those for which the correct response is "different", trends indicated in Table 9

TABLE 7

Analysis of Variance - Plain-Palatalized 4-Concepts x Subjects

Experiment II

| Source of Variance | d.f. | Mean Square | F |
|---|---|---|---|
| 4-Concepts | 5 | 8.27 | $6.59^{+}$ |
| Subjects | 19 | 4.48 | $3.57^{+}$ |
| 4-Concepts x Subjects | 95 | 1.26 | - |

$^{+}p < .001$

TABLE 8

Vowels x Consonants x Subjects Analysis of
Variance in Total Errors on Lists 1-6. Experiment II

| Source of Variance | d.f. | Mean Square | F |
|---|---|---|---|
| Vowels | 4 | 169.16 | 53.33⁺⁺ |
| Consonants | 2 | 63.74 | 20.10⁺⁺ |
| Subjects | 19 | 53.31 | 16.81⁺⁺ |
| Vowels x Consonants | 8 | 17.60 | 5.55⁺⁺ |
| Vowels x Subjects | 76 | 5.84 | 1.84⁺ |
| Consonants x Subjects | 38 | 15.56 | 4.90⁺⁺ |
| Vowels x Consonants x Subjects | 152 | 3.17 | - |

$^{+}p < .001$

$^{++}p < .001$

emerge. As one would anticipate from the preceding analysis, the decline in errors is rather slight. For the s pairs, the proportion of errors fell from .081 in Lists 1-6 to .071 in Lists 7-24. On the corresponding lists, the proportion of errors on d pairs dropped from .223 to .186. In agreement with Experiment I, most of the improvement occurs on the d pairs,

---

Insert Table 9 about here

---

even though there is more room for improvement on s pairs in the present experiment than in Experiment I. To compare the performance on the s and d pairs which appeared in all lists used in the experiment, two sign tests were run. First of all, when the number of errors by a given subject on the s and d items were paired, (yielding twenty pairs), the d pairs proved to be significantly more difficult. For Lists 1-6, and again for Lists 7-24, the difference was significant at the .01 level. To ascertain whether there was any significant improvement on Lists 7-24 from Lists 1-6, a sign test was run on the s pairs, and another on the d pairs. Each subject's proportion correct in the earlier lists was paired with his proportion in the later lists. For the s pairs, the difference in proportion correct between Lists 1-6 and Lists 7-24 was not significant. For the d pairs, 17 differences were in one direction, indicating significant improvement ($p < .01$).

Proportions of errors for pairs classified by consonants or by vowel. By comparing the sixth and seventh columns on Table 6, it may be seen that improvement occurred on all those consonant contrasts which appeared on all 24 lists, except for the /z : zj/ pairs. Likewise, comparison of the next to last row with the preceding row reveals improvement on pairs containing vowels other than /a/. When the proportions are computed over only the /p : b/ and /s : ts/ pairs (to allow for differential elimination of certain vowels in the selection of pairs for Lists 7-24), improvement is indicated for the /a/ pairs also. Figure 2 gives the plot of the proportion of errors in sets of six successive lists for the pairs which appeared in all lists. Each data point is based on 2,400 observations for the /s : ts/ and /p : b/ pairs, and on 480 observations for

---

Insert Fig. 2 about here

---

the /k : x/ and /z : zj/ pairs. A comparison of the /s : ts/ and /p : b/ curves reveals no difference in the initial level of learning (Lists 1-6) but a wide difference in learning rate.

Table 10 presents a more detailed breakdown of the learning data. The /k : x/ pairs were excluded from this tabulation of the proportion of errors over sets of six successive lists, since the proportions were negligible for these pairs. The proportions for the /z : zj/ pairs are based on 120 observations. The proportions for the other pairs are based

---

Insert Table 10 about here

---

on 600 observations each. Among the pairs involving /p : b/ and /s : ts/ it is clear that the s pairs are easier than the d pairs. In the absence

TABLE 9

Proportion of Errors for s and d Pairs

| Type | All Items in Exp. Lists 1-6 | 7-24 | 1-24 | Items appearing in all lists 1-6 |
|---|---|---|---|---|
| s | .081 | .071 | .075 | .085 |
| d | .223 | .186 | .201 | .259 |

TABLE 10

Proportions of Errors on Various Types of Pairs on Sets of Six Successive Lists

| Type | Lists 1-6 | 7-12 | 13-18 | 19-24 |
|---|---|---|---|---|
| /b : b/ | .068 | .037 | .047 | .032 |
| /b : p/ | .447 | .313 | .287 | .158 |
| /p : b/ | .210 | .183 | .173 | .090 |
| /p : p/ | .162 | .143 | .133 | .080 |
| All /b, p/ | .222 | .169 | .160 | .090 |
| /s : s/ | .075 | .085 | .108 | .102 |
| /s : ts/ | .337 | .368 | .275 | .262 |
| /ts : s/ | .200 | .188 | .137 | .118 |
| /ts : ts/ | .058 | .043 | .045 | .033 |
| All /s, ts/ | .168 | .171 | .141 | .129 |
| /zi : zi/ | .067 | .067 | .175 | .100 |
| /zi : zji/ | .033 | .100 | .042 | .008 |
| /zji : zi/ | .125 | .125 | .233 | .075 |
| /zji : zji/ | .042 | .050 | .092 | .033 |
| All /z, zj/ | .067 | .085 | .135 | .054 |
| All items | .183 | .163 | .149 | .104 |

The /k : g/ and /t : d/ items were presented only in Lists 1-6. The proportions of errors were /g : g/ - .039, /g : k/ - .272, /k : g/ - .097, /k : k/ - .057, /d : d/ - .068, /d : t/ - .460, /t : d/ - .215, and /t : t/ - .137.

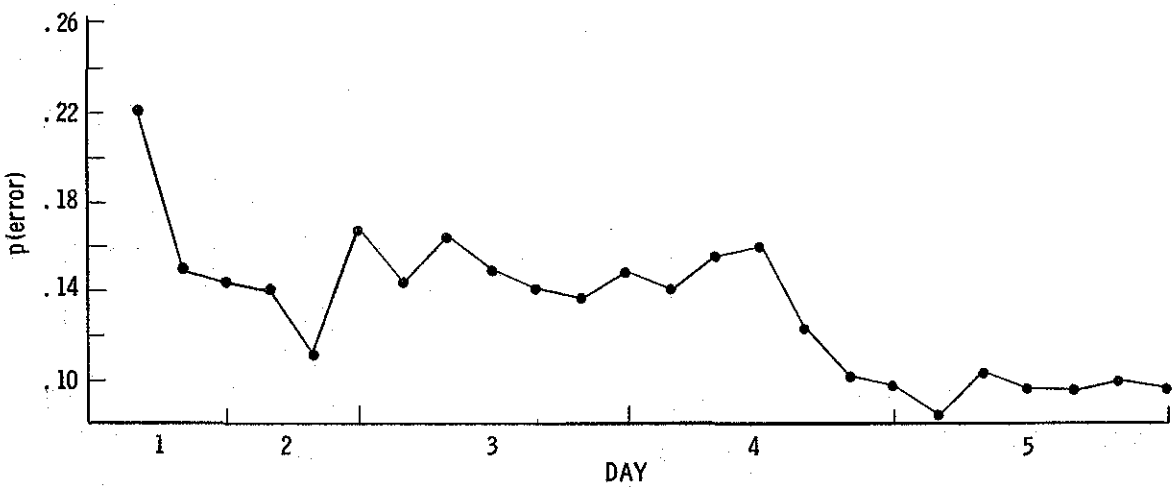



Fig. 1 Proportion of errors, calculated over all subjects and items, for each list on each day of the experiment.

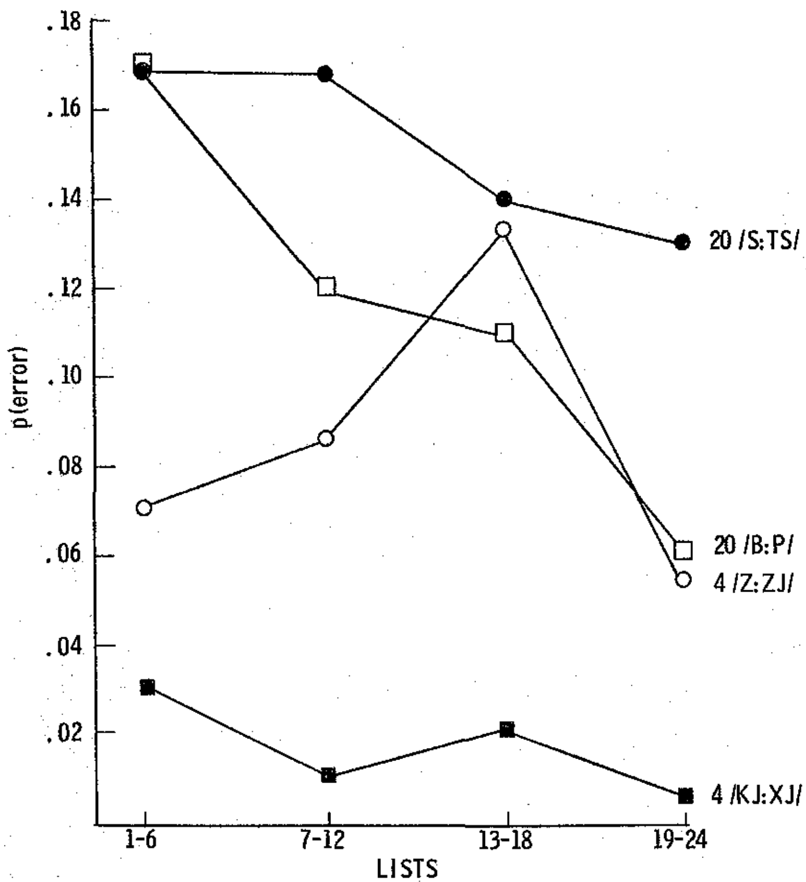
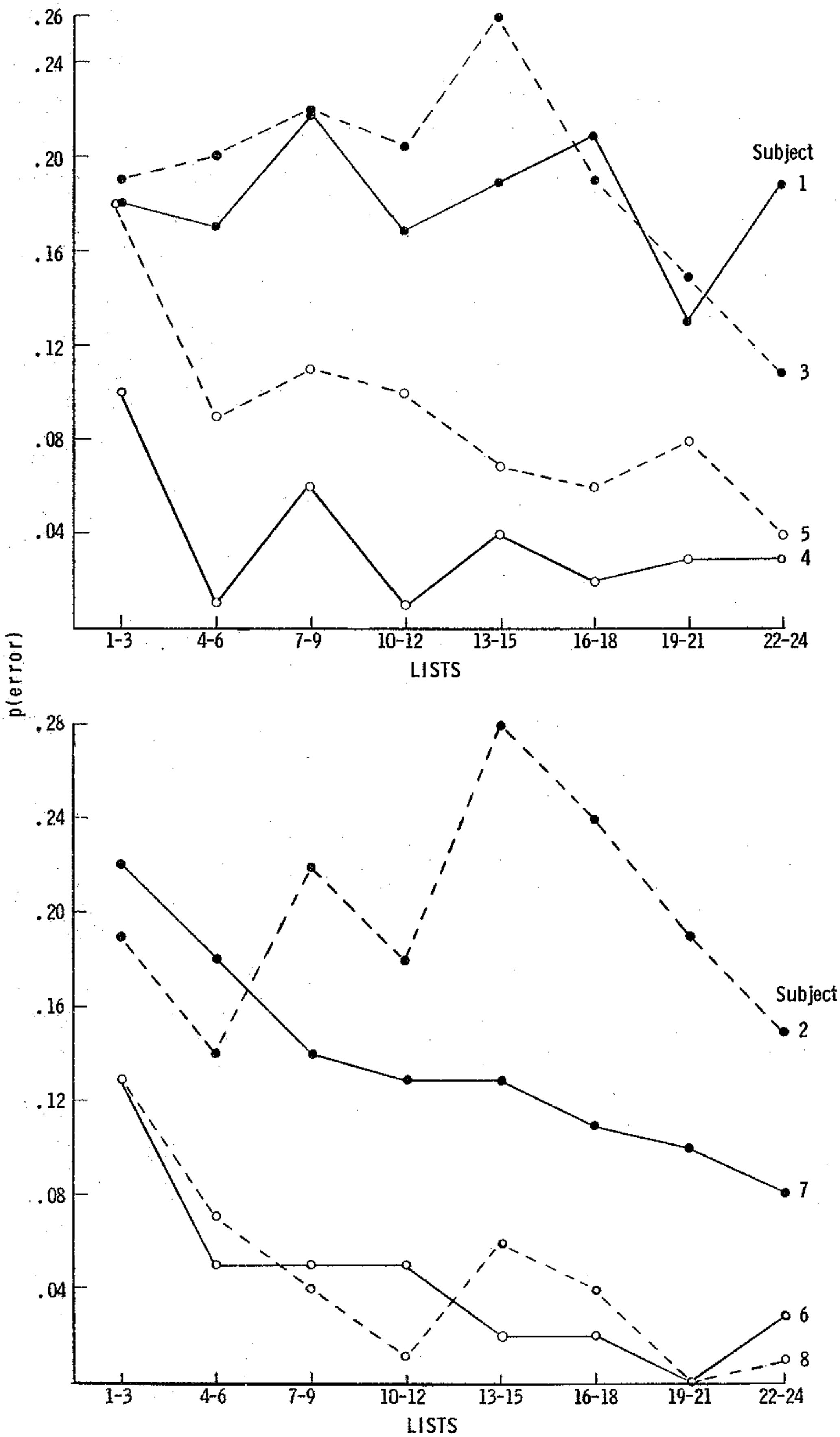
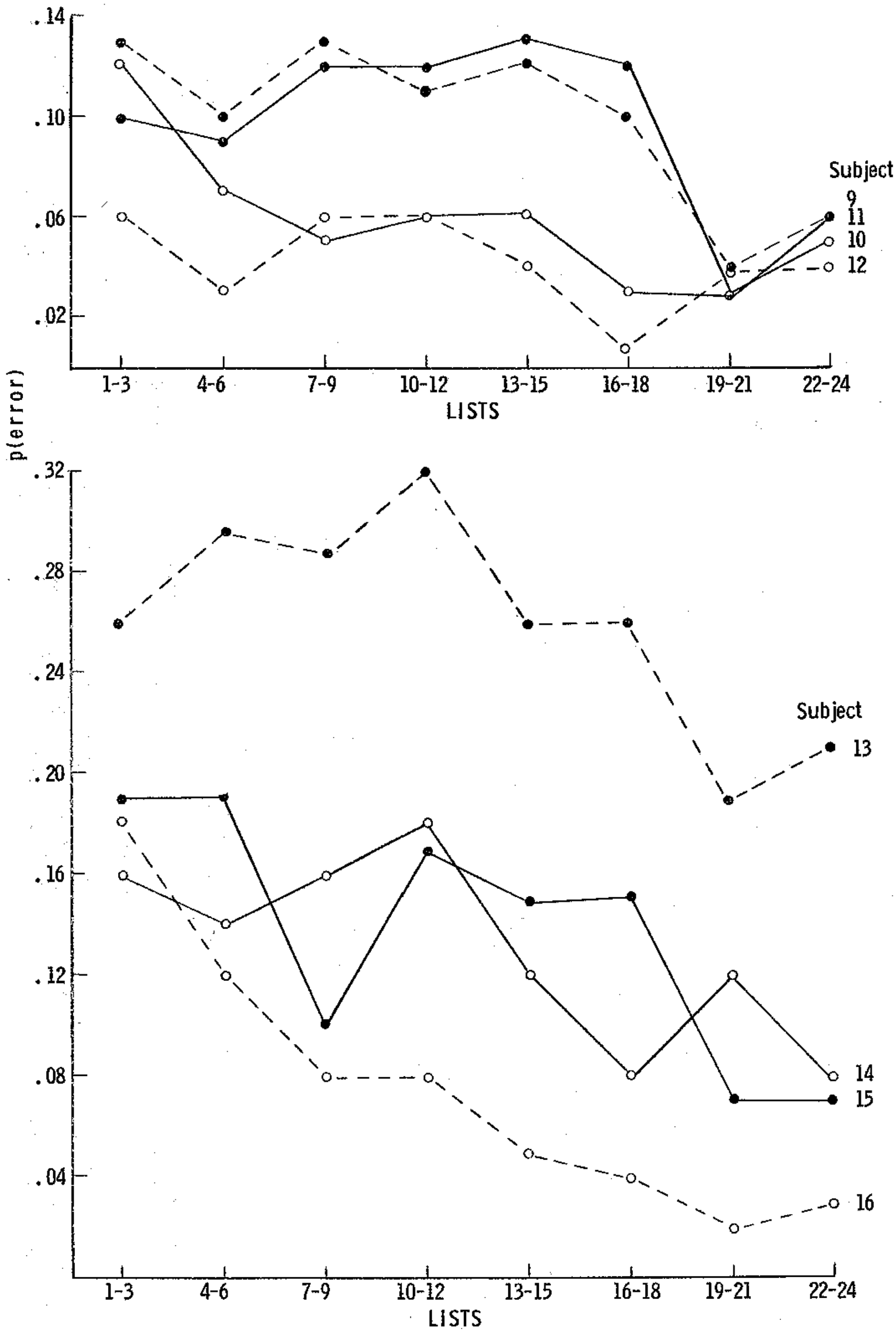



Fig. 2 Proportion of errors for each of the item types which appeared in all experimental lists.




Figs. 3, 4 Proportion of errors by individual subjects on each set of three successive lists.




Figs. 5, 6 Proportion of errors by individual subjects on each set of three successive lists.

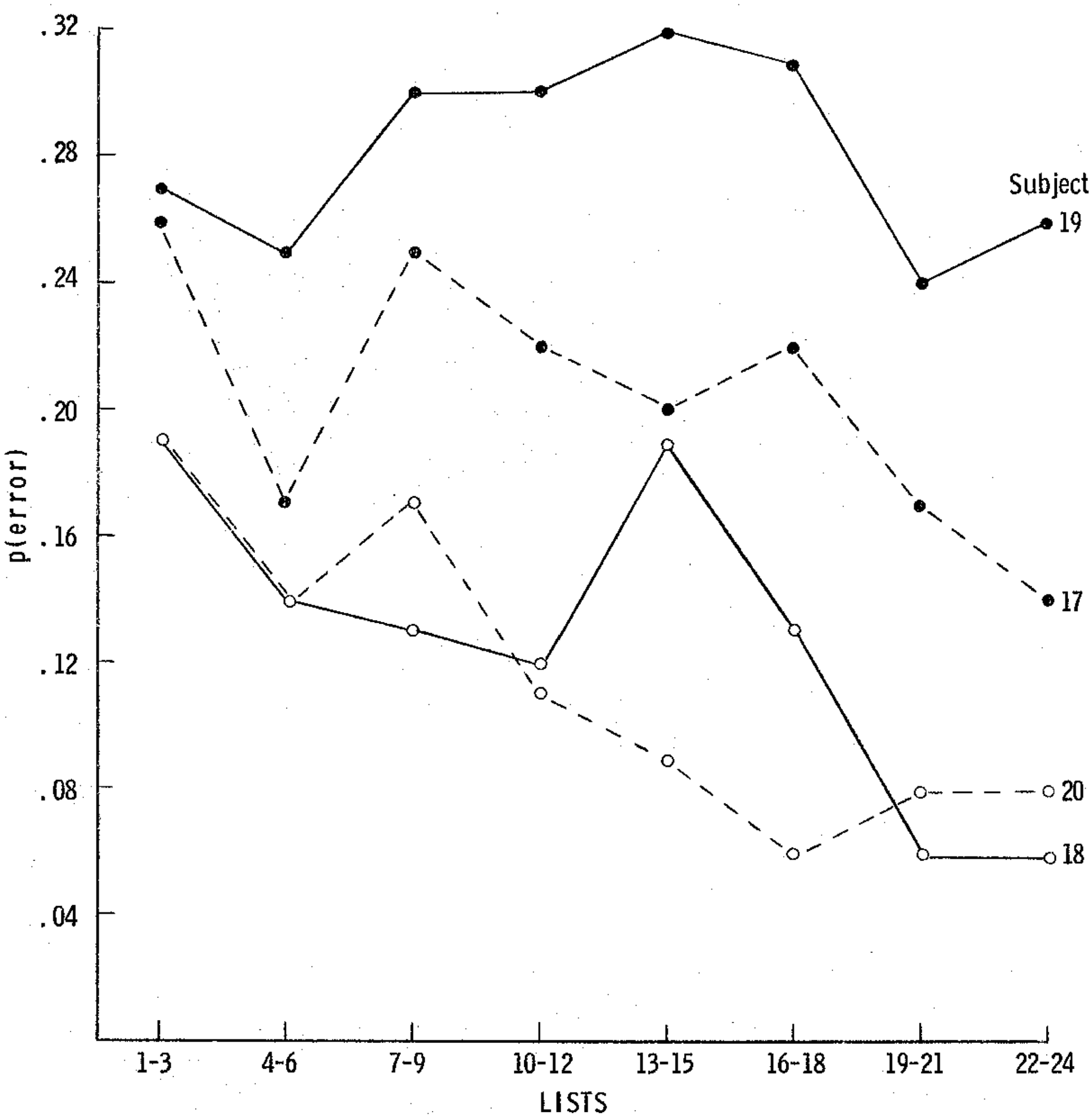


Fig. 7. Proportion of errors by individual subjects on each set of three successive lists.

of statistical analyses, we may roughly say that learning is seen clearly on all four pair types involving /p : b/, only on the /s : ts/ d pairs, and it is not appreciable on the /z : zj/ pairs.

Individual learning curves. Figures 3-7 display the proportion of errors computed over sets of three successive lists for each subject. From inspection, it appears that there are considerable inter-subject differences in the forms of the learning curves. A systematic interpretation of these data will be deferred until after the mathematical models have been presented in the next section. Table 11 gives the proportion of errors, computed across all lists, for all subjects. It is clear

---

Insert Figs. 3-7 about here

---

that there are substantial individual differences in discrimination

---

Insert Tables 11 and 12 about here

---

proficiency. The proportion of errors for individual subjects ranged from .041 to .278. Table 12 gives the results of the computations of mean and variance in total errors for subjects in 25th, 50th, 75th and 100th percentiles in total errors.

Tests of response independence. The first analysis sought to ascertain if the probability of a correct response was independent of the correctness of the response to the immediately preceding pair. Therefore, the probability of a correct response to a pair, given an incorrect response on the immediately preceding pair, was compared with the probability of a correct response, given a correct response to the preceding pair. One shortcoming of this independence test is that the proportion of correct following incorrect is computed largely from the slow learners, difficult pairs, and early stages of learning, while the proportion of correct after correct is based largely on the fast learners, easy pairs, and later stages of learning. An attempt to minimize this bias was made by computing the proportions separately for each subject and for each quartile of trials before the trial of last error. Appendix D gives the individual conditional proportions for each quartile, as well as the means over subjects. The p(correct|correct) entries are based on an average N of 50.5 while the average N for p(correct|incorrect) is 328. The mean proportion of correct following correct exceeds the mean proportion of correct following incorrect by .048, .007, .005 and -.004 in the first, second, third and fourth quartiles, respectively. After pairing the two conditional proportions for each subject, a sign test was run on the data of each quartile. The difference is significant at the .01 level for the first quartile and not significant thereafter. Thus, the probability of a correct response appears to be independent of the correctness of the preceding response after the first quartile.

The hypothesis that the response is independent of the immediately preceding pair type (s or d) was tested next. Since the reinforcement after each response informs the subject as to whether an s or d item had been presented, we in effect tested the assumption that the response is independent of the preceding reinforcement. The data from all subjects

TABLE 11

Overall Proportion of Errors for Each Subject in Experiment II

| Subject | p(Error) | Subject | p(Error) |
|---|---|---|---|
| 19 | .278 | 20 | .124 |
| 13 | .267 | 9 | .101 |
| 17 | .207 | 11 | .098 |
| 2 | .194 | 5 | .098 |
| 3 | .193 | 16 | .089 |
| 1 | .180 | 10 | .067 |
| 7 | .148 | 8 | .055 |
| 15 | .146 | 6 | .052 |
| 18 | .135 | 12 | .042 |
| 14 | .133 | 4 | .041 |

TABLE 12

Total Errors by Subjects at Different Performance Levels

| Subjects' Percentile in Total Errors | Total Errors | |
|---|---|---|
| | Mean | Variance |
| 0-100[1] | 159.6 | 8338.50 |
| 75-100 | 56.4 | 100.68 |
| 50-75 | 116.6 | 217.04 |
| 25-50 | 172.0 | 553.20 |
| 0-25 | 293.4 | 1934.64 |

[1] This percentile range includes all subjects.

were pooled and four $\chi^2$s were computed. The first two were chi-square independence tests computed for the case where the pair on the present (not preceding) trial was an s pair, one from the Lists 1-6 data and another from the Lists 7-24 data. In like manner, two $\chi^2$s were computed for the case where the pair on the present trial was a d pair. Table 13 gives the chi-square values obtained under the four conditions. Neither of the $\chi^2$s on the Lists 1-6 data are statistically significant, although both approach significance. On the other hand, for Lists 7-24,

---

Insert Table 13 about here

---

responses to the s pairs are dependent on the pair type presented on the preceding trial. ($\chi^2 = 6.173$, d.f. = 1, $.01 < p < .02$). A comparison of the observed and theoretical frequencies from which the significant $\chi^2$ was computed revealed that the observed frequency of correct response on an s pair, given an s pair on the preceding trial, exceeded the predicted frequency. Hence, by necessity the observed frequency of correct responses on an s pair, given a d pair on the preceding trial, fell short of the predicted frequency.

Analysis of variance of item and subject differences. First we ask whether the variance in total errors is due primarily to inter-subject differences or to differences in difficulty of various 4-concepts. To answer this question, a 20 subjects x 12 "hard" 4-concept analysis of variance was run, in which each cell entry represented the total errors on a given 4-concept by a given subject. As Table 14 shows, the variance due to each source is significant ($p < .001$). Therefore both the inter-subject and inter-concept differences are considerable. A second analysis represents one of the preliminary attempts to determine whether the linguistically defined concepts are responded to "as units". That is, if the /p : b/ pairs, for example, contain common cues which are a basis

---

Insert Table 14 about here

---

for including all of them in the same concept, one might expect some "transfer" between learning one subset of /p : b/ pairs and learning of another /p : b/ subset. It seems natural to choose as the subsets of each 20-concept the 4-concepts included in the particular 20-concept. (We recall that there are five 4-concepts included in the /p : b/ concept, since the /p : b/ pairs may appear with any of the five vowel phonemes. Likewise, there are five 4-concepts included in the /s : ts/ concept. The data for the /k : x/ and /z : zj/ pairs are less appropriate to the analysis, since each type includes only the 4-concept involving the vowel /i/). It seems that a rough index of "transfer" across 4-concepts within the same 20-concept may be obtained by comparing the correlations between total errors by a subject on one 4-concept and another. A correlation coefficient for each pair of 4-concepts was computed by matching each subject's total errors on one 4-concept with his total errors on the other 4-concept.

Table 15 presents the correlations between the number of errors on each pair of 4-concepts, and the mean and standard deviation of the

TABLE 13

Values of $X^2$ Obtained in Tests of Hypothesis that the Response on Trial n is Independent of the Type of Item Presented on Trial n - 1

| Item Type on Trial n | Lists 1-6 | Lists 7-24 |
|---|---|---|
| s | $3.53^{+}$ | $6.17^{++}$ |
| d | $3.66^{+}$ | 0.92 |

$^{+}$ $.05 < p < .10$

$^{++}$ $.01 < p < .02$

TABLE 14

Analysis of Variance - Hard 4-Concepts x Subjects; Lists 1-24

| Source of Variation | d.f. | Mean Square | F |
|---|---|---|---|
| 4-Concepts | 11 | 530.34 | 10.36+ |
| Subjects | 19 | 731.45 | 14.29+ |
| 4-Concepts x Subjects | 209 | 51.17 | |

+p < .001

number of errors on each concept. The coefficients range between .135 and .967. The entries enclosed by the same triangle represent correlations between pairs of 4-concepts included in the same 20-concept. The table

---

Insert Table 15 about here

---

reveals that, without exception, the correlations between total errors on pairs of 4-concepts are higher when the 4-concepts are included in the same 20-concept than when the two 4-concepts are from different 20 concepts. For the /s : ts/ pairs, the intercorrelations are remarkably high (.792-967), whereas the maximum correlation between an /s : ts/ 4-concept and a non-/s : ts/ 4-concept is .608. These high intercorrelations within a 20-concept contrast with the much lower correlations between total errors on /p : b/ and /s : ts/ 4-concepts which involve the same vowel phoneme (e.g., the correlation between /pa : ba/ and /sa : tsa/ errors is .357). The intercorrelations within a 20-concept indicate some learning of the general concept (e.g., /p : b/), but the lack of perfect correlation indicates that each 4-concept also presents unique stimuli to the subject.

A more sophisticated way of studying transfer across 4-concepts within a 20-concept is to examine the consequence of assuming that transfer is perfect, i.e., that all presentations of a given 20-concept represent repeated trials on that concept. This way of looking at the problem is developed later in connection with the application of the one-element model.

## 4. LINGUISTIC INTERPRETATION OF RESULTS

In this section, the results which seem particularly pertinent to the application of linguistics to teaching of a second language are summarized and interpreted.

Several limitations of the study should be borne in mind at the outset: (a) auditory discrimination was at issue exclusively, a much narrower field than the usual dimension in language learning; (b) the recording and playback apparatus although adequate, was not of professional quality; (c) a single native speaker was used throughout; (d) the subjects were homogeneous only in that they were students at Stanford University, and they did not know Russian. With these reservations, we now note the results which may have wider implications for second language learning.

Learning. As a result of about 125 minutes' exposure to the stimulus material and reinforcements of Experiment I, a fair amount of learning occurred. Under the schedule of one 25 minute session daily for five consecutive days, the overall proportion of errors dropped from $11^{o}/o$ on List 1 to $2^{o}/o$ on List 24. For Experiment II, with approximately the same exposure time, the drop was from $22^{o}/o$ to $10^{o}/o$. (See Fig. 1). Experiment I suggests that additional presentations of the voiced-voiceless contrasts might result in nearly perfect identification of them. On the other hand the failure to attain perfect performance in Experiment II suggests a not-surprising limitation in the pedagogic effectiveness of the experimental design. One possible technique for improving learning would be to abandon the random presentation order in favor of repeated trials on the same concept.

The fact that in Experiment II the reduction of errors occurred

TABLE 15

Intercorrelations between Number of Errors on Pairs of 4-Concepts

| Contrast - | /p:b/ | /p:b/ | /p:b/ | /p:b/ | /p:b/ | /s:ts/ | /s:ts/ | /s:ts/ | /s:ts/ | /s:ts/ | /kj:xj/ | /z:zj/ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Vowel - | /a/ | /e/ | /i/ | /o/ | /u/ | /a/ | /e/ | /i/ | /o/ | /u/ | /i/ | /i/ |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| 1 | | .752 | .673 | .863 | .686 | .357 | .329 | .301 | .258 | .286 | .301 | .290 |
| 2 | | | .705 | .905 | .762 | .529 | .502 | .591 | .426 | .438 | .137 | .258 |
| 3 | | | | .718 | .779 | .486 | .463 | .544 | .408 | .488 | .296 | .276 |
| 4 | | | | | .861 | .508 | .431 | .505 | .365 | .371 | .256 | .230 |
| 5 | | | | | | .407 | .374 | .528 | .323 | .407 | .239 | .135 |
| 6 | | | | | | | .967 | .837 | .949 | .900 | .553 | .526 |
| 7 | | | | | | | | .886 | .955 | .938 | .608 | .572 |
| 8 | | | | | | | | | .792 | .811 | .568 | .371 |
| 9 | | | | | | | | | | .943 | .539 | .601 |
| 10 | | | | | | | | | | | .574 | .547 |
| 11 | | | | | | | | | | | | .318 |
| 12 | | | | | | | | | | | | |
| Total Errors | | | | | | | | | | | | |
| Mean | 11.40 | 13.45 | 20.30 | 17.15 | 14.45 | 13.60 | 15.90 | 19.75 | 14.05 | 9.65 | 1.70 | 8.20 |
| S.D. | 8.18 | 8.66 | 9.63 | 11.90 | 9.34 | 12.69 | 13.98 | 11.63 | 12.25 | 9.91 | 2.10 | 4.65 |

primarily within Lists 1-6 and Lists 15-24 points to the plateau-type of language acquisition rather than continuous learning.

*s and d pairs.* When the two members of a pair are the same, it is easier to identify them as such than it is to identify as different the members of a *d* pair. In Experiment I, the overall error rate on *d* items was $5^{o}/o$, and dropped from $7.5^{o}/o$ on Lists 1-6 to $3.4^{o}/o$ on Lists 7-24. (See Table 4). By contrast, the overall error rate for *s* items was $3.4^{o}/o$, and the drop was from $4^{o}/o$ to $3^{o}/o$. The same findings were true for Experiment II, where the over-all error rate for *s* pairs was $7.5^{o}/o$ (with a drop from $8.1^{o}/o$ on Lists 1-6 to $7.1^{o}/o$, on Lists 7-24), and for *d* pairs $20^{o}/o$ (with a drop from $22^{o}/o$ on Lists 1-6 to $18^{o}/o$ on Lists 7-24.) This suggests that it is useful to present the material in the 4-concept approach of the experimental design in order to take advantage of the lower error rate of *s* pairs and the higher learning rate of *d* pairs. However, *d* pairs should also be presented at a higher ratio than *s* pairs.

That phonemes should be presented in pairs rather than individually, was not tested in view of experimental literature available, e.g., Pollack's (1952) findings on comparative versus individually presented sounds, which showed that a great many more sounds could be distinguished when presented in comparison.

It is also interesting to note that, as intended, learning proceeded in terms of phonemes and not allophones, and that over-discrimination of consonant allophones in *s* pairs and of vowel allophones in *d* pairs did not seem to occur.

*Consonant difficulty.* The consonants exhibited a definite order of difficulty. (See Table 1). Generally, fricatives were more readily discriminated than stops. This is due in part to the random noise characteristic of the former, usually more easily recognized than the complete absence of energy in the pre-released portions of the stops; but perhaps even more important here is the phonetic similarity of Russian and English fricatives. This is not true of the stop phonemes where the Russian voiceless stops are not highly aspirate as their English analogues are, and the voiced ones are fully voiced. Hence, due to their own linguistic background, the subjects had difficulty in discriminating between voiced and voiceless stops in Russian. The order of difficulty within the stops was unexpected: discrimination of labials, /p, b/, proved more difficult than that of dentals and velars, /t,d,k,g/, contrary to acoustic tests on burst perception.

The fricatives of Experiment I, /f, v, s, z, sh, zh/, were not included in Experiment II, whereas the stops were, with the addition of the pairs /d : dj, k : x, l : lj, n : nj, z : zj, s : sj, s : ts/. (See Table 6). There is a striking difference, between Experiments I and II, in the error rate on those items which were presented in both experiments. The difference in total exposures, as listed, although considerable, probably is not solely the cause of the difference: (The entries are the presentation frequencies for a given subject.)

| | /p,b/ | /t,d/ | /k,g/ |
|---|---|---|---|
| Exp I | 480 | 480 | 288 |
| Exp II | 300 | 120 | 72 |

We also ruled out the possibility that the subjects in Experiment I were more sophisticated linguistically than those in Experiment II. The questionaire data on their prior language training indicates that the two groups were comparable. It seems most likely that the inter-experiment differences in error rate on items common to both experiments is due to effects of the items unique to each experiment. In Experiment I, although both fricatives and stops were used, they were contrasted only within each of the two categories, and not across categories. Furthermore, the discrimination of the fricatives was quite easy so that the subjects could focus their attention on the stops. In Experiment II, on the other hand, in addition to the stops of the previous experiment, more difficult Russian consonant phonemes in relation to the English phonemic system were introduced, and contrasts were presented across categories, e.g., /k : x/, stop : fricative. Thus, the construction of lists for Experiment II was more intricate by far, and each pair required the subject to make a number of decisions in discrimination.

Another finding can best be interpreted in terms of a feature analysis of the phonemes, namely, the error rate of d pairs beginning with a voiceless consonant was consistently lower than that for pairs beginning with a voiced consonant. For example, many more errors were made on pairs of the /ba : pa/ type than on those of the /pa : ba/ type. We could therefore say that the addition of the feature of voicing to the second member of the pair made for better discrimination than the presentation of this added feature with the first member. In regard to fricative : affricate, however, the higher proportion of errors occurred consistently with the fricative as the first member, rather than the affricate. This is consistent with the interpretation of the affricate as a strident consonant rather than as a stop plus constrictive, a fact demonstrated by the lack of any intervening intensity minimum when the speech wave is analyzed as a function of time. (Jakobson, 1952).

*Vowel difficulty*. The data on the error rates of vowels in the two experiments are not quite comparable because only plain consonants were used in Experiment I, and plain and palatalized consonants in Experiment II, which require different vowel allophones. For Experiment I, the low central /a/ affects judging of consonants that precede it the least. The back vowels /u/ and /o/ affect judgement to some degree, about equally for both vowels. The front vowels have the most marked effect on discrimination, with /i/ causing much greater difficulty than /e/. (See Table 1).

Since the allophones of /i/, spelled "И" and "bl" respectively, after palatalized and plain consonants are quite different phonetically, the Experiment II items may be divided into two sets: one, plain contrasts only, and two, plain and palatalized contrasts. If we consider only the latter part of the experiment, in addition to the small absolute number of occurrences of /i/, the importance of the high error rate becomes apparent. This is not surprising in view of the nature of the English vowel system which causes the subjects to perceive the high front and high central allophones of the Russian /i/ as two separate vowels, a fact not true of the allophones of the other four Russian vowels. A correlate can be found in Halle's (1959) formant frequencies, computed from stationary portions of sonagrams, or if no such portions were available, from the middle point of the formant (p. 115):

| | | i | e | a | o | u |
|---|---|---|---|---|---|---|
| /p/ | $F_1$ | 200 | 400 | 700 | 400 | 275 |
| | $F_2$ | 1475 | 1875 | 1250 | 700 | 550 |
| | $F_3$ | 2125 | 2500 | 2200 | 2125 | 2150 |
| /pj/ | $F_1$ | 150 | 425 | 700 | 500 | 300 |
| | $F_2$ | 2150 | 1900 | 1375 | 1000 | 575 |
| | $F_3$ | 3000 | 2625 | 2250 | 2200 | 2500 |

The phoneme /i/ has by far the greatest discrepancy of formants, a fact with which our findings correlate.

In summary, except for (i) the discrepancies from a predicted order of difficulty which was based on discrimination and production rather than only on discrimination, and (ii) the relative difficulty of the discrimination of bilabials, the experimental linguistic findings fulfill most of the expectations resulting from a contrastive phonemic analysis of Russian and English.

## 5. QUANTITATIVE APPLICATION OF STIMULUS-RESPONSE THEORY

In the preceding sections, we have attempted to present in considerable detail the empirical results of the experiment, with little emphasis on a psychological interpretation of these results. In this concluding section of the Report, we turn to a quantitative analysis of the experimental data in terms of one fundamental stimulus-response theory of learning. The basic theory we apply originates with Estes' paper (1950); a large number of other investigators have contributed to the development of the theory in the past decade.

In a highly simplified form, the basic ideas are as follows. The organism is presented with a sequence of trials, on each of which he makes a response that is one of several possible choices. In any particular experiment it is assumed that there is a set of stimuli from which the organism draws a sample at the beginning of each trial; it is also assumed that on each trial each stimulus is conditioned to at most one response. The probability of making a given response on any trial is postulated to be simply the proportion of sampled stimuli conditioned to that response. However, if there are no conditioned stimuli in the sample, it is postulated that there is a "guessing" probability for each response, and this guessing probability is independent of the trial number and the past sequence of events. Learning takes place in the following way. At the end of the trial, a reinforcing event occurs identifying that one of the possible responses which was correct. With some fixed probability the sampled stimuli become conditioned to this response if they are not so already, and the organism begins another trial in a new state of conditioning. The sequence of events postulated to occur on a given trial may be illustrated by the following diagram:

state of conditioning → stimuli sampled → response made →

reinforcement occurs → reconditioning of sampled stimuli → new state of conditioning

Note that the trial begins with a certain kind of conditioning and ends with a new state of conditioning. This change of conditioning represents the most essential part of the learning process. (A more explicit formulation of these ideas is to be found in Suppes and Atkinson (1960)).

The four basic models we wish to describe here may be viewed as special cases of this general theory. Roughly speaking, they correspond to assuming that different numbers of stimuli are available for sampling on every trial. In this sense the different models correspond to postulating that a different number of stimulus components or patterns are sampled from the CV pairs presented to the subject on each trial in the present experiments.

<u>One-element model</u>. A simple model, and one that has proved empirically highly satisfactory in a wide range of experiments, is the one for which it is postulated that there is exactly one stimulus element which is available and sampled on each trial by the subject. A mathematical model that arises from this simple one-element assumption can be described in the following way. On every trial the subject is in one of two states: either the single element is conditioned (state C) to the correct response, in this case the verbal responses "same" or "different", or it is unconditioned (state U). We formulate the mathematical background of the

model in such a way that the subject's behavior forms a Markov process in these two states with the transition matrix indicated below.

| | | state on trial n | |
|---|---|---|---|
| | | C | U |
| state on trial n-1 | C | 1 | 0 |
| | U | c | 1-c |

The meaning of this matrix is simple. When the subject is in the unconditioned state there is a probability $c$ that he will move to the conditioned state. Once he becomes conditioned he remains so as indicated by the probability $1$. Secondly, we postulate that the subject guesses the correct response with probability $g$ when he is in the unconditioned state and responds correctly with probability $1$ when he is in the conditioned state.

From a psychological standpoint the simple one-element model represents conditioning as an all-or-none process. The assumption of a constant guessing probability on each trial before conditioning implies that there is a binomial distribution with parameter $g$ of responses prior to the last error. This observation has important consequences for the analysis of experimental data, the most important one being that the mean learning curve, when estimated over responses prior to the last error for each subject, should be a horizontal line. This is because on all trials prior to the last error the subject must be in the guessing state. Therefore his probability of making a correct response is constant (and equal to $g$) to these trials.

The observation that according to the model responses prior to the last error have a binomial distribution, suggests the consideration of a number of goodness-of-fit tests. The virtue of these tests is that they permit a genuine statistical evaluation of the null hypothesis that the model fits the data. Following the more detailed discussion in Suppes and Ginsberg (1961) there are four tests that are appropriate to apply. The statistical properties of these four tests are well known in the literature and do not need to be discussed here.

<u>Stationarity</u>. The first and most important test concerns the property already mentioned, namely that the mean learning curve when estimated over the responses prior to the last error is a horizontal line. The appropriate test in this case is the statistical test for stationarity, formulated in terms of the null hypothesis that there is no change in the proportion of correct responses over trials prior to the last error. Letting the variable $t$ run over blocks of trials the appropriate chi square test is as follows:

$$\chi^2 = \sum_{t,\, i} N\,(t) \left( \frac{n_i(t)}{n\,(t)} - \frac{n_i}{N} \right)^2 \Big/ \frac{n_i}{N} ,$$

where $i = 0, 1$, $n_i(t)$ is the number of correct $(i = 1)$ or incorrect $(i = 0)$ responses in block $t$, $n(t)$ is the total number of responses in block $t$, $n_i$ is the number of correct (or incorrect) responses summed over all blocks, and $N$ is the total number of responses summed over all blocks. The $\chi^2$ statistic has the usual limiting distribution with $T - 1$ degrees of freedom, where $T$ is the number of blocks of trials. (If there are $m > 2$ responses, the number of degrees of freedom is

$(m - 1)(T - 1)$.) Under the restriction to two responses, the expression for $\chi^2$ may be simplified to

$$\chi^2 = \sum_t \left[Nn_1(t) - n_1 n(t)\right]^2 \Big/ n_1 n_2 n(t) \quad ,$$

thus eliminating the summation over $i$ .

Order. The second test concerns the null hypothesis that the sequence of responses do indeed form a sequence of Bernoulli trials, i.e., that responses are statistically independent from one trial to another. The alternative hypothesis is that there is a first order dependence. The appropriate formulation of the chi square test is as follows

$$\chi^2 = \sum_{i,j} n_i \left(\frac{n_{ij}}{n_i} - \frac{n_j}{N}\right)^2 \Big/ \frac{n_j}{N} \quad ,$$

where $j$ as well as $i$ is $0$ or $1$, $n_{ij}$ is the number of transitions from state $i$ to state $j$, $n_i = \sum_j n_{ij}$ , $n_j = \sum_i n_{ij}$ , and $N$ is the total number of responses, as before. Again, $\chi^2$ has the usual limiting distribution with $(m - 1)^2$ degrees of freedom, where $m$ is the number of states; here, $m = 2$ . Acceptance of the null hypothesis has the strong implication that we cannot predict responses better if we know whether the preceding response was correct or incorrect.

Distribution of Responses. The third test concerns the question whether responses do indeed exhibit a binomial distribution. Because the number of responses prior to the last error varies from subject to subject and because, unless the number of subjects is very large, insufficient data will be obtained by grouping subjects together, the practical way to test this hypothesis is to consider blocks of trials in some given length, say four. On the null hypothesis that responses are statistically independent a standard chi square test for goodness of fit of the empirical histogram is appropriate.

Distribution of sequence of responses. We may go beyond the binomial distribution of responses to the more detailed question of the distribution of sequences of responses. Again we look at blocks of a given length, say, four, and in this case ask if the sixteen possible sequences of four responses exhibit the appropriate distribution. A chi square test may again be applied in exactly the manner appropriate to the distribution of responses themselves.

It also may be remarked that the distribution of last errors may be examined from a statistical standpoint but unfortunately, in the present experiments, the number of subjects reaching criterion was too small to provide adequate data.

The four tests just described were applied to the group data for Experiment II. However, one important point of interpretation for application of the model needs to be mentioned. We may apply the one-element model at different concept levels. We mean by this the following: in a variety of experiments the one-element model has been successfully interpreted in terms of a conditioning association between a concept and the correct response, e.g., the concept of a geometrical form like a quadrilateral or an abstract concept like that of identity of sets. The association need not be between a particular stimulus and the correct response. In the present experiments, as the preceding analysis has already indicated, it is possible to identify several levels of concepts. We shall indeed apply the model to the 1-concepts, 4-concepts, and 20-concepts already defined.

<u>Two-element model</u>. Because the one-element model does not adequately fit the data of Experiment II, it is necessary to consider additional, more complicated models that may be derived from the fundamental theory. The next step beyond postulating that conditioning is an all-or-none process is to postulate that learning takes place in two stages. In particular, we assume that associated with each situation are two stimulus elements and, therefore, that the learning proceeds in two stages of all-or-none conditioning. Each of these two elements is conditioned on an all-or-none basis but the two parameters of conditioning, one for each element, may be adjusted to produce various incremental effects on the response probabilities. Let $\sigma$ and $\tau$ be the two elements. The basic learning process may be represented by the following four-state Markov process where the four states $(\sigma,\tau)$, $\sigma$, $\tau$, and $0$ represent the possible states of conditioning of the two-stimulus elements.

| | $(\sigma,\tau)$ | $\sigma$ | $\tau$ | $0$ |
|---|---|---|---|---|
| $(\sigma,\tau)$ | 1 | 0 | 0 | 0 |
| $\sigma$ | $b'/2$ | $1-b'/2$ | 0 | 0 |
| $\tau$ | $b'/2$ | 0 | $1-b'/2$ | 0 |
| $0$ | 0 | $a/2$ | $a/2$ | $1-a$ |

Because we do not attempt experimentally to identify the stimuli $\sigma$ and

$\tau$, this Markov process may be collapsed into a three-state process, whose states are simply the <u>number</u> of stimuli conditioned to the correct response. In the matrix shown above $a$ is the probability of conditioning at the second stage. The division by 1/2 in the matrix simply represents the equal probability of sampling one of the two elements. If we consider only the number of stimuli, it is convenient to replace $\frac{b'}{2}$ by $b$ and we obtain the transition matrix shown below:

| | 2 | 1 | 0 |
|---|---|---|---|
| 2 | 1 | 0 | 0 |
| 1 | b | 1-b | 0 |
| 0 | 0 | a | 1-a |

To complete the description of the process we associate with the states 0 and 1 the guessing probabilities $g_0$ and $g_1$. This means that we now have a process with four free parameters, the conditioning parameters $a$ and $b$ and the guessing probabilities $g_0$ and $g_1$. In actual fact, in terms of the methods we shall use for analyzing data, these four parameters are reduced to three, because we shall only consider response data prior to the last error. This means that the second conditioning parameter $b$ will not enter into the analysis of data, for the subjects must be in state 0 or 1 prior to the last error. Necessarily a transition from state 1 to 2 cannot have occurred at this stage.

Estimation of the three parameters can be approached in a number of ways. In the application considered below, we shall restrict ourselves to a consideration of data from individual subjects, that is, the estimation of parameters shall be for individual subjects and not for group data. This introduces a considerable simplification both in the estimation of parameters and the analysis of goodness of fit. It has the particularly desirable feature of eliminating any problems concerning homogeneity of parameters across subjects.

When learning data for individual subjects are considered, it is apparent what sort of learning curve is predicted by the two-element model. The learning curve is simply a step function with the first step being at levels $g_0$, the second at level $g_1$ and the third at level 1, corresponding to the probability 1 of a correct response in state 2. From analysis of individual data it is of course impossible to tell exactly when a subject passes from state 0 to state 1. What we have done is to apply a method of least squares in the following manner. We divide the data for each subject into octiles preceding the last error. On the assumption that the transition from state 0 to state 1 occurred at the $j^{th}$ octile we fit $g_0$ and $g_1$ by the method of least squares. This estimation is performed for each octile. We then select as the point of transition from state 0 to state 1 the octile which has the minimum least squares deviation. The equation for the least squares function $f(j)$ for the $j^{th}$ octile is as follows:

$$f(j) = \sum_{i=i}^{j} (x_i - \hat{g}_0)^2 + \sum_{i=j+1}^{8} (x_i - \hat{g}_1)^2 \tag{1}$$

where $x_i$ is the observed proportion of correct responses in the $j^{th}$ octile. Taking partial derivatives with respect to $g_0$ and $g_1$, we then obtain the following two equations, which were used to estimate $g_0$ and $g_1$:

(2) $$\hat{g}_0 = \frac{\sum_{i=1}^{j} x_i}{j}$$

(3) $$\hat{g}_1 = \sum_{i=j+1}^{8} \frac{x_i}{8-j}$$

The computations are done for $j = 1, 2, \ldots, 8$. The case $j = 8$ means that only state 0 occurs and hence is equivalent to the one-element model. It should be realized in passing from the one-element to the two-element model that any simple operational identification of the two elements is not possible. It is a common question to ask what the two elements correspond to in the stimulus material heard by the subject. Various psychological interpretations of the two elements can be given, but at the present stage of research it does not seem possible to identify them psychologically in any experimentally definite manner. Perhaps the most suggestive way to think about these two elements is that they correspond to the two most important aspects or properties of the stimulus material.

Because of the theoretical character of the two elements postulated in the two-stage model, there is no real reason to restrict the analysis to two elements. In other experimental situations (see, for example, Chapter 10 of Suppes and Atkinson (1960)) the number of stimuli has been estimated for the data. Because of the relatively small number of observations for individual subjects, we have not attempted this extension in the analysis of Experiment II. This would be possible if the data from subjects were combined, but we feel that in the present experiment the heterogeneity of individual subject behavior is sufficiently great to argue against this approach.

Linear model. Another alternative model that we wish to consider is the linear incremental model with a single operator. The intuitive idea of this model is precisely the opposite of the all-or-none conditioning model. The supposition is that learning proceeds on an incremental basis. Let $q_n$ be the probability of an error on trial $n$. Then the model is formulated by the following recursive equation:

$$q_{n+1} = (1-\theta)q_n , \tag{4}$$

where $0 < \theta \leq 1$. It is simple to show but somewhat surprising that this purely incremental model has precisely the same mean learning curve as the all-or-none model if we set $c = \theta$. (To obtain this identity of the learning curves we must, of course, consider all responses and not simply responses prior to the last error.) The incremental model does differ sharply from the all-or-none model in the kind of learning curve predicted for responses prior to the last error, as is evident from equation (4).

The estimation of $q_1$, the initial probability of an error and $\theta$, the learning rate, was performed as in the case of the two-element model, by minimizing the sum, over octiles of the squared deviation between the predicted and observed frequencies of correct response. The equation used was

$$(5) \qquad f(q_1, \theta) = \sum_{i=i}^{8} \left\{ F_i - \sum_{n \in O_i} \left[ 1-(1-\theta)^{n-1} q_1 \right] \right\}^2 ,$$

where $F_i$ is the observed frequency of correct responses in the $i^{th}$ octile, and the inside summation is over all trials in that octile. The parameter estimation consisted in arbitrarily fixing $\hat{\theta}$ at a predetermined value (the range $.00 \leqq \hat{\theta} \leqq .06$ proved suitable and was explored in small increments of $\hat{\theta}$) and then computing the $q_1$ value which minimized $f(q_1, \hat{\theta})$. The $\hat{q}_1$ and $\hat{\theta}$ yielding $\min_{\hat{q}_1, \hat{\theta}} f(\hat{q}_1, \hat{\theta})$ were selected as the parameter estimates.

Concerning the psychological interpretation of the linear model, it may be remarked that it corresponds to assuming that there is a very large population of stimulus elements and that a fixed proportion of these elements are sampled on every trial (or equivalently, that each element is sampled with an independent probability $\theta$). Prior to detailed empirical investigation of goodness of fit, it is a plausible hypothesis that for material as perceptually complicated as the linguistic stimuli used in the present experiment the linear model would fit better than the simple all-or-none one or two-element models. In this case the assumption that the population of stimuli is very large corresponds psychologically to assuming that the subjects are responding to a very large number of aspects or properties of the stimulus material.

We turn now to the experimental comparison of the various models. We first consider the chi-square tests for stationarity, order and binomial distribution of responses outlined above in connection with the one-element model.

Stationarity tests. Table 16 gives the results of stationarity chi-square tests at each level of concept analysis. Each $\chi^2$ was computed in

---

Insert Table 16 about here

---

the following manner. The data were analyzed in sets of four successive responses in each protocol. Therefore, n(t), the number of responses in block t, was four multiplied by the number of protocols. Whenever the responses in a given protocol met the criterion of thirty successive correct, that protocol was eliminated from the computation. The computation was terminated after reaching the highest block number such that fewer than half of the protocols had been eliminated. The initial number of protocols was equal to the product of the number of subjects (20) and the number of concepts at the given level of analysis (e.g., 48 for the 1-concept, 12 for the 4-concept, one for the /p : b/ 20-concept). So the initial block size was 3840 for the 1-concept, 960 for the 4-concept, 80 for the /p : b/ 20-concept, etc. For the 24 s pairs (24 concept), the 1-concept, and the /k : x/ 4-concept, the results show significant non-stationarity at the .05, .02, and .01 levels of confidence, respectively. In all other cases, the stationarity hypothesis is rejected at the .001 level or beyond. When the number of degrees of freedom (one less than the number of blocks) for a given $\chi^2$ exceeded 30, the normal approximation $z = \sqrt{2\chi^2} - \sqrt{2m-1}$ was used. We conclude that this analysis shows that there seems to be no obvious classification of pairs into linguistically defined concepts such that stationarity tests on group data yield the result demanded by the one-element model.

Order tests. Table 17 presents the results of the chi-square tests of the hypothesis that responses on successive presentations of a given concept are independent of each other. The tests indicate independence

---

Insert Table 17 about here

---

only for the /kj : xj/, /z : zj/ and s pairs. It should be noted that these are precisely the pairs which exhibit the lowest error rates (Table 6). For the 4-concept, non-independence of successive responses is indicated at the .05 level. For the remaining concepts, the hypothesis of response independence is rejected at beyond the .001 level. It is interesting to contrast these findings with those in Section 3. In that analysis, the sequence of responses examined for independence was the subject's original sequence of responses in the order that they occurred, regardless of the concept type. In the present analysis "successive" refers to instances of the same concept rather than to the entire sequence of responses. The analyses indicate independence of successive responses, but not of responses to successive presentations of the same concept.

Tests for binomial distribution of responses. Table 18 presents the results of the chi-square tests of the hypothesis that the distribution of responses prior to learning is binomial. We consider blocks of trials of length four, and take for each subject the highest multiple of the block length equal to or less than the total number of responses prior to last error. We then sum over subjects the total number of such blocks and construct the histogram of the frequency of 0, 1, 2, 3, or 4 errors. For the 1-, 4-, /p : b/, /s : ts/, voiced-voiceless, and d concepts, the departures from the binomial distribution are significant at beyond the .001 level. The only concepts whose response distributions

---

Insert Table 18 about here

---

are not significantly different from the binomial distribution are those which consistently exhibit the lowest error rates. The remark made in conjunction with the order tests, viz., that tests based on lower proportions of errors are less likely to reveal departures from predicted properties, applies here also.

Tests for binomial distribution of response sequences. Here we look at sequences of responses, such as "error-correct-correct-error" within blocks of four successive trials. There are $2^4 = 16$ such sequences, and we wish to compose the observed and predicted frequency of each sequence. Proceeding as with the distribution of responses, we perform a chi-square test of the goodness of fit of the empirical histograms. As we would anticipate from the results of the preceding tests, the tests of the hypothesis that the frequencies of the possible sequences of four successive responses are binomially distributed indicate significant departure from that distribution (Table 19). Even the concepts (except the s

---

Insert Table 19 about here

---

concept) which yielded non-significant $\chi^2$s on the response distribution tests yielded significant $\chi^2$s on this response sequence distribution test.

TABLE 16

Results of Stationarity Tests

| Level of Analysis | Number of Members of Concept | $X^2$ | d.f. | p |
|---|---|---|---|---|
| Single Pair[1] | 1 | 7.83 | 2 | <.02 |
| Pairs with Same Contrast and Same Vowel[1] (4-concept) | 4 | 52.34 | 16 | <.001 |
| /p : b/ | 20 | 485.42 | 89 | <.001 |
| /s : ts/ | 20 | 341.21 | 117 | <.001 |
| /kj : xj/ | 4 | 32.24 | 14 | <.01 |
| All Voiced-Voiceless Stops | 52 | 649.48 | 137 | <.001 |
| s Pairs[1] | 24 | 16.30 | 7 | <.05 |
| d Pairs[1] | 24 | 267.33 | 100 | <.001 |

1 Includes only pairs which appeared in all lists.

TABLE 17

Results of Order Tests

| Level of Analysis | Number of Members of Concept | $X^2$ | d.f. | p |
|---|---|---|---|---|
| Single Pair[1] | 1 | 941.94 | 1 | $<.001$ |
| Pairs with Same Contrast and Same Vowel[1] (4-concept) | 4 | 4.93 | 1 | $<.05$ |
| /p : b/ | 20 | 11.89 | 1 | $<.001$ |
| /s : ts/ | 20 | 31.16 | 1 | $<.001$ |
| /kj : xj/ | 4 | .75 | 1 | $<.50$ |
| /z : zj/ | 4 | .04 | 1 | $<.90$ |
| All Voiced-Voiceless Stops | 52 | 28.03 | 1 | $<.001$ |
| s Pairs[1] | 24 | .18 | 1 | $<.70$ |
| d Pairs[1] | 24 | 41.76 | 1 | $<.001$ |

[1] Includes only pairs which appeared in all lists.

TABLE 18

Results of Distribution of Responses Tests

| Level of Analysis | Number of Members of Concept | $\chi^2$ | d.f. | p |
|---|---|---|---|---|
| Single Pair[1] | 1 | 782.92 | 3 | <.001 |
| Pairs with Same Contrast and Same Vowel[1] (4-concept) | 4 | 33.03 | 2 | <.001 |
| /p : b/ | 20 | 31.85 | 2 | <.001 |
| /s : ts/ | 20 | 40.07 | 2 | <.001 |
| /kj : xj/ | 4 | .12 | 1 | <.80 |
| All Voiced-Voiceless Stops | 52 | 36.31 | 3 | <.001 |
| s Pairs[1] | 24 | .01 | 1 | <.95 |
| d Pairs | 24 | 45.94 | 3 | <.001 |

[1] Includes only pairs which appeared in all lists.

TABLE 19

Results of Distribution of Response Sequences Tests

| Level of Analysis | Number of Members of Concept | $X^2$ | d.f. | p |
|---|---|---|---|---|
| Single Pair[1] | 1 | 991.35 | 13 | <.001 |
| Pairs with Same Contrast and Same Vowel[1] (4-concept) | 4 | 47.82 | 14 | <.001 |
| /p : b/ | 20 | 59.39 | 14 | <.001 |
| /s : ts/ | 20 | 78.50 | 14 | <.001 |
| /kj : xj/ | 4 | 6.10 | 1 | <.02 |
| All Voiced-Voiceless Stops | 52 | 75.51 | 14 | <.001 |
| s Pairs[1] | 24 | 10.12 | 5 | <.10 |
| d Pairs[1] | 24 | 59.54 | 14 | <.001 |

1 Includes only pairs which appeared in all lists.

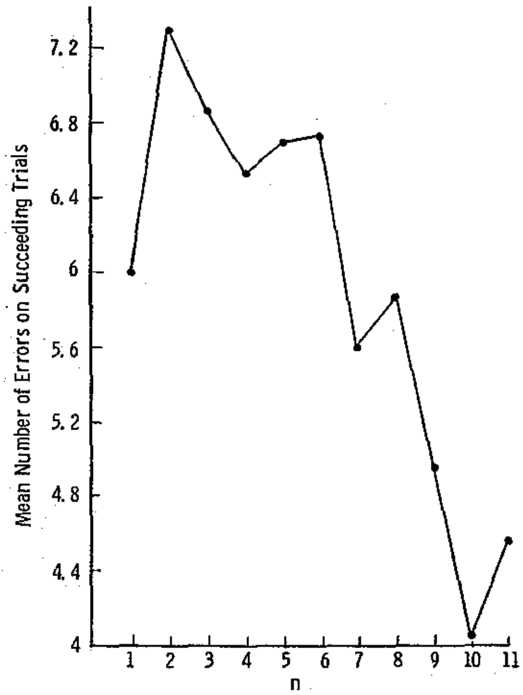



Fig. 8 1-concept. Mean number of errors following an error on trial n. The computation is over subject-items missed on trial n.

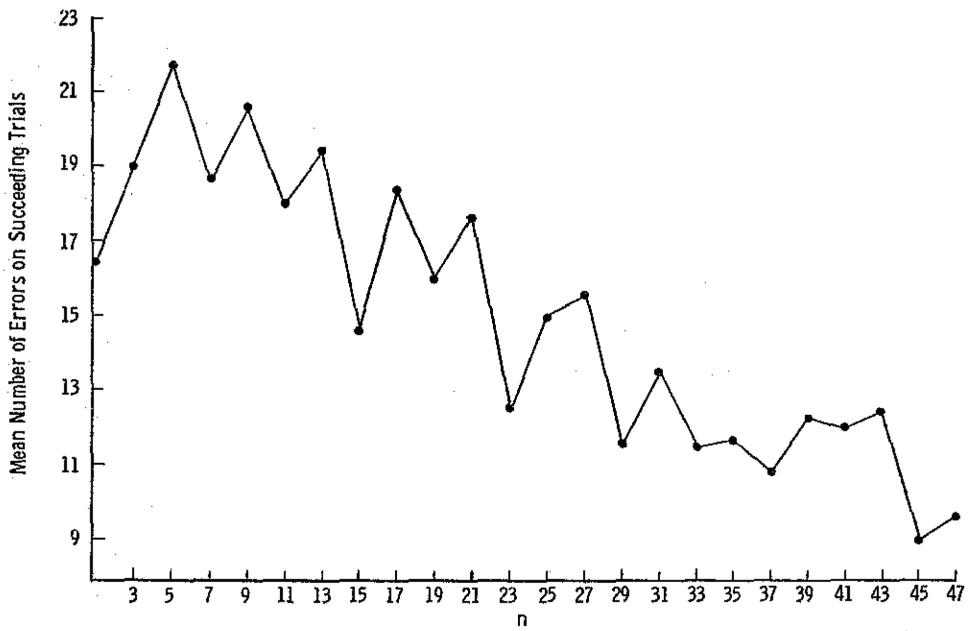



Fig. 9 4-concept. Mean number of errors following an error on trial n. The computation is over subject-items missed on trial n.

Another stationarity prediction (Bower, 1961) based on the binomial properties of the one-element model involves the analysis of sequences of responses by a given subject to a given concept. The prediction is that, given an error on the $n^{th}$ presentation of a certain concept, the number of times that the subject misses that concept on subsequent presentations should be independent of n. This is because an error, regardless of when it occurs, is assumed to imply that the concept is completely unlearned. Hence the expected number of subsequent errors on that concept is independent of n. Figures 8 and 9 show the curves of the number of errors plotted against n for the 1- and 4-concepts. The data are plotted

---

Insert Figs. 8 and 9 about here

---

for the first half of the trials (minus one) at each of the two levels of analysis. For both the 1-concept and the 4-concept the curves generally decline, instead of remaining horizontal as predicted by the one-element model.

Vincent curves of group data. The Vincent curves for a given pair type were plotted by dividing the trials prior to last error on that type into quartiles for each subject, finding the number of errors per quartile, and adding over subjects. Table 20 gives the number of errors and the

---

Insert Table 20 about here

---

number of responses per quartile for each concept analyzed. From these data, the proportions of correct responses per quartile were graphed (Figure 10). These data were also used to compute chi-square tests of

---

Insert Fig. 10 about here

---

the hypothesis that the number of errors per quartile is stationary for a given concept. Except for the /k : x/ concept, the obtained chi-square values indicate significant departure from stationarity (Table 21). This finding agrees with the corresponding analysis of the non-Vincentized

---

Insert Table 21 about here

---

data (Table 16). Figure 10 shows exactly what patterns of non-stationarity occurred. Discounting the less interesting types /kj : xj/ and /z : zj/, the general trend is an increasing, negatively accelerated curve through the first three quartiles, and an upward "spurt" in the fourth quartile. This "midplateau" has been found by Zeaman et al. in studies of discrimination learning by retarded children (1961). It occurs between quartiles 2 and 3 in five of the six curves (of course, this is consistent with the group learning curve of Figure 1). However, there are several reasons why we do not wish to emphasize the "midplateau". In the first place, our evidence would be more convincing if the curves were based on independent observations of different sets of items. Secondly, the effect did not appear when the data of Experiment I were plotted in Vincentized form. There, a monotonically increasing, negatively accelerated curve appeared

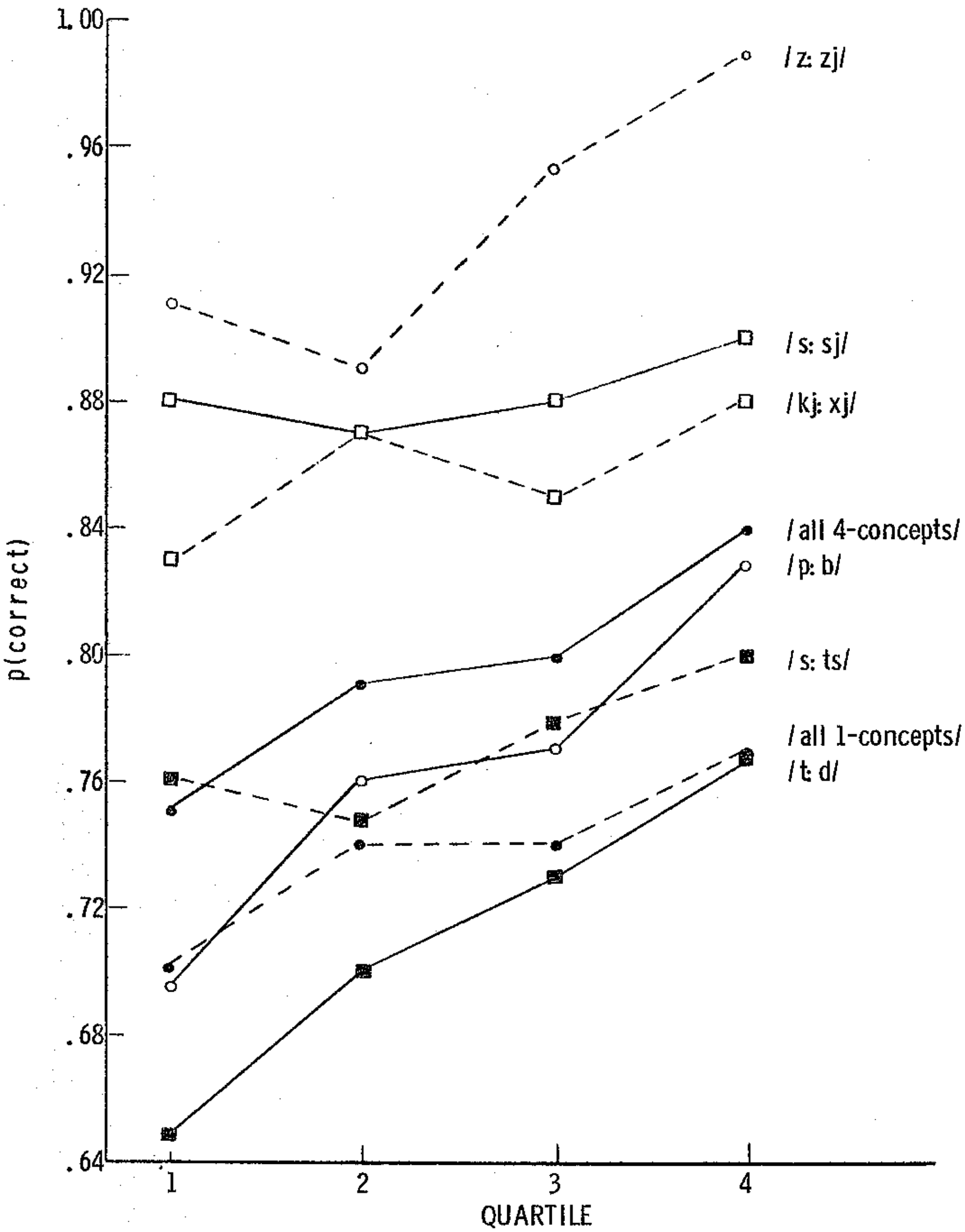



Fig. 10 Observed proportion of correct responses in each quartile of Vincentized group data. The curves are displayed separately for each level of concept analysis.

TABLE 20

Number of Errors n(i) in Quartile i (i = 1, 2, 3, 4) and Number of Responses per Quartile (N) at Each Level of Concept Analysis

| Concept | Number of Members of Concept | Number of Members per Instance | N | n(1) | n(2) | n(3) | n(4) |
|---|---|---|---|---|---|---|---|
| One[1] | 48 | 1 | 2182 | 635 | 569 | 539 | 498 |
| Four[1] | 12 | 4 | 3320 | 819 | 710 | 690 | 541 |
| /p : b/ | 1 | 20 | 1415 | 418 | 334 | 318 | 244 |
| /s : ts/ | 1 | 20 | 1339 | 321 | 341 | 295 | 268 |
| /kj : xj/ | 1 | 4 | 223 | 26 | 29 | 34 | 27 |
| /z : zj/ | 1 | 4 | 35 | 3 | 4 | 2 | 1 |
| All Voiced-Voiceless Stops[2] | 1 | 52 | 2210 | 588 | 489 | 497 | 391 |
| s[1] | 1 | 24 | 427 | 74 | 55 | 53 | 44 |
| d[1] | 1 | 24 | 1797 | 628 | 542 | 516 | 407 |

1 Includes only pairs which appeared in all lists.

2 Includes the 12 /k : g/ pairs and 20 /t : d/ pairs present in Lists 1-6, as well as the 20 /p : b/ pairs present in all lists.

TABLE 21

Results of Stationarity Tests on Vincentized Data

| Level of Analysis | Number of Members of Concept | $\chi^2$ | d.f. | p |
|---|---|---|---|---|
| Single Pair[1] | 1 | 23.99 | 3 | <.001 |
| Pairs with Same Contrast and Same Vowel[1] (4-concept) | 4 | 71.79 | 3 | <.001 |
| /p : b/ | 20 | 60.62 | 3 | <.001 |
| /s : ts/ | 20 | 12.76 | 3 | <.01 |
| /kj : xj/ | 4 | 1.51 | 3 | <.70 |
| /z : zj/ | 4 | 2.15 | 3 | <.70 |
| All Voiced-Voiceless Stops | 52, 20 | 50.91 | 3 | <.001 |
| s Pairs[1] | 24 | 9.73 | 3 | <.05 |
| d Pairs[1] | 24 | 67.11 | 3 | <.001 |

[1] Includes only pairs which appeared in all lists.

for each concept analyzed.

Individual Vincent curves. We next considered several refinements in the analyses of Vincent curves: (a) the Vincent curves were plotted for individual subjects; (b) to permit closer evaluation of the learning models, the trials prior to last error were divided into octiles rather than into quartiles; (c) two breakdowns, each into four sets, of the more difficult pairs were studied; the sets are indicated below;

| Pair type | First Classification | Second Classification |
|---|---|---|
| /p : b/ | s vs. d | /p/ vs. /b/ as first member of pair |
| /s : ts/ | s vs. d | /s/ vs. /ts/ as first member of pair. |

Each set represents ten pairs (e.g., the /s : ts/ s class includes one /s : s/ and one /ts : ts/ pair with each of the five vowels). (d) The preceding steps greatly reduced the number of responses and errors per octile, so the analysis was restricted to those sets of pairs which contained enough errors to provide worthwhile tests of the models. One important fact about the Vincent curves for these "hard" sets must be noted; viz., that the learning criterion was not met in these cases. Hence, only the initial portion of the prelearning trials, rather than all of them, has been divided into octiles. The criterion for "enough" errors in a given response sequence was more than fifteen in at least one octile. (An octile could include up to 30 responses). Fifty-four such sets of octile data met this criterion. The first two columns of Table 22 list the subject and pair type which contributed each of these sets of data.

Insert Table 22 about here

Tests of the two-element and linear models.

Goodness of fit of predicted Vincent curves. Each of three models was applied to each of the 54 sets of octile data. For the slope-intercept and linear models, two parameters were estimated separately for each of the sets of data. Three parameters were estimated in each case for the two-element model. The manner of application of the models will be discussed next.

Two-element model. As mentioned earlier, this model assumes that, prior to the trial of last error on a given concept, the two-element stimulus set representing the concept may pass from the initial state in which neither element is conditioned to the correct response to the intermediate state in which one of the two elements is conditioned. Therefore, before predictions regarding the data before last error can be made, it is necessary to estimate the guessing probabilities $g_0$ and $g_1$, and the trial on which transfer from state 0 to state 1 occurred. The exact manner of estimating these quantities is given above (see equation (1)-(3)).

It is of some interest to note the distribution of the frequency of the various j estimates (Figure 11). According to the two-element model,

Insert Fig. 11 about here

the passage from state 0 to state 1 occurs most often at octiles 1, 2, 6,

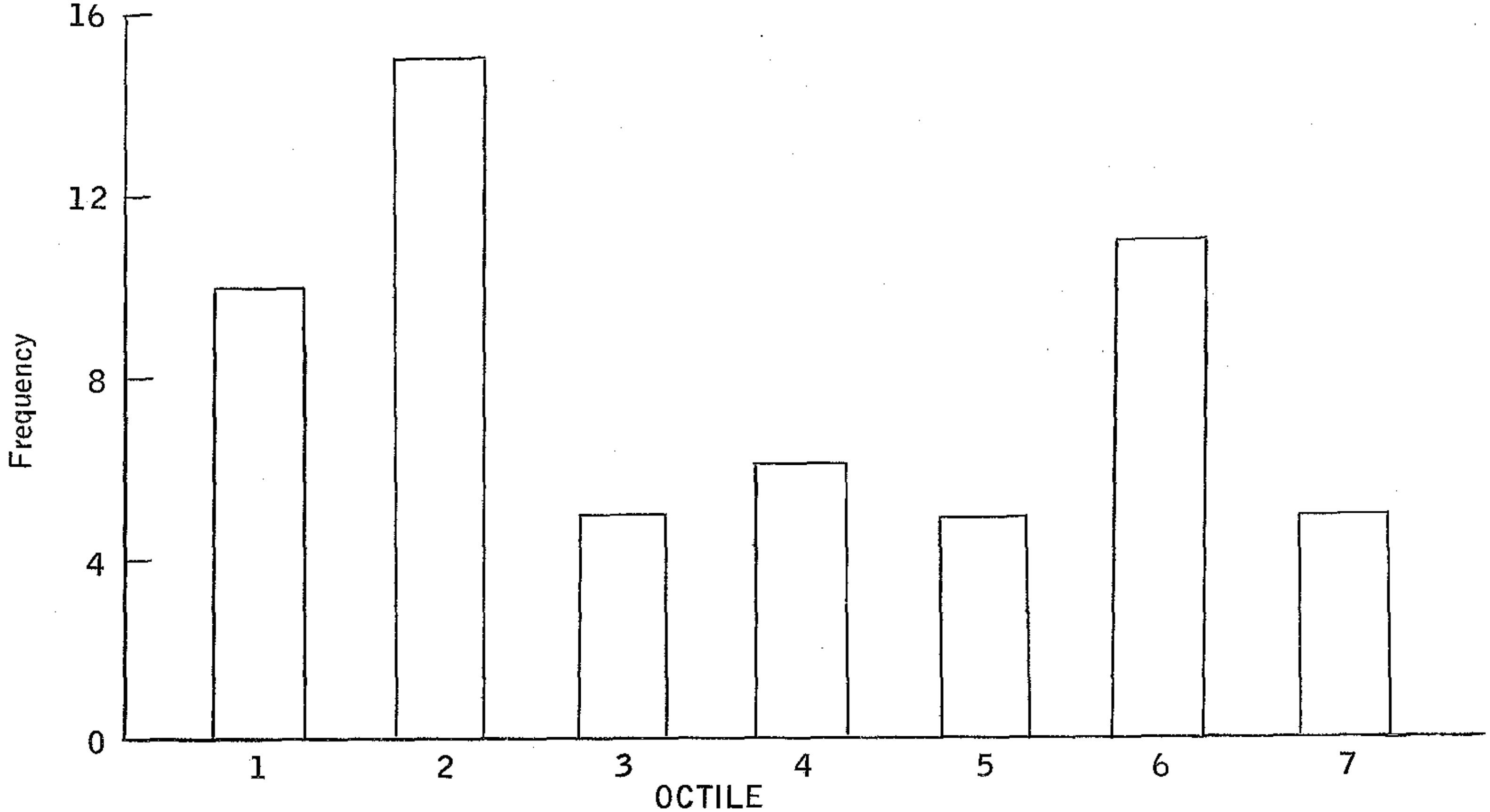



Fig. 11 Frequency of each estimate, $\hat{j}$, of the transition octile (two-element model).

TABLE 22

Parameter Estimates and Goodness-of-Fit of Two-Element and Linear Models

| Subject | Set of Pairs[1] | Responses per Octile | Two-Element Model: j | $\hat{g}_0$ | $\hat{g}_1$ | $\chi^2$ | d.f. | Linear Model: $\theta$ | $\hat{q}_1$ | $\chi^2$ | d.f. |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 29 | 6 | .603 | .707 | 12.354* | 5 | .0004 | .611 | 14.010* | 6 |
| 1 | 4 | 29 | 3 | .586 | .676 | 2.590 | 5 | .0010 | .598 | 3.811 | 6 |
| 1 | 5 | 29 | 1 | .655 | .749 | 6.900 | 5 | .0000 | .737 | 7.805 | 6 |
| 1 | 7 | 29 | 4 | .707 | .767 | 1.377 | 5 | .0010 | .704 | 2.012 | 6 |
| 2 | 2 | 27 | 1 | .593 | .815 | 8.273 | 5 | .0014 | .752 | 13.733* | 6 |
| 2 | 4 | 29 | 3 | .322 | .400 | 2.296 | 5 | .0008 | .309 | 194.400* | 5 |
| 2 | 5 | 26 | 5 | .838 | .705 | 11.726* | 3 | .0000 | .788 | 18.444* | 6 |
| 2 | 6 | 27 | 7 | .884 | .963 | 0.815 | 2 | .0040 | .839 | 0.022 | 2 |
| 2 | 7 | 29 | 7 | .611 | .552 | 1.937 | 5 | .0000 | .603 | 2.305 | 6 |
| 3 | 2 | 29 | 6 | .374 | .586 | 2.456 | 5 | .0018 | .294 | 4.767 | 6 |
| 3 | 4 | 28 | 6 | .607 | .786 | 3.722 | 5 | .0010 | .610 | 9.097 | 6 |
| 3 | 5 | 29 | 5 | .559 | .598 | 2.697 | 5 | .0002 | .563 | 2.458 | 6 |

[1] The numbering is: 1 - /p : b/, s; 2 - /p : b/, d; 3 - /s : ts/, s; 4 - /s : ts/, d; 5 - /p : b/, /b/ first; 6 - /p : b/, /p/ first; 7 - /s : ts/, /s/ first; 8 - /s : ts/, /ts/ first.

TABLE 22 (continued)

| Subject | Set of Pairs[1] | Responses per Octile | j | Two-Element Model $g_0$ | $g_1$ | $\chi^2$ | d.f. | Linear Model $\theta$ | $q_1$ | $\chi^2$ | d.f. |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 3 | 6 | 28 | 2 | .714 | .833 | 3.951 | 5 | .0040 | .699 | 3.044 | 6 |
| 3 | 7 | 28 | 6 | .685 | .804 | 2.834 | 5 | .0012 | .673 | 5.038 | 6 |
| 5 | 4 | 28 | 3 | .738 | .921 | 3.814 | 2 | .0100 | .626 | 0.860 | 2 |
| 5 | 7 | 29 | 2 | .724 | .879 | 4.060 | 2 | .0060 | .705 | 5.041 | 2 |
| 7 | 2 | 29 | 1 | .448 | .714 | 1.943 | 5 | .0030 | .553 | 5.103 | 6 |
| 7 | 5 | 29 | 1 | .517 | .828 | 3.709 | 5 | .0050 | .643 | 8.083 | 4 |
| 7 | 6 | 29 | 4 | .647 | .784 | 4.196 | 5 | .0020 | .641 | 7.198 | 6 |
| 9 | 3 | 28 | 2 | .893 | .804 | 3.836 | 4 | .0000 | .826 | 6.175 | 6 |
| 9 | 7 | 29 | 2 | .849 | .718 | 12.853* | 4 | .0000 | .750 | 17.563* | 6 |
| 11 | 2 | 26 | 2 | .615 | .865 | 8.842* | 2 | .0080 | .582 | 8.772 | 4 |
| 11 | 7 | 29 | 2 | .897 | .741 | 7.888 | 4 | .0000 | .780 | 14.448* | 6 |
| 13 | 2 | 29 | 6 | .511 | .828 | 7.789 | 5 | .0050 | .294 | 5.113 | 6 |
| 13 | 4 | 29 | 2 | .121 | .345 | 4.808 | 4 | .0014 | .164 | 10.831 | 6 |
| 13 | 5 | 29 | 4 | .569 | .767 | 4.654 | 5 | .0050 | .432 | 1.783 | 6 |

TABLE 22 (continued)

| Subject | Set of Pairs[1] | Responses per Octile | j | Two-Element Model $g_0$ | $g_1$ | $\chi^2$ | d.f. | Linear Model $\theta$ | $q_1$ | $\chi^2$ | d.f. |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 13 | 6 | 29 | 7 | .759 | .966 | 4.002 | 3 | .0020 | .729 | 7.180 | 6 |
| 13 | 7 | 29 | 7 | .586 | .414 | 1.723 | 5 | .0000 | .565 | 4.752 | 6 |
| 13 | 8 | 29 | 2 | .500 | .672 | 2.964 | 5 | .0030 | .484 | 2.784 | 6 |
| 14 | 2 | 29 | 5 | .490 | .747 | 3.776 | 5 | .0030 | .424 | 11.961 | 6 |
| 14 | 5 | 29 | 4 | .641 | .805 | 4.811 | 5 | .0030 | .582 | 4.445 | 6 |
| 14 | 6 | 28 | 6 | .760 | .643 | 3.024 | 5 | .0000 | .746 | 4.683 | 6 |
| 15 | 1 | 29 | 2 | .690 | .822 | 5.845 | 5 | .0030 | .702 | 5.942 | 5 |
| 15 | 2 | 28 | 6 | .673 | .982 | 15.232* | 4 | .0050 | .579 | 33.307* | 5 |
| 15 | 5 | 28 | 6 | .720 | .947 | 6.524 | 3 | .0050 | .625 | 15.269* | 5 |
| 15 | 6 | 29 | 6 | .724 | .914 | 7.111 | 4 | .0030 | .679 | 12.752* | 5 |
| 16 | 2 | 26 | 2 | .404 | .859 | 14.248* | 2 | .0150 | .161 | 2.381 | 3 |
| 16 | 5 | 22 | 2 | .591 | .886 | 4.642 | 2 | .0130 | .503 | 2.578 | 3 |
| 16 | 6 | 27 | 2 | .722 | .889 | 7.182* | 2 | .0070 | .700 | 5.137 | 3 |
| 17 | 2 | 29 | 1 | .517 | .744 | 3.964 | 5 | .0040 | .558 | 2.910 | 6 |

TABLE 22 (continued)

| Subject | Set of Pairs[1] | Responses per Octile | j | Two-Element Model $g_0$ | $g_1$ | $\chi^2$ | d.f. | Linear Model $\theta$ | $q_1$ | $\chi^2$ | d.f. |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 17 | 4 | 29 | 3 | .253 | .476 | 7.723 | 5 | .0020 | .234 | 7.523 | 6 |
| 17 | 7 | 29 | 1 | .448 | .576 | 1.627 | 5 | .0004 | .540 | 3.196 | 6 |
| 17 | 8 | 29 | 3 | .609 | .793 | 2.187 | 5 | .0040 | .573 | 3.386 | 6 |
| 18 | 2 | 24 | 5 | .858 | .708 | 18.202* | 3 | .0000 | .802 | 28.208* | 6 |
| 18 | 6 | 25 | 7 | .806 | .960 | 7.775 | 3 | .0000 | .825 | 14.372* | 6 |
| 18 | 7 | 29 | 4 | .716 | .793 | 3.834 | 5 | .0006 | .736 | 5.565 | 6 |
| 19 | 2 | 29 | 6 | .379 | .603 | 0.342 | 5 | .0016 | .321 | 5.092 | 6 |
| 19 | 4 | 29 | 2 | .414 | .299 | 5.157 | 5 | .0000 | .328 | 7.515 | 6 |
| 19 | 5 | 29 | 6 | .661 | .776 | 6.864 | 5 | .0014 | .636 | 8.664 | 6 |
| 19 | 6 | 29 | 1 | .690 | .621 | 4.437 | 5 | .0000 | .629 | 2.587 | 6 |
| 19 | 7 | 29 | 1 | .672 | .517 | 1.244 | 5 | .0000 | .556 | 5.430 | 6 |
| 19 | 8 | 29 | 5 | .710 | .770 | 2.265 | 5 | .0006 | .713 | 3.983 | 6 |
| 20 | 2 | 29 | 4 | .681 | .914 | 4.234 | 3 | .0100 | .479 | 2.301 | 3 |
| 20 | 5 | 29 | 2 | .655 | .891 | 8.091* | 2 | .0100 | .555 | 1.528 | 3 |
| SUM | | | | | | 291.350 | 228 | | | 587.327 | 288 |

* $p < .05$

and 7. Of course, this agrees with the "midplateau" finding mentioned before. We note that for the first $\hat{j}$ octiles, this model predicts that the proportion of correct responses will equal $g_0$, and for the last $8-\hat{j}$ octiles, the proportion correct should be $g_1$. After $g_0$, $g_1$, and j were estimated for each of the sets of octile data, the predicted learning curves were plotted. The graphs of the theoretical proportion correct per octile are compared with the observed proportions in Figures 12-43.

---

Insert Figs. 12-43 about here

---

Slope-intercept model. It is possible that the plot of the proportion correct against the octile number would be more adequately described by a straight line of non-zero slope than by the pair of horizontal line segments (Figures 12-43) which the two-element model requires. Although we had no fundamental grounds for preferring the slope-intercept model, it is worthwhile to determine its fit to the present data. By so doing we shall have something against which to compare the fit of the two-element model. The slope and intercept parameters were computed, also by the method of least squares and the minimum sum of squared deviations was computed for each of the 54 sets of octile data. As an index of the relative accuracy of the two-element and slope-intercept models, we may compare the sum of squared deviations between predicted and observed frequencies for the two models. Of the 54 comparisons, the summed squared deviations were lower for the two-element model than for the slope-intercept model in 47 cases. Using the normal approximation to the binomial distribution, the hypothesis that there is no difference between the models in the summed squared deviations is rejected at the .0001 significance level ($z = 5.38$). Of the 54 comparisons between the linear and slope-intercept models, the summed squared deviations were lower for the linear model about half the time.

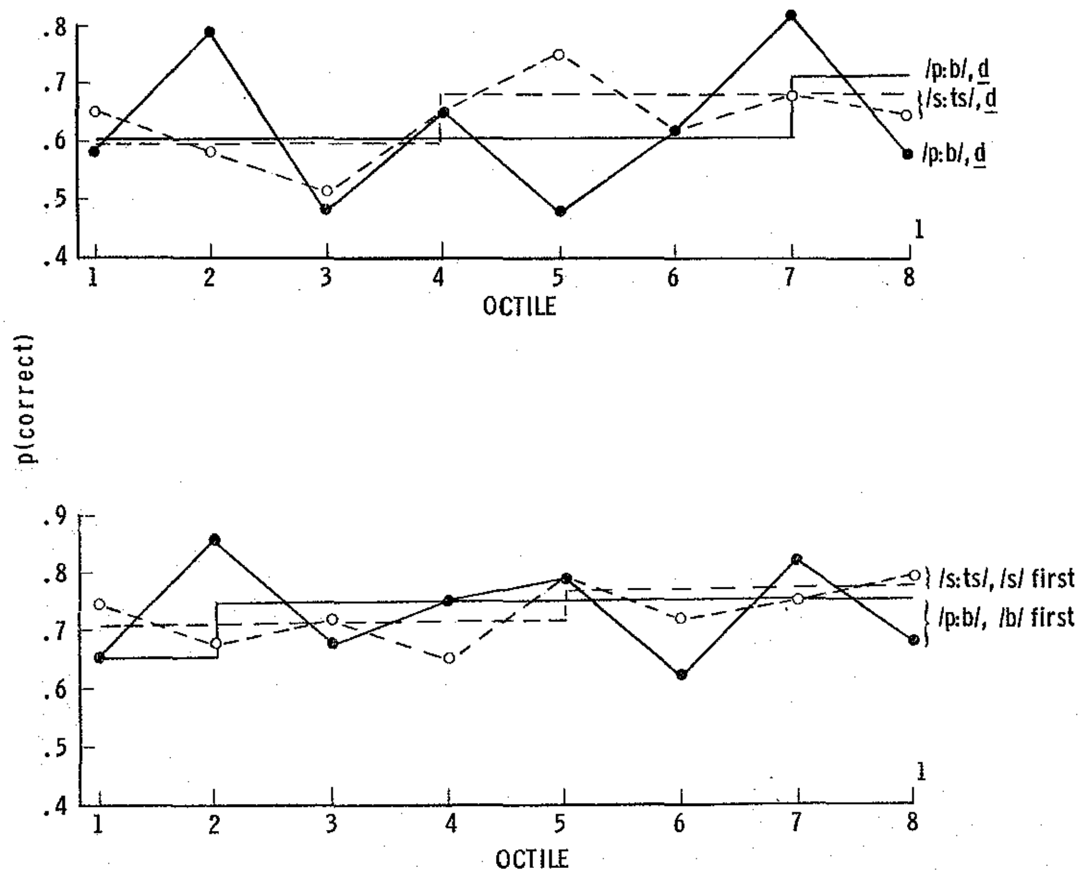
Linear model. The predictions of the linear model were determined for each of the sets of octile data. The first step was to estimate $q_1$, the initial probability of an error, and $\theta$, the learning rate. As with the other models, the estimation was performed by minimizing the sum, over octiles, of the squared deviations between the predicted and observed frequencies of correct response (see equation (4)). However, unlike the situations for the other two models, an explicit algebraic solution for $q_1$ and $\theta$ in terms of the observed quantities was prohibitively difficult. This problem was met by exploring $\theta$ in increments of .0002 from .0000 to .0020 (the range in which over half of the best estimates of $\theta$ actually fell), in increments of .0010 to .020, and in increments of .010 thereafter. For each of these values of $\theta$, the $q_1$ which produced the least sum of squared deviations was found. Then that single pair ($\hat{q}_1$, $\hat{\theta}$) which yielded the lowest sum of squared deviations was selected as the estimates of $q_1$ and $\theta$. Using equation (4) the goodness-of-fit was computed for each set of octile data. These chi-square values play a major role in our evaluation and comparison of the two-element and linear models.

Evaluation of two-element and linear models. Table 22 gives the parameter estimates for the two-element and linear models and the results of the goodness-of-fit tests. The number of degrees of freedom takes into account the pooling of adjacent octiles which was required to yield sufficient theoretical observations per cell. There appears to be no consistent relation between $\hat{g}_0$ and $\hat{g}_1$ in the two-element model. The very small $\theta$ values in the linear model express the fact that the learning proceeded very slowly. The wide inter-subject differences in $\hat{g}_0$ and $\hat{q}_1$ suggest large individual variation in initial ability.

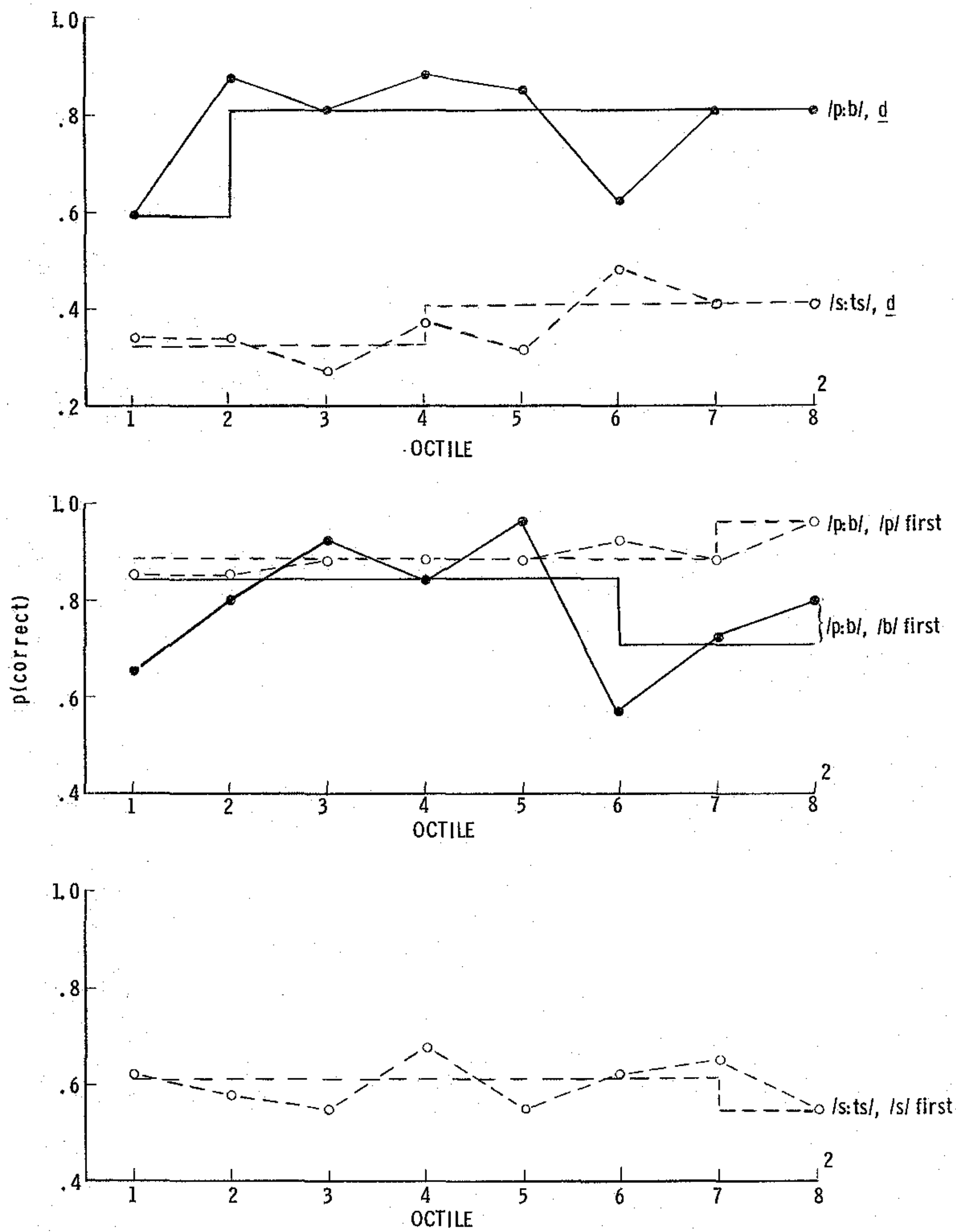
With 5 degrees of freedom a $\chi^2$ of 11.1 is required for significance at the .05 level. With 6 degrees of freedom, the corresponding figure is 12.6. Of the 54 $\chi^2$ values, nine are significant at the .05 level for the two-element model, and eleven for the linear model. According to this comparison, the two-element model is slightly superior. An overall evaluation of the models was also performed in the following manner. For each model, the goodness of fit was determined by summing the $\chi^2$ over all sets of data in Table 22. The sum was 291.350 for the two-element model (228 degrees of freedom) and 587.327 for the linear model (288 degrees of freedom). The normal approximation yielded $z = 2.808$, $p < .005$ for the two-element model, and $z = 10.297$, $p < .0001$ for the linear model. If subject 2, Set 4 is omitted from the linear model calculations, $\chi^2 = 392.972$, d.f. = 283, $z = 4.265$, and again $p < .0001$. Therefore, it may be concluded that the deviations between either model and the data are highly significant. Owing to the large number of observations included in the analysis, this fact is not surprising. A more informative measure of the adequacy of the models consists in comparing the sum of their $\chi^2$s. On this basis, the two-element model is clearly superior.

Some tentative conclusions. As the results just given indicate, the overall comparison of the two-element and linear models is favorable to the two-element model. Also, both of these models fare better than the simple one-element conditioning model. The not unexpected superiority of the two-element model to the one-element model follows from the fact that $\hat{j}$ in the former model was never equal to eight, the value which reduces to the one-element model. The linear model is more adequate than the one-element model, since the case $\hat{\theta}$ equals zero (which reduces to the one-element model for our pre-learning data) rarely obtained. It is somewhat surprising that the two-element model turns out to be superior to the linear model, for as remarked earlier, the complexity of the auditory stimulus material used in the experiment could easily have led to results favoring the linear model. On the other hand, the goodness of fit of the two-element model to the 54 cases of individual data is not close enough to warrant the drawing of any decisive inferences concerning the number of aspects or properties of the auditory stimulus material which determine the response conditioning of subjects. That a stimulus sampling model with a small number of elements works fairly well is encouraging. In future work, we hope to pursue more deeply the identification of those aspects of the stimulus material that are most important in determing responses, but we also realize that it is very likely the case that models of greater formal complexity will be necessary adequately to account for all major aspects of the data. We certainly do not feel that the present application of mathematical learning theory to learning in a linguistic context is to be regarded as other than a tentative first step.

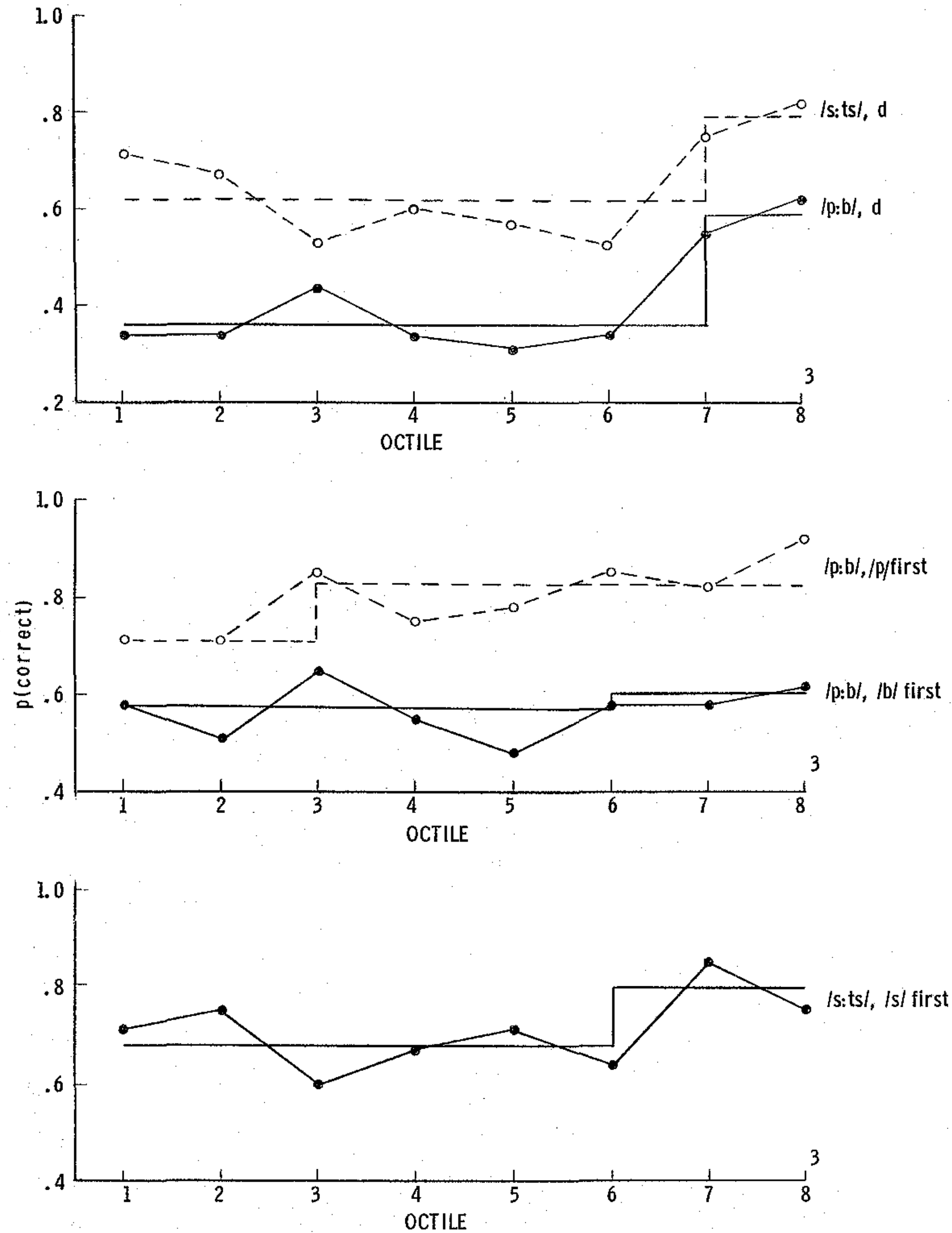
Figures 12-44 Observed and predicted (two-element model) proportions correct per octile for the indicated pairs. Subject number appears in lower right corner.



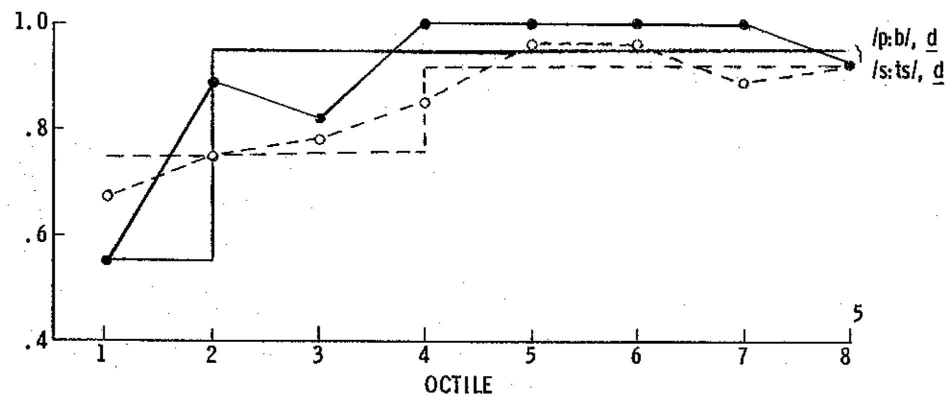

Figs. 12, 13

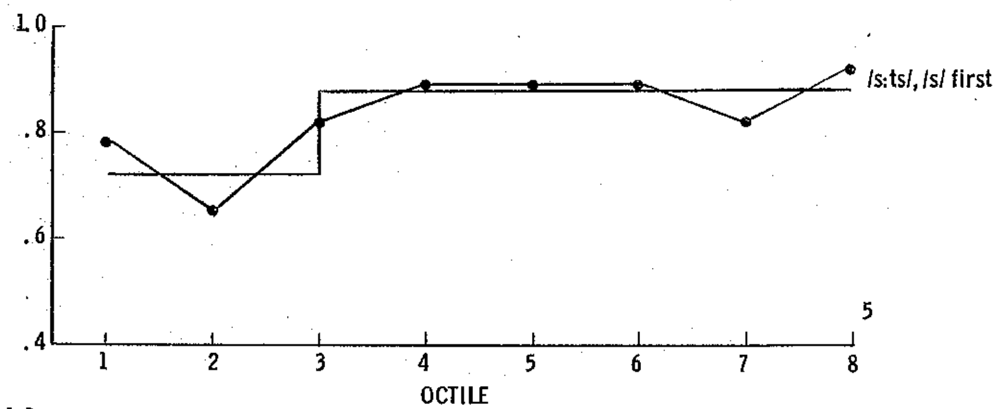

1.0
.8
.6
.4
.2
/p:b/, d
/s:ts/, d
2
1 2 3 4 5 6 7 8
OCTILE
p(correct)
/p:b/, /p/ first
/p:b/, /b/ first
OCTILE
/s:ts/, /s/ first
OCTILE

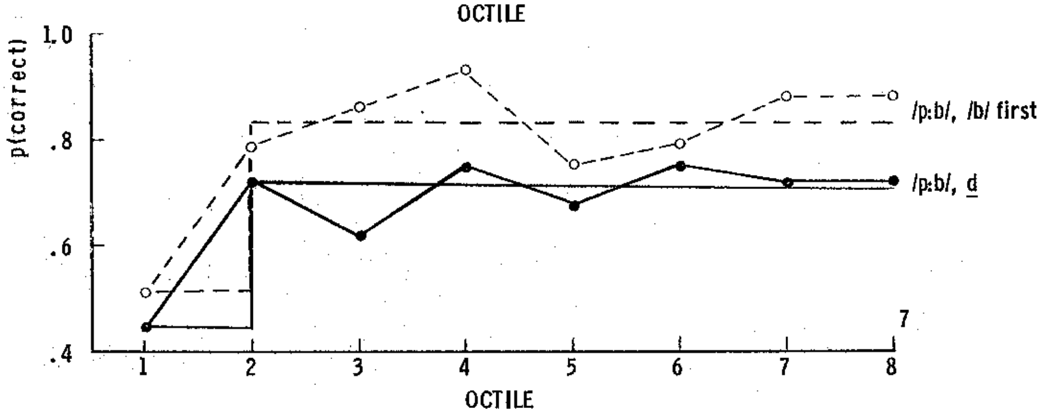
Figs. 14, 15, 16

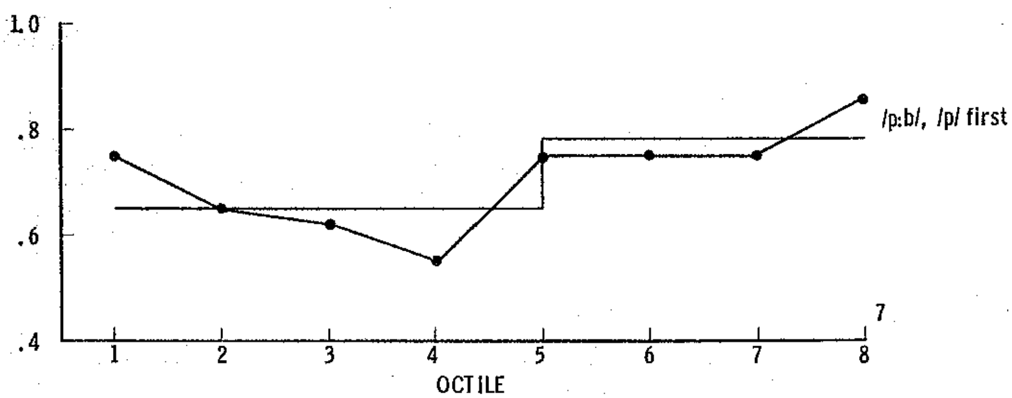

1.0
.8
.6
.4
.2
1
2
3
4
5
6
7
8
OCTILE
/s:ts/, d
/p:b/, d
3
p(correct)
/p:b/, /p/first
/p:b/, /b/ first
/s:ts/, /s/ first

Figs. 17, 18, 19

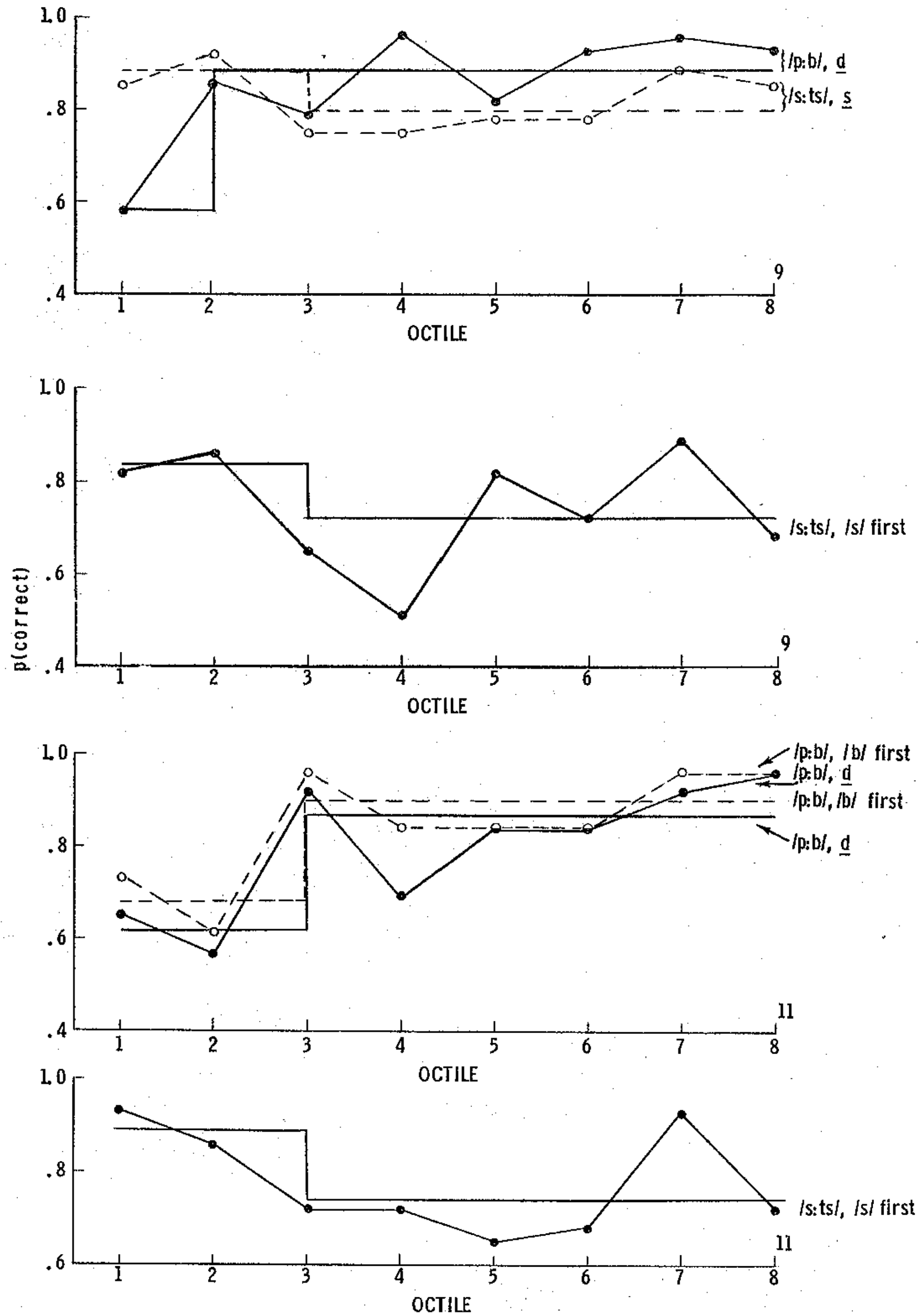


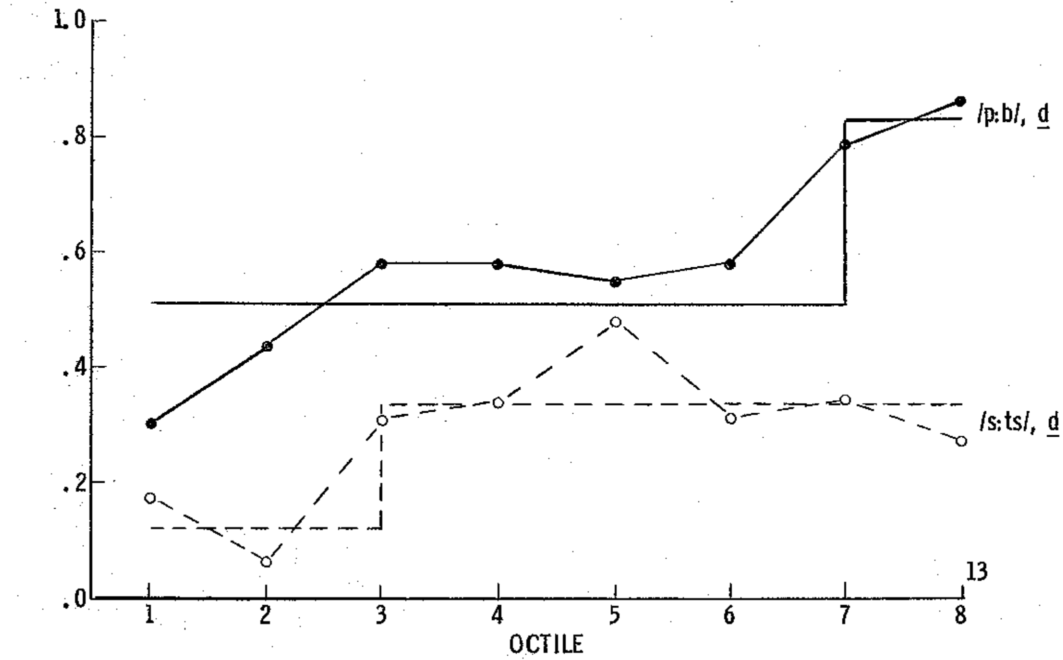



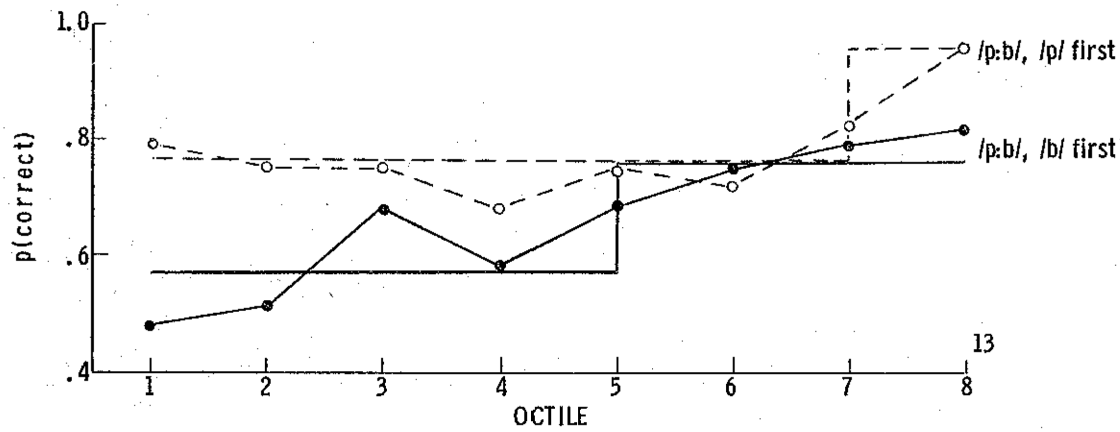



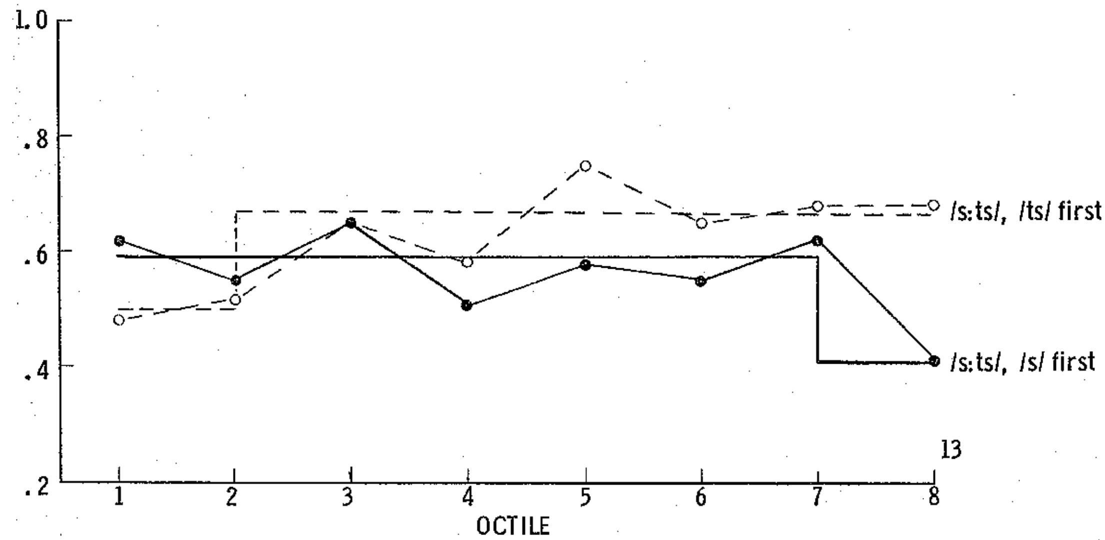

Figs. 20, 21, 22, 23



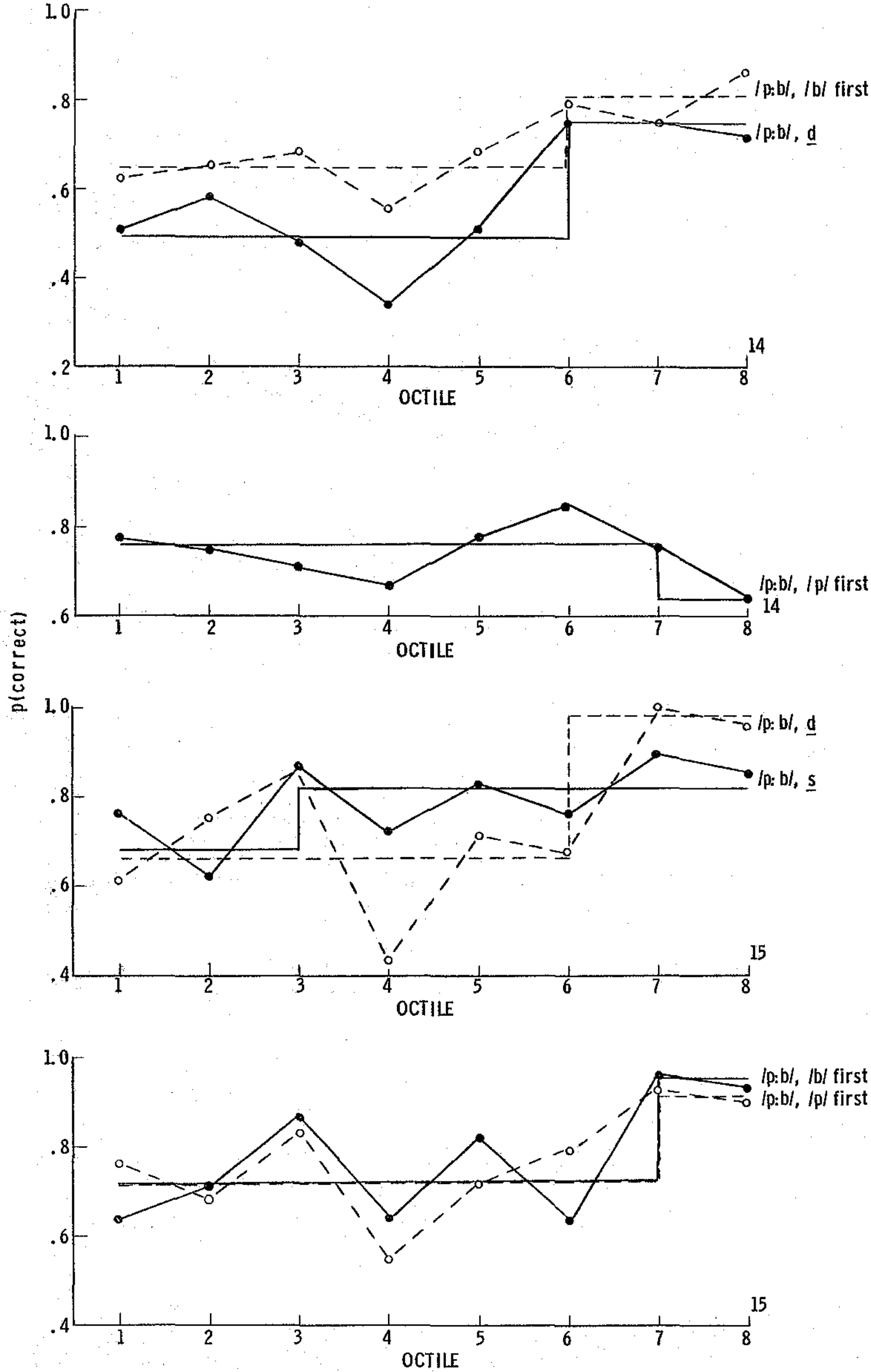

Figs. 24, 25, 26, 27


1.0
.8
.6
.4
.2
.0
/p:b/, d
/s:ts/, d
13
1 2 3 4 5 6 7 8
OCTILE

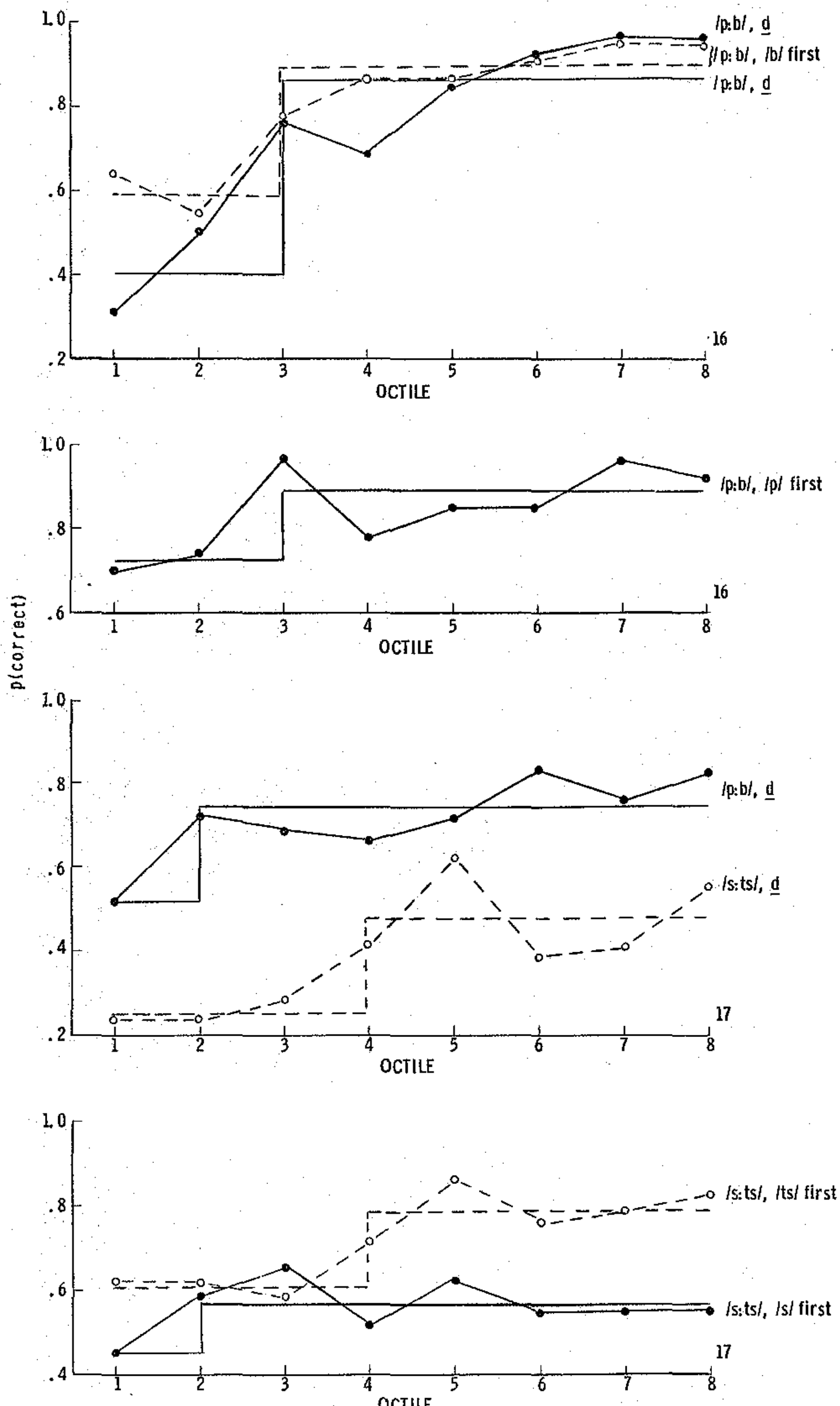

p(correct)
1.0
.8
.6
.4
/p:b/, /p/ first
/p:b/, /b/ first
13
1 2 3 4 5 6 7 8
OCTILE


1.0
.8
.6
.4
.2
/s:ts/, /ts/ first
/s:ts/, /s/ first
13
1 2 3 4 5 6 7 8
OCTILE

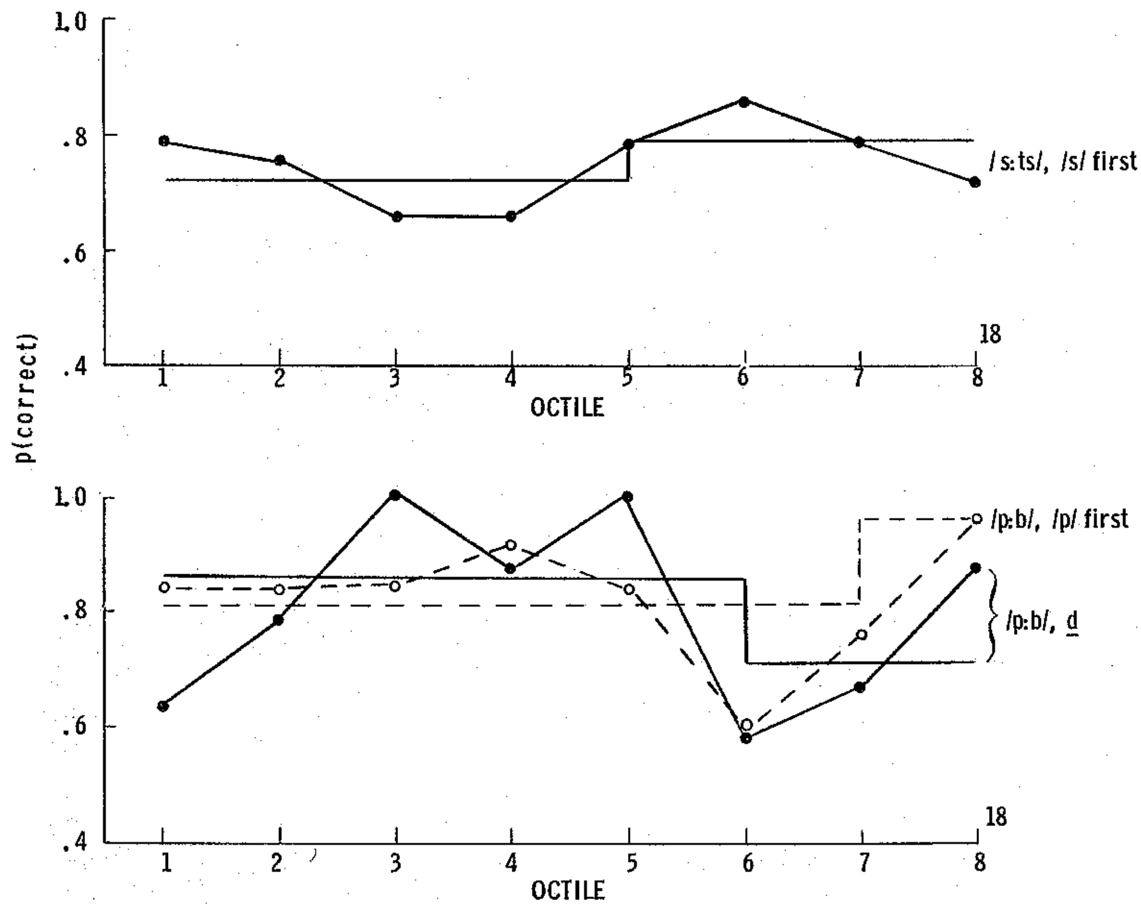
Figs. 28, 29, 30



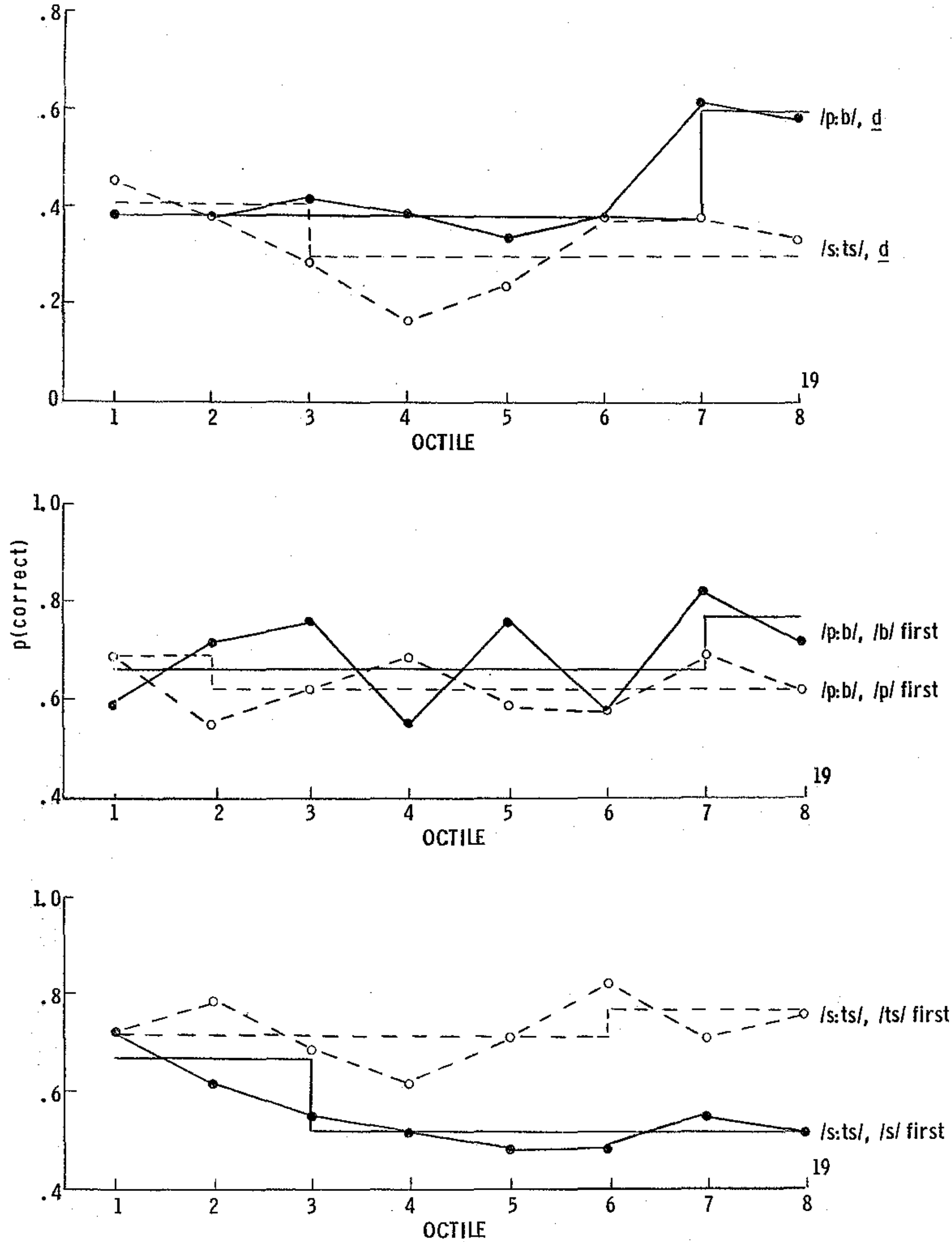

Figs. 31, 32, 33, 34




Figs. 35, 36, 37, 38




Figs. 39, 40


.8
.6
.4
.2
0
/p:b/, d
/s:ts/, d
19
1
2
3
4
5
6
7
8
OCTILE
1.0
.8
.6
.4
p(correct)
/p:b/, /b/ first
/p:b/, /p/ first
19
OCTILE
1.0
.8
.6
.4
/s:ts/, /ts/ first
/s:ts/, /s/ first
19
OCTILE

Figs. 41, 42, 43

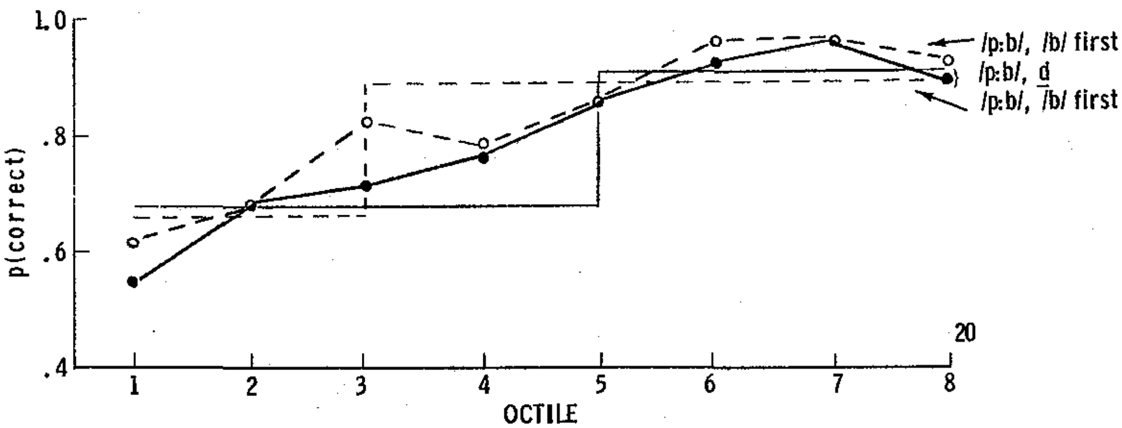



Fig. 44

## References


Bush, R. R., and Estes, W. K., Studies in Mathematical Learning Theory, Stanford, California: Stanford University Press, 1959.

Bush, R. R., and Mosteller, F., Stochastic Models for Learning, New York: Wiley, 1955.

Estes, W. K., "Toward a statistical theory of learning", Psychological Review, Vol. 57, (1950), pp. 94-107.

Estes, W. K., "Growth and function of mathematical models for learning", Current Trends in Psychological Theory, Pittsburgh, Pennsylvania: University of Pittsburgh Press, 1961.

Fischer-Jørgensen, E., "Acoustic analysis of stop consonants" Miscellanea Phonetica, Vol. 2 (1955), pp. 42-49.

Gleason, H. A., Jr., An Introduction to Descriptive Linguistics, Revised edition, New York, 1961.

Halle, M., The Sound Patterns of Russian, 's - Gravenhage, 1959.

Jakobson, R., Fant, C. G. M., and Halle, M., Preliminaries to Speech Analysis, Cambridge: M.I.T., 1952.

McNemar, Q., Psychological Statistics, New York: Wiley, 1962.

Miller, G. A., and Nicely, P. E., "An analysis of perceptual confusion among some English consonants", in Psycholinguistics, S. Saporta, ed., New York: Holt, Rinehart and Winston, 1961, pp. 153-175.

Pollack, I., "The information of elementary auditory displays", JASA, Vol. 24 (1952), pp. 745-769.

Suppes, P., and Atkinson, R. C., Markov Learning Models for Multiperson Interactions, Stanford, California: Stanford University Press, 1960.

Suppes, P., and Ginsberg, R., "A fundamental property of all-or-none models, binomial distribution of responses prior to conditioning, with application to concept formation in children." Technical Report No. 39, Institute for Mathematical Studies in the Social Sciences, Stanford University, Stanford, California, 1961. (To appear in Psychol. Rev.).

Zeaman, D., and House, B. J., "Role of attention in retardate discrimination learning", N. R. Ellis (ed.), Handbook in Mental Deficiency: Psychological Theory and Research, New York: McGraw-Hill, 1962, (in press).

Appendix A

Russian Minimal Pairs

1 fa-va
2 fa-fa
3 va-va
4 va-fa
5 fo-vo
6 fo-fo
7 vo-vo
8 vo-fo
9 fu-vu
10 fu-fu

11 vu-vu
12 vu-fu
13 sa-za
14 sa-sa
15 za-za
16 za-sa
17 so-zo
18 so-so
19 zo-zo
20 zo-so

21 su-zu
22 su-su
23 zu-zu
24 zu-su
25 se-ze
26 se-se
27 ze-ze
28 ze-se
29 sa-za
30 sa-sa

31 za-za
32 za-sa
33 so-zo
34 so-so
35 zo-zo
36 zo-so
37 su-zu
38 su-su
39 zu-zu
40 zu-su

41 se-ze
42 se-se
43 ze-ze
44 ze-se
45 si-zi
46 si-si
47 zi-zi
48 zi-si
49 si-zi
50 si-si

51 zi-zi
52 zi-si
53 pa-ba
54 pa-pa
55 ba-ba
56 ba-pa
57 po-bo
58 po-po
59 bo-bo
60 bo-po

61 pu-bu
62 pu-pu
63 bu-bu
64 bu-pu
65 pe-be
66 pe-pe
67 be-be
68 be-pe
69 ka-ga
70 ka-ka

71 ga-ga
72 ga-ka
73 ko-go
74 ko-ko
75 go-go
76 go-ko
77 ku-gu
78 ku-ku
79 gu-gu
80 gu-ku

81 ta-da
82 ta-ta
83 da-da
84 da-ta
85 to-do
86 to-to
87 do-do
88 do-to
89 tu-du
90 tu-tu

91 du-du
92 du-tu
93 te-de
94 te-te
95 de-de
96 de-te
97 pi-bi
98 pi-pi
99 bi-bi
100 bi-pi

101 ti-di
102 ti-ti
103 di-di
104 di-ti

## Appendix B

## CV Pairs Used in Pilot Studies

Pilot Study 1

kji-kji
kji-xji
xji-kji
xji-xji

sa-sa
sa-tsa
tsa-sa
tsa-tsa
se-se
se-tse
tse-se
tse-tse
si-si
si-tsi
tsi-si
tsi-tsi
su-su
su-tsu
tsu-su
tsu-tsu

sha-sha
sha-šča
šča-sha
šča-šča
shi-shi
shi-šči
šči-shi
šči-šči
sho-sho
sho-ščo
ščo-sho
ščo-ščo
shu-shu
shu-šču
šču-shu
šču-šču

cha-cha
cha-tja
tja-cha
tja-tja
chi-chi
chi-tji
tji-chi
tji-tji
cho-cho
cho-tjo
tjo-cho
tjo-tjo
chu-chu
chu-tju
tju-chu
tju-tju

kje-kje
kje-tje
tje-kje
tje-tje
kji-kji
kji-tji
tji-kji
tji-tji

ka-ka
ka-xa
xa-ka
xa-xa
ko-ko
ko-xo
xo-ko
xo-xo
ku-ku
ku-xu
xu-ku
xu-xu

sja-sja
sja-tsa
tsa-sja
tsa-tsa
sje-sje
sje-tse
tse-sje
tse-tse
sji-sji
sji-tsi
tsi-sji
tsi-tsi
sjo-sjo
sjo-tso
tso-sjo
tso-tso

cha-cha
cha-tsa
tsa-cha
tsa-tsa
chi-chi
chi-tsi
tsi-chi
tsi-tsi
cho-cho
cho-tso
tso-cho
tso-tso
chu-chu
chu-tsu
tsu-chu
tsu-tsu

Appendix B (continued)

Pilot Study 2

pi-pi
pi-pji
pji-pi
pji-pji
bi-bi
bi-bji
bji-bi
bji-bji

ti-ti
ti-tji
tji-ti
tji-tji
di-di
di-dji
dji-di
dji-dji

si-si
si-sji
sji-si
sji-sji
zi-zi
zi-zji
zji-zi
zji-zji

mi-mi
mi-mji
mji-mi
mji-mji
ni-ni
ni-nji
nji-ni
nji-nji

li-li
li-lji
lji-li
lji-lji
ri-ri
ri-rji
rji-ri
rji-rji

la-la
la-lja
lja-la
lja-lja
ra-ra
ra-rja
rja-ra
rja-rja

tje-tje
tje-tse
tse-tje
tse-tse
tja-tja
tja-tsa
tsa-tja
tsa-tsa
tsi-tsi
tsi-tji
tji-tsi
tji-tji
tjo-tjo
tjo-tso
tso-tjo
tso-tso
tju-tju
tju-tsu
tsu-tju
tsu-tsu

sje-sje
sje-tse
tse-sje
tse-tse
sju-sju
sju-tsu
tsu-sju
tsu-tsu

so-so
so-tso
tso-so
tso-tso

## Appendix C

### Proportions of Errors in Pilot Studies of Consonant Phoneme Discrimination

Pilot Study 1

| Contrast | Vowel | No. of Items | p(error) |
|---|---|---|---|
| /kj : xj/ | i | 4 | .14 |
| /s : ts/ | a, e, i, u | 16 | .13 |
| /sh : šč/ | a, i, o, u | 16 | .06 |
| /ch : tj/ | a, i, o, u | 16 | .05 |
| /kj : tj/ | e, i | 8 | .03 |
| /k : x/ | a, o, u | 12 | .02 |
| /sj : ts/ | a, e, i, o | 16 | .01 |
| /ch : ts/ | a, i, o, u | 16 | .00 |

Appendix C (continued)

Pilot Study 2 (Each contrast involved four items)

| Contrast | Vowel | p(error) |
|---|---|---|
| /s : ts/ | o | .28 |
| /z : zj/ | i | .17 |
| /l : lj/ | i | .11 |
| /s : sj/ | i | .10 |
| /n : nj/ | i | .09 |
| /l : lj/ | a | .08 |
| /sj : ts/ | e | .08 |
| /tj : ts/ | u | .07 |
| /p : pj/ | i | .06 |
| /d : dj/ | i | .06 |
| /r : rj/ | a | .06 |
| /r : rj/ | i | .06 |
| /sj : ts/ | u | .05 |
| /ts : tj/ | e | .04 |
| /t : tj/ | i | .04 |
| /m : mj/ | i | .03 |
| /tj : ts/ | a | .02 |
| /tj : ts/ | o | .01 |
| /tj : ts/ | i | .00 |
| /b : bj/ | i | .00 |

Appendix D

Response Dependency Analysis

1 = error    0 = correct

Quartile

| Subject | p(1\|0) | p(1\|1) | p(1\|0) | p(1\|1) | p(1\|0) | p(1\|1) | p(1\|0) | p(1\|1) |
|---|---|---|---|---|---|---|---|---|
| 1. | .162 | .261 | .185 | .183 | .171 | .217 | .166 | .156 |
| 2. | .169 | .221 | .185 | .162 | .248 | .217 | .194 | .090 |
| 3. | .185 | .230 | .211 | .250 | .225 | .258 | .133 | .120 |
| 4. | .062 | .241 | .038 | .067 | .024 | .000 | .030 | .000 |
| 5. | .156 | .180 | .092 | .086 | .076 | .069 | .047 | .056 |
| 6. | .109 | .182 | .051 | .000 | .033 | .000 | .016 | .000 |
| 7. | .217 | .217 | .168 | .115 | .119 | .167 | .077 | .206 |
| 8. | .109 | .163 | .060 | .000 | .042 | .063 | .011 | .000 |
| 9. | .129 | .104 | .100 | .190 | .106 | .122 | .056 | .087 |
| 10. | .119 | .130 | .053 | .095 | .055 | .047 | .035 | .133 |
| 11. | .090 | .162 | .112 | .116 | .116 | .152 | .056 | .091 |
| 12. | .058 | .045 | .038 | .067 | .041 | .000 | .033 | .077 |
| 13. | .250 | .314 | .320 | .261 | .276 | .290 | .221 | .152 |
| 14. | .140 | .196 | .142 | .237 | .146 | .145 | .093 | .059 |
| 15. | .178 | .233 | .186 | .150 | .172 | .113 | .076 | .036 |
| 16. | .163 | .190 | .098 | .154 | .066 | .080 | .027 | .000 |
| 17. | .223 | .286 | .217 | .154 | .228 | .190 | .156 | .194 |
| 18. | .168 | .209 | .138 | .182 | .152 | .183 | .073 | .000 |
| 19. | .247 | .292 | .273 | .299 | .288 | .356 | .261 | .275 |
| 20. | .157 | .206 | .153 | .197 | .096 | .108 | .080 | .034 |
| Mean Proportion | .155 | .203 | .141 | .148 | .134 | .139 | .092 | .088 |