R842-09 ERIC REPORT RESUME

ED 010 236 1-31-67 48 (REV)
STUDIES IN LANGUAGE AND LANGUAGE BEHAVIOR, PROGRESS REPORT NO. 1.
LANE, HARLAN * AND OTHERS
KUM37771 UNIVERSITY OF MICHIGAN, ANN ARBOR
BR-5-1278-B
15-OCT-65 OEC-5-14-036
EDRS PRICE MF-$0.45 HC-$10.88 272P.

*LANGUAGES, *LINGUISTICS, *PROGRAMED INSTRUCTION,
*EDUCATIONAL RESEARCH, *LANGUAGE RESEARCH, GRAMMAR,
WORD ASSOCIATIONS (READING), HIGHER EDUCATION, LANGUAGE INSTRUCTION,
LANGUAGE ENRICHMENT, ANN ARBOR, MICHIGAN

THE UNIVERSITY OF MICHIGAN'S CENTER FOR RESEARCH ON LANGUAGE AND LANGUAGE BEHAVIOR PUBLISHED PROGRESS REPORT 1 WHICH CONTAINS 14 SEPARATE RESEARCH ARTICLES AND REPORTS AND 19 LISTED STUDIES IN PROGRESS. THE STUDIES WERE ALL RELEVANT TO LANGUAGE AND LANGUAGE BEHAVIOR. ALSO INCLUDED IS A LIST OF AFFILIATED PERSONNEL CONDUCTING STUDIES IN THIS AREA OF CONCENTRATION FOR THE CENTER. (HB)

5-1278-B

THE UNIVERSITY OF MICHIGAN

Studies in Language and Language Behavior

Progress Report
No. 1
Contract OE 5-14-036

CENTER FOR RESEARCH ON LANGUAGE AND LANGUAGE BEHAVIOR

THE UNIVERSITY OF MICHIGAN

CENTER FOR RESEARCH

ON

LANGUAGE AND LANGUAGE BEHAVIOR

Title of Project: Studies in Language and Language Behavior

Contract Number: Language Research OE 5-14-036 (5-1278)

Project Director: Harlan Lane

Harlan L. Lane
Associate Professor of Psychology
The University of Michigan

Name of Institution: Center for Research on Language and Language Behavior
The University of Michigan

Purpose of Report: Progress Report No. 1

Date of Report: 15 October 1965

Submitted to: Language Section
Division of Higher Education Research
U. S. Office of Education
Washington, D. C.

TABLE OF CONTENTS

Generalization Decrement or Discrimination?

The Problem in the Light of Some Recent Data

James L. Kopp

Stimulus generalization is a behavioral phenomenon that has been known to psychologists and their predecessors in philosophy and physiology for many years. The early associationists, such as John Stuart Mill and Herbert Spencer, gave it a place in their systems under the name "Law of Similarity" (Razran, 1949). I. P. Pavlov took the first steps toward a quantitative description of generalization with his development of techniques for laboratory demonstrations of conditional control of reflex behavior (1927, p. 25).

Stimulus generalization refers to the empirical observation that a given property of behavior (amplitude, latency, rate, etc.) is invariant in spite of changes in the controlling stimulus. If, during generalization testing, a decrease in the response measure is observed on the occasion of the presentation of a new stimulus value, generalization decrement is said to occur. Generalization decrement, when it is observed, is characteristically a systematic function of the physical difference between the original controlling stimulus and other stimulus values. This functional relation is called a generalization gradient.

Definitions like the above are typically found in the beginning paragraphs of most modern discussions of stimulus generalization (Shepard, 1957; Mednick & Freedman, 1960; Prokasy & Hall, 1963; Baron, 1965).

Stimulus discrimination is a phenomenon with an equally long history in psychology. It is typically viewed as a distinctly different phenomenon from generalization decrement and, as Prokasy and Hall have observed (1963), most integrating texts on learning and conditioning give the two phenomena separate treatment. This is puzzling in view of the fact that empirical definitions of discrimination typically involve the same terms as definitions of generalization decrement. In discussing discrimination, Spence (1937) talks about differential responding to differences in stimulus degree. English and English (1958) define discrimination as an organism's reacting differently to different objects (stimuli). Hebb (1958) speaks of differential responses to two stimuli, and Dember (1960) has recently described discrimination as a response (presumably contrasting with no response) to stimulus changes. These definitions, like the definitions of generalization decrement, appose changes in some response measure to changes in the physical stimulus.

Clearly, empirical definitions of generalization decrement and discrimination are quite similar, if not identical, and this fact has been thoroughly amplified by Brown (1965). In addition, recent experimentation has made the distinction between generalization decrement and discrimination appear to be rather specious.

If generalization decrement and discrimination are identical, one is led to ask how it is that the two have been given separate treatment. This paper will examine the problem in the light of the literature. Specifically, it will first examine the evolution of the problem--the theoretical and emiprical bases for making the distinction between generalization decrement and discrimination and, second, discuss the experimental evidence that concerns the usefulness of such a distinction.

Evolution of the Problem

Formal distinctions

Pavlov first characterized generalization decrement as a phenomenon distinct from discrimination when he stated that there was a gradual decrement in the eliciting power of generalized stimuli as a function of their physical distance from the conditioned stimulus:

> "...if a tone of 1000 d.v. is established as a conditioned stimulus, many other tones spontaneously acquire similar properties, such properties diminishing proportionally to the intervals of these tones from the one of 1000 d.v." (italics added) (1927, p. 113).

Initially, Pavlov catalogued this response decrement as discrimination; but he then observed that the decrement could never be made greater by simply repeating the CS in the presence of reinforcement. On the other hand, only a few presentations of another stimulus without reinforcement resulted in large decrements of responding for stimuli other than the original CS. The discontinuity between the original decrement and the much larger decrement obtained by contrasting two stimuli with respect to reinforcement led Pavlov to conclude that the original decrement was independent of training, and inherent in the acquisition of a conditioned reflex. He therefore concluded that it was a phenomenon distinct from discrimination, and explained it in terms of irradiation of sensory excitation over the surface of the cerebral cortex.

Pavlov's theory is weak in maintaining the distinction between generalization decrement and differential responding. In Pavlov's formulation, there are at least three distinct mechanisms mediating differential responding:

irradiation, internal inhibition, and external inhibition. Any given observation of differential responding might be explained by Pavlov as resulting from either or all of the three mechanisms. It is therefore often very difficult to extricate true generalization decrement from the other types of differential responding in any given situation. An example of this is given in Pavlov's description of the organism's behavior in the course of discrimination training:

> "It was noticed that when, after a conditioned reflex to a definite stimulus (e.g., a definite musical tone) had been firmly established, the effect of another closely allied stimulus (a neighboring musical tone) was tried for the first time, the conditioned reflex which resulted from the new stimulus was frequently much weaker than that obtained with the original conditioned stimulus...."

Now this would seem to be a straightforward example of generalization decrement, according to the original definition, but Pavlov then goes on to note some variations:

> "On repetition of the stimulus of the neighboring tone, always, of course, without reinforcement, the secretory effect increased until it became equal to that given by the originally established stimulus, but subsequently on further repetition began to diminish, falling finally to a permanent zero. Thus it appeared that at first the two closely allied stimuli were discriminated straight away but that later this discrimination for some reason disappeared, only gradually to re-establish itself and finally to become absolute" (1927, p. 118).

The original response decrease, which one might expect to be called generalization decrement, and which Pavlov says appears to be discrimination, is attributed in fact to external inhibition resulting from the action of an "investigatory reflex" (1927, p. 118). This external inhibition was quickly replaced by 100 per cent generalization, only to give way finally to internal inhibition, the basis for true discrimination. But might not external inhibition be just as valid an explanation for generalization decrement alone? Further, at what point, and why, in the course of the above observations do the effects of external inhibition disappear, giving way to total generalization and, finally, discrimination? Pavlov never seems to make explicit how generalization decrement is to be observed as an isolated phenomenon.

Perhaps in view of this confusion, writers in Pavlov's wake tried to be more explicit in their interpretation of the nature of generalization decrement and discrimination. Eventually, a controversy arose in which the protagonists were the theorists Hull and Spence, on the one hand, and Lashley and Wade, on the other. The theories of several later writers (Razran, 1949; Mednick & Freedman, 1960) have chiefly been attempts to reconcile these two positions.

According to Hull's theory (Hull, 1943; Spence, 1937) there is no such thing as discrimination, except as an epiphenomenon of generalization. Generalization decrement was interpreted by Hull, in the tradition of Pavlov, as an innate property of animal behavior. In place of Pavlov's rather strict anatomical "irradiation" interpretation, however, Hull substituted the notion of a behavioral dynamism by which response potential is apportioned to stimuli along certain physical dimensions, the amount of such potential diminishing in a systematic fashion as a function of the psychological distance between the training stimulus and the test stimulus. There were two types of response potential--excitation potential and inhibition potential--both of which generalized to an equal and opposite extent. Discrimination resulted from the algebraic summation of these mutually antagonistic gradients of excitatory and inhibitory potential (Hull, 1943, p. 265). Thus, according to the Hullian system, there is only generalization. Discrimination exists simply as an interface between two antagonistic processes: generalization of excitation and generalization of inhibition.

This point of view was diametrically opposed by Lashley and Wade (1946) who argued that there is no such thing as generalization; that no unique mechanism or dynamism exists which mediates spread of response potential, let alone systematic decrements in response potential. These writers argued intuitively that where 100 per cent generalization is observed there is actually a lack of attention on the part of the organism. Because the animal is not attending along some physical continuum, changes in the controlling stimulus along that continuum do not at all constitute changes for the animal. Attention to stimulus changes along a continuum results from the organism's associating the response with that aspect of the stimulus represented by the continuum. The association is determined before generalization by "...innate tendencies to perceptual organization, the past experience of the organism, and emphasis on one or another attribute given by the experimental situation" (1946, p. 82). Further, according to Lashley and Wade, when the organism does associate (is attending), no generalization should occur beyond that which results from confusion at stimulus values in the zone of minimal sensory acuity (discriminability). The terms "association" and "attention" seem to be used synonymously to mean "discrimination" for Lashley and Wade state that Pavlovian differentiation (discrimination) "...involves the redirection of attention to new aspects of the stimuli and the formation of new association with these" (1946, p. 74). The phenomenon of the generalization gradient, according to the Lashley and Wade theory, can be attributed to two factors:

1. habits of relational thinking (in studies using human subjects, the stimulus series usually represents familiar relational sequences) and
2. fluctuating difference thresholds (in studies using animals, testing is usually carried out over a continuum of small stimulus differences under distracting test conditions).

In sum, Lashley and Wade argue for the existence of 100 per cent generalization when the stimulus dimension is new or not attended by the organism and, in all other cases, complete discrimination limited only by the sensory capability of the organism. The so-called generalization gradient will occur to the extent that the animal's ability to discriminate absolutely is impaired, or the animal has a repertoire of "habits of relational thinking." According to this analysis there is nothing but discrimination. Generalization is simply a complete lack of discrimination and generalization decrement results from measuring discrimination under a very dubious set of experimental conditions.

Obviously the Hullian and Lashley-Wade viewpoints conflict; either one would settle the question of how to distinguish between generalization decrement and discrimination: it would simply eliminate one of them. An obstacle to accepting either viewpoint, and thus resolving the issue, is the lack of independent evidence for the postulated mediating processes, "inhibitory and excitatory potential," "association," "attention," and the like. In view of this fact it is necessary to look elsewhere for a satisfactory means of contrasting the two phenomena.

Nonformal distinctions

What basis is there for separating the concepts of generalization decrement and discrimination without appealing to hypothetical processes? If one looks at the two phenomena on the basis of what the individual scientist emphasizes when he wants it clearly understood by a colleague just what kind of experiment he is doing, it is apparent that a distinction is easily made. Conventionally, the training period preceding the demonstration of a generalization gradient involves the presentation of a single stimulus value in whose presence a response is reinforced. The training period preceding the demonstration of a discrimination function, on the other hand, always involves the alternate presentation of two stimuli, with reinforcement being administered differentially, depending upon which stimulus is present when a response is observed. Because explicit differential reinforcement is absent in the former case, even though differential responding is observed in subsequent testing, the experiment is understood to be a generalization experiment, and generalization decrement, rather than discrimination, is said to occur. This way of distinguishing the two phenomena is accepted by the majority of invesitgators, according to Brown (1965).

This consensual distinction is a convenient means for separating generalization experiments from discrimination experiments. But it is an arbitrary convention and should not lead us to conclude that discriminative conditioning (differential reinforcement) is not involved at all in the genesis of generalization decrement, or that generalization decrement is a unique sort of phenomenon embedded in a dynamism or process by which training effects spread spontaneously to previously neutral stimuli. An organism's behavior is subject to discriminative control from a wide variety of sources, and such control cannot be discounted unless it is assured that appropriate steps were taken to exclude these collateral sources from the situation. Recent evidence points to an interpretation of generalization decrement in exactly these terms. The following review of studies indicates that the observation of generalization decrement is quite contingent upon events which are conventionally viewed as preconditions for the development of discrimination.

Experimental Studies

One of the conditions upon which discrimination obviously depends is that the distance between the controlling stimuli is greater than the distance defining the resolving power (discriminability or difference threshold) of the sensory system. Interestingly enough, it has been shown that, given stimulus differences greater than the difference threshold, and a relatively small amount of training time, discrimination is still not perfect and, in fact, varies directly with the distance between the positive and negative stimulus (Frick, 1948; Raben, 1949). Ekman (1955) has further found that the discriminability function for visual wavelength can be deduced from the amounts of differential responding to stimuli separated by distances much greater than the difference threshold. It is also true that the classic difference threshold is actually defined by a discrimination function in which differential responding ranges from 0 to 100 per cent over some stimulus range, and the experimental situation is supposedly arranged to eliminate factors other than stimulus differences per se which might influence discrimination, such as amount of training, motivational parameters, etc. Such facts have led many investigators to suggest that discriminability (defined as the just-noticeable difference) is simply a very special case of discrimination (Stevens, 1939, 1965; Brown, Bilodeau, & Baron, 1955; Guttman, 1963).

Given the relationship between discriminability and discrimination, a direct relation between generalization decrement and discrimination can be demonstrated by showing a direct relation between generalization decrement and discriminability. A technique for putting this notion to test was devised by Guttman and Kalish (1956). These investigators used keypeck responses in the pigeon in an operant conditioning situation to obtain conventional peaked bi-directional gradients at selected points on the visual wavelength spectrum. Over the stimulus range used, the discriminability function is repeatedly nonmonotonic, describing three areas of high discriminability and two areas of low discriminability. Each training stimulus was selected so that discriminability was better on one side of it than on the other. The slopes of the gradients between a pair of stimulus values equidistant from, but on opposite sides of, the training stimulus were compared to the average Weber fractions for the segments bounded by each point and the training stimulus. These comparisons were expected to reveal a direct relationship between generalization and discriminability, such that when the Weber fraction was larger on one side of the training stimulus (indicating an area of poorer discriminability) the slope of the gradient would be smaller (less generalization decrement), and vice versa. The existance of this covariation between generalization decrement and discriminability would be taken as evidence for a direct relationship between generalization decrement and discrimination, according to the syllogism stated above.

The Guttman and Kalish study did not yield significant relationships in this regard. Although the gradients were not perfectly symmetrical, their slopes were not significantly different statistically and it was thus concluded that no statement could be made concerning how generalization decrement covaried with the discriminability function. A partial replication of the Guttman and Kalish experiment by another investigator (Blough, 1961) also yielded a gradient for which the hypothesis of asymmetry could not be confirmed. Kalish (1958), however, used the original Guttman and Kalish apparatus with human subjects and a verbal labeling response, and he did obtain the desired relationship. Recently, Ganz (1962), and Yarczower and Bitterman (1965) have claimed positive results, using operant conditioning techniques with monkeys and goldfish, respectively.

In the light of the positive results from these last three investigations, it is rather puzzling that Guttman and Kalish, as well as Blough, obtained the results that they did. Recent work by Shepard (1965) suggests that the negative findings of these studies may have been the result of relying on an analysis which is rather insensitive to data trends that are in the desired direction, but not of sufficient magnitude to allow rejection of the null hypothesis at accepted levels of statistical significance. Shepard has shown that if several bi-directional gradients are obtained at different points on the same continuum there is a unique transformation of the physical stimulus values that converts all gradients to a uniform shape. The resultant spacing of the physical stimuli can be taken to reflect psychological distance in j.n.d. units, according to which stimuli spaced close together are highly discriminable and stimuli spaced far apart are less discriminable. Shepard re-analyzed the data from Guttman and Kalish, and Blough in this fashion and found that analysis yielded stimulus spacings which fit quite well those spacings given by the discriminability function for the pigeon over the same range of wavelengths. By converting the latter spacings to units of relative discriminability and plotting them cumulatively, the form of the discriminability function can be compared to that implied by the Shepard transformation (Fig. 1). It can be seen that the two functions are remarkably similar in form, being accelerated at the extreme wavelengths and somewhat flattened in the areas of 530 and 590 millimicrons (areas of maximum and minimum discriminability, respectively).

For reference, the wavelength discriminability function for humans, which is generally considered comparable to that for pigeons, is also plotted in Fig. 1.

We may conclude from the studies cited above that generalization decrement is directly related to stimulus discriminability, at least in pigeons, and for stimuli on the visual spectrum. To the extent that the assertion that discrimination and discriminability are directly related is valid, these data can be taken as evidence for a direct relation between generalization decrement and discrimination.

Another approach to establishing an identity relation between generalization decrement and discrimination comes from comparing the slope of the generalization gradient to some index of discrimination difficulty. In an experiment of this type, a pair of stimuli for which there is little generalization decrement should also be difficult to discriminate in a subsequent task requiring successive discrimination. Hanson (1959) collected data on the generalization gradients of pigeons on a visual wavelength continuum, then trained them in a discrimination involving one of the generalization test stimuli as S^{Δ}, with the original training stimulus as S^{D}. A plot of the number of training periods required for criterion discrimination as a function of the S^{Δ} values yielded an ease of discrimination gradient which could be compared to the original generalization gradient. Hanson's discrimination gradient showed an initial steep slope followed by a plateau over the same range of stimuli for which there was a gradual increase in generalization decrement. Hanson concluded that the amount of generalization decrement was not a good index for predicting ease of discrimination. Recently, however, Haber and Kalish (1963) have performed a more systematic study, using response measures which make the generalization and discrimination indices more comparable. These investigators obtained generalization gradients from three groups of pigeons over a segment of the visual wavelength continuum. Each group was trained under a different schedule of reinforcement (VI 15 sec, VI 1 min and VI 4 min). The slopes of the gradients varied inversely with reinforcement density, the VI 15 sec group yielding the steepest gradient. Haber and Kalish then divided each group in half and trained each in a successive discrimination on a VI 1 min schedule to a criterion level of discriminative performance. One group received the original training stimulus as S^D and the next adjacent generalization stimulus as S^Δ. The other group received this latter stimulus as S^D and the next adjacent generalization test stimulus as S^Δ. In order to compare generalization decrement and discrimination, these investigators compared the ratio of responses made to the S^D and S^Δ in the discrimination experiment to the ratio of responses made to these same two stimuli in the generalization experiment. They thus had identical measures of differential responding in both situations. The results showed that response ratios were nearly identical for the same pair of stimuli in both experiments. Thus, when the decrement in responding from the training stimulus to an adjacent stimulus was 10 per cent during generalization testing, the decrement was also 10 per cent when these stimuli were the S^D and S^Δ, respectively, in a successive discrimination. Likewise, an animal that emitted about 20 per cent as many responses to the third stimulus as it did to the second during generalization testing, also emitted about 20 per cent as many responses to the third stimulus as it did to the second when they were the S^Δ and S^D, respectively, in a successive discrimination. The authors conclude: "The results of these experiments present rather striking evidence that, at least in certain instances, the relationship between generalization and discrimination may be regarded as inverse" (1963, p. 413).

The demonstration of direct covariation between generalization decrement and discrimination leaves open, however, the question of why generalization decrement occurs in the first place, since there is no explicit differential reinforcement involved during training. For an answer to this question we turn to a set of studies intended to determine the extent to which behavior in the generalization situation can be shown to be under discriminative control from sources other than the explicit differential reinforcement administered by the experimenter.

The first set of studies addressed to this question examines the extent to which there will be generalization decrement when all differential reinforcement with respect to the generalization stimuli is precluded from the subject's history. If generalization decrement appears under these conditions, then there are grounds for assuming that it results from an innate process, thus revitalizing the Pavlovian-Hullian formulation, and the distinction between generalization and discrimination. If, on the other hand, generalization does not occur under these conditions, then we may conclude that generalization decrement does require prior differential reinforcement, thus supporting the case for its identity with discrimination and the view that it deserves no unique treatment in the analysis of behavior.

Peterson (1962) raised ducklings in a monochromatic environment, then conditioned them to emit a key peck response for food in the presence of a light of the same wavelength. Subsequent generalization testing along the visual spectrum yielded flat (100%) generalization gradients. A control group reared under normal laboratory conditions (room light) showed the usual amount of generalization decrement. Consequently, the observation of generalization decrement for the control group was attributed to the existence of a discriminative history with respect to the stimuli used in the experiment, albeit not a history controlled by the experimenter. Additional evidence for this position comes from Ganz and Riesen (1962) who raised monkeys in total darkness (except for brief exposures to diffuse white light to prevent retinal deterioration), then conditioned the animals to emit an operant response in the presence of light of a single wavelength. Generalization testing along the wavelength continuum revealed no generalization decrement. Again, a group of control animals showed generalization decrement and yielded conventionally-shaped gradients.

These two studies provide evidence that generalization decrement will occur only when the organism has some history of discrimination with respect to the stimulus set in question. The Peterson, and Ganz and Riesen data show quite conclusively that generalization decrement will not be observed without some discriminative experience prior to the generalization experiment. Incidentally, the Lashley and Wade notion of "familiar habits of relational thinking" leading to generalization decrement is in a sense confirmed in these data; behavior under the control of a discriminative history could be considered the operational counterpart of such "habits," and "relational thinking" could be operationalized in terms of differential responding to stimulus differences regardless of their absolute values. Indeed, Honig and Day (1962) have shown that relative stimulus differences do constitute an orderly continuum for stimulus generalization.

Although the above results show that a history of differential reinforcement is a necessary condition for the observation of generalization decrement, experiments have shown that it is not a sufficient condition. Given that the organism has been provided a discriminative history involving the relevant stimulus dimension, it is also necessary to provide immediate discriminative control over the response by the training stimulus.[1] Such control is the result of correlating the training stimulus with reinforcement, and its absence (or the presence of a stimulus on some dimension other that the final generalization continuum) with nonreinforcement.[2]

Jenkins and Harrison (1960) have shown that normally-reared animals, who have no opportunity to emit unreinforced responses in the presence of an S^{Δ} during training, will yield flat (100%) generalization gradients during subsequent testing. These investigators presented a pure tone during a training period in which pigeons emitted a key peck for reinforcement. Blackout periods were interpolated in order to duplicate the conditions under which conventional wavelength generalization experiments are carried out. These periods were accompanied by removal of the S^D but did not constitute S^Δ periods, since pigeons reportedly do not peck in darkness; there was no opportunity, therefore, for differential reinforcement. When extinction was subsequently carried out, while the frequency of the tone was varied systematically, 100 percent generalization was observed. A group of control animals, whose training was identical, except that the tone was removed during S^Δ intervals without the simultaneous blackout condition, yielded conventionally-shaped gradients of generalization. Jenkins and Harrison conclude, "The present results suggest that the failure to obtain...a gradient may be due to a lack of control of the response by the experimental stimulus." We may conclude from the findings of Jenkins and Harrison that discriminative control, as well as a differential history, is required before generalization decrement can be observed during testing.

It may be objected that several studies have demonstrated generalization decrement following an experimental history, without differential reinforcement--studies in which there was no S^Δ and responses were reinforced in the continual presence of some single stimulus value. A study conducted by Heinemann and Rudolph (1963), however, allows us to maintain that the existence of generalization decrement depends on differential reinforcement in the training situation. These investigators show that stimulus control resulting from inadvertent differential reinforcement may be involved in many generalization studies that purportedly involved only reinforcement for responses emitted in the continual presence of the training stimulus (Brown, 1942; Schlosberg & Solomon, 1943; Grice, 1948; Brush et al., 1952; Guttman & Kalish, 1956; Jenkins et al., 1958; Blough, 1959). In all of these studies, the stimulus was visual and localized in space. Heinemann and Rudolph contend that in these situations responses emitted when the animal was not oriented toward the stimulus would not be reinforced, and that the discriminative control necessary for generalization to be observed could be obtained in this way, albeit fortuitously. In order to demonstrate the existence of such "inadvertent differential reinforcement," Heinemann and Rudolph used a cardboard patch surrounding a response key as the training stimulus. Each of three groups of normally-reared pigeons was trained to peck at the key when the surround was 5.5, 6.5, or 192.0 sq. in., respectively (but of constant reflectance, luminance = 10 ml.). Every group was then divided into three sub-groups and extinguished for 10 minutes in the presence of a surround of the original size but of one of three reflectances (luminances = 1.5, 10, 15 ml.). It was thought that the groups trained and tested in the presence of the larger surround would show less generalization decrement because there was less opportunity for inadvertent discrimination training and therefore less discriminative control. This expectation was confirmed. Systematically greater amounts of decrement were observed as the size of the surround decreased. The authors conclude, "If this analysis is correct, it would appear that some discriminative training must inevitably occur in any situation in which the stimulus is a fairly small visual area and the required response is a movement that is directed with respect to the stimulus, such as pecking at the stimulus, or walking or jumping through a door upon which the stimulus appears, or turning at a choice point marked by the stimulus" (1963, p.657).

Stimulus control resulting from inadvertent discrimination training may, by the same token, account for observed generalization decrement in those studies where a normally reared organism is confined in the experimental situation for a period of time but given discrete-trial training prior to generalization testing (Pavlov, 1927; Bass & Hull, 1934; Gibson, 1939; Razran, 1949; Brown, Bilodeau, & Baron, 1951; Rosenbaum, 1953; Mednick, 1955). In these situations, where there is some probability that unreinforced responses are emitted or elicited during the inter-trial interval, there exists, according to the results of Heinemann and Rudolph, the possibility that discrimination training is inadvertently taking place. This possibility has also been noted by Perkins (1953).

The data from the Peterson, Ganz and Riesen, Jenkins and Harrison, and Heinemann and Rudolph studies make it quite evident that the existence of generalization decrement is dependent upon discriminative control, the result of differential reinforcement. The sources for such control, when it is not explicitly initiated and maintained by the experimenter, are 1) the organism's history of extra-experimental differential reinforcement with respect to the stimulus continuum, and 2) inadvertent differential reinforcement in the experimental situation.[3]

Summary and Conclusions

Although generalization decrement and discrimination have identical empirical definitions, they have not been treated as equivalent phenomena in the study of behavior. The distinction seems to have been based in large part on the observation of generalization decrement over a range of stimulus values despite the absence of differential reinforcement administered by the experimenter. This observation has led some theorists to postulate mechanisms other than discrimination to account for generalization decrement--mechanisms such as those implied in Pavlov's "irradiation" hypothesis or Hull's concept of "spread of effect." Although there has been some resistance to the tendency to treat generalization as a special phenomenon, notably the writings of Lashley and Wade, the alternative interpretations are cast in such mentalistic terms as "attention" and "associations," which, as they stand, are no less nebulous than the mechanisms postulated by Hull and Pavlov. Lacking, on the one hand, a firm theoretical basis for distinguishing generalization decrement and discrimination, and, on the other, experimental evidence that they are in fact the same, most investigators have come to adopt the convention of describing as examples of generalization decrement those data in which differential responding is observed, but in which explicit differential reinforcement is lacking. By this convention the term discrimination is reserved for data from experiments in which two stimuli are obviously presented under contrasting reinforcement conditions.

However, studies have recently been performed that make this distinction spurious. It has been shown that generalization decrement varies directly with both discrimination and its more refined cousin, discriminability. In addition, studies have shown that a history of differential reinforcement with respect to the generalization stimuli, as well as immediate discriminative control of the response during training (the result of differential reinforcement), are necessary and sufficient conditions for the observation of generalization decrement. Finally, the controverting evidence from generalization studies where decrements were obtained without explicit differential reinforcement during training, has been found suspect and may be an artifact of inadvertent differential reinforcement, using stimuli localized in space and/or discrete-trial training procedures with animals that have had some discriminative history with respect to the stimulus continuum.

In the light of the evidence presented, then, it is concluded that generalization decrement and discrimination are identical phenomena reflecting the range of discriminative control for behavior along some stimulus continuum. There seems little merit in maintaining a distinction on the grounds that explicit differential reinforcement is lacking in the training phase of experiments designed to demonstrate generalization decrement. In this case the distinction is between two experimental situations in which discrimination can be observed and not between two different experimental situations in which two kinds of behavior, mediated by two different psychological processes, is observed. Generalization decrement, like discrimination, is differential responding that results from differential reinforcement.

References

Baron, M. R. Stimuli, stimulus control, and stimulus generalization. In D. Mostofsky (Ed.), Stimulus generalization. Stanford: Stanford University Press, 1965.

Bass, M. J., & Hull, C. L. The irradiation of a tactile conditioned reflex in man. J. comp. Psychol., 1934, 17, 47-65.

Blough, D. Generalization and preference on a stimulus intensity continuum. J. exp. Anal. Behav., 1959, 2, 307-315.

Blough, D. The shape of some wavelength generalization gradients. J. exp. Anal. Behav., 1961, 4, 31-40.

Brown, J. S. The generalization of approach responses as a function of stimulus intensity and strength of motivation. J. comp. physiol. Psychol., 1942, 33, 209-226.

Brown, J. S. Generalization and discrimination. In D. Mostofsky (Ed.), Stimulus generalization. Stanford: Stanford University Press, 1965.

Brown, J. S., Bilodeau, E. A., & Baron, M. R. Bi-directional gradients in the strength of a generalized voluntary response to stimuli on a visual-spatial dimension. J. exp. Psychol., 1951, 41, 52-61.

Brush, F. R., Bush, R. R., Jenkins, W. O., John, W. F., & White, J. W. M. Stimulus generalization after extinction and punishment: an experimental study of displacement. J. abnorm. soc. Psychol., 1952, 47, 633-640.

Dember, W. N. The psychology of perception. New York: Holt, 1960.

Ekman, G. Contributions to the psychophysics of color vision. Studium Generale, 1963, 16, 54-64.

English, H. B., & English, A. C. _A comprehensive dictionary of psychological and psychoanalytical terms_. New York: Longmans, Green & Co., 1958.

Frick, F. C. An analysis of an operant discrimination. _J. Psychol._, 1948, 26, 93-123.

Ganz, L. Hue generalization and hue discriminability in macaca mulatta. _J. exp. Psychol._, 1962, 64, 142-150.

Ganz, L., & Riesen, A. H. Stimulus generalization to hue in the dark reared Macaque. _J. com. physiol. Psychol._, 1962, 55, 92-99.

Gibson, E. J. Sensory generalization with voluntary reactions. _J. exp. Psychol._, 1939, 24, 237-253.

Grice, G. R. The acquisition of a visual discrimination habit following response to a single stimulus. _J. exp. Psychol._, 1948, 38, 633-642.

Guttman, N. The pigeon and the spectrum and other perplexities. _Psychol. Rep._, 1956, 2, 449-460.

Guttman, N. Laws of behavior and facts of perception. In S. Koch (Ed.), _Psychology: a study of a science_, Vol. 5. New York: McGraw-Hill, 1963, 114-178.

Guttman, N., & Kalish, H. I. Discriminability and stimulus generalization. _J. exp. Psychol._, 1956, 51, 19-88.

Haber, A., & Kalish, H. I. Prediction of discrimination from generalization after variations in schedule of reinforcement. _Science_, 1963, 142, 412-413.

Hanson, H. M. Effects of discrimination training on stimulus generalization. _J. exp. Psychol._, 1959, 58, 321-334.

Hebb, D. O. _A textbook of psychology_. Philadelphia: Saunders, 1958.

Heineman, E. G., & Rudolph, R. L. The effect of discriminative training on the gradient of stimulus generalization. Amer. J. Psychol., 1963, 76, 653-658.

Honig, W. K., & Day, R. W. Discrimination and generalization on a dimension of stimulus differences. Science, 1962, 138, 29-31.

Hull, C. L. Principles of behavior. New York: Appleton-Century-Crofts, 1943.

Jenkins, H. M., & Harrision, R. H. Effect of discrimination training on auditory generalization. J. exp. Psychol., 1960, 59, 246-253.

Jenkins, W. O., Pascal, G. R., & Walker, R. W., Jr. Deprivation and generalization. J. exp. Psychol., 1958, 56, 274-277.

Kalish, H. I. The relationship between discriminability and generalization: A re-evaluation. J. exp. Psychol., 1958, 55, 637-644.

Lashley, K. S., & Wade, M. The Pavlovian theory of generalization. Psychol. Rev., 1946, 53, 72-87.

Mednick, S. A. Distortions in the gradient of stimulus generalization related to cortical brain damage and schizophrenia. J. abnorm. soc. Psychol., 1955, 51, 536-542.

Mednick, S. A., & Freedman, J. L. Stimulus generalization. Psychol. Bull., 1960, 57, 169-200.

Pavlov, I. P. Conditioned reflexes. Oxford: Oxford University Press, 1927.

Perkins, C. C., Jr. The relation between conditioned stimulus intensity and response strength. J. exp. Psychol., 1953, 43, 225-236.

Peterson, N. Effects of monochromatic rearing on the control of responding by wavelength. Science, 1962, 136, 774-775.

Prokasy, W. F., & Hall, J. F. Primary stimulus generalization. Psychol. Rev. 1963, 70, 310-322.

Raben, M. W. The white rat's discrimination of differences in intensity of illumination measured by a running response. J. comp. physiol. Psychol., 1949, 42, 254-272.

Razran, G. H. S. Stimulus generalization of conditioned responses. Psychol. Bull., 1949, 46, 337-365.

Rosenbaum, G. Stimulus generalization as a function of level of experimentally induced anxiety. J. exp. Psychol., 1953, 45, 35-43.

Schlosberg, H., & Solomon, R. L. Latency of response in a choice discrimination. J. exp. Psychol., 1943, 33, 22-39.

Shepard, R. N. Approximation to uniform gradients of generalization by monotone transformations of scale. In D. Mostofsky (Ed.), Stimulus Generalization. Stanford: Stanford University Press, 1965.

Spence, K. W. The differential response in animals to stimuli varying within a single dimension. Psychol. Rev., 1937, 44, 430-444.

Stevens, S. S. Psychology and the science of science. Psychol. Bull., 1939, 36, 221-263.

Stevens, S. S. On the uses of poikilitic functions. In D. Mostofsky (Ed.), Stimulus Generalization. Stanford: Stanford University Press, 1965.

Yarczower, M., & Bitterman, M. E. Stimulus generalization in the goldfish. In D. Mostofsky (Ed.), Stimulus Generalization. Stanford: Stanford University Press, 1965.

Footnotes

1. It should be noted that in both the Peterson, and Ganz and Riesen studies differential reinforcement was probably present during training. Ganz and Riesen punished responses during a blackout period (S^{Δ}) and Peterson used a transilluminated response key which probably led to inadvertent differential reinforcement, as Heinemann and Rudolph define it (see below).

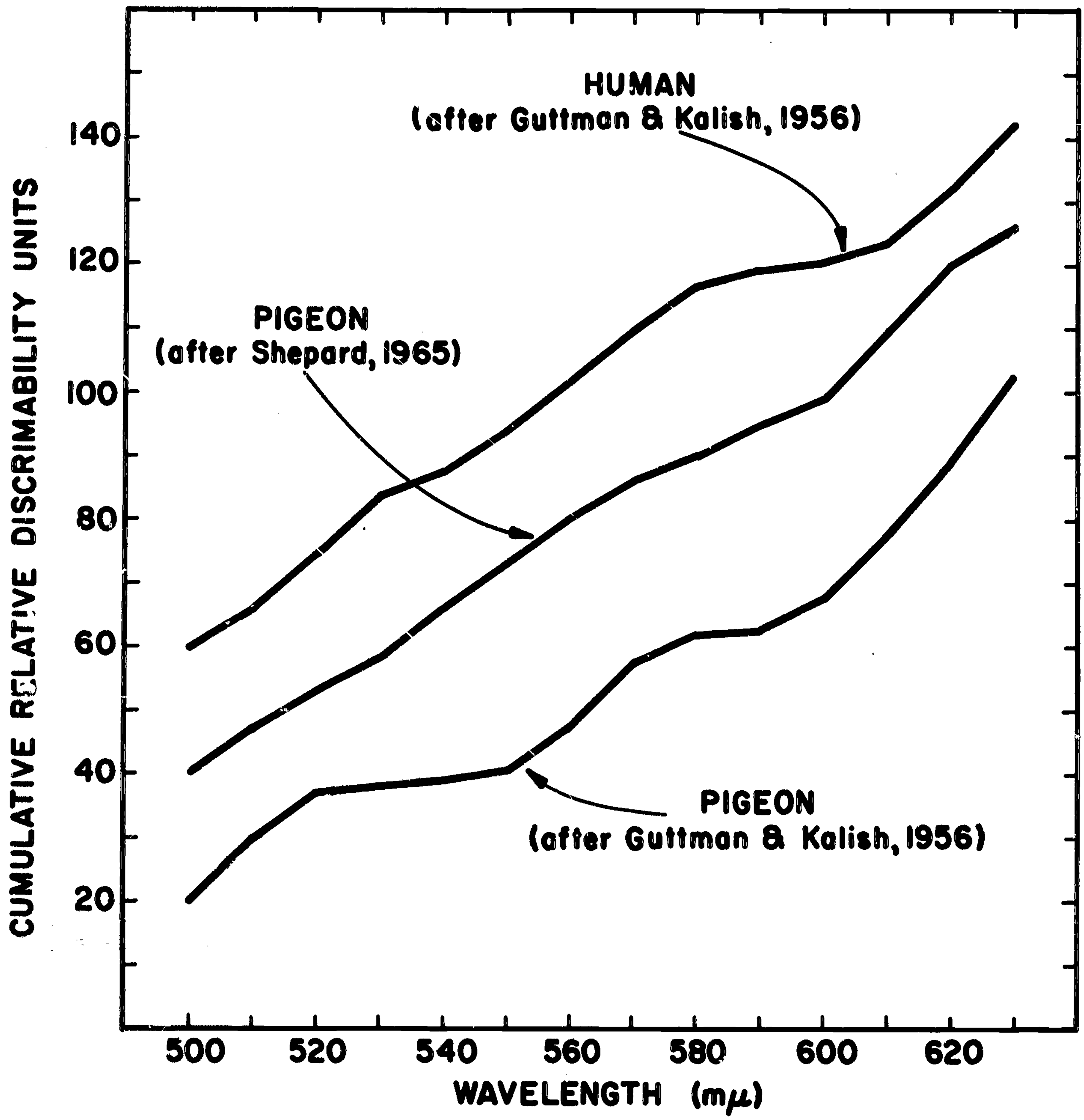
2. Of course such control would be obtained using a negative stimulus on the generalization continuum but then we would be measuring discrimination rather than generalization decrement as it is conventionally defined.

3. It may be apparent that the experimental evidence cited here supports some of the Lashley and Wade criticisms of the Pavlovian-Hullian formulation. However, the data also contravene several of Lashley and Wade's contentions. First, these authors allow the "habits of relational thinking" argument to apply to human subjects only; they attempt to invalidate the studies using animals on other grounds (e.g., the use of threshold stimulus values and noisy or distracting test conditions). However, the research of Peterson with ducklings may be considered to operationalize and then confirm the Lashley and Wade construct. In addition, the Guttman and Kalish experiment with pigeons clearly demonstrates the existence of a gradient over nearly the entire visible spectrum with stimulus control so precise that Guttman likens the effect on behavior of changes in the physical wavelength to turning on and off a water faucet (1956, p. 451). Again, Lashley and Wade argue that attention to the pertinent aspect of the stimulus, resulting from emphasis on one or another attribute given by the experimental situation among other things, is necessary for the observation of generalization decrement; a

statement in line with the findings of Jenkins and Harrison, and Heinemann and Rudolph. Lashley and Wade, however, had in mind the passive presentation of stimuli on the generalization continuum whereas these latter studies showed conclusively that differential reinforcement in the presence of two or more stimuli is necessary, and, furthermore, that the negative stimulus need not be on the generalization continuum. In retrospect, Lashley and Wade were correct in opposing Hull's insistence on a unique process to explain the generalization gradient, but by attacking Hull with Hull's own weapons (hypothetical constructs, rather than empirical generalizations) they added to the confusion in an already confused area.

Figure Captions

Fig. 1. The relationship between cumulated units of discriminability and wavelength for pigeons and humans. The function for human Ss and the lower function for pigeons are taken from Guttman and Kalish (1956) who based their plots on the data of Jones (1917) and Hamilton and Coleman (1933). The middle function is based on data derived by Shepard (1965) from the generalization gradients from pigeons by Guttman and Kalish (1956) and Blough (1961). This function represents the differential discriminability of the stimuli according to their spacing under a transformation that rendered the original generalization gradients uniform in shape. The absolute discriminability values (delta lambda) for each function were adjusted proportionately so that each function spans the same range.



HUMAN
(after Guttman & Kalish, 1956)
PIGEON
(after Shepard, 1965)
PIGEON
(after Guttman & Kalish, 1956)
CUMULATIVE RELATIVE DISCRIMABILITY UNITS
20
40
60
80
100
120
140
500
520
540
560
580
600
620
WAVELENGTH (mμ)

Fig. 1

Attentional Factors in Categorical Judgments

of Pure Tone Similarity

David V. Cross

According to Shepard (1964, p. 59) the Euclidean assumptions underlying multidimensional scaling methods would have gone largely unquestioned if it were not for Attneave's finding in 1950 that, when subjects are required to "rate" the overall similarity of pairs of geometrical figures, their ratings conform more to the city block metric than to the Euclidean. Indeed, in most discussions of scaling theory Attneave's finding is cited as the one anomalous exception to the generally satisfactory solutions achieved with the Euclidean algorithm. In an attempt to explain this finding, Torgerson (1958, pp. 254, 292) conjectured that Attneave's stimuli differed with respect to "obvious and compelling" dimensions, so that in judging the similarity of two figures a subject may behave as though he were saying "these two stimuli differ this much with respect to brightness, plus this much with respect to size, plus this much with respect to shape." Torgerson added that Attneave's special selection of stimuli presented very favorable conditions for obtaining the "simple sum" type of judgment: they were selected so as to form a "plus" when plotted as points in physical coordinates--stimuli that varied simultaneously in two dimensions were avoided. Torgerson believes that this arrangement would serve to direct the subject's attention to the "given" dimensions whereas with a more or less random scattering of points the dimensions of comparison might be less obvious and "the subject might be more likely to judge the over-all difference directly." Thus, according to Torgerson, the form of the metric may depend on whether or not the dimensions of stimulus variation are obvious and compelling to the subject. Similarly, Shepard (1964, p. 80) proposed that perhaps a distinction should be made between two types of stimuli: "those that are reacted to as homogeneous, unitary wholes, and those that tend to be analyzed into perceptually distinct components or properties." Stimuli of the first type, color samples for example, tend not to be analyzed into separate components but are reacted to directly. A different dimension of comparison is likely to be invoked by each stimulus pair and the basis of judgment may be influenced by the context of other stimuli presented in the experiment. On the other hand, the over-all difference between a pair of stimuli of the second type tends to be analyzed into the component differences with respect to each of the separate fixed dimensions of the space.

The question raised here appears to be to what extent do the stimulus properties that exert behavioral control in a given experiment depend upon the direction and magnitude of prevailing stimulus variation with respect to these properties; in other words, do the effective dimensions of control for a given stimulus domain represent a unique and stable frame of reference for perception or do they arise in the experimental situation as a consequence of the investigator's special selection of stimuli? In terms of a metric representation for the stimuli, the absence of fixed directions of control suggests that psychological space is Euclidean in nature and its axes are subject to essentially arbitrary rotation. It has been noted (Cross, 1965), in a discussion of the effects of selective attention on the metric structure of pooled data, that measures of stimulus control based on averaged data may appear to conform to a Euclidean metric when in fact the directions of control for individual <u>Ss</u> are unique and fixed but differ in orientation in the space for different subjects, or differ for the same subject upon different presentations of a given stimulus.

Shepard contended that the component attributes of analyzable stimuli may be selectively attended by some subjects and "subjects might very well divide their attention in some fluctuating manner between the component dimensions," but that averaged data collected under these conditions would not permit a spatial (metric) representation of the stimuli because the condition of triangle inequality would be violated. The results of a recent study (Cross, 1965) show this is not necessarily the case; although the component dimensions were selectively and exclusively attended on each stimulus presentation, the pooled data exhibited simple metric structure. It is appropriate, in view of this outcome, to question the assumption that Ss take all relevant aspects into account when rating the similarity-dissimilarity of complex stimuli or that the conformity of averaged ratings to a city block metric necessarily implies that in a single judgment S "simply sums" the perceived unidimensional differences. It seems not at all unlikely that the component properties of complex stimuli may be as selectively effective in the control of rating behavior as in the control of simple discriminative responses and that control may be exerted by different stimulus properties for different subjects or for the same Ss upon different presentations of a given stimulus pair. A city-block outcome may indicate that the effective dimensions of control are in fact the dimensions manipulated by the investigator, and that the additive structure occurs because, in single judgments, Ss selectively attend one or another dimension exclusively. On the other hand, the emergence of alternative dimensions of control that differ for individual Ss in directions that cut across the reference dimensions (the dimensions involved in the prediction of multidimensional distances) may result in measures of stimulus control that conform more to the Euclidean than to the city block metric.

With the hypothesis that the multidimensional structure of behavioral data arises from the pooling of individual unidimensional observations, the following study examines selective attention to single dimensions of stimulus variation in categorical judgments of the similarity-difference relations among two-dimensional stimuli. Subjects are required to rate pairs of pure tones with respect to their apparent difference under conditions in which (a) each stimulus pair differs with respect to frequency alone or intensity alone, and (b) the stimuli differ simultaneously in both properties. The particular selection of stimuli was based on generally accepted psychophysical evidence that the physical properties of frequency and intensity interact in the control of discriminative behavior and also on the basis of the findings of a preliminary study by this investigator that individual Ss differ widely both with respect to the nature and the degree of this interaction. In the study in question several comparison tones differing from a fixed standard in frequency alone, intensity alone, and in both properties simultaneously were presented to 40 Ss half of whom were required to judge whether the comparison tone was greater than, equal to, or less than the standard in "loudness." The other half were required to base their judgments on "pitch." The findings for some Ss showed no evidence for interaction (for the range of values involved); that is to say, the control exerted by a given intensity change for a "loudness" comparison was independent of frequency level, and similarly for comparisons based on "pitch." The findings for several other Ss, on the other hand, indicated that the control exerted by frequency or intensity on comparative judgments of the correlated attribute depended markedly on the level of the other property and that the direction of this dependence differed between Ss. This finding indicates that dimensionalization of the tonal space differs for individual Ss. A second objective of the following study, therefore, is to examine the metric structure of the obtained data under the hypothesis that the interaction of physical continua in the control of individual similarity ratings will give rise to effective dimensions of control that differ in direction in the tonal space for individual Ss and therefore favor a Euclidean embedding.

Method

Subjects

Fourteen University students served as Ss in individual sessions lasting approximately 80 min. They were paid at the rate of $1.25/hour. None had previous experience as a participant in a psychophysical experiment.

Stimuli and Apparatus

The stimuli were 250 msec pure tones varying in frequency and intensity. They were prerecorded on magnetic tape as a sequence of pairs with 2 sec intervals between members of a pair and 6 sec intervals separating successive pairs. The first member of each pair was always the same tone with a frequency of 205 cps and an intensity level of 90 db re: .0002 dynes/cm^2 (as measured upon playback at S's headphone). This was designated as the standard. The second tone in each pair was a comparison tone which was varied in eleven 1-cps steps from 200 to 210 cps and/or in eleven 1 db steps from 85 to 95 db SPL.

The signal source was a precision audio oscillator (Hewlett Packard 241A). The desired tonal frequency could be rapidly and accurately selected by push-button. The output of the oscillator was passed through a variable attenuator (Hewlett Packard 350D), the settings of which determined the relative intensities of the tones, and an electronic switch (Grason-Stadler 829D). The switch, controlled by an interval timer (Grason-Stadler 471-1), passed signals, with the desired duration and interstimulus intervals, to a tape recorder (Ampex 351-2U).

Two stimulus sets were constructed so the relation in frequency-intensity coordinates, between the array of comparison stimuli and the standard was different. The frequency-intensity combinations comprising the stimuli in each set are shown schematically in Fig. 1. Set A stimuli are represented by the spots and are seen to form a "plus" in the coordinate system; that is, they form two arrays perpendicular to the axes and intersecting at the standard stimulus. Set B stimuli, represented by crosses, consisted of the standard tone plus an equal number of comparison tones scattered in each quadrant. The selection of frequency-intensity combinations for set B tones was quasi-random, subject to the constraint that for each stimulus in a given quadrant there resulted a stimulus in each adjacent quadrant with the same frequency or the same intensity. Thus, each stimulus set contained 21 different comparison stimuli (including the standard which was also paired, as a comparison tone, with itself).

The experimental space was a sound-attenuating room containing only a table and chair. The reproducing equipment used to present stimuli to S occupied an adjoining room. The tones were reproduced by a tape-recorder (Ampex 601) and presented binaurally to S through calibrated earphones (Grason-Stadler TDH-35). Interposed between the tape recorder and the earphones was a band-pass filter (Allison 2BR) which was adjusted to attenuate random noise.

Procedure

The Ss were seated individually in the experimental chamber and the following instructions were read:

You will be presented a sequence of pairs of tones. The first tone in each pair is referred to as the standard and it will always be the same. The second tone in each pair is presented for comparison with the first. It may be the same or different. Your task is to estimate how much the second tone in each pair is like the first tone by rating it on a scale from 1 to 5 according to the following criteria:

1. no difference
2. slightly different
3. moderately different
4. quite different
5. extremely different.

Each S was presented with both stimulus sets in 10 randomly permuted orders for a total of 10 presentations of each stimulus pair. Seven Ss received set A stimuli first, then set B, and the other 7 received the sets in the opposite order. A 10 min rest period was provided between the two presentations which required approximately 30 min each.

Results

The mean category ratings corresponding to frequency changes (solid lines) and intensity changes (dashed lines) for set A stimuli are shown for each S separately in Fig. 2. The Ss who rated set A stimuli first have primed numbers. The others rated set A stimuli after rating set B. It is indicated that individual Ss differ considerably with respect to the relative importance of the two stimulus properties in the control of rating behavior and in the overall range of scale values obtained. On the one xtreme, some Ss (3, 5, 9 & 12) attended only frequency differences; intensity differences were all rated the same as no difference at all. On the other hand, the ratings of four other Ss (1, 4, 6 & 8) appear to be predominated by intensity changes; frequency differences had slight or no effect on mean category rating. Of the remaining Ss, four appear to have been attending both properties, one subject (11) rated nearly all stimuli as "not different" from the standard, and the category ratings of another subject (2) show no systematic relation to stimulus change at all.

The distributions of response frequency over rating categories for the pooled data of 13 Ss (S2 was omitted) are shown in Table 1. Each comparison stimulus is denoted by an ordered pair ($^{\Delta}f$, $^{\Delta}I$) indicating the difference between it and the standard stimulus in both frequency (cps) and intensity (db). The number of responses falling in each rating category over the maximum total of 130 presentations of each stimulus (10 to each S) is given.

For the most part the distributions of responses over rating categories disguise the fact demonstrated in Fig. 2 that several Ss were attending different properties of the stimulus. The distributions tend to flatten the more the comparison stimulus differs from the standard but only the distributions of ratings for (+0, +5) and (+0, +4) depart markedly from unimodality.

The mean category assignment for each stimulus pair (S_{oo} , S_{ij}), where i and j index the physical difference between the standard and comparison stimulus with regard to frequency and intensity, respectively, is denoted by $g(S_{ij})$ and is presented in the indicated columns in Table 1. Separate category scales were thus obtained for the two stimulus sets. The distance between the standard stimulus and each comparison stimulus in sets A and B was taken as the difference between the mean category assignment for the given stimulus and that for the standard stimulus of the corresponding set. This change of origin for the two scales is dictated by the requirement that the distance of a point from itself is zero. In Fig. 3 these presumptive distances (multiplied by a factor of 10) are shown as a function of frequency with intensity constant, $\bar{g}(S_{io})$, and as a function of intensity with frequency constant, $\bar{g}(S_{oj})$.

The distances obtained for the comparison stimuli in set B (tones that differ from the standard in both frequency and intensity) are presented in Fig. 4 as a function of the predictions, $\hat{X}$, arising from four alternative rules governing the combination of component unidimensional distances, $X_f = \bar{g}(S_{io})$ and $X_I = \bar{g}(S_{oj})$ obtained for the corresponding set A stimuli. Each panel depicts the goodness of fit of the indicated model in the form of a scatter plot relating obtained and predicted quantities. The dashed lines drawn at 45^o to the axes in each panel represent perfect prediction. The four models compared are: (1) the dominance model, $\hat{X} = \max \{X_f, X_I\}$, (2) the Euclidean model, $\hat{X} = [X_f^2 + X_I^2]^{1/2}$, (3) the city block model, $\hat{X} = X_f + X_I$, and (4) a multiple linear regression model which optimally weights the component unidimensional distances in an additive rule of combination and provides an additive constant which adjusts the origin of the scale.

Of the three spatial models the Euclidean appears to achieve the closest direct fit to the data with a mean algebraic difference between predicted and obtained distances of -0.1 (-0.01 in original category scale units). The product moment correlation coefficient between obtained and predicted measures is 0.967 (which, coincidentally, is identical to the correlation reported by Attneave between obtained inter-stimulus distances and fitted quantities obtained from a multiple linear regression solution to his data).

The dominance combination rule clearly under-predicts obtained distances, with a mean algebraic difference between predicted and obtained of -1.6. The city block model consistently over-predicts with a mean algebraic difference between predicted and obtained of +3.8. Curiously, although of the three spatial models the city block metric provides the poorest direct fit to the data, the quantities predicted by this model show the highest correlation with the obtained measures, with a coefficient of 0.986. The multiple linear regression model provides an excellent direct fit to the data. The mean algebraic difference between predicted and obtained distances is, of course, zero and a multiple correlation coefficient of 0.987 was obtained for the two measures, indicating that 97.5% of the total variance in obtained distance measures is accounted for by the reconstruction.

Discussion

The findings of this study are consistent with the notion that the multidimensional structure of behavioral data arises from the indiscriminate pooling of individual unidimensional observations. It is clearly indicated that, in rating the apparent differences among two-dimensional stimuli, Ss selectively attend single dimensions of stimulus variation, since individual Ss differ with respect to the dimension attended, however, averaged data tend to disguise this fact and give rise, instead, to the erroneous inference that, in single judgments, Ss somehow integrate preceived differences along all relevant dimensions of stimulus change.

It was demonstrated that the practice of combining the responses of individuals without regard to their state of attention does not necessarily result in data that violate the assumptions of a metric embedding. In fact, a Euclidean metric provides a satisfactory representation of the relation between two-dimensional stimulus differences and their component unidimensional differences.

The appropriateness of the Euclidean solution is interpreted as resulting from the fact that the physical properties of the stimuli interact in the control of behavior, giving rise to effective dimensions of control that may be peculiar to individual <u>Ss</u> and that are not collinear with the physical dimensions of the stimulus space.

Table 1

Pooled data of 13 Ss: total number of responses falling in each rating category and mean category assignments for each stimulus.

Set A:							Set B:						
Stimulus	Rating Category						Stimulus	Rating Category					
(Δf, ΔI)	(1)	(2)	(3)	(4)	(5)	$g(S_{ij})$	(Δf, ΔI)	(1)	(2)	(3)	(4)	(5)	$g(S_{ij})$
(-5, +0)	18	50	31	18	13	2.68	(-5, +2)	18	57	23	24	8	2.59
(-4, +0)	18	44	46	15	7	2.61	(-5, -5)	12	24	46	20	28	3.21
(-3, +0)	32	59	26	10	3	2.18	(-4, +4)	12	28	36	34	20	3.17
(-2, +0)	44	57	18	11	0	1.97	(-4, -1)	37	51	21	14	6	2.25
(-1, +0)	70	48	8	3	1	1.59	(-3, +5)	15	28	40	30	17	3.05
(+1, +0)	72	52	5	0	1	1.51	(-3, -4)	27	44	36	18	4	2.44
(+2, +0)	37	65	22	5	1	1.98	(-2, +1)	68	39	14	7	2	1.74
(+3, +0)	35	49	38	7	1	2.15	(-2, -2)	62	36	21	9	2	1.87
(+4, +0)	27	33	48	19	3	2.52	(-1, +3)	54	26	34	13	3	2.12
(+5, +0)	10	34	57	21	8	2.87	(-1, -3)	55	43	21	9	1	1.90
(+0, +0)	91	27	11	0	1	1.41	(+0, +0)	97	30	1	1	0	1.27
(+0, +5)	37	18	29	13	33	2.90	(+1, +1)	84	40	5	1	0	1.41
(+0, +4)	39	20	27	29	15	2.70	(+1, -2)	64	55	9	1	1	1.62
(+0, +3)	48	35	18	22	7	2.27	(+2, +5)	23	31	28	27	21	2.94
(+0, +2)	55	40	16	15	4	2.02	(+2, -4)	23	51	42	13	1	2.37
(+0, +1)	77	34	14	4	1	1.60	(+3, +3)	23	46	35	20	6	2.54
(+0, -1)	81	43	6	0	0	1.42	(+3, -3)	18	61	44	7	0	2.31
(+0, -2)	66	39	16	9	0	1.75	(+4, +2)	26	36	50	14	4	2.49
(+0, -3)	60	38	22	7	3	1.88	(+4, -5)	5	41	60	16	8	2.85
(+0, -4)	33	48	28	12	9	2.35	(+5, +4)	14	10	47	43	16	3.28
(+0, -5)	38	30	33	16	12	2.49	(+5, -1)	25	48	43	14	0	2.35

References

Attneave, F. Dimensions of similarity. *Amer. J. Psychol.*, 1950, 63, 516-556.

Cross, D. V. Metric properties of multidimensional stimulus control. Unpublished doctoral dissertation, The University of Michigan, 1965.

Shepard, R. N. Attention and the metric structure of the stimulus space. *J. math. Psychol.*, 1964, 1, 54-87.

Torgerson, W. S. *Theory and methods of scaling*. New York: Wiley, 1958.

Figure Captions

Fig. 1. The parameter values of two sets of stimuli (e, x), varying in frequency and intensity, used in an experiment on judgments of similarity.

Fig. 2. The mean category ratings corresponding to frequency changes (solid lines) and intensity changes (dashed lines) for set A stimuli for each S.

Fig. 3. The change in mean category assignment as a function of the change in frequency (intensity constant) or the change in intensity (frequency constant).

Fig. 4. The accuracy of each of four rules for combining unidimensional distances obtained with the set A stimuli in predicting two dimensional distances obtained with the set B stimuli. The distance between the standard stimulus and each comparison stimulus is the difference between the corresponding mean category assignments. The rules of combination are those entailed by (L. to R., Top to Bot.): the dominance model, the Euclidean model, the city-block model, and a multiple linear regression model.

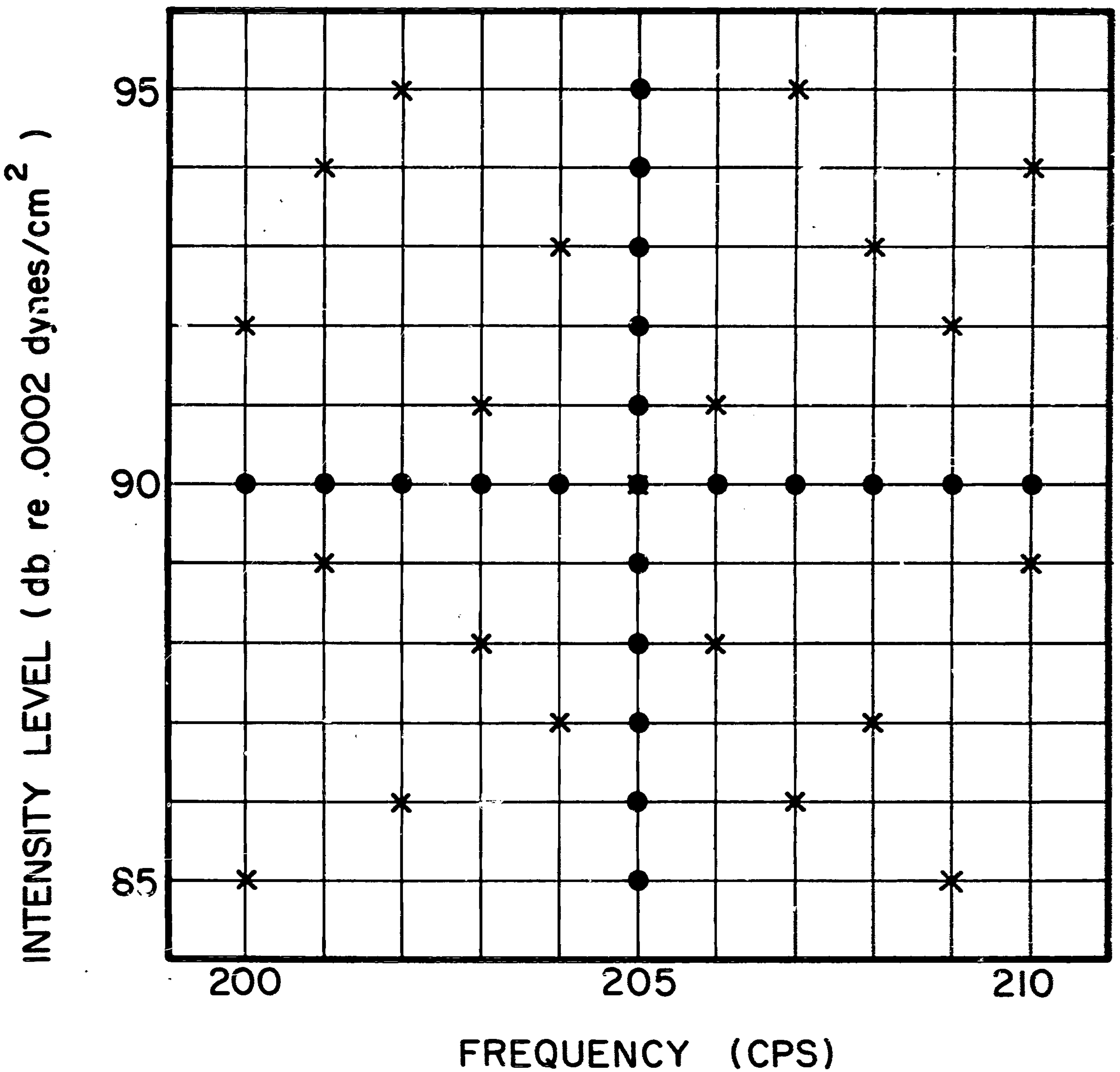
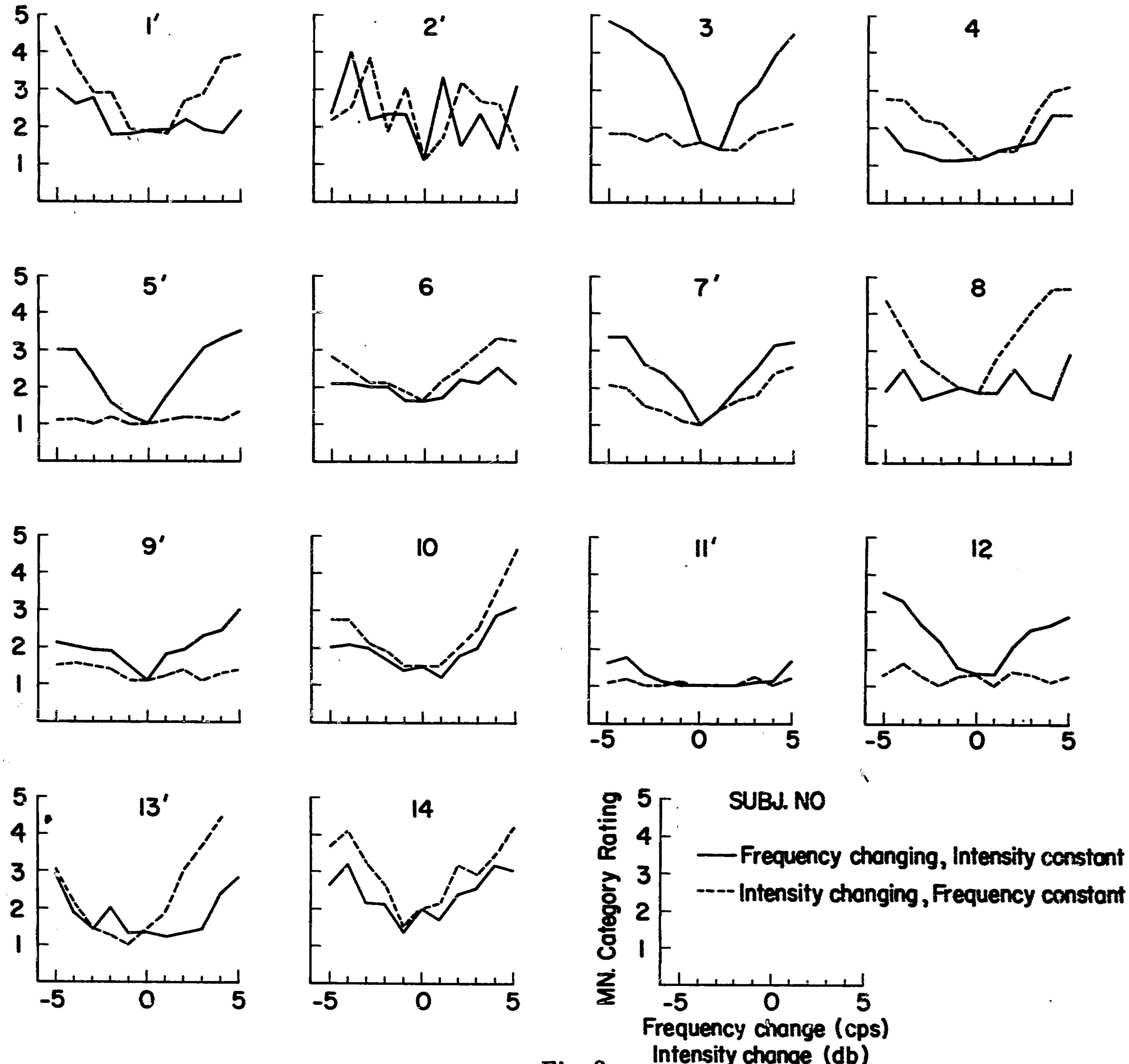
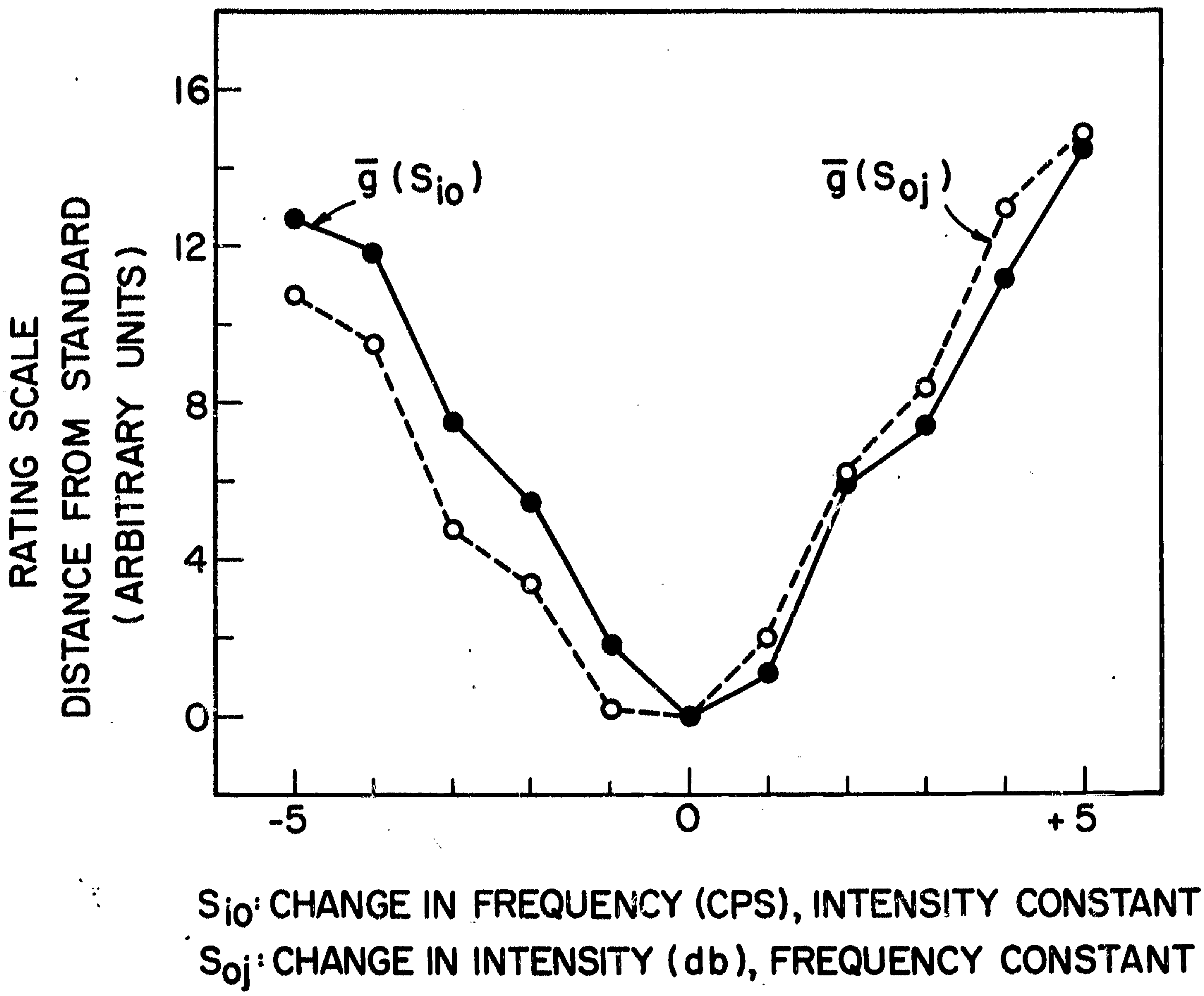

95
90
85
INTENSITY LEVEL (db re .0002 dynes/cm2)
200
205
210
FREQUENCY (CPS)

Fig. 1



1'
2'
3
4
5'
6
7'
8
9'
10
11'
12
13'
14
SUBJ. NO
Frequency changing, Intensity constant
Intensity changing, Frequency constant
MN. Category Rating
Frequency change (cps)
Intensity change (db)

Fig. 2



16
12
8
4
0
RATING SCALE
DISTANCE FROM STANDARD
(ARBITRARY UNITS)
$\bar{g}(S_{io})$
$\bar{g}(S_{oj})$
-5
0
+5
S_{io}: CHANGE IN FREQUENCY (CPS), INTENSITY CONSTANT
S_{oj}: CHANGE IN INTENSITY (db), FREQUENCY CONSTANT

Fig. 3

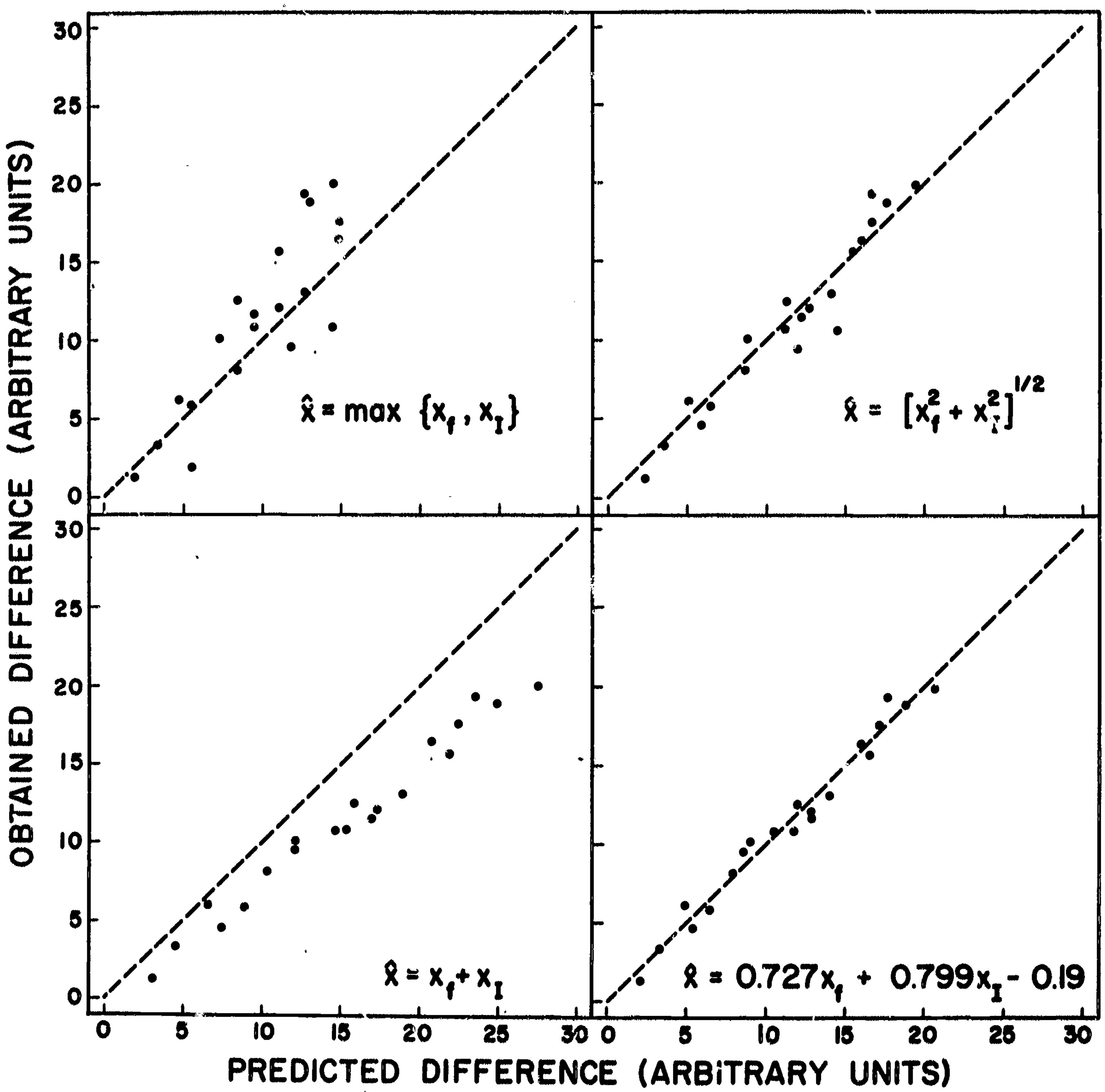

OBTAINED DIFFERENCE (ARBITRARY UNITS)
PREDICTED DIFFERENCE (ARBITRARY UNITS)
$\hat{x} = \max\{x_f, x_I\}$
$\hat{x} = [x_f^2 + x_I^2]^{1/2}$
$\hat{x} = x_f + x_I$
$\hat{x} = 0.727x_f + 0.799x_I - 0.19$

Fig. 4

Loudness of Pure Tones as a Function of Frequency,

Intensity, and Middle-Ear Mechanics

Part I

Matching Functions for Loudness

Strange Ross

Loudness of pure tones is a complex function of many variables. These variables are associated partly with the stimulating tone, partly with the receiving organism. When only pure tones are considered, the relevant stimulus variables can be exhaustively described by stating the frequency, the sound-pressure level at the eardrum, and the duration of the stimulating tone. The organismic variables are harder to classify and state exhaustively. A partial list of these variables would include the stimulation history of the organism, the particular psychophysical method employed in collecting data, and a host of variables relating more specifically to a given organism. Of these latter variables, the static as well as the dynamic properties of the middle ear as a sound-conducting pathway plays an obvious role. In addition, inner-ear processes as well as "higher-level" variables, including possibly irreducible "individual differences," act to further complicate the total picture.

Previous research has explored the relationship between loudness and most of the variables mentioned. Of direct interest to the present investigation the following areas of study may be mentioned. Many studies (e.g., Kingsbury, 1927; Fletcher & Munson, 1933; Churcher & King, 1937; Feldtkeller & Zwicker, 1956; Robinson & Dadson, 1956) have examined the role of frequency and intensity in the perception of loudness, resulting in sets of so-called "equal-loudness contours." Likewise, the growth of loudness with intensity at a given frequency has been studied experimentally as well as theoretically (e.g., Garner, 1954; Stevens, 1955; Zwicker, 1958, 1963; Zwislocki, 1965). And recently, the development of new experimental tools has resulted in an upsurge of inquiries into the acoustical properties of the middle ear (e.g., Zwislocki, 1962, 1963; Dallos, 1964).

The present research attempts to integrate a number of these previous approaches, and at the same time to remedy some of their shortcomings. Some of these shortcomings, in the opinion of the present author, are the following: 1) Very often only group data have been reported. This prevents a satisfactory analysis of individual invariances as well as the recognition of apparently irreducible and genuine individual differences. 2) In the collection of data for the construction of equal-loudness contours, traditionally a standard tone of fixed frequency has been employed. This procedure would seem to favor an undesirable bias, and also introduces undue variability in the data for comparision frequencies widely separated from that of the standard. 3) Most studies fail to give a satisfactory specification of the intensity of the stimulating tones. The practice of employing earphones calibrated on an artificial ear has long been recognized to provide only a very approximate indication of the actual sound-pressure level produced at the eardrum. 4) Practically all studies of changes in acoustic impedance at the eardrum, from which the conductive properties of the middle ear may be inferred, relate these changes to the intensity of a tone stimulating the <u>opposite</u> ear. Thus, previous studies provide no data on the dynamic properties of the middle-ear mechanism as a direct function of the stimulating tone. 5) Due to technical difficulties, only a rather limited frequency range has been employed in exploring the sound-conducting properties of the middle ear. 6) Finally, no studies are available which attempt to relate frequency and intensity of the stimulating tone as well as the properties of the middle ear to the perception of loudness.

The present research attempts to overcome the shortcomings listed above in the following way. For three subjects, loudness balances were obtained for a great number of combinations of 13 different frequencies, extending from 20 cps to 5000 cps. In so doing, the sound-pressure level of the tones was measured very close to the eardrum. From the data collected from each subject, a symmetric and transitive subset was selected for further analysis. Growth of loudness with intensity was examined by analyzing the form of the matching functions, specifically by comparing the data with curves describing the theoretical form of a matching function derived from specified assumptions concerning the growth of loudness with intensity. Further, measures of acoustic impedance at the eardrum as a function of frequency and intensity of a stimulating tone were collected from each subject. From these data, the static as well as the dynamic properties of each middle ear will be inferred. These latter results will be applied to an analysis of the equal-loudness contours constructed for each subject from the above-mentioned subset of symmetric and transitive matches. When this analysis is completed, it is hoped that equal-loudness contours referred to volume velocity at the oval window can be drawn for each subject. This set of curves would clearly indicate the role played by the inner ear in determining the relationship between loudness on the one hand, and frequency and intensity of a stimulating tone on the other hand.

Part I of this report, presented here, describes apparatus and procedure employed in collecting loudness balances. Also, the resulting matching functions are analyzed in terms of a rationale briefly outlined. A more thorough discussion of the significance of the findings, however, will be postponed to later parts of this report.

Method

Apparatus

The functions required of the experimental set-up for obtaining loudness balances (matches) were the following: 1) to present alternately two tones of different frequencies with no silent interval between the tones, and with no audible switching transients; 2) to allow the subject to adjust continuously the intensity of one of the two alternating tones (the comparison tone); 3) following a signal from the subject, indicating the completion of a loudness match, to make a permanent record of the intensities of the two tones; 4) after each adjustment, to increase the intensity of the standard tone by a predetermined amount; 5) to make a corresponding increase in the level of the comparison tone, in order to maintain an appropriate intensity range of this tone; 6) after each completed adjustment, to introduce various amounts of attenuation, randomly programmed, of the comparison tone in order to eliminate possible cues from the position of the subject's potentiometer; 7) to present the two tones, at all intensity levels, with a low degree of distortion; 8) to maintain, at all intensity levels, a high signal-to-noise ratio for the two signals; 9) to ensure a high degree of reproducibility of the sound pressure levels presented to the subject; 10) to permit an absolute determination of the sound pressure levels presented, as measured at the subject's eardrum.

These objectives were accomplished in the following way (see Fig. 1).

Insert Figure 1 about here

The standard tone was generated by an audio oscillator (General Radio 1304-B).

The signal was amplified (General Radio 1206-B) and then fed through two attenuators to one channel of an electronic switch (Grason-Stadler 829D), set at a switching rate of 2.5 secs/cycle. One of the attenuators mentioned served to adjust the general signal level in order to obtain the highest signal-to-noise ratio possible, while the other attenuator was specially constructed to permit the automatic increase in intensity of the standard tone by a predetermined multiple of .5 db. This "stepping" attenuator consisted of two 10-section ladder attenuators in steps of .5 db and 5 db, respectively. Each ladder was terminated by a cathode follower in order to obtain the high terminating impedance required for this type of attenuator, and the 10 sections of each ladder were connected to the terminals of one of two sections of an electro-mechanical counter with electrical read-out (Neuron Counter 7005, Data Instruments). Electro-mechanical devices (Grason-Stadler) generated, following the subject's signal, the number of pulses required to raise the intensity of the standard tone by a predetermined amount (in multiples of .5 db).

The comparison tone was generated by a second oscillator (Bruel & Kjaer 1014) which fed the signal through a "sone" potentiometer, controlled by the subject, and three attenuators to the other channel of the electronic switch. One of the three attenuators mentioned was a fixed attenuator which served to control the signal-to-noise ratio. The second was a "stepping" attenuator identical to the one described above. And the third attenuator was a relay-operated attenuator with four fixed values of 0, 2.5, 5.0, and 7.5 db; a stepping switch (Grason-Stadler E3126B) was used to select--according to a random schedule--one of the four possible attenuations after the completion of each adjustment.

In order to keep the distortion (arising from the cathode followers of the "stepping" attenuators and from the electronic switch) at a minimum, each output of the electronic switch was passed through a band-pass filter

(Krohn-Hite 310-ABR) set at ±15 per cent of the respective frequencies, and the outputs of the two filters were combined in a matching network. The output of this network was amplified (General Radio 1304-B, output section), and finally fed through an attenuator (General Radio 1450-TBR). The distortion of the final output was always below 1 percent.

Depression of the subject's key initiated the following operations: 1) The pen of a graphic level recorder (Bruel & Kjaer 2305) was momentarily lowered to make a dot on the recording paper, indicating the momentary voltage level at the output of the "stepping" attenuator; (through the calibration of the system, this voltage level was converted to the corresponding voltage level over the telephone). 2) The two "stepping" attenuators were advanced 5.0 db. 3) The recording paper was advanced by an amount corresponding to the increment of the intensity of the standard tone. 4) The "random" attenuator was adjusted to its next value in a random series.

The signals from the final attenuator were applied to the subject's ear by means of a small electro-dynamical telephone (Beyer DT-508), mounted in an ear insert moulded to fit the external meatus of the individual subject. The ear insert contained a canal for the conduction of the sound, a steel tube through which a probe microphone could be inserted, and a narrow steel tube for equalizing the static pressure of the air on both sides of the ear insert. When a probe microphone was not inserted, a dummy with the same acoustical properties took its place. This technique for applying the stimulus (which has been described in greater detail elsewhere [Ross, 1963]) affords the following advantages: 1) Due to the individual moulding of each ear insert, the position of the sound-transducing elements can be closely replicated; this, in turn, means that from one experimental session to the next, the same sound pressure levels are always produced at the eardrum for given stimulus parameters. 2) These sound pressure levels can be measured, thereby providing a conversion from the electric parameters of the signals to absolute acoustic measures. 3) Due to the tight acoustic coupling between the telephone membrane and the enclosed small volume of air (totalling approximately 2cc), the distortion of the acoustic signal is kept very low; for all frequencies employed, and for sound pressure levels up to 120 db SPL, the distortion due to the telephone was found to be less than 1 percent.

Procedure

At the outset of each experiment, four threshold determinations were made for each of the two frequencies involved in the series of matches to follow. The test tone was presented through the comparison channel, and the standard channel was disconnected. The test tone thus was on for 1.25 secs., off for 1.25 secs., etc. The subject was requested to adjust the tone to "just not audible" or to "just audible," whatever he preferred, and to press his key when the adjustment was completed.

Following these threshold determinations, the intensity of the standard tone was adjusted to approximately 5-10 db below the subject's threshold as estimated from the four threshold adjustments, and the subject was requested to adjust the intensity of the comparison tone to the same loudness as that of the standard tone, and to indicate a completed adjustment by pressing his key. After each key press the intensity of the standard tone was automatically increased by 5 db.

During the first few trials, the subject would not be able to hear the standard tone, and according to the instructions he would indicate this by turning down the intensity of the comparison tone all the way before pressing his key. In this way loudness matches from levels very close to the thresholds were secured. The subject continued to match the two tones until he came close to an uncomfortable loudness level. Subjects were able to make 4-6 complete runs a day. After each complete run the subjects rested for about fifteen minutes.

Thirteen frequencies extending from 20 cps to 5000 cps, and spaced approximately 1/5 decade apart, were employed; (the calibration procedure of the sound transducer as outlined above prevented the inclusion of frequencies higher than 5000 cps). All (57) frequency combinations up to and including a spacing of 6/5 of a decade were employed, except identical frequencies. For any frequency combination, each frequency served once as standard and once as comparison tone. If the matches resulting from two complete runs exhibited systematic differences, one of the two constellations was repeated. In case these latter results departed from the matches with reversed frequencies, the other constellation was repeated. As a general rule, no more runs involving the given frequency combination were carried out after these repetitions. The total number of complete runs for the subjects JR, LL, and FP was 236, 200, and 176, respectively. On the average, each complete run covered an intensity range close to 100 db, giving, for each subject, a total of approximately 4700, 4000, and 3500 matches, respectively.

Treatment of Data

For each complete run of loudness matches, the graphical record of the level recorder was transcribed to regular graph paper. All runs involving the same two frequencies were plotted on the same graph, using different signatures to identify the individual sets of data.

As mentioned earlier, it was desired for each subject to select a subset of data that satisfied the demands of symmetry and transitivity. This objective was accomplished in the following way. On each graph depicting all matches of a given pair of frequencies, smooth curves were drawn by hand through those regions for which at least one set of matches showed no systematic departures from at least one other set of matches with the opposite frequency configuration (that is, standard and comparison frequencies were interchanged for these two sets of matches). Thus, these smooth curves represent a subset of matches that is symmetric with respect to standard and comparison frequencies.

Next, the transitivity of this subset of matches has to be tested, that is, it must be ascertained whether the intensity I_1 of a tone with frequency f_1 that matches tone (f_2, I_2) can be successfully predicted from two other matches, viz., the match of (f_2, I_2) to a third tone (f_3, I_3), and the match between (f_3, I_3) and (f_1, I_1). This test was accomplished by devising a technique for averaging the subset of data described by the smooth curves mentioned above, and then checking these average values against the values of the subset. Since the averaging technique is based on an assumption of transitivity, a significant departure of average values from observed values would indicate that the demands of transitivity were not fulfilled.

The averaging process is carried out for as many loudness levels as desired; in the present case a spacing of 10 db of the corresponding intensities of a 320 cps tone was employed. Table 1 illustrates the process.

Insert Table 1 about here

A tone of 320 cps and, say, 30 db SPL, was selected as a starting point. From all graphs containing 320 cps as one frequency, the intensity of the matching tone was read off from the smooth curve drawn; if no such curve was drawn in the given region, the corresponding entry was omitted. The resulting set of values describes a first approximation to an equal-loudness contour through a point corresponding to the originating tone.

As th next step, this process was repeated, this time employing one of the tones from the first set as a starting point. Thus, e.g., all tones matching the 200 cps tone that matched the original 320 cps tone are read off the smooth curves from the appropriate graphs. This step produces a second set of matching intensities of the given frequencies, i.e., another approximation to an equal-loudness contour through the originating point.

This process is then repeated for a third frequency, say, 500 cps. (Checks indicated that the final outcome of the averaging process was almost completely independent of the employed sequence of frequencies.) This time, however, two matching intensities are available for this frequency, one for each of the preceding steps (assuming that smooth curves in both cases were drawn in the regions under consideration). Since none of these intensities has any priority with respect to the other, the average intensity is employed.

When this process has been repeated for all 13 frequencies employed, the average intensity for each frequency is calculated. These values, then, represent a composite of all matches performed at the given loudness level.

This averaging process is repeated for as many loudness levels as desired (in this case for a 10 db spacing of the originating 320 cps tone), and the resulting average matching intensities are plotted on the corresponding graphs and compared to the location of the smooth curves. If no systematic and significant departures are observed, the subset of matches represented by the average matching points can be regarded as transitive as well as symmetrical.

Results and Preliminary Discussion

The kinds of data obtained from the three subjects are illustrated by the fairly representative samples given in Figs 2 - 4. These data were all obtained from matches of 80 cps re 1250 cps, and vice versa. (In some cases

Insert Figures 2 - 4 about here

significantly higher degrees of dispersion were observed; these cases will be discussed later.) These examples illustrate the criteria for repeating runs, and for fitting smooth curves through the data points.

Average matches were computed for all three subjects according to the procedure described in the preceding section. For subjects JR and FP, no average matching point deviated more than 5 db from the smooth curves (measured as the smallest horizontal or vertical distance from point to curve), and most points deviated much less. The average matches therefore were accepted for these two subjects as representing a symmetric and transitive subset of loudness matches. For subject LL, a number of deviations larger than 5 db between average matches and smooth curves were observed. Consequently, the smooth curves were revised by omitting those sections for which deviations exceeding 5 db were observed. On the basis of these revised smooth curves the whole process of averaging was repeated, and the resulting set of average matches was accepted as representing a symmetric and transitive subset of loudness matches. In Figs. 2 - 4 the average matches are indicated by crosses. Their close proximity to the smooth curves is fairly representative of the data as a whole. Tables 2-4 present the average matches together with the average thresholds for each subject.

Insert Tables 2-4 about here

These average matches are plotted re frequency for each subject in Figs. 5 - 7, and points corresponding to a given loudness level have been connected

---

Insert Figures 5 - 7 about here

---

by straight lines. The resulting graphs thus represent an approximation to the equal-loudness contours for the given subject, (an approximation because data are not available for intermediate frequency values). A discussion of these equal-loudness contours will be presented in Part II of this report, which will be concerned with the dependency of loudness on frequency and middle-ear mechanics.

For each frequency combination, a plot of the corresponding average matches indicates the form of the matching function in question. The following is concerned with an analysis of the mathematical form of these matching functions, and with the associated theoretical implications.

Zwislocki (1965) has recently developed a theory for the perception of loudness, partly based on the work by Zwicker (e.g., 1958, 1963). In the following we shall give a brief description of this theory and show how it can be applied to the present results.

The theory describes the formation of the loudness of a signal imbedded in noise. Limiting ourselves at the outset to one CB (critical band), the following events are assumed to happen. 1) The total stimulating acoustic energy, that is, the sum of the energy (P_S^2) of the signal with sound pressure P_S and the energy (P_N^2) of the noise with sound pressure P_N, is converted through a power transformation to a corresponding excitation or loudness L_{S+N}:

(1) $$L_{S+N} = K(P_S^2 + P_N^2)^{\theta}.$$

2) When the loudness of the <u>signal</u> only is to be judged, the noise is regarded as causing an irrelevant loudness component which must be <u>subtracted</u> from the

total loudness L_{S+N}. The loudness L_N of the noise above follows from the same power transformation as above:

(2) $$L_N = K(P_N^2)^\theta .$$

3) The loudness of the signal L_S then follows as the difference between L_{S+N} and L_N:

(3) $$\begin{aligned} L_S &= L_{S+N} - L_N \\ &= K(P_S^2 + P_N^2)^\theta - K(P_N^2)^\theta \\ &= K[(P_S^2 + P_N^2)^\theta - P_N^{2\theta}] . \end{aligned}$$

(It must be accepted, but at the same time noted as a curious fact, that in the mechanism described the loudness of the irrelevant signal is available as an operative entity, while the loudness of the relevant signal only appears indirectly as the difference between two other loudnesses. If the loudness of the noise was to be judged, rather than the loudness of the signal, the loudness of the noise--according to the theory--suddenly would no longer be directly available, but now would appear as the difference between the total loudness (available, presumably, at all times) and the now directly available loudness of the signal.)

When the irrelevant signal extends over more than one CB, the argument runs as follows: The noise components outside that CB which contains the signal give rise to a certain loudness L_N'. Since the total loudness of acoustic energies located in different CB's is assumed to equal the sum of the loudnesses from each CB, the total loudness of signal and noise L_{S+N}' equals the sum of the previously considered total loudness L_{S+N} (for noise and signal within one CB) and L_N'. However, the loudness of the irrelevant signal (i.e., the noise) likewise is increased by L_N', which means that the difference between total and irrelevant loudness remains the same as above, that is,

(4) $$L_S = K[(P_S^2 + P_N^2)^\theta - P_N^{2\theta}],$$

where P_N indicates the sound pressure of the noise within that CB in which the signal is centrally located.

In case no external noise is applied, it is assumed that a certain residual, internal noise with an equivalent sound pressure P_{NI} is present. This noise is assumed to be responsible, in part, for the shape of the absolute threshold curve. For a given frequency, we may express the equivalent acoustic energy of this internal noise in terms of the acoustic energy at the absolute threshold for the same frequency:

(5) $$P_{NI}^2 = kP_T^2,$$

which leads to the following expression for the loudness of a signal of limited frequency extension with no external noise present:

(6) $$L = K[(P^2 + kP_T^2)^\theta - kP_T^{2\theta}],$$

or, expressed in terms of sensation level (SL) P/P_T:

(7) $$L = KP_T^{2\theta}\ [((P/P_T)^2 + k)^\theta - k^\theta].$$

For large values of P, this expression converges toward:

(8) $$L = KP^{2\theta}.$$

From a preliminary analysis of the matching data (with special consideration of sound pressures below the level at which a change in acoustic impedance at the eardrum was observed), two characteristics of the data as a whole were apparent: 1) for higher intensity levels, the matching functions for all frequency combinations approached a slope of 1.0; 2) the "matching point" obtained by plotting the threshold intensities against each other seemed to belong to the matching function for any combination of frequencies. These two observations form the basis for two assumptions which will be utilized in deriving a theoretical expression for the matching function for a given pair of frequencies. The first observation leads to the assumption that the exponent θ in (7) is independent of frequency (because only under this assumption will the slope of the matching function equal 1.0 for higher intensities). The second observation leads to assuming that the loudness at threshold is independent of frequency.

The assumption of a frequency-independent value of θ entails that the loudness L_x of a tone of frequency x can be expressed as:

$$L_x = K_x P_{Tx}^{2\theta} [((P_x/P_{Tx})^2 + k_x)^\theta - k_x^\theta]. \tag{9}$$

Loudness at threshold, L_{Tx}, is then given by

$$\begin{aligned} L_{Tx} &= K_x P_{Tx}^{2\theta} [((P_{Tx}/P_{Tx})^2 + k_x)^\theta - k_x^\theta] \\ &= K_x P_{Tx}^{2\theta} [(1 + k_x)^\theta - k_x^\theta]. \end{aligned} \tag{10}$$

The assumption that loudness at threshold L_o is independent of frequency means that

$$L_o = L_{Tx} = K_x P_{Tx}^{2\theta} [(1 + k_x)^\theta - k_x^\theta], \tag{11}$$

that is,

(12) $$K_x = \frac{L_o}{P_{Tx}^{2\theta}\,[(1 + k_x)^\theta - k_x^\theta]}.$$

Substituting (12) in (9) gives:

(13) $$L_x = \frac{L_o}{(1 + k_x)^\theta - k_x^\theta}\,[((P_x/P_{Tx})^2 + k_x)^\theta - k_x^\theta].$$

The matching function between two frequencies, x and y, is described by $L_x = L_y$, that is,

(14) $$\frac{L_o}{(1 + k_y)^\theta - k_y^\theta}\,[((P_y/P_{Ty})^2 + k_y)^\theta - k_y^\theta]$$

$$= \frac{L_o}{(1 + k_x)^\theta - k_x^\theta}\,[((P_x/P_{Tx})^2 + k_x)^\theta - k_x^\theta];$$

cancelling L_o and rearranging terms yield:

(15) $$[(P_y/P_{Ty})^2 + k_y]^\theta$$

$$= \frac{(1 + k_y)^\theta - k_y^\theta}{(1 + k_x)^\theta - k_x^\theta}\,[(P_x/P_{Tx})^2 + k_x]^\theta - \frac{(1 + k_y)^\theta - k_y^\theta}{(1 + k_x)^\theta - k_x^\theta}\,k_x^\theta - k_y^\theta.$$

For $(P_y/P_{Ty})^2 >> k_y$ and $(P_x/P_{Tx})^2 >> k_x$, (15) approximates the expression

(16) $$(P_y/P_{Ty})^{2\theta} = \frac{(1 + k_y)^\theta - k_y^\theta}{(1 + k_x)^\theta - k_x^\theta}\,(P_x/P_{Tx})^{2\theta},$$

that is, for larger sensation levels P_y and P_x are linearly related (yielding a straight line with slope 1.0 in a db-db plot).

In order to test whether theoretical matching functions as given by (15) would fit the obtained data, a digital computer (PDP-4) was programmed to compute values of P_y/P_{Ty} for 1 db increments of P_x/P_{Tx} for given values of the parameters θ, k_x, and k_y.[1] The resulting values were plotted by means of a graphic recorder (Honeywell Visicorder).

Zwislocki assumes θ = .27, and k_x = 2.5 for 1000 cps. Families of theoretical matching functions were obtained for bracketed values of these parameters, for varying k_y. The following values of θ were selected: .15, .20, .27, .35, and .50. For each of these θ-values, a range of k_x-values was explored. For each k_x-value selected, k_y was systematically varied from very small to very large values. Due to the relative difficulty of obtaining such families of curves, a greater number of these are reproduced in Figs. 8a - 8r.

Insert Figures 8a - 8r about here

In order to render the matching data in a form permitting a comparison with the theoretical curves, average matches were plotted relative to the median frequency of 320 cps for each subject. (Since the average matches constitute a transitive subset, the choice of reference frequency is of no consequence.) These plotted distributions of average matches were compared to the families of theoretical matching functions, and the best-fitting family was selected for each subject. Figures 9 - 11 show for each subject the average matches relative to 320 cps, superimposed on the best-fitting family of curves.

Insert Figures 9 - 11 about here

In selecting the best-fitting family of curves, only matches involving intensities below those causing changes of the acoustic impedance at the eardrum were considered. It should be pointed out, however, that this selection was not highly critical. The neighboring values of k_x provided almost as good a fit for the same exponent, and the neighboring exponents usually produced sets of curves for some other value of k_x which were usable. In the final selection of the best-fitting family of curves, however, no preference was given <u>a priori</u> to $\theta = .27$; the choice of this exponent for all three subjects was made entirely on the basis of best over-all fit between average matches and curves.

The rather close similarity, within the intensity range considered, between empirical values and theoretical curves in Figs. 9 - 11 lends at least preliminary support to the theory of loudness put forth by Zwislocki and Zwicker. Part II of this report will present an attempt to test whether the matches for the higher intensities also conform to this theory when possible changes in the conductive properties of the middle ear as a function of sound intensity are taken into account.

From each of Figs. 9 - 11 the value of k_x for each frequency was estimated (see Table 5). According to equation (5) we find:

(17) $$k = P_{NI}^2/P_T^2,$$

that is,

(18) $$k(db) = 10 \log k = P_{NI}(db) - P_T(db).$$

Insert Table 5 about here

In other words, 10 times the logarithm of the estimated k-values gives the number of decibels by which the equivalent sound pressure of the assumed internal noise exceeds the threshold sound pressure. These k(db)-values are plotted in Fig. 12 as a function of frequency for each subject.

Insert Figure 12 about here

Two points should be noted concerning these estimated k-values. 1) The estimate of k, of course, is a function of the location of the distribution of average matches for a given frequency combination relative to the theoretical curves. This relative location is a direct function of the values of the thresholds for the two frequencies involved. Since the procedure for collecting thresholds in the present research was not very sophisticated, some uncertainty is inherent in this relative location, and hence in the estimate of k. 2) Since the selected value of k_y was not highly critical, the same applies to the estimated k_x-values. Thus, since k_y may vary within a factor of from 1/2 to 2, the <u>level</u> of the k(db)-values in Fig. 12 may vary within ±3 db. This latter source of uncertainty, however, does not affect the <u>shape</u> of the k(db) vs. frequency functions of Fig. 12. Possible interpretations of these functions will be discussed in Part II of this report.

As mentioned in the beginning of this section, not all sets of matches collected exhibited the same rather small dispersion as illustrated in Figs. 2 - 4. In many cases the matches appeared to fall in two or three quite distinct and separate patterns; one such case is illustrated in Fig. 13. For subject JR, practically all instances of this

---

Insert Figure 13 about here

---

kind were characterized by the variable tone being adjusted to higher intensities than were to be expected from the matches with reversed frequency combination. The two other subjects, however, did not show this regularity, the variable being adjusted to higher and to lower values than expected in about an equal number of cases. These split cases did not seem to be very clearly correlated with the frequencies of the tones involved, nor with the frequency separation between these tones.

It was observed, however, that in practically all cases of "split matches " the matching functions (up to the level of impedance change) could be described quite satisfactorily by the same set of theoretical curves used in describing the average matching functions. This is illustrated in Fig. 13, where a set of "split matches" for subject JR has been superimposed on the set of theoretical curves which was applied to the same subject's average matches in Fig. 9. It will be seen that the data are distributed in two patterns, each of which bears a reasonable resemblance to one theoretical curve up to a certain intensity level. (Since no acoustic impedance measurements are available for 2000 cps, it is not possible to state exactly at which intensity level impedance changes occur for one or both of the tones to be matched.)

A preliminary analysis of the cases of "split matches" has indicated that theoretical matching functions with an exponent of .27 afford the best fit to these data. However, no clear relationship between these data and the parameters k_x and k_y has yet been found, that is, it has not been possible to determine which particular family of theoretical matching functions (with $\theta = .27$) affords the best fit to a given set of "split matches." This lack of specificity is due to two factors: 1) the scatter of the data points does not permit a unique fitting of any theoretical curve to a given set of data points; 2) as yet no sufficiently clear rationale concerning the functional role of the parameters k_x and k_y is available to permit the direct testing of well-founded hypotheses. In spite of this theoretically unsatisfactory state of affairs, it may still be justified to claim that the similarity between "split matches" and theoretical curves indicates that the same basic mechanism is at work for "split matches" and for "regular matches."

References

Churcher, B. G., & King, A. J. The performance of noise meters in terms of the primary standard. J. Inst. Electrical Engrs.,1937, 81, 57-90.

Dallos, P. J. Dynamics of the acoustic reflex: phenomenological aspects. J. acoust. Soc. Amer., 1964, 36, 2175-2183.

Feldtkeller, R., & Zwicker, E. Das Ohr Als Nachrichtenempfänger. Stuttgart: Hirzel, 1956.

Fletcher, H., & Munson, W. A. Loudness, its definition, measurement, and calculation. J. acoust. Soc. Amer., 1933, 5, 82-108.

Garner, W. R. A technique and scale for loudness measurement. J. acoust. Soc. Amer., 1954, 26, 73-88.

Kingsbury, B. A. A direct comparison of the loudness of pure tones. Phys. Rev., 1927, 29, 588-600.

Robinson, D. W., & Dadson, R. S. A redetermination of the equal-loudness relations for pure tones. Brit. J. appl. Physiol., 1956, 7, 166-181.

Ross, S. A technique for producing known sound pressures at the eardrum. Progress Report No. IV, November 1963, Ann Arbor: Behavior Analysis Lab., University of Michigan.

Stevens, S. S. The measurement of loudness. J. acoust. Soc. Amer., 1955, 27, 815-829.

Zwicker, E. Ueber psychologische und metodische Grundlagen der Lautheit. Acustica, 1958, 8, 237 -

Zwicker, E. Ueber die Lautheit von gedrosselten und ungedrosselten Schallen. Acustica, 1963, 13, 194 -

Zwislocki, J. Analysis of the middle-ear function. Part I: Input impedance. J. acoust. Soc. Amer., 1962, 34, 1514-1523.

Zwislocki, J. Analysis of the middle-ear function. Part II: Guinea-pig ear. J. acoust. Soc. Amer., 1963, 35, 1034-1040.

Zwislocki, J. Analysis of some auditory characteristics. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), Handbook of mathematical psychology. Vol. 3. New York: Wiley, 1965.

Footnotes

[1]The program for the PDP-4 computer (Digital Equipment Corporation) was written by Sally Ginet and David Vander Yacht. A large portion of the data was collected by Jette Ross. My sincere appreciation of these services is hereby expressed.

Table 1

20	32	50	80	125	200	320	500	800	1250	2000	3200	5000 cps
92.5	73	62	48	43	34	*30*	26	26	15	21	22.5	10
94	-	-	49.5	43	*34*	30	27.5	24.5	16	22.5	27.5	/
/	74.5	62	47.5	-	33.5	31	*27*	23.5	14	21.5	25	15
95.5	75	65.5	52.5	*43*	34	30	-	29.5	17.5	22.5	/	/
/	/	71	51.5	40	36.5	30	29	*26*	16	23.5	25	13.5
92	70.5	63	*50*	40.5	34.5	32	29.5	24.5	15	/	/	/
/	/	/	51	41	33.5	30.5	28.5	25.5	*15.5*	24.5	31	15
89	72.5	*64.5*	51	42	34	32.5	29	22.5	/	/	/	/
/	/	/	/	43	34	31	28	25.5	14	*22.5*	28	12.5
91.5	*73*	65	52	42	-	30	26	/	/	/	/	/
/	/	/	/	/	33	34	28.5	27.5	12.5	21.5	*26.5*	-
92.5	74.5	68	50.5	41	32.5	30	/	/	/	/	/	/
/	/	/	/	/	/	32.5	25.5	25.5	14.5	23	-	*13*
92.5	73.5	65	50.5	41.5	34	31	28	25.5	15	22.5	26.5	13

Sample of tabulation for computing average matching values for checking transitivity, and for the construction of equal-loudness contours.

(Entries in each row represent the intensity in db SPL of a tone with frequency as given by the column heading that matches the tone indicated by an italicized entry. An italicized entry equals the average of the entries appearing above it in the given column. Matches were determined from smooth curves drawn through those data regions that contain symmetrical matches. The bottom row contains the average values of the non-italicized entries for each column. Data from subject JR.)

Table 2

Freq.	20	32	50	80	125	200	320	500	800	1250	2000	3200	5000 cps
Level													
thr.	82	61.5	54	42.5	35	24.5	23	18.5	17.5	5.5	13	16	1.5
1	92.5	73.5	65	50.5	41.5	34	31	28	25.5	15	22.5	26.5	13
2	102.5	86	78	63.5	53.5	46	41.5	39.5	37.5	27	35.5	39	28
3	109	(92.5)	86	73	64	56	51	49.5	48	38	47	52	39
4	114	98.5	92.5	81	71	65.5	59	57	56	47.5	57.5	62	49.5
5	121	107.5	101.5	93.5	84	78.5	69.5	68	67	60.5	71	76	63
6	127	115.5	108	102	92.5	87	79	77.5	76	70	78.5	85.5	72.5
7	(129.5)	125	115.5	109.5	101.5	94	88	84.5	87	79.5	88	92	80
8	-	(133.5)	128	121	112	104.5	99	95	96.5	88	94.5	(95.5)	(91)
9	-	-	134	(131.5)	119	113	107	101	101.5	96.5	(102.5)	101	(93)
10	-	-	(136.5)	(136)	124	117	(112)	102.5	105	100	(110)	(104.5)	(90)
11	-	-	-	-	(131)	(126.5)	-	(112)	(114)	(107.5)	-	(108.5)	-

Tables 2 - 4: Average matches computed from smooth curves drawn through symmetrical data points. (Subjects JR, LL, and FP, respectively. The upper row contains average thresholds. All entries are in db SPL. The entries in parentheses are based on less than three values.)

Table 3

Freq. Level	20	32	50	80	125	200	320	500	800	1250	2000	3200	5000 cps
thr.	98	84.5	65.5	45.5	33	23.5	17.5	12	7.5	16	18	29	30.5
1	(107)	88.5	73	54	42.5	33.5	27.5	25.5	19.5	29	30	39.5	44
2	112	92	77.5	61	50.5	41	33.5	33.5	28.5	37.5	36.5	49	53
3	117.5	99	83	69.5	(62.5)	50	44	45.5	36	47	46	(56)	(63)
4	120	106.5	88	75	68.5	60	53	54.5	45.5	53.5	53.5	64	68.5
5	124	114.5	96	82	75.5	70	64.5	69	60	65	65	77.5	82.5
6	133.5	122.5	104.5	94.5	87.5	82	77	78	73	79	79	93.5	95.5
7	-	131.5	117	108	100	93	86.5	89.5	84.5	92	97	105	107.5
8	-	139.5	127.5	119	109.5	102	99	100	95.5	101.5	105.5	112	115.5
9	-	-	138	130	120	111.5	110.5	110	106	108.5	112	(117.5)	121
10	-	-	(141)	137	127	120	118	115.5	113.5	115	117.5	125	126.5
11	-	-	-	(141.5)	136	129	126.5	123	123.5	121.5	124.5	(131.5)	133
12	-	-	-	-	(139)	(134.5)	131	(125)	(127)	123.5	126.5	(133)	(132.5)

Table 4

Freq.	20	32	50	80	125	200	320	500	800	1250	2000	3200	5000 cps
Level													
thr.	88.5	61	48	35.5	27	17.5	11	9.5	7.5	6	12.5	24.5	5.5
1	96.5	70	56.5	45	34.5	27.5	21	21	19	17.5	23	36	17
2	103.5	77	67.5	55	43.5	37.5	29.5	31	31	27.5	33	46	27.5
3	109	87.5	74.5	65.5	56.5	46.5	40.5	44	45.5	40.5	45	(62.5)	40
4	113	94	82.5	73.5	63.5	56.5	49.5	52	56	49.5	55.5	(68)	50
5	(116.5)	101	88.5	81.5	71	63	60	61	64	61.5	68.5	77	59
6	(120)	110.5	96	90	80.5	74	70.5	71.5	72	72	80	90	67.5
7	(126)	117	103	98	91.5	84	79.5	81	80.5	82.5	87	98	74.5
8	-	123.5	(110)	108	102	94.5	89	89.5	91.5	96	(97)	105	(83.5)
9	-	(130.5)	123.5	115	108	101	100	98	102.5	102	(102.5)	110.5	(91)
10	-	-	-	(134.5)	(126)	(120.5)	(110)	(113.5)	116.5	(114.5)	-	(120)	(101.5)

Table 5

Subject	JR		LL		FP	
f	k	k(db)	k	k(db)	k	k(db)
20	-	-	-	-	-	-
32	(10)	(10.0)	-	-	22	13.4
50	9	9.5	10	10.0	20	13.0
80	7	8.5	3.5	5.4	15	11.8
125	8.5	9.3	2	3.0	13	11.1
200	5	7.0	1.25	.95	11	10.4
320	10	10.0	1.25	.95	10	10.0
500	8	9.0	.45	-3.5	7	8.4
800	8	9.0	.6	-2.2	4	6.0
1250	5	7.0	.6	-2.2	5	7.0
2000	3.5	5.4	1.0	0	5	7.8
3200	3	4.8	1.0	0	5	7.0
5000	2.5	4.0	.6	-2.2	5	7.0

Values of k_x estimated from Figs. 9 - 11 for each frequency and each subject. (The value for 32 cps for subject JR is based on the medium intensities only. k(db) equals 10 log k.)

Figure Captions

Fig. 1. . Experimental set-up for obtaining loudness balances. See text for details.

Figs. 2 - 4. Representative samples of loudness balances for subjects JR, LL, and FP, respectively. Filled and unfilled symbols represent matches for 80 cps as standard and variable, respectively. Crosses represent average matches as computed for each subject from all symmetric matches at the given loudness level. (For subject LL, plus-signs represent average matches based on a selected set of symmetric matches.) Arrows indicate approximate sound pressure levels at which changes in acoustic impedance at the eardrum began to occur; (no such information is available at 1250 cps for subjects JR and LL). Smooth curves have been visually fitted to those data regions for which at least one frequency constellation and its opposite produced data with no systematic and significant mutual departures.

Figs. 5 - 7. . Average matches and absolute thresholds in db SPL from Tables 2-4 plotted re frequency for subjects JR, LL, and FP, respectively. Neighboring points, one or both of which are based on less than three values, are connected by a broken line. Lower set of points are thresholds. The curves represent an approximation to the equal-loudness contours.

Figs. 8a - 8r. . Families of curves generated by the general expression (15) given in the text.

Figs. 9 - 11. . Average matches in db SL relative to 320 cps for subjects JR, LL, and FP, respectively, superimposed on the best-fitting family of curves generated by the general expression . (15) given in the text.

Fig. 12. Estimate of equivalent sound pressure of assumed internal noise relative to sound pressure at threshold, expressed in db, as a function

of frequency for the three subjects. This db-difference equals 10 times the logarithm of k for the given frequency, estimated from Figs. 9 - 11.

Fig. 13. A case of "split matches" from the data of subject JR, superimposed on the family of theoretical matching functions affording the best fit to the average matches of the same subject, (illustrated in Fig. 9). This particular family of curves was not selected as a best fit to the "split matches," and is presented only to indicate the general resemblance between observed data and theoretical functions.

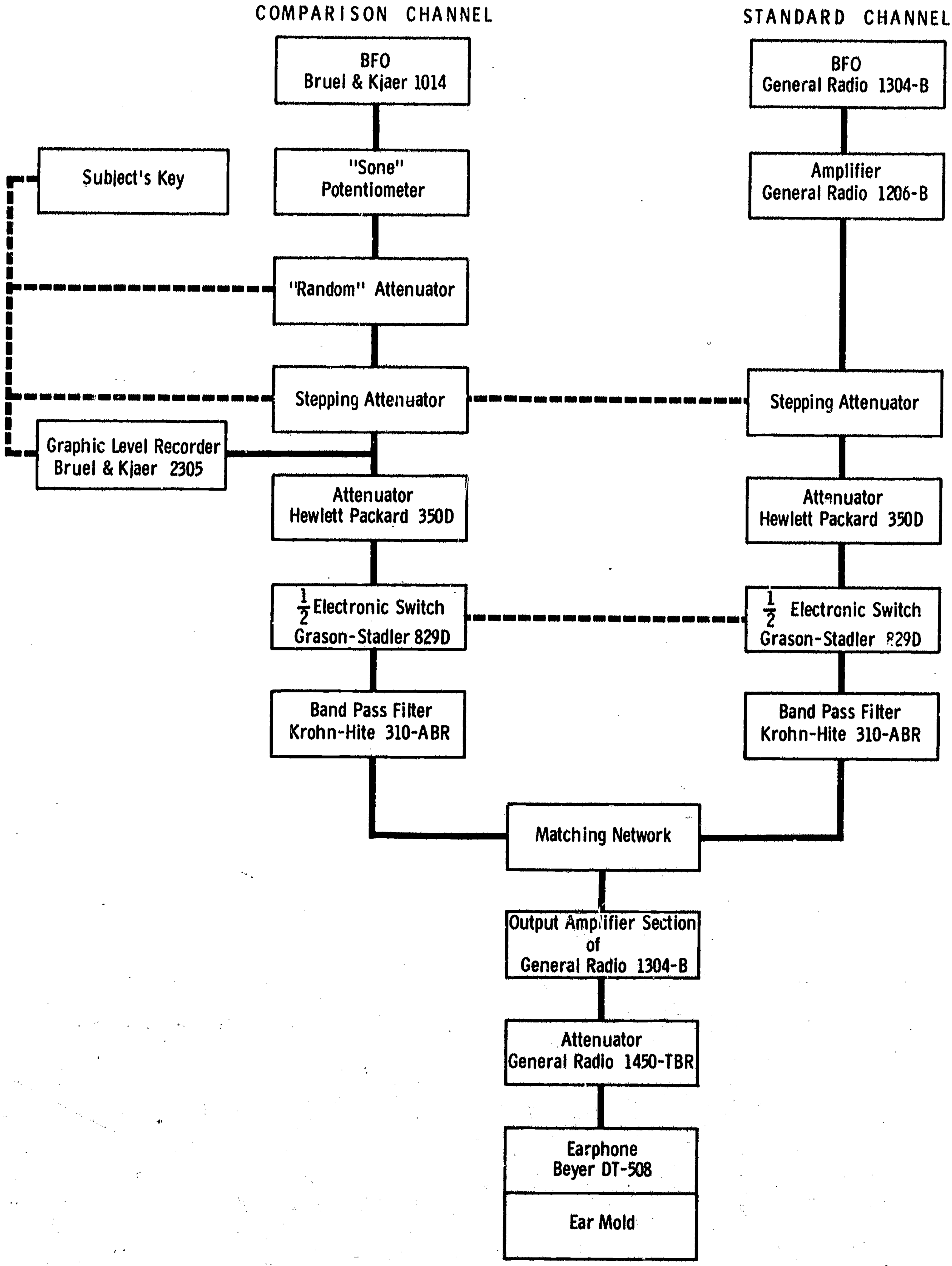

COMPARISON CHANNEL
STANDARD CHANNEL
BFO
Bruel & Kjaer 1014
BFO
General Radio 1304-B
Subject's Key
"Sone"
Potentiometer
Amplifier
General Radio 1206-B
"Random" Attenuator
Stepping Attenuator
Stepping Attenuator
Graphic Level Recorder
Bruel & Kjaer 2305
Attenuator
Hewlett Packard 350D
Attenuator
Hewlett Packard 350D
1/2 Electronic Switch
Grason-Stadler 829D
1/2 Electronic Switch
Grason-Stadler 829D
Band Pass Filter
Krohn-Hite 310-ABR
Band Pass Filter
Krohn-Hite 310-ABR
Matching Network
Output Amplifier Section
of
General Radio 1304-B
Attenuator
General Radio 1450-TBR
Earphone
Beyer DT-508
Ear Mold

Fig. 1

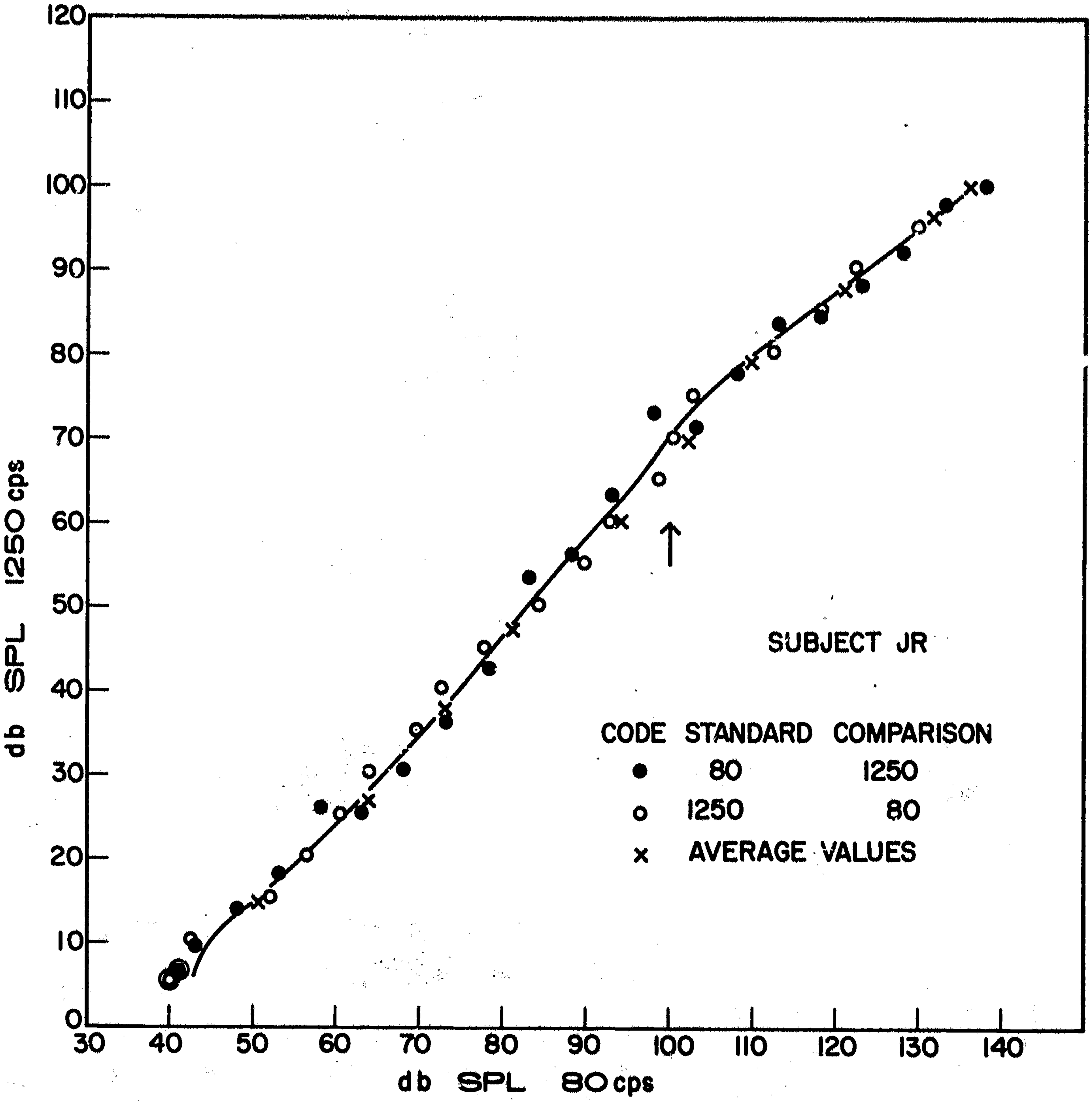
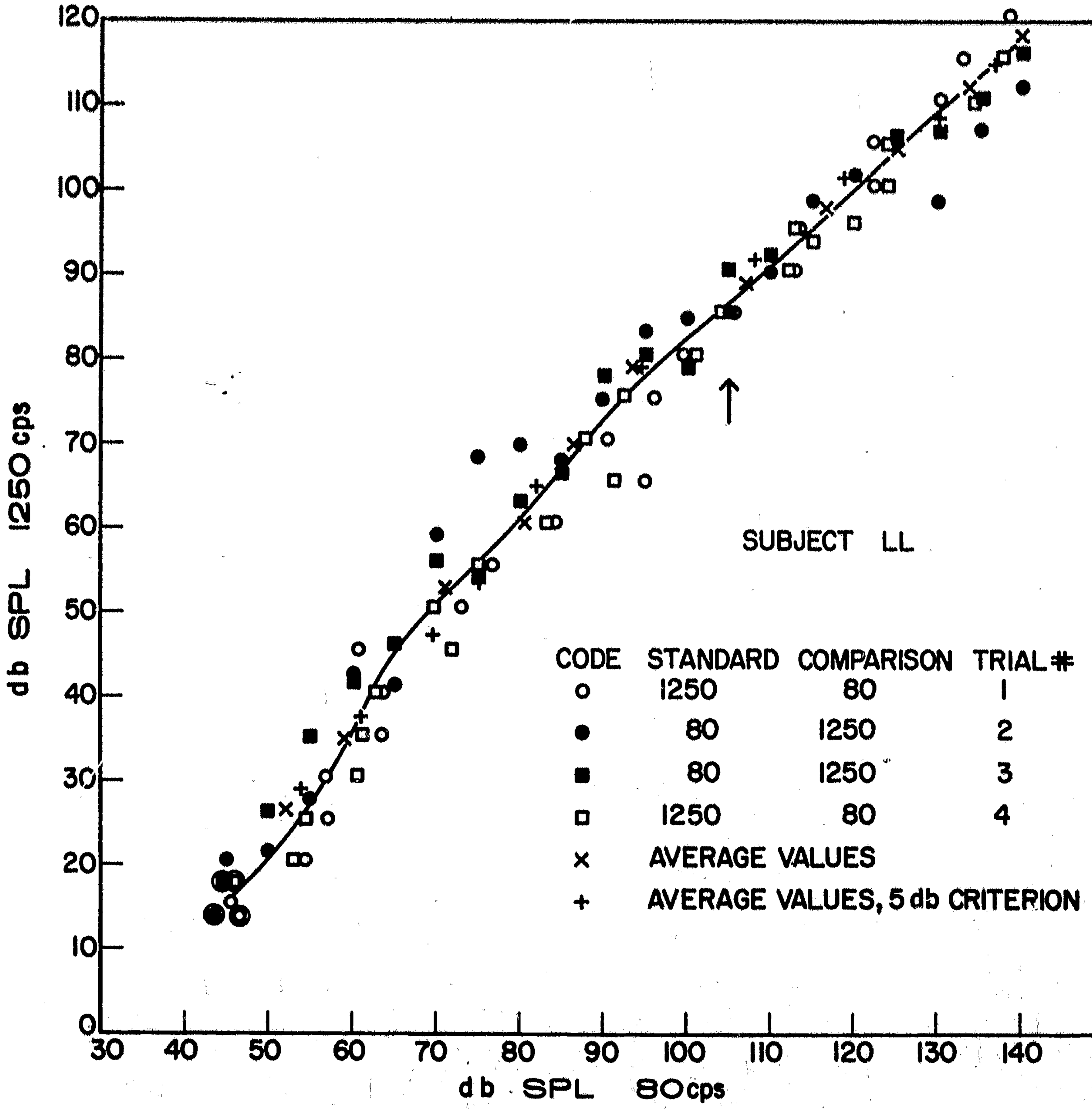
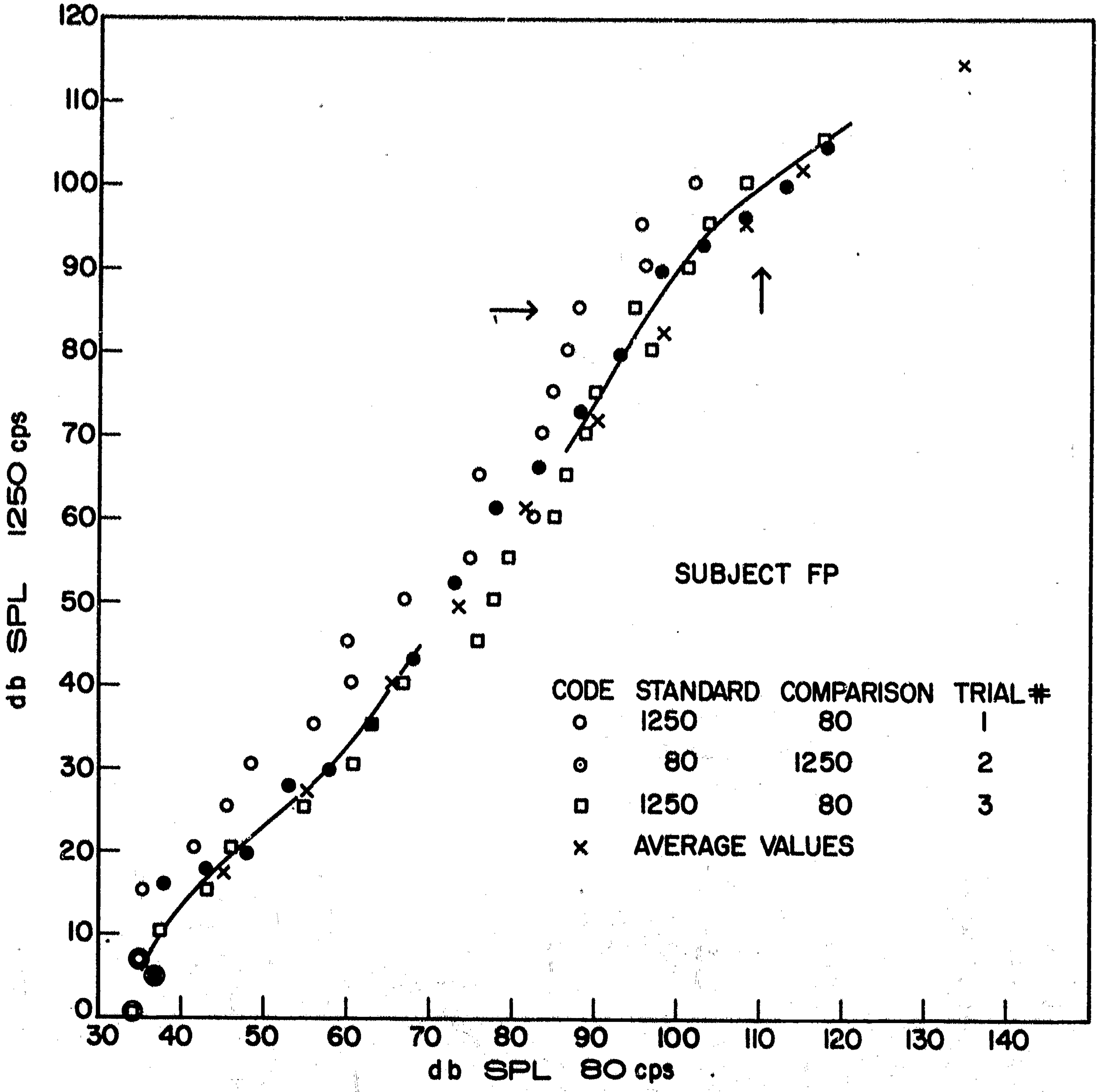

SUBJECT JR
CODE STANDARD COMPARISON
● 80 1250
o 1250 80
x AVERAGE VALUES
db SPL 1250 cps
db SPL 80 cps
0
10
20
30
40
50
60
70
80
90
100
110
120
30
40
50
60
70
80
90
100
110
120
130
140

Fig. 2



SUBJECT LL
CODE STANDARD COMPARISON TRIAL #
○ 1250 80 1
● 80 1250 2
■ 80 1250 3
□ 1250 80 4
× AVERAGE VALUES
+ AVERAGE VALUES, 5 db CRITERION
db SPL 1250 cps
db SPL 80cps
0
10
20
30
40
50
60
70
80
90
100
110
120
30
40
50
60
70
80
90
100
110
120
130
140

Fig. 3



SUBJECT FP
CODE STANDARD COMPARISON TRIAL #
○ 1250 80 1
⊙ 80 1250 2
□ 1250 80 3
× AVERAGE VALUES
db SPL 1250 cps
db SPL 80 cps
0
10
20
30
40
50
60
70
80
90
100
110
120
30
40
50
60
70
80
90
100
110
120
130
140

Fig. 4

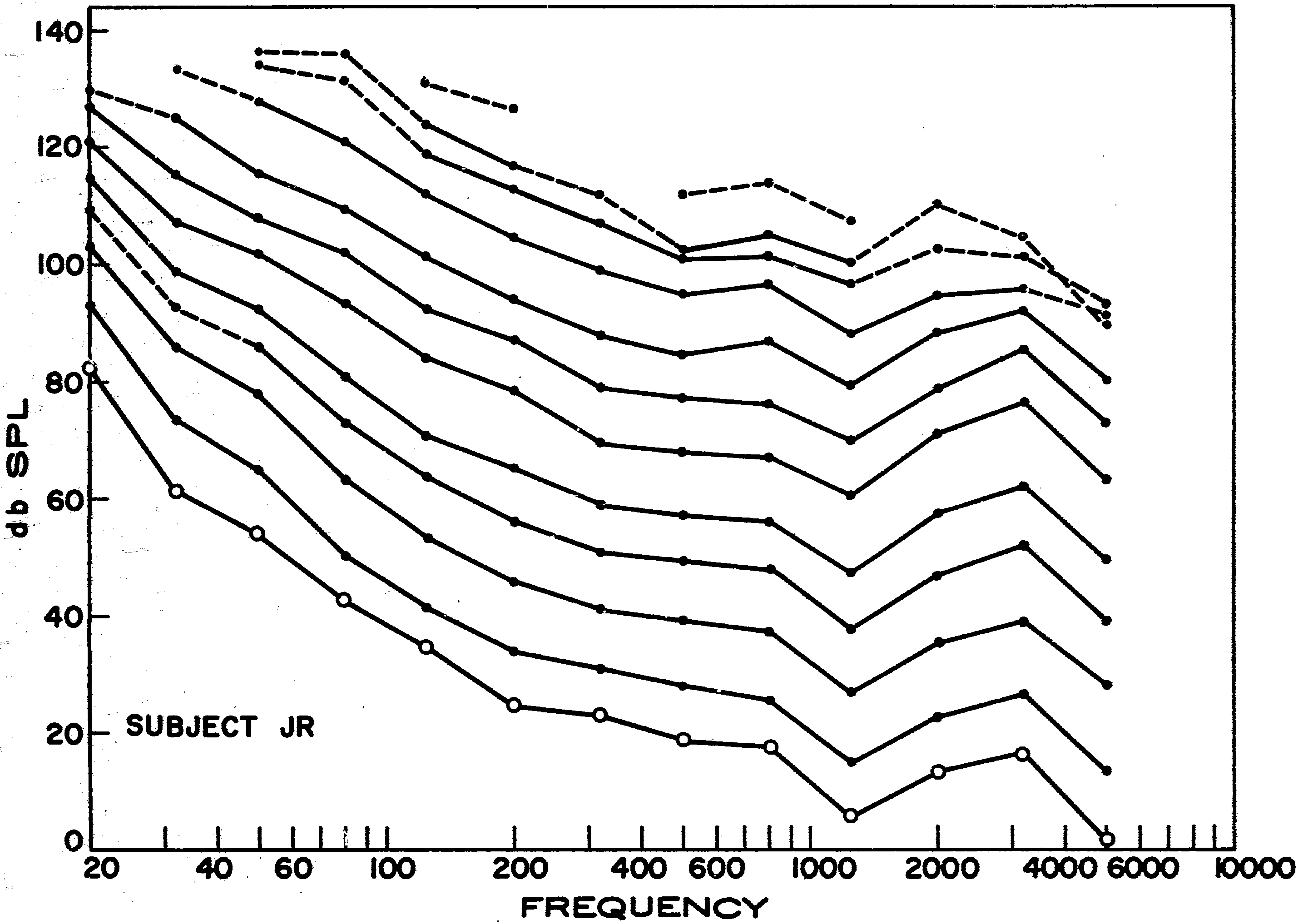
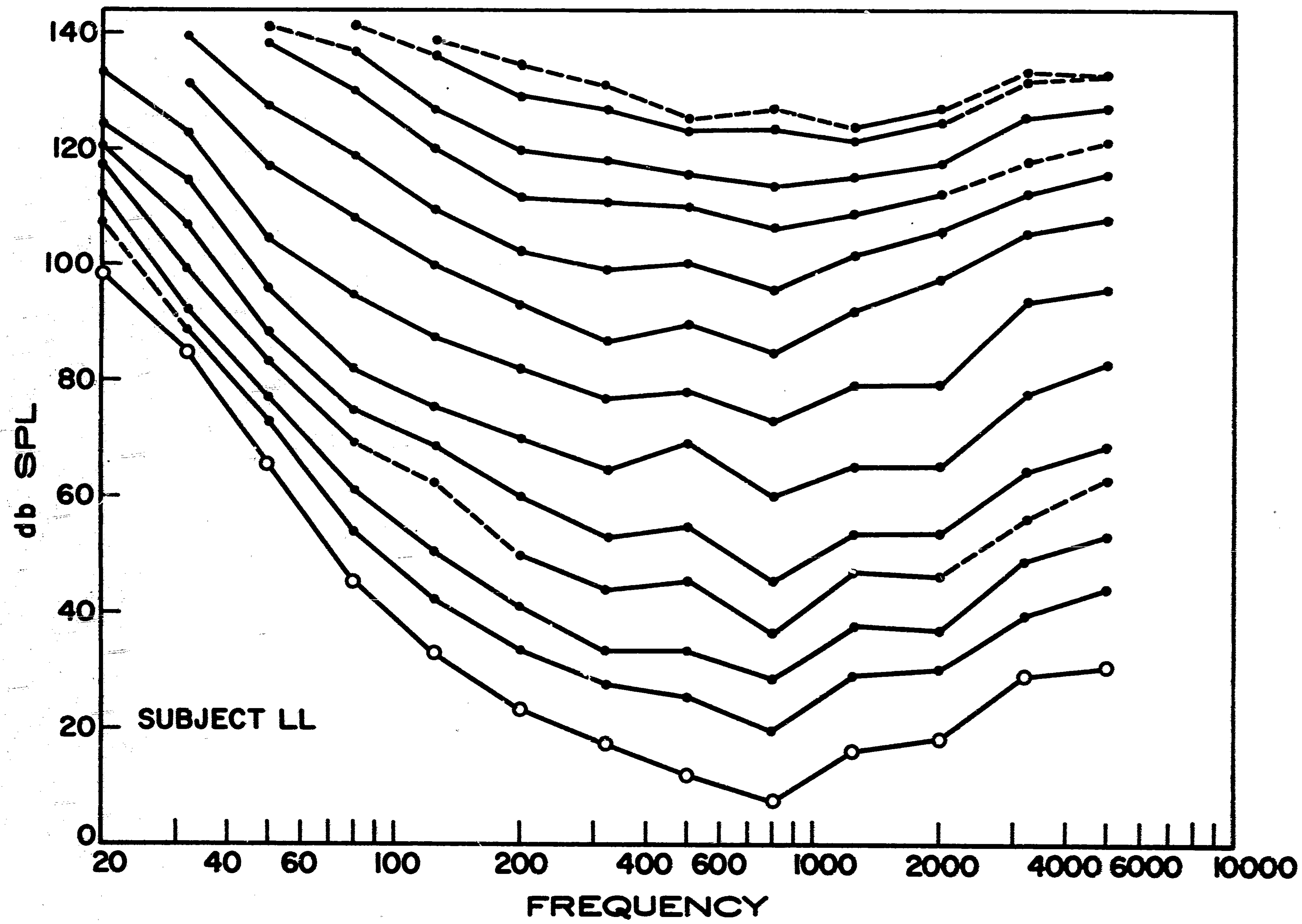

140
120
100
80
60
40
20
0
db SPL
SUBJECT JR
20
40
60
100
200
400
600
1000
2000
4000
6000
10000
FREQUENCY

Fig. 5



140
120
100
80
60
40
20
0
db SPL
SUBJECT LL
20
40
60
100
200
400
600
1000
2000
4000
6000
10000
FREQUENCY

Fig. 6

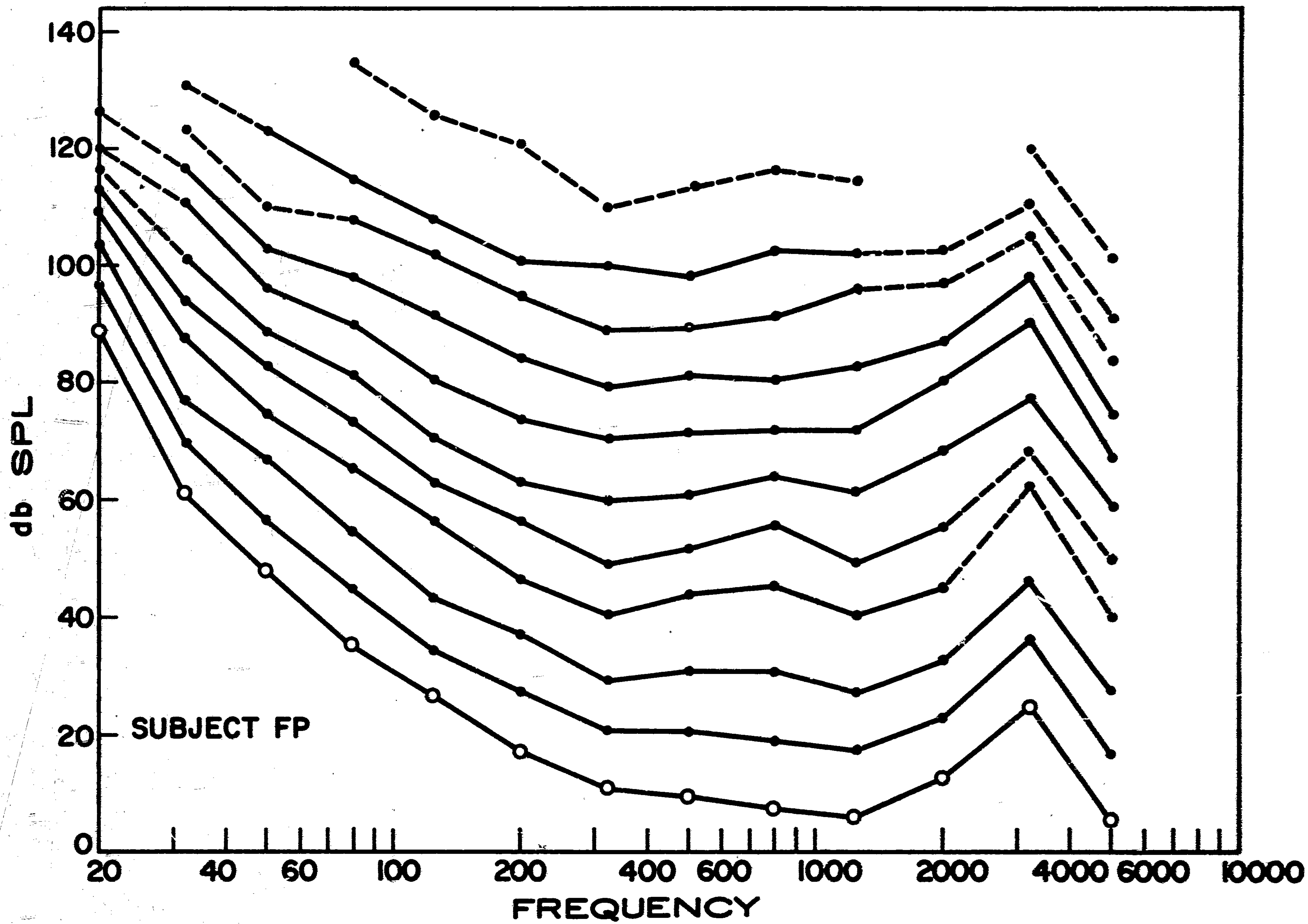
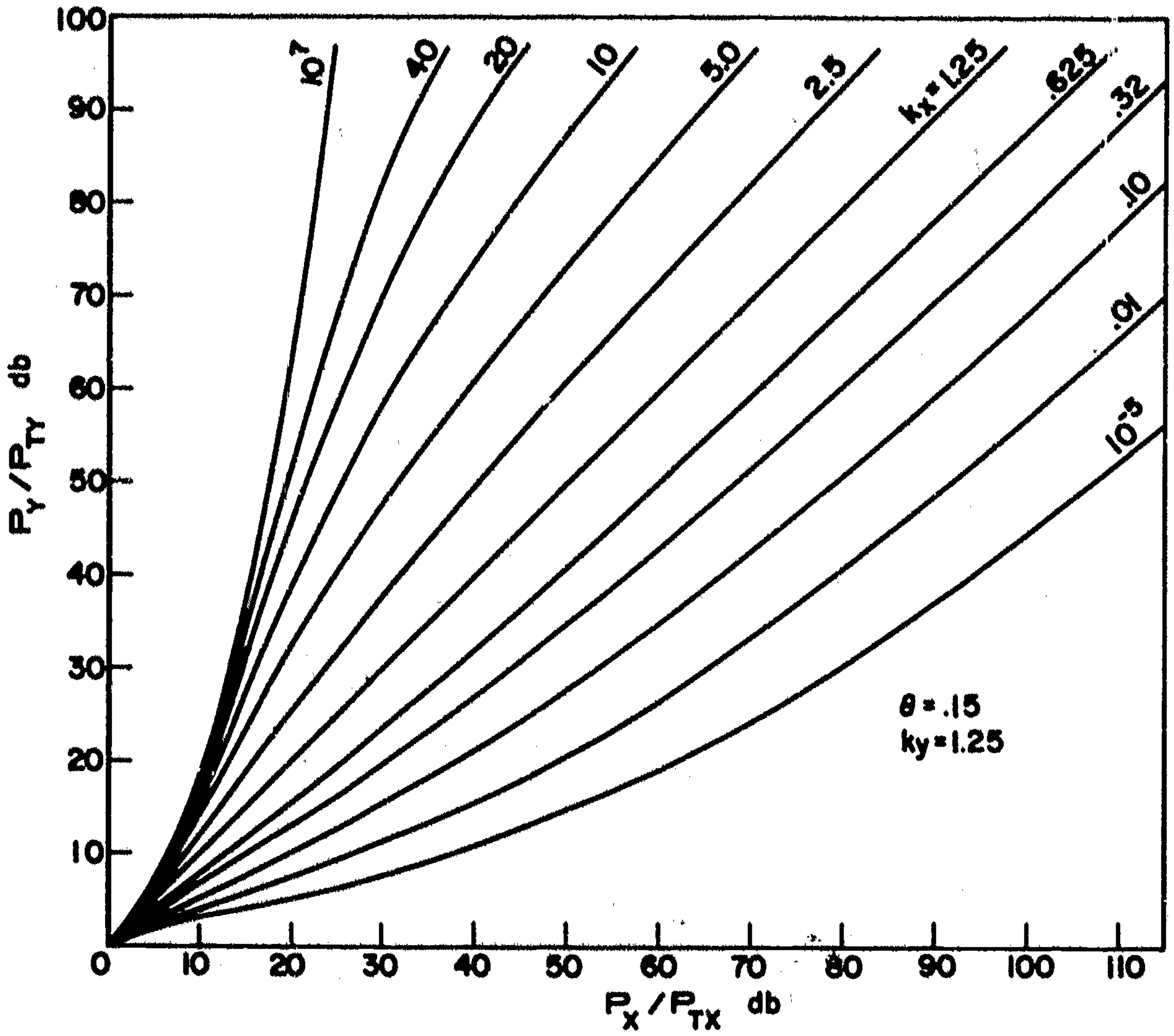
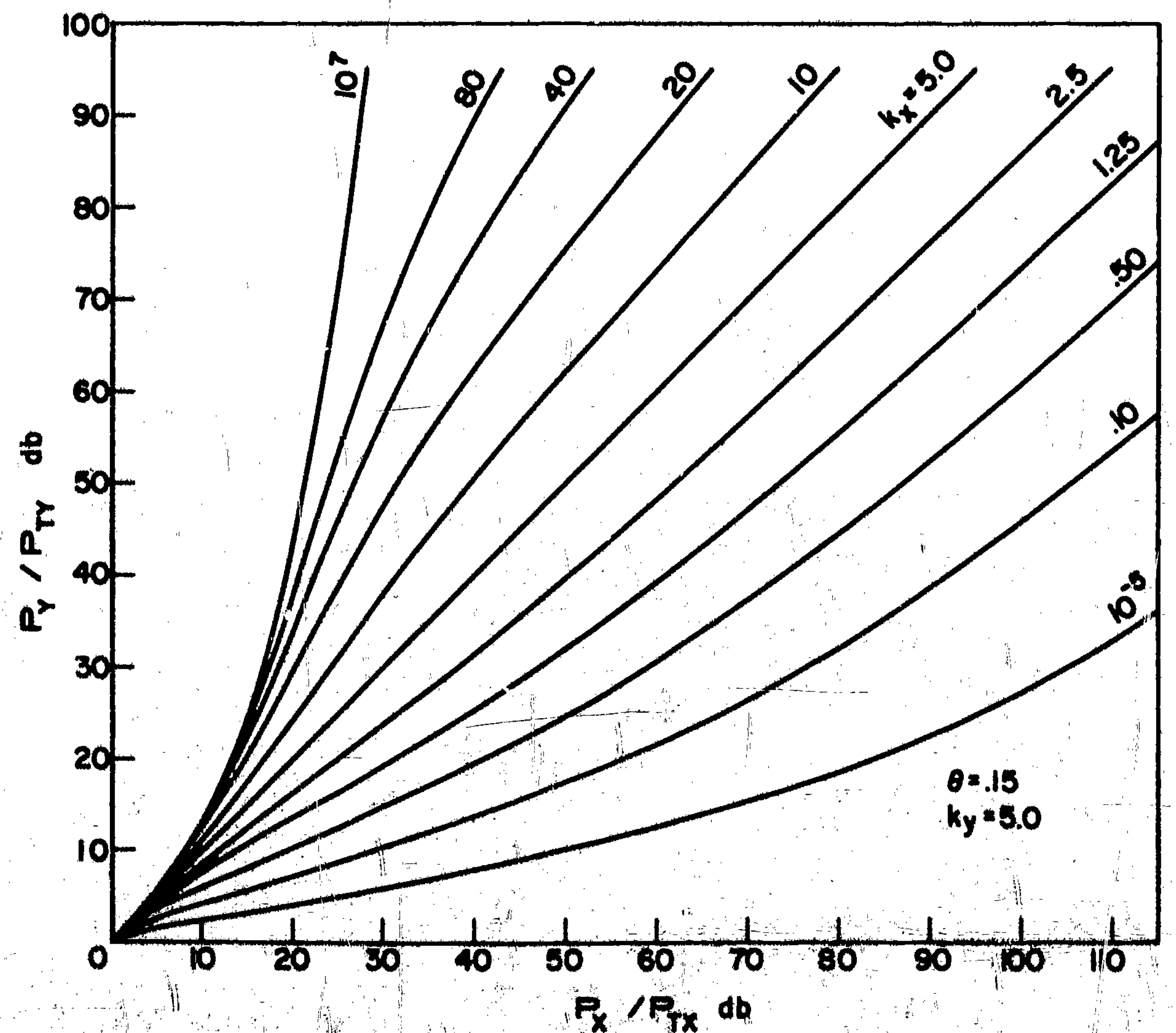

140
120
100
80
60
40
20
0
db SPL
SUBJECT FP
20
40
60
100
200
400
600
1000
2000
4000
6000
10000
FREQUENCY

Fig. 7



10^7
40
20
10
5.0
2.5
k_x=1.25
.625
.32
.10
.01
10^{-5}
θ=.15
ky=1.25
P_Y/P_{TY} db
P_X/P_{TX} db

Fig. 8A



10^7
80
40
20
10
k_x=5.0
2.5
1.25
.50
.10
10^{-5}
θ=.15
ky=5.0
P_Y/P_{TY} db
P_X/P_{TX} db

Fig. 8B

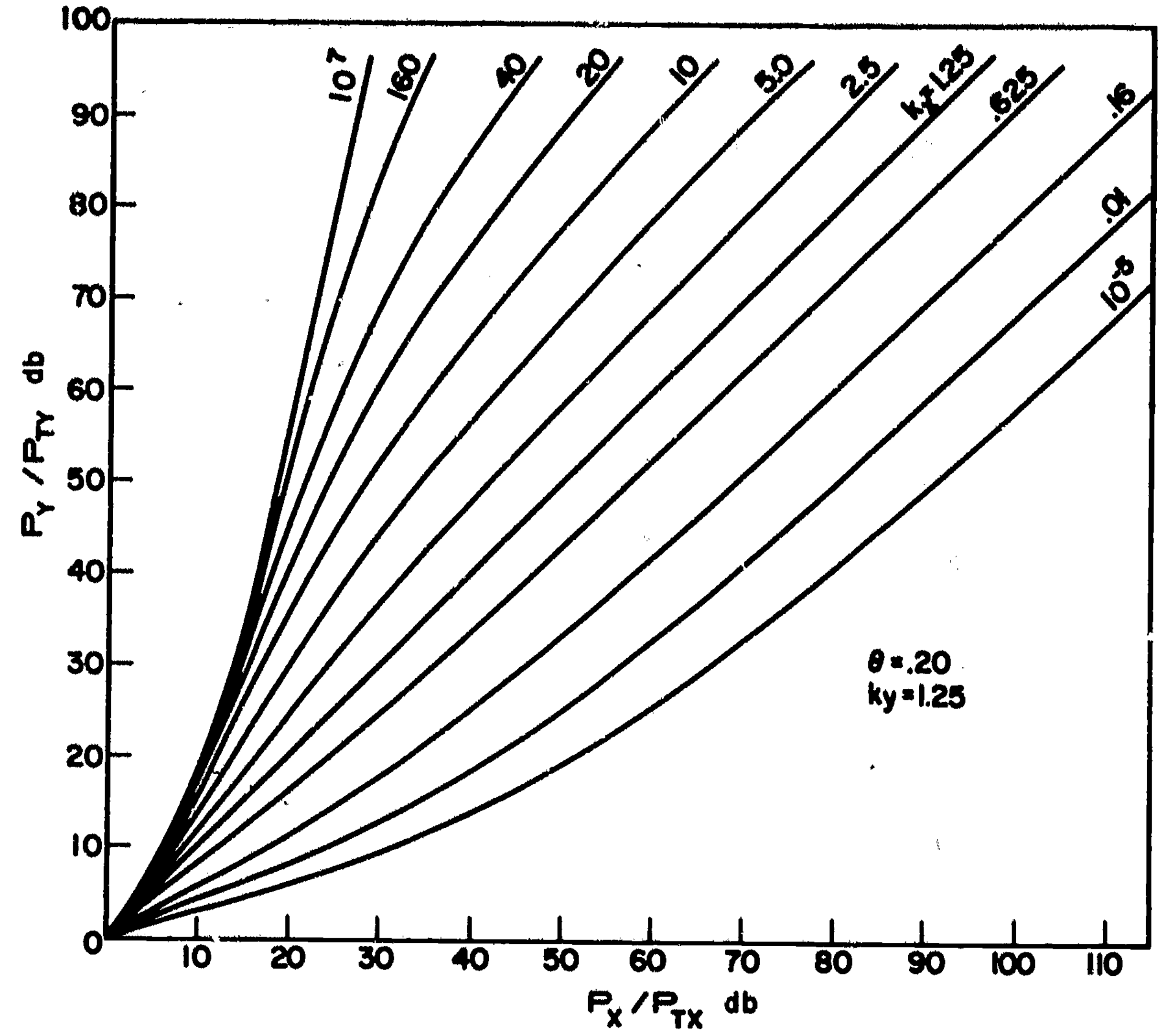

100
90
80
70
60
50
40
30
20
10
0
10
20
30
40
50
60
70
80
90
100
110
P_Y / P_{TY} db
P_X / P_{TX} db
10^7
160
40
20
10
5.0
2.5
k_x=1.25
.625
.16
.01
10^{-5}
θ=.20
ky=1.25

Fig. 8C

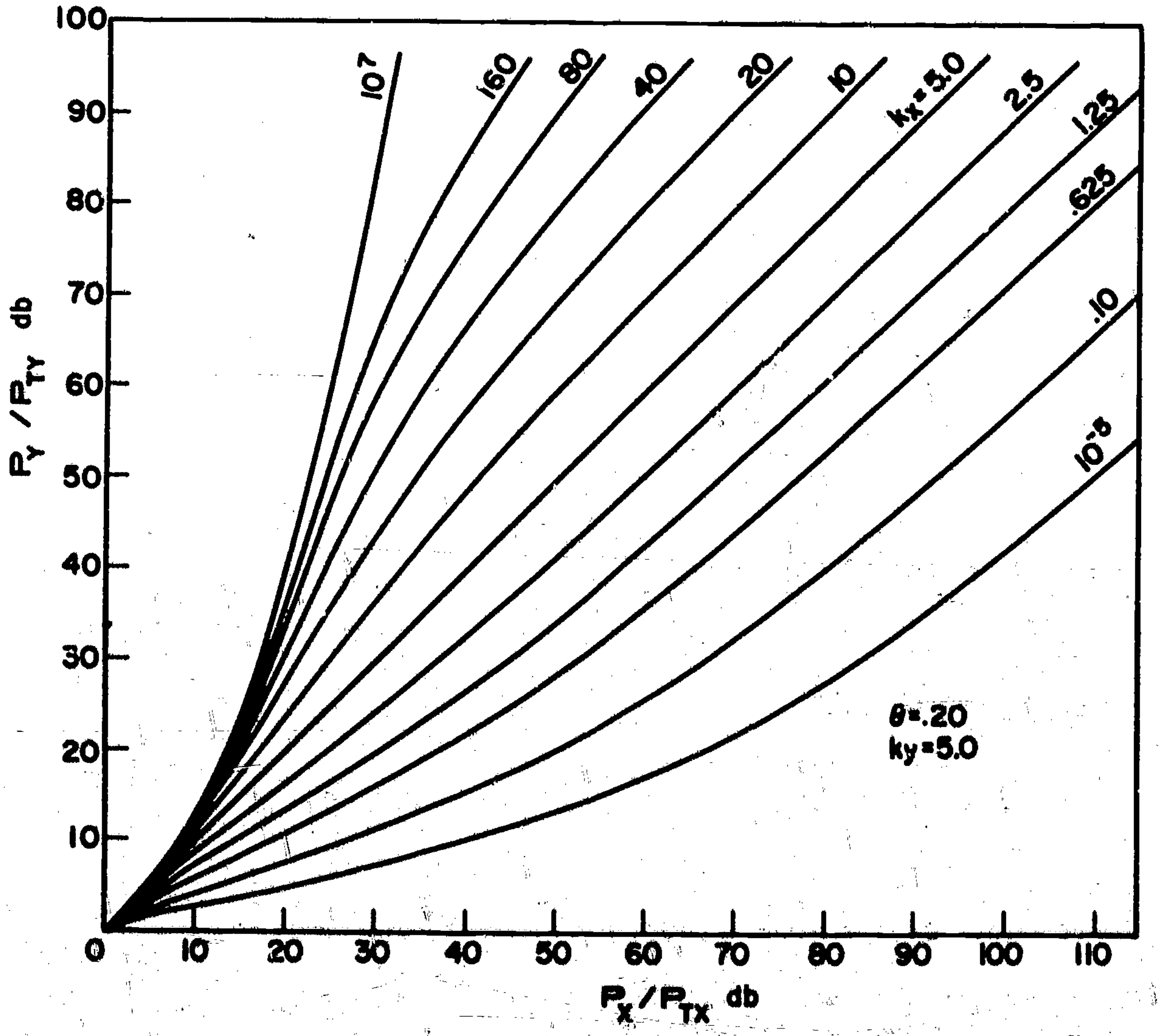
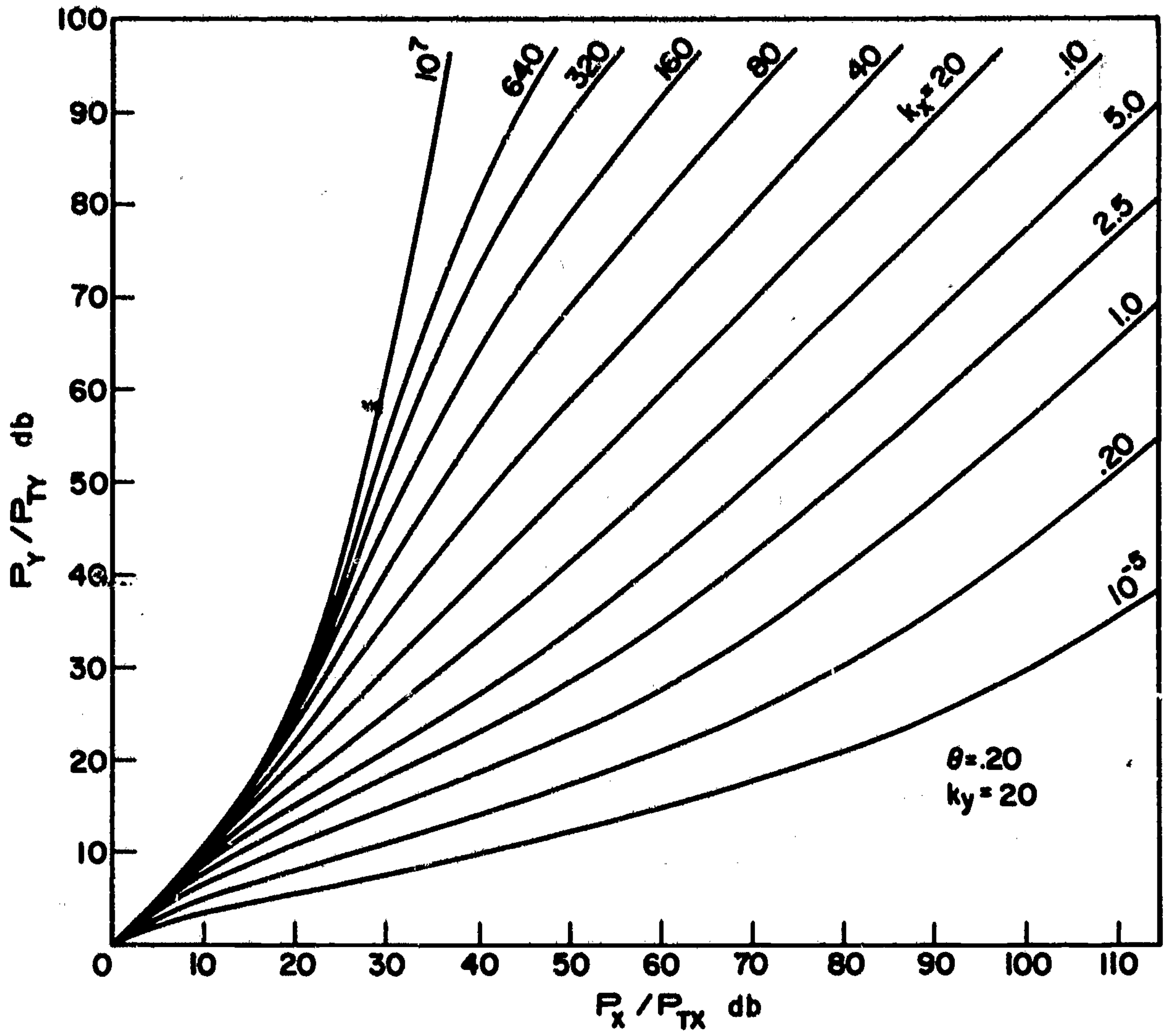
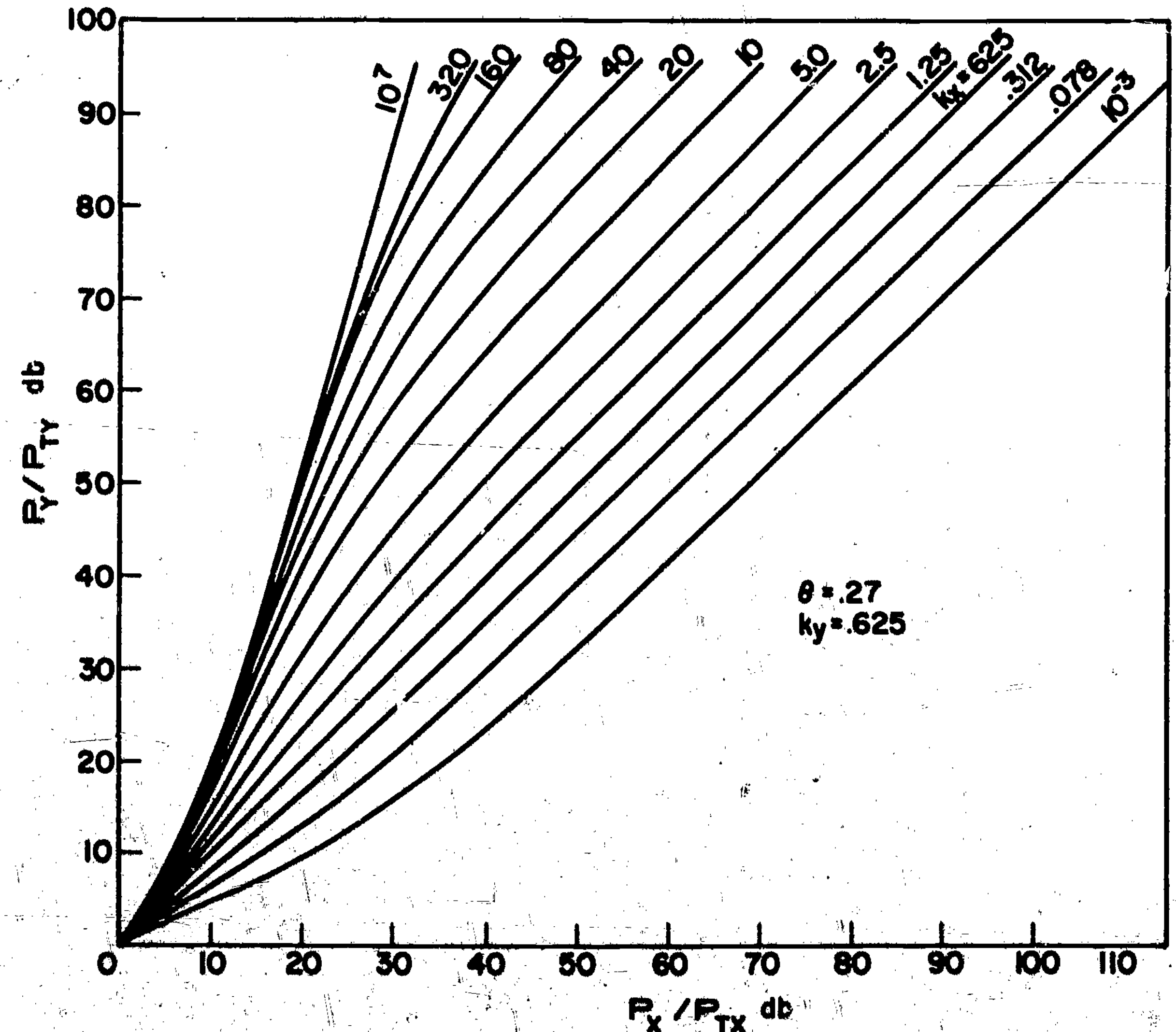

100
90
80
70
60
50
40
30
20
10
0
10
20
30
40
50
60
70
80
90
100
110
P_Y / P_{TY} db
P_X / P_{TX} db
10^7
160
80
40
20
10
k_x=5.0
2.5
1.25
.625
.10
10^{-5}
θ=.20
ky=5.0

Fig. 8D



P_Y/P_{TY} db
P_X/P_{TX} db
10^7
640
320
160
80
40
k_x=20
10
5.0
2.5
1.0
.20
10^{-3}
θ=.20
ky=20

Fig. 8E



P_Y/P_{TY} db
P_X/P_{TX} db
10^7
320
160
80
40
20
10
5.0
2.5
1.25
k_x=.625
.312
.078
10^{-3}
θ=.27
ky=.625

Fig. 8F

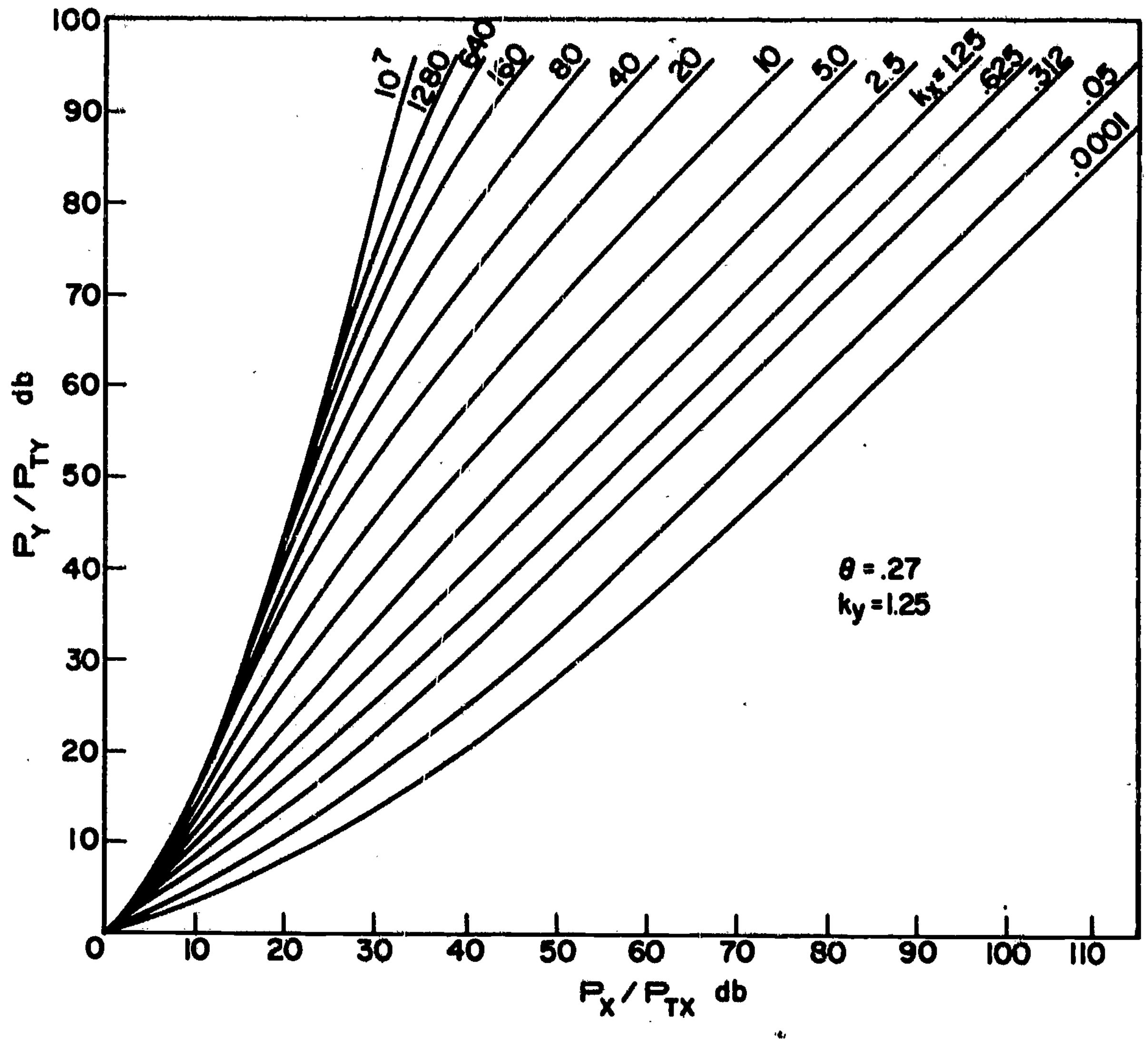


Fig. 8G

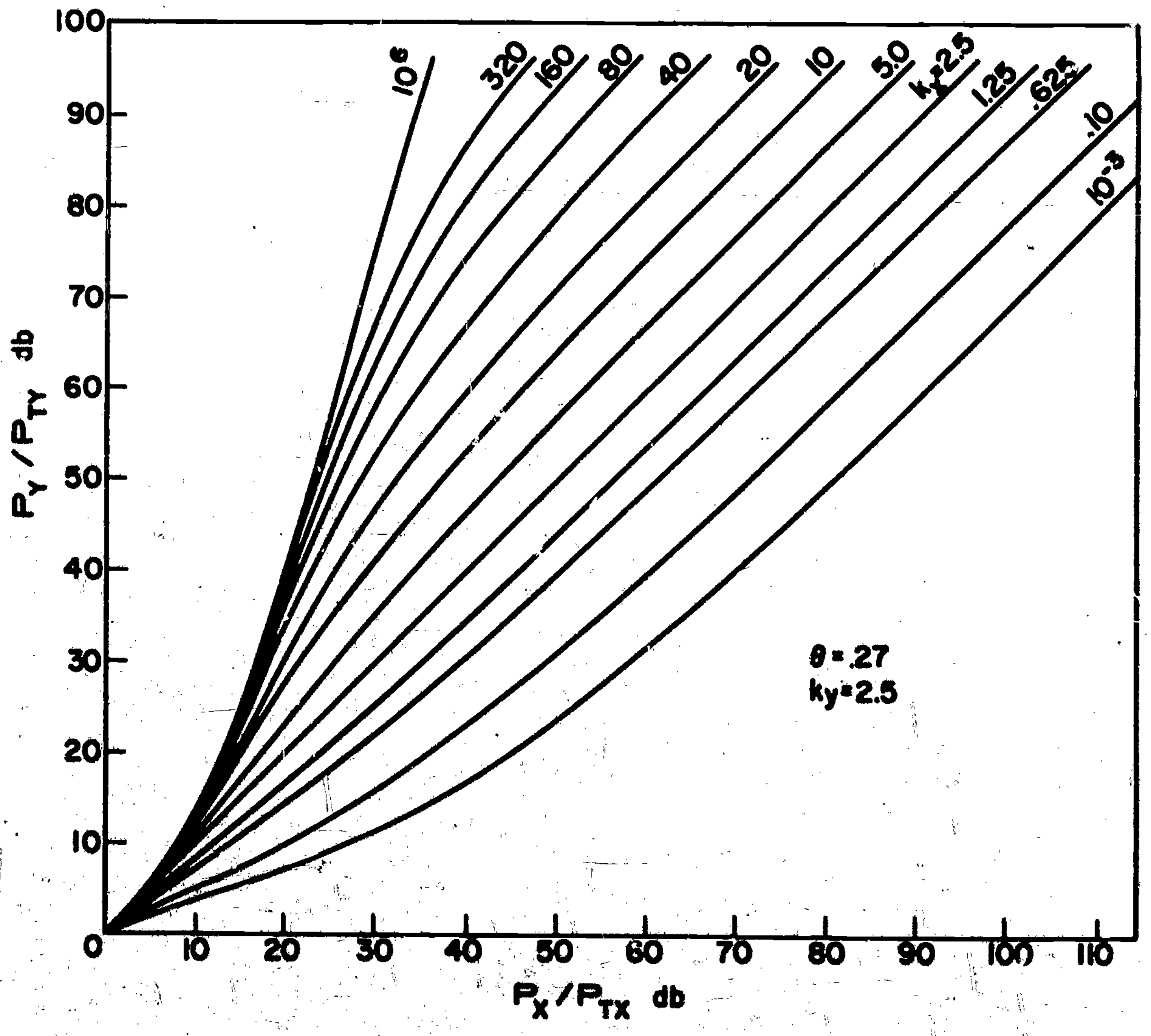
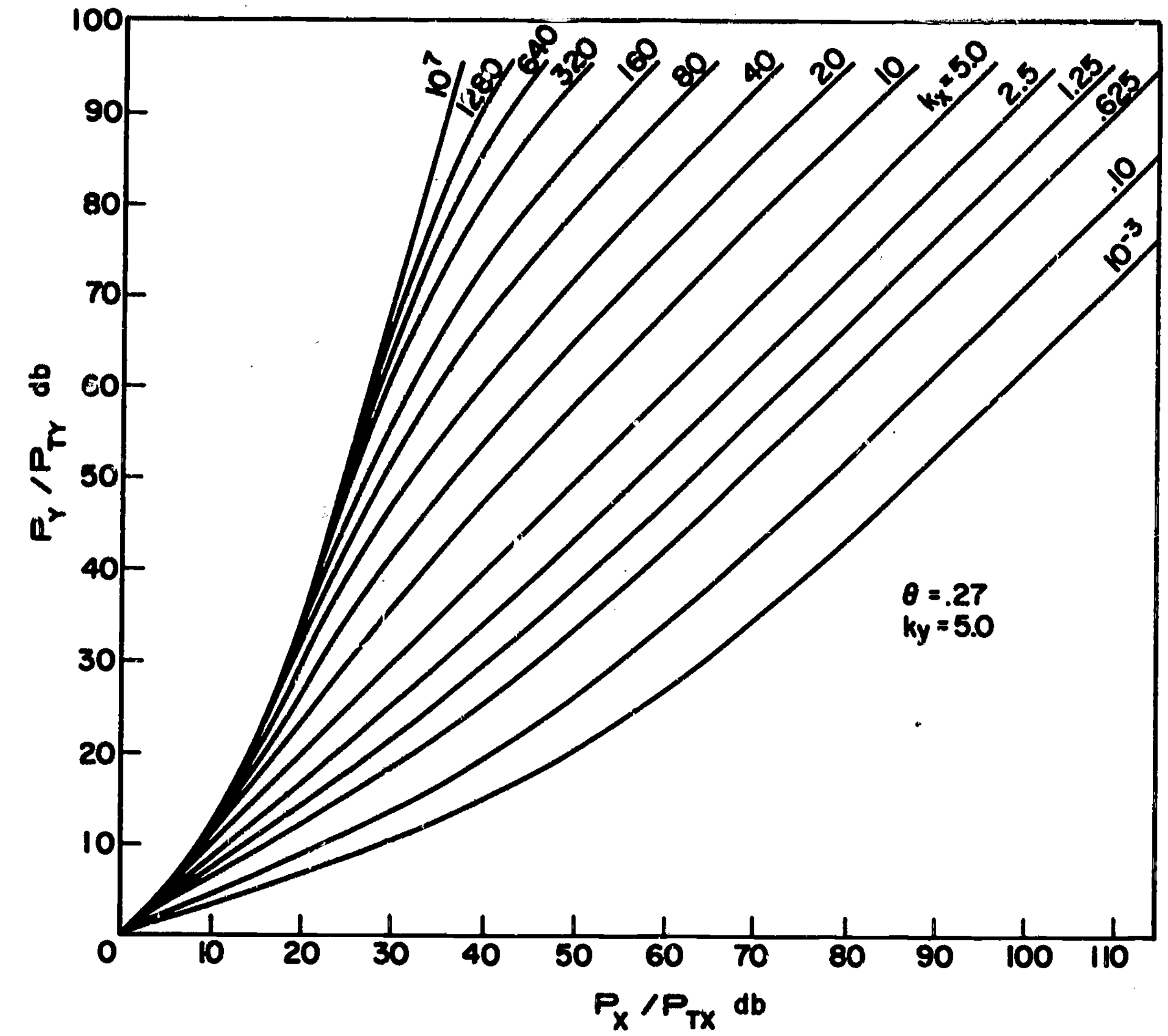
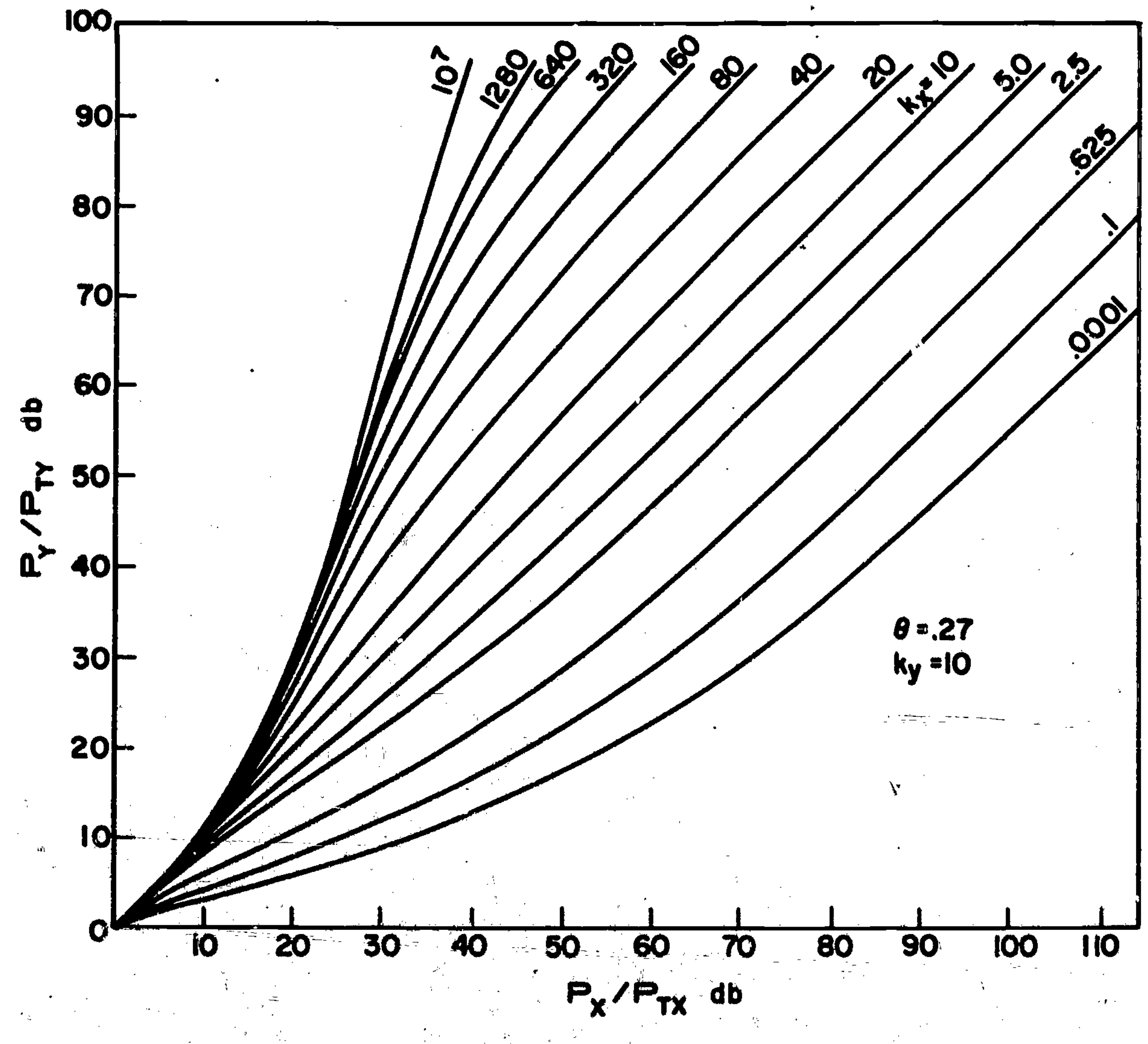


Fig. 8H



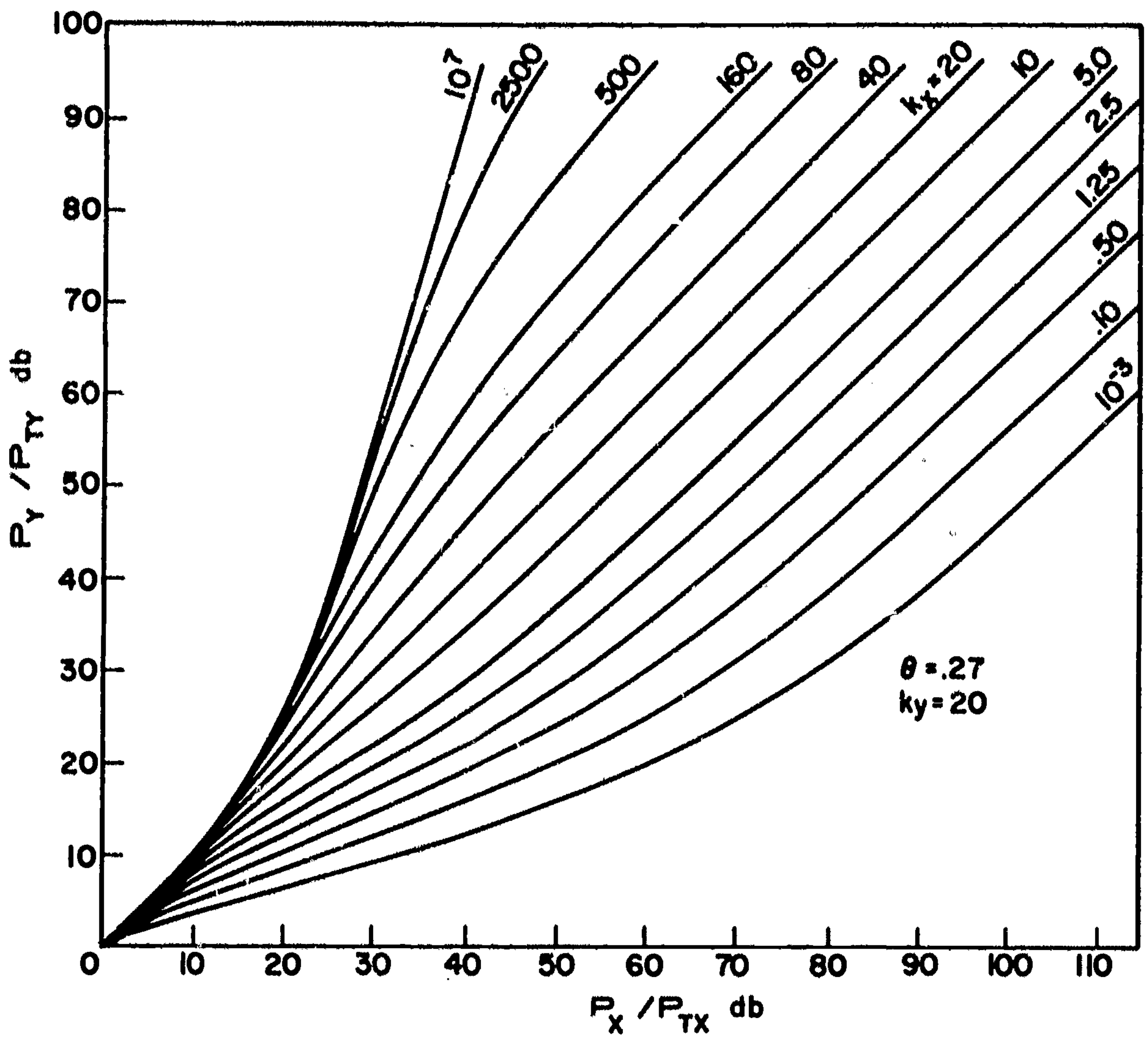
10^7
1280
640
320
160
80
40
20
10
k_x=5.0
2.5
1.25
.625
.10
10^{-3}
θ =.27
ky = 5.0
P_Y / P_{TY} db
P_X / P_{TX} db

Fig. 8I



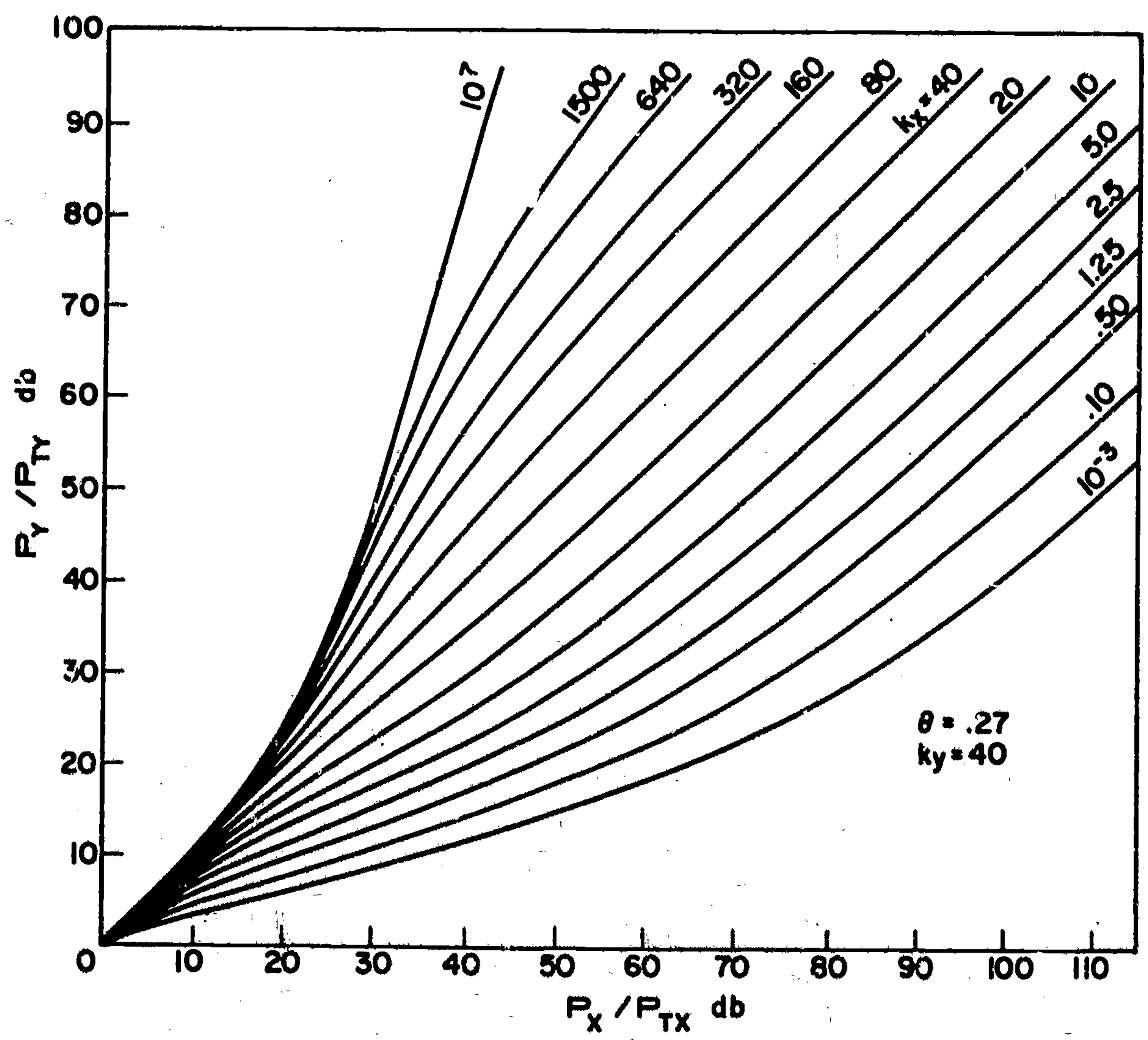
10^7
1280
640
320
160
80
40
20
k_x= 10
5.0
2.5
.625
.1
.0001
θ =.27
ky =10
P_Y / P_{TY} db
P_X / P_{TX} db

Fig. 8J



P_Y / P_{TY} db
P_X / P_{TX} db
10^7
2500
500
160
80
40
$k_x = 20$
10
5.0
2.5
1.25
.50
.10
10^{-3}
θ = .27
ky = 20

Fig. 8K



P_Y / P_{TY} db
P_X / P_{TX} db
10^7
1500
640
320
160
80
$k_x = 40$
20
10
5.0
2.5
1.25
.50
.10
10^{-3}
θ = .27
ky = 40

Fig. 8L

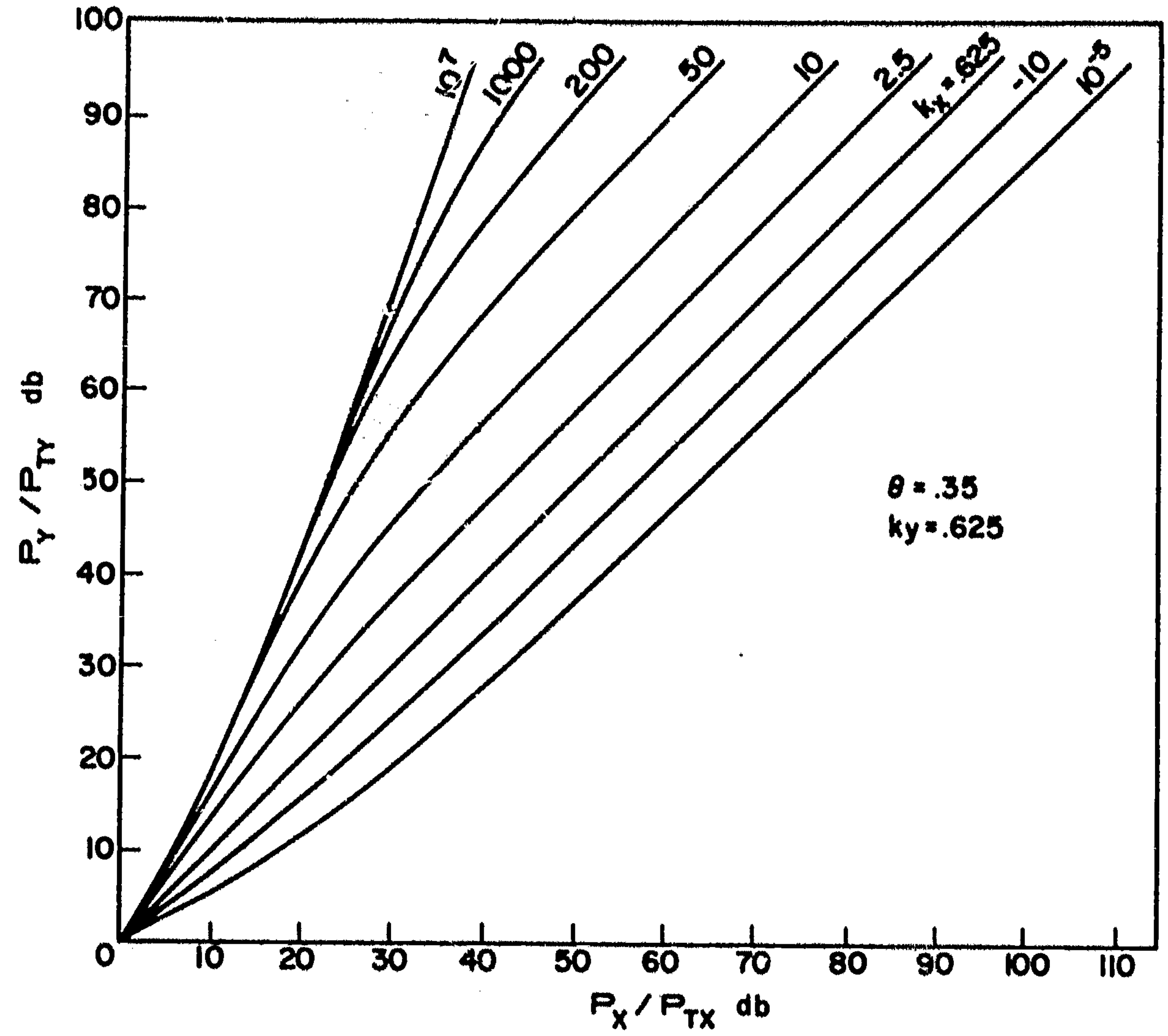


Fig. 8M

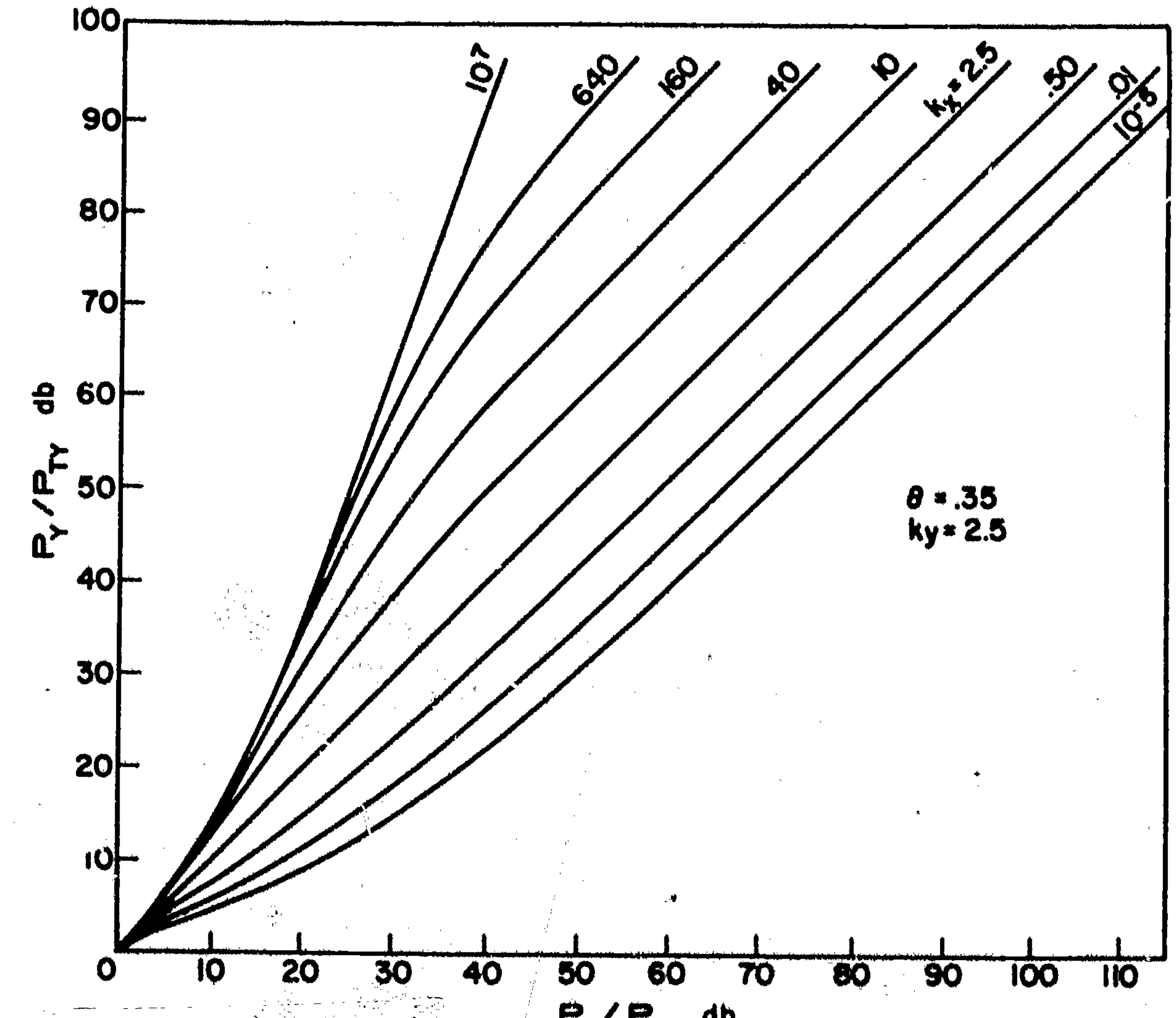
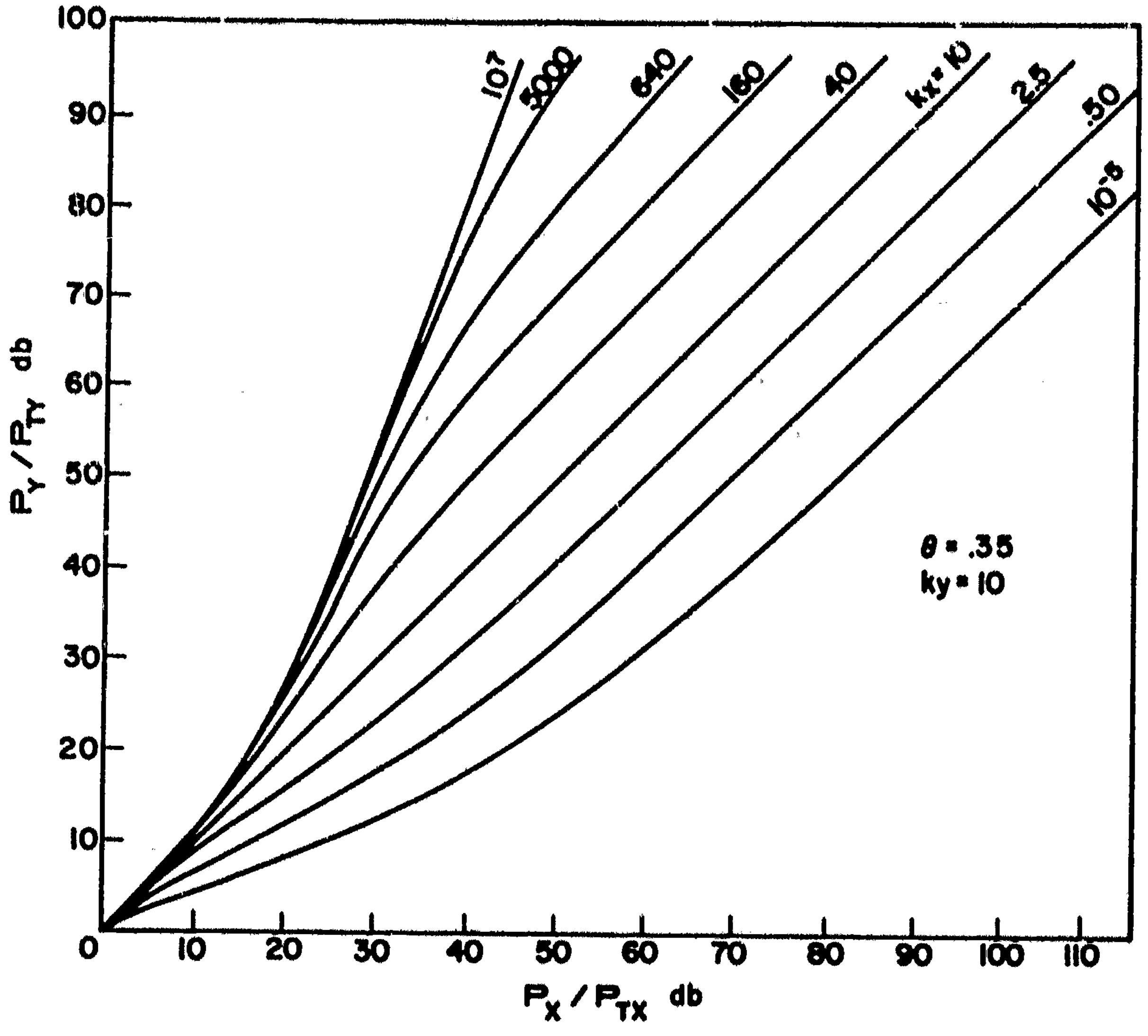
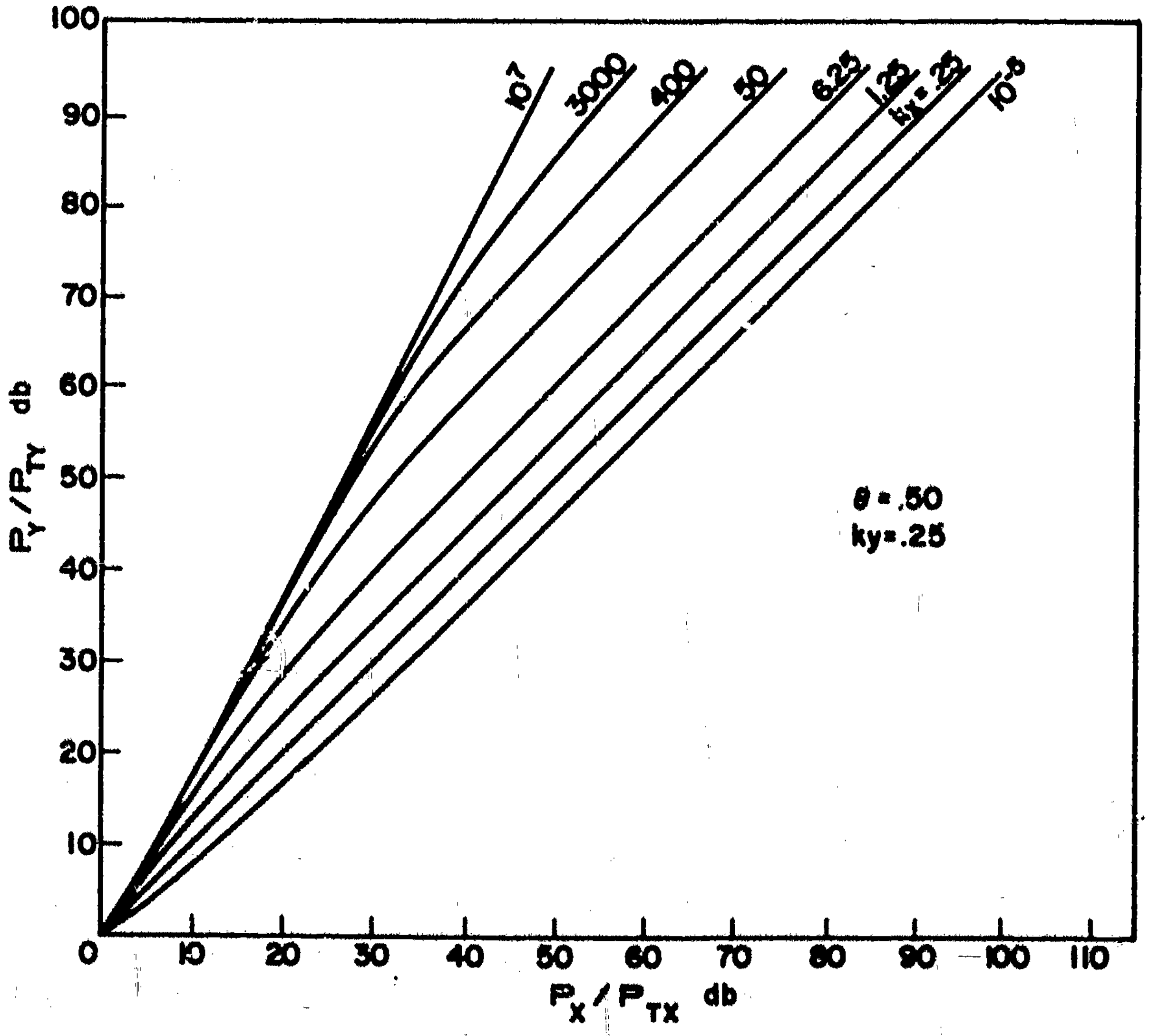


Fig. 8N



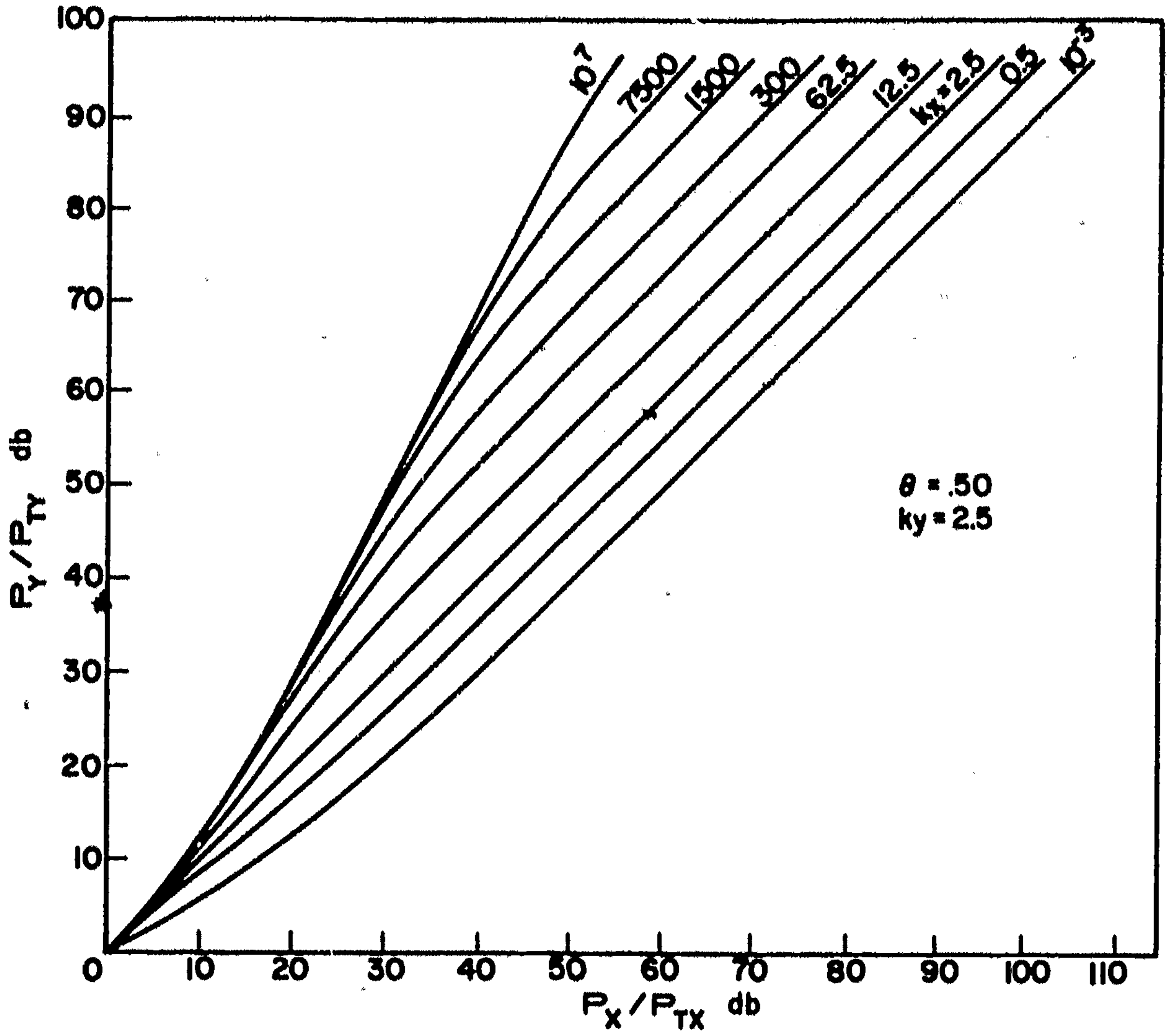
100
90
80
70
60
50
40
30
20
10
0
10
20
30
40
50
60
70
80
90
100
110
P_Y / P_{TY} db
P_X / P_{TX} db
10^7
5000
640
160
40
kx = 10
2.5
.50
10^{-8}
θ = .35
ky = 10

Fig. 8 O



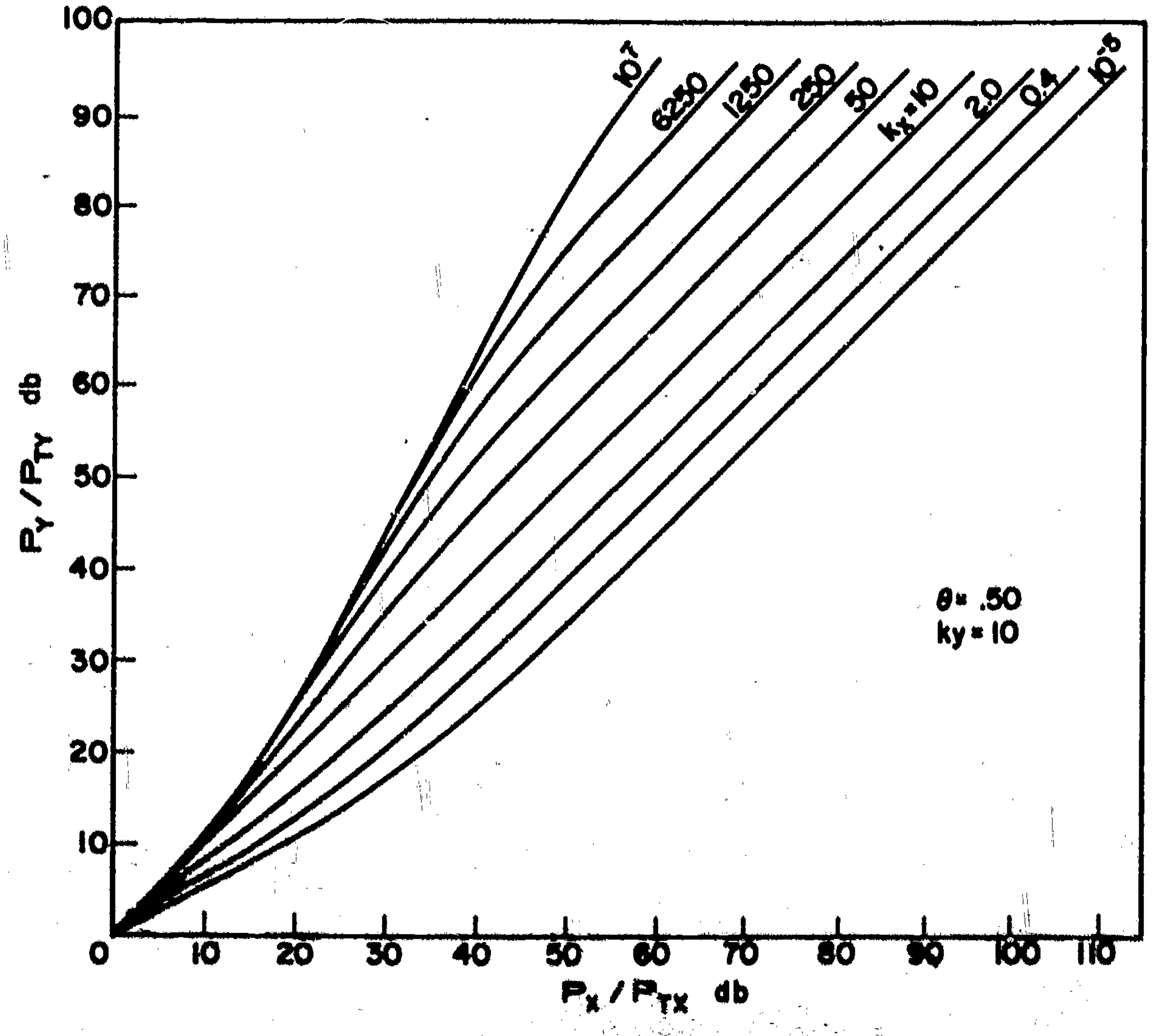
100
90
80
70
60
50
40
30
20
10
0
10
20
30
40
50
60
70
80
90
100
110
P_Y / P_{TY} db
P_X / P_{TX} db
10^7
3000
400
50
6.25
1.25
kx = .25
10^{-8}
θ = .50
ky = .25

Fig. 8P



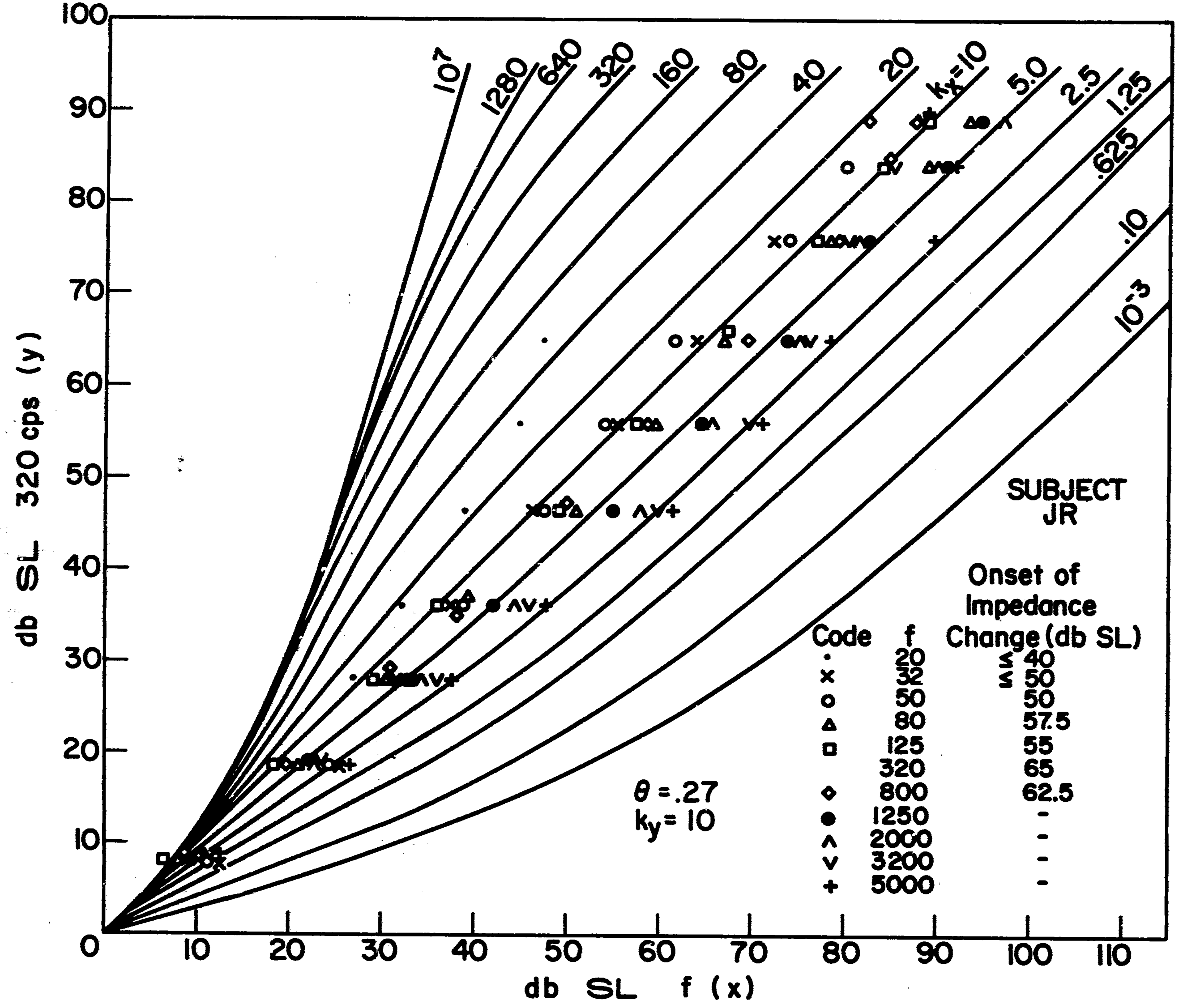
10^7
7500
1500
300
62.5
12.5
k_x=2.5
0.5
10^{-3}
θ = .50
ky = 2.5
P_Y/P_{TY} db
P_X/P_{TX} db

Fig. 8Q



10^7
6250
1250
250
50
k_x=10
2.0
0.4
10^{-3}
θ = .50
ky = 10
P_Y/P_{TY} db
P_X/P_{TX} db

Fig. 8R



SUBJECT
JR
Onset of
Impedance
Code f Change (db SL)
· 20 ≤ 40
× 32 ≤ 50
○ 50 50
△ 80 57.5
□ 125 55
320 65
◇ 800 62.5
● 1250 -
∧ 2000 -
∨ 3200 -
+ 5000 -
θ = .27
k_y = 10
10^7
1280
640
320
160
80
40
20
k_x=10
5.0
2.5
1.25
.625
.10
10^{-3}
db SL 320 cps (y)
db SL f (x)
0
10
20
30
40
50
60
70
80
90
100
110

Fig. 9

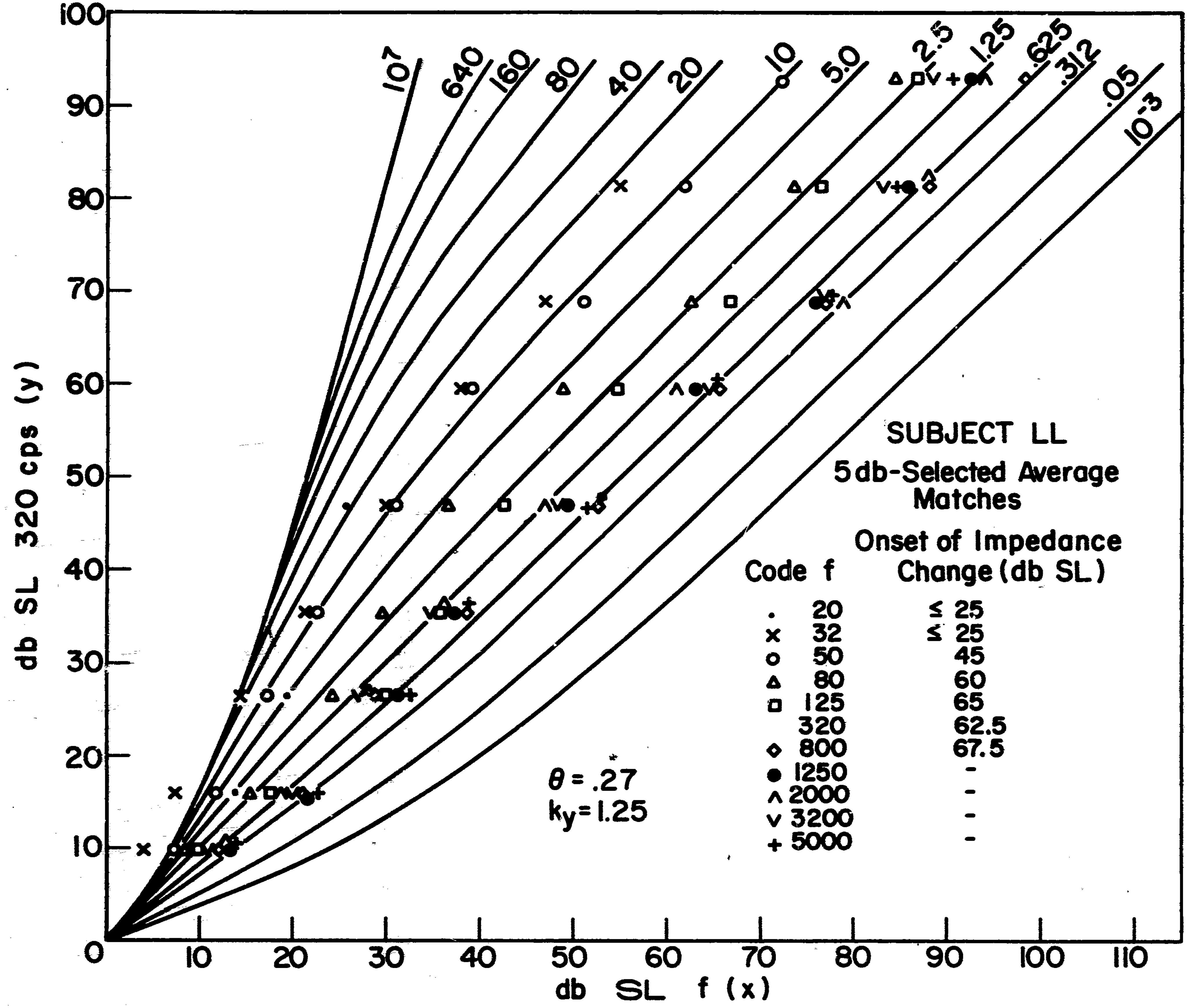
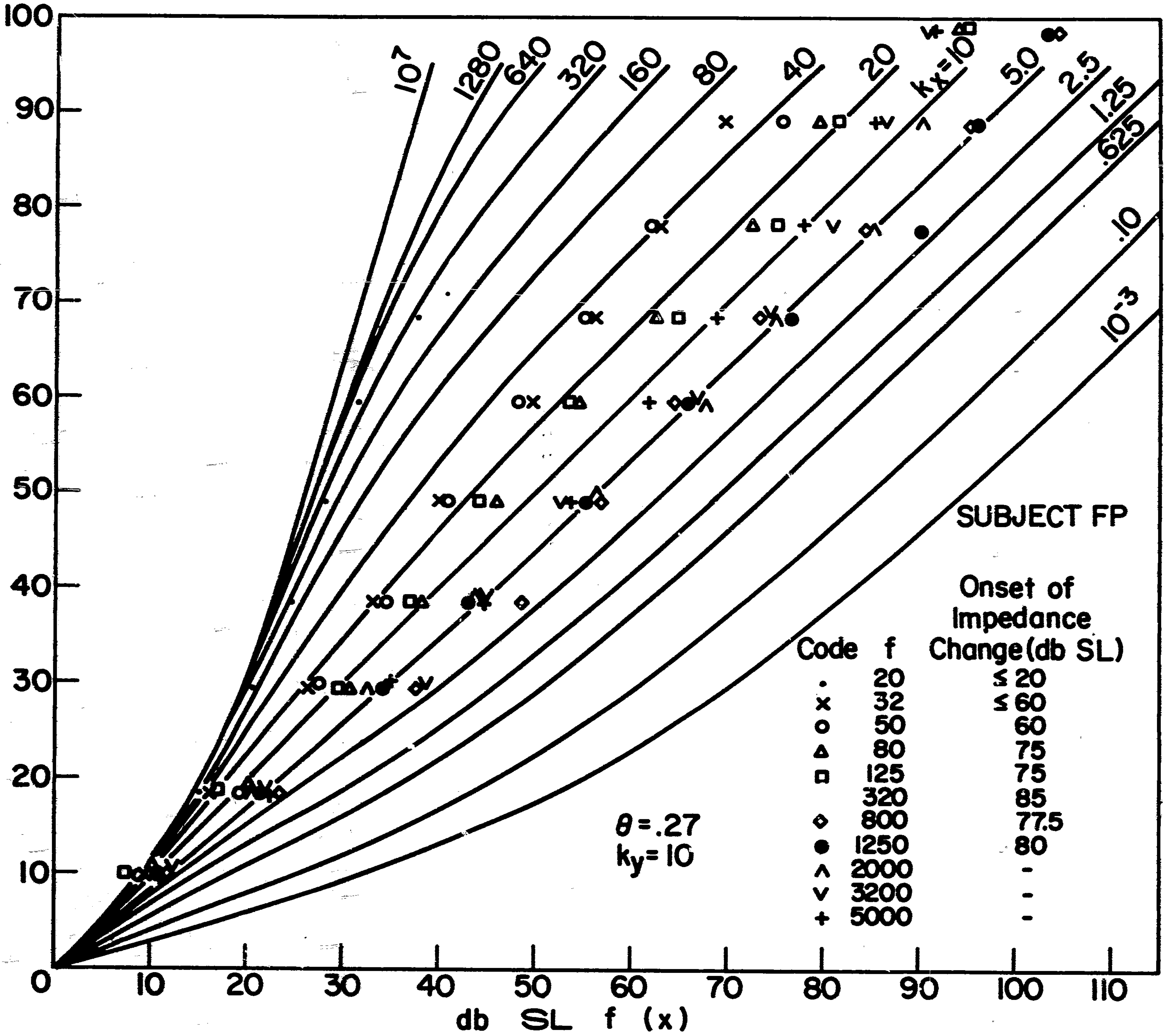

SUBJECT LL
5 db-Selected Average Matches
Code f
Onset of Impedance Change (db SL)
· 20 ≤ 25
× 32 ≤ 25
○ 50 45
△ 80 60
□ 125 65
320 62.5
◇ 800 67.5
● 1250 -
∧ 2000 -
∨ 3200 -
+ 5000 -
θ = .27
k_y = 1.25
10^7
640
160
80
40
20
10
5.0
2.5
1.25
.625
.312
.05
10^-3
db SL 320 cps (y)
db SL f (x)
0
10
20
30
40
50
60
70
80
90
100
110

Fig. 10



SUBJECT FP
Code f Onset of Impedance Change (db SL)
• 20 ≤ 20
× 32 ≤ 60
○ 50 60
△ 80 75
□ 125 75
320 85
◇ 800 77.5
● 1250 80
∧ 2000 -
∨ 3200 -
+ 5000 -
θ = .27
k_y = 10
k_x = 10
10^7
1280
640
320
160
80
40
20
5.0
2.5
1.25
.625
.10
10^-3
db SL f (x)

Fig. 11

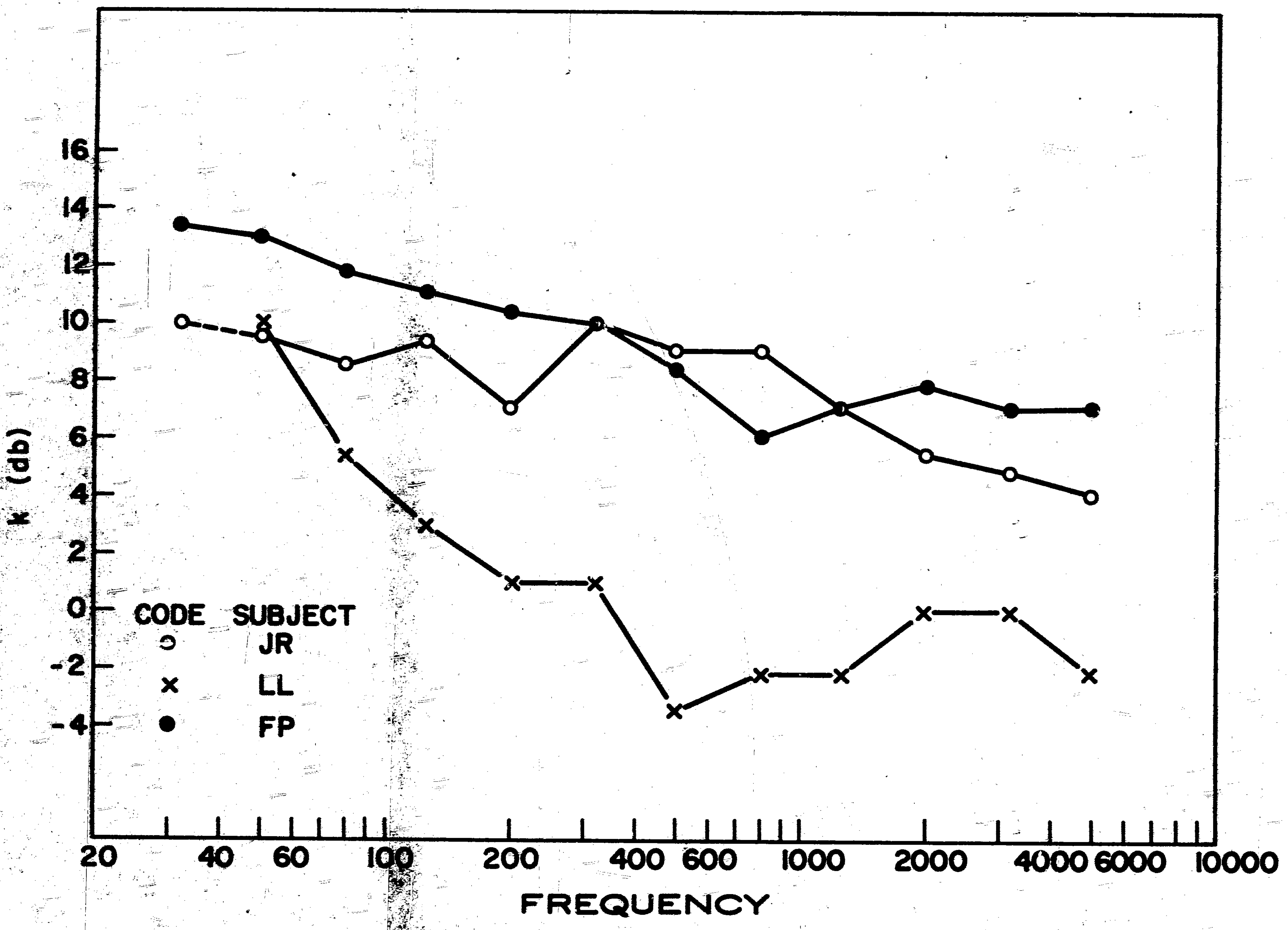
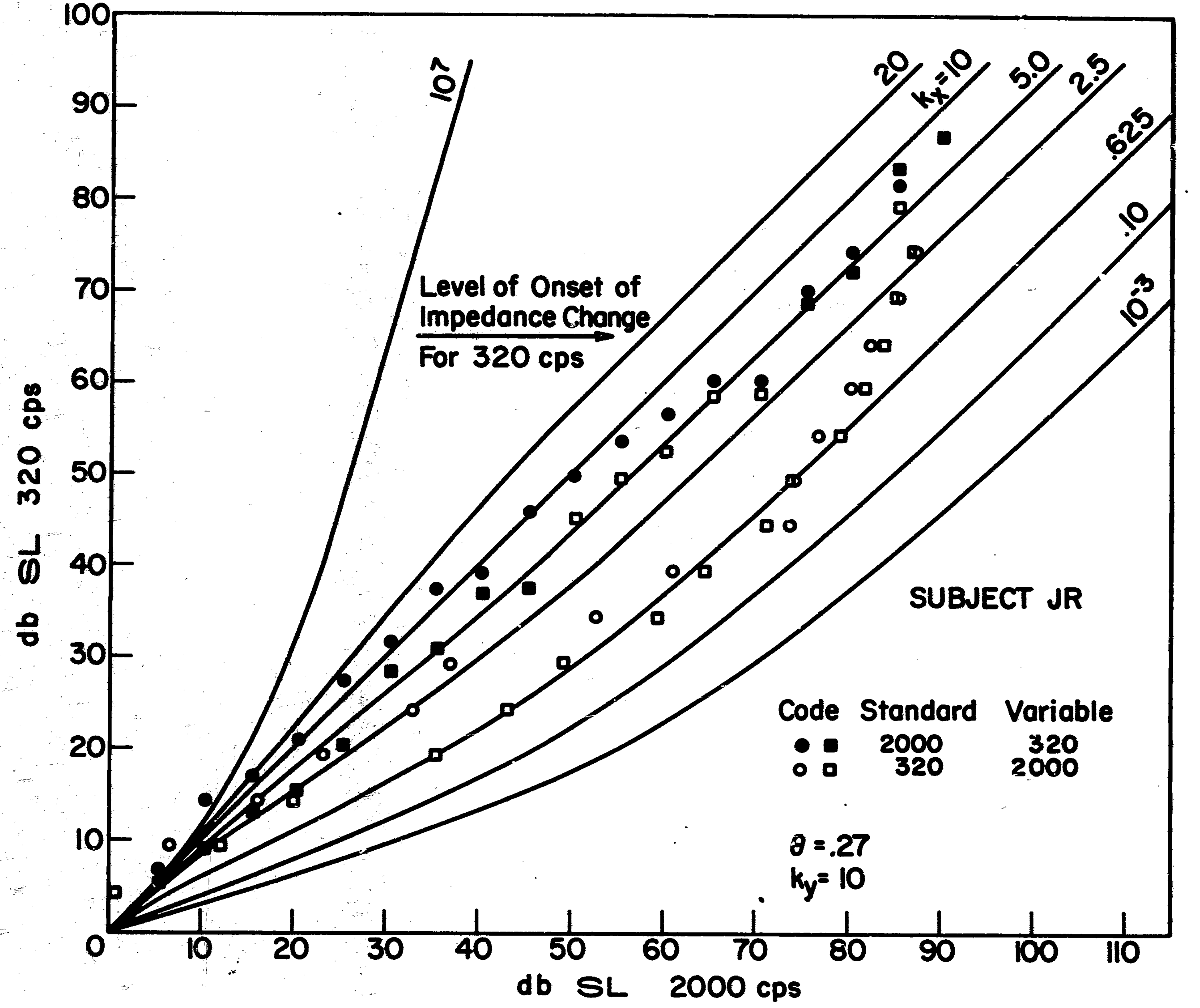

16
14
12
10
8
6
4
2
0
-2
-4
k (db)
CODE SUBJECT
○ JR
× LL
● FP
20
40
60
100
200
400
600
1000
2000
4000
6000
10000
FREQUENCY

Fig. 12



db SL 320 cps
db SL 2000 cps
Level of Onset of
Impedance Change
For 320 cps
10^7
20
k_x=10
5.0
2.5
.625
.10
10^-3
SUBJECT JR
Code Standard Variable
2000 320
320 2000
θ = .27
k_y= 10

Fig. 13

Ocular Movement and Graphic Pattern Recognition

Carl C. Semmelroth and Donald E. P. Smith

Micro-movements during apparent fixations in reading

Saccadic eye movements, the ballistic motion and apparent stops or fixations during reading, have been studied extensively (Dodge, 1905; Judd & Buswell, 1922; Taylor, Frackenpohl, & Pettee, 1960). Observation of the oral reading errors of problem readers led to the hypothesis that small, extremely rapid visual excursions to adjacent lines of print must occur during or between fixations, resulting in the processing of extraneous letters or words. The Mackworth Optiscan camera (1962) was modified to provide film records of the reading behaviors of normal and problem readers.

The Optiscan consists of a miniature light source, a periscope, and a movie camera. Its purpose is to record the scene as viewed by the observer (O), along with a small eye-position marker (spot of light), superimposed to indicate where O is looking (fixating) from moment to moment. The image of a miniature lamp is reflected from the cornea of the left eye into the periscope which then superimposes this marker onto the scene image. Thus, the eye movements are recorded from the left eye, while only the right eye is "seeing."

Both head movement and displacement of the rotational axis of the eye produce distortion in the record. These were reduced by modifying the recording and analysis procedures. The O sat in a specially-constructed chair, topped by a camera platform. His head was secured to a headrest and a bite-board provided a rest for his upper jaw, while still allowing oral reading. The visual display, a paragraph from an oral reading test (Gray, 1955), appeared on a screen at eye-level, 116" from O, and print size was equated to near-point reading. Before and after reading, _O_ fixated coordinates superimposed on the display. The locus of the coordinates on the finished film allowed determination of the amount and direction of error due to head movement. Since the task was brief, several records were made and those with least head movement selected for analysis. Distortion due to shifts in the rotational axes of the eye was reduced by using a small field (less than 6° of arc in radius) and by adjusting the size of the image on playback so that the distance between coordinates remained the same, as in the display condition. Thus, lateral distortion was minimized by using a small field and reduced further through counter-distortion in the analysis procedure.

A frame-by-frame analysis provided the results appearing in Fig. 1. Line A is the record of a college student reading at a high third-grade level; line B is his record after perceptual training (Geake & Smith, 1962); line C is the record of a normal reader. Each circle has a latency of .07 sec or less. Squares represent three or more frames too close together to allow differentiation. The following points are worthy of note:

1. Fixations in the traditional sense (latencies of .24 to .33 sec., Judd & Buswell, 1922; Taylor,et al., 1960) consist of three or more micro-fixations. The slow speed of the camera (14 frames/sec) makes latencies less than .07 sec difficult to detect. Exposure time of the film is about .03 sec per frame. Movements within a .03 sec interval are not observed except for an occasional light trace.

2. The amplitude of micro-movements within a letter ranges from 5 to 15 minutes of arc on the size of print used (the maximum visual angle subtended by an individual letter was 26 minutes of arc).
3. Micro-movements during a fixation appear to follow the contours of letters. (Stimuli consisted of black letters on a white field.)
4. Excursions occur when a similar word, letter or, perhaps, identifying characteristic of a letter is adjacent to the fixated word (note and and queen).
5. After perceptual training, excursions continue to occur but with less frequency and shorter latency. Total time on line A is 7.0 sec, on line B, 2.45 sec.
6. Excursions of extremely brief latency occur also in the record of the normal reader (total time: 2.52 sec).

To provide a more accurate analysis of single-letter preception, the width of the field was reduced to 4^o of arc in radius, and the size of the print was increased by about 60%. Brightness of the point source of light was decreased and a faster film was used. The pattern was superimposed on the print with the result appearing in Fig. 2. The following points are notable:

1. The amplitude of micro-movements within letters of this size ranges from 5 to 25 minutes of arc (maximum visual angle subtended by an individual letter was 39 minutes of arc).
2. Fixations tend to occur on the middle letters, whereas, in Fig. 1., they tend to occur on beginning and ending letters of words. This difference may be a function of the familiarity of the material.
3. Attack usually begins on the right side and midway in the letter (note no and little).

4. Movement follows the contours, with curves, verticals and terminals receiving equivalent attention. Note the excursion on the e of the.

A possible relationship between the micro-movements observed during reading in this study and the slower "jerky" forms of physiological nystagmus should be considered. Ratliff and Riggs (1950) reported motions with amplitudes between one and five minutes of arc occurring at the rate of two to five cycles per sec during fixation of a small black dot. They concluded that the shape of the fixation object had no effect on the nature of the physiological nystagmus in their study. However, their data were not obtained during an active perceptual search task such as reading.

Further formulation of problem and report on instrumentation

Do brief idiosyncratic ocular excursions during reading contribute to some characteristic errors in problem readers? This question was prompted by observation of problem readers. Preliminary work with a modified Mackworth Optiscan has led to the more general question: Are there stimulus-bound ocular movements in the size and frequency range of physiological nystagmoid movements which are significant contributors to graphic pattern recognition?

The present objective is to construct a device superior to the Optiscan which has been employed previously. Ocular activity during pattern recognition will be recorded with sufficient accuracy to permit analysis of movements in the nystagmoid range.

The major components and characteristics of the apparatus are, in brief:

1. Two photocells positioned in a horizontal plane below eye level.
2. Two other photocells positioned very close together, one over the other, with a lens arranged so that a blurred corneal reflection falls on both of them.

3. A small infrared source furnishes corneal reflection.

4. Horizontal movements of the eye are detected and quantified by voltage differentials between the outputs of the two photocells positioned in a horizontal plane. These voltage changes come principally from the movements of the iris-sclera boundary in relation to the photocells. This movement causes a differential change in the amount of light reaching the two photocells because there is a difference in reflectance between the iris and the sclera.

5. Vertical movements of the eye are detected and quantified by differences in voltage between the outputs of the two photocells that receive a blurred image of the corneal reflection. Since these two photocells are placed one over the other, as the eye changes its position in a vertical plane, the blurred corneal reflection occupies a larger portion of one of the cells while vacating the other.

6. The photocell outputs are fed to an analog computer which applies corrections needed for linearity of output as well as removing correlation between the vertical and horizontal vectors.

7. The computer outputs may be displayed or stored in any one of a number of ways, e.g.:

   a. It may be converted to digital form for computer analysis.

   b. It may be displayed and photographed on an oscilloscope.

   c. It may be fed to a closed-circuit television system and mixed with a picture of the field being viewed by the eye.

References

Dodge, R. The illusion of clear vision during eye movement. *Psychol. Bull.*, 1905, 2, 193-199.

Geake, R. R., & Smith, D. E. P. *Visual tracking*. Ann Arbor: Ann Arbor Publishers, 1962.

Gray, W. S. *Oral reading paragraphs*. Indianapolis: Bobbs-Merrill, 1955.

Judd, C. H., & Buswell, G. T. Silent reading: a study of various types, *Suppl. educ. Monogr.*, 1922, 23.

Mackworth, N. H., & Thomas, E. L. Head-mounted eye-marker camera. *J. Opt. Soc. Amer.*, 1962, 52, 713-716.

Ratliff, F., & Riggs, L. A. Involuntary motions of the eye during monocular fixation. *J. exp. Psychol.*, 1950, 40, 687-701.

Taylor, S. E., Frackenpohl, Helen, & Pettee, J. L. No. 3, *Educ. Dev. Labs*, 1960.

Figure Captions

Fig. 1. Micro-movements of the eye during reading as recorded by a corneal reflection technique. Circles are of .07 sec. duration or less; squares indicate three or more circles. Lines A and B are the records of a college student with a reading disability, before and after perceptual training. Line C is the record of a normal reader. Stimulus material consisted of solid black letters on a white field; only outlines of letters are shown here.

Fig. 2. Pattern of micro-fixations of a normal reader while reading familiar material. For clarity, dotted lines between groups of micro-fixations are not included. Filled circles are initial fixations; latencies are similar to those in Fig. 1. Curved dotted lines represent excursions inferred from light traces on single frames of film.

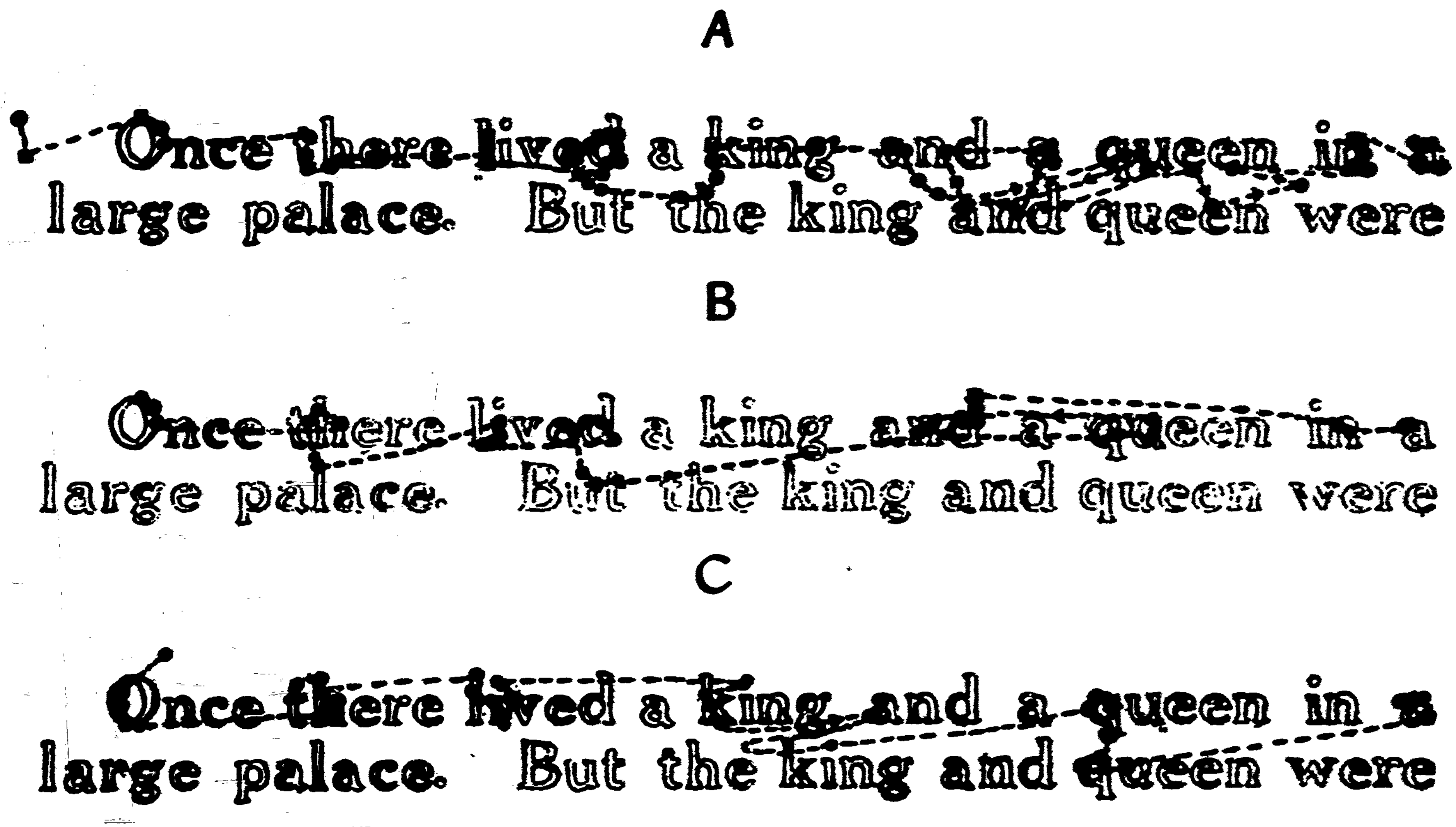

A
Once there lived a king and a queen in a
large palace. But the king and queen were
B
Once there lived a king and a queen in a
large palace. But the king and queen were
C
Once there lived a king and a queen in a
large palace. But the king and queen were

Fig. 1

Fig. 2

Bibliography of Research

Using the Technique of Matching-to-Sample

George L. Geis

Dwight Spenser

In teaching discriminative or conceptual behaviors a common procedure is matching-to-sample. The subject is faced with a sample, or example, and a series of alternatives, one of which matches the sample. He must choose from among the alternatives; a correct choice yields reinforcement. This procedure is used in many educational situations, such as the teaching of letter- and word- recognition. Recently, psychologists have shown renewed interest in using the technique in the laboratory as well--for example, in studying discrimination learning.

In view of the proven usefulness of and renewed interest in matching-to-sample and the lack of a satisfactory guide to studies using this technique, the annotated bibliography presented here may prove helpful. In its preparation a search was made of the psychological literature, and fourteen investigators in the area provided information on the current status of their research. This version of the bibliography will be circulated to interested investigators and a revision will be prepared for final dissemination.

The following abbreviations are used in this article:

m/s matching-to-sample,

S subject,

ST standard stimulus, or, sample and

CO comparison, matching, or alternative stimulus.

In order to limit the scope of the literature search, we have defined the m/s procedure as one in which:

(1) An ST is presented,

(2) a set of COs, one of which is identical with the ST, is presented, and

(3) S is required to make a response to the COs, with reinforcement contingent upon S's selecting the CO which is identical with the ST. Although exceptions to these definitional constraints, also included are studies of oddity- and symbolic-matching. In oddity-matching, S is presented with two COs, one of which physically matches ST, the other of which does not; reinforcement is contingent upon selection of the latter CO. In symbolic-matching, neither CO is physically the same as ST; the CO upon whose selection reinforcement is contingent is arbitrarily designated by the experimenter. In oddity-matching, therefore, the "incorrect" CO must remain constant for any given sample; in symbolic matching the "correct" CO must remain constant.

Excluded from the bibliography by this definition are psychophysical studies in which m/s procedures are employed but reinforcement is not contingent upon a correct response; references to test materials which include examples of m/s procedures (e.g., concept-formation tests involving card sorting); and a large literature on discrimination-learning, using some variation of the Wisconsin General Test Apparatus and procedures similar to those described above.

The Bibliography

Baumeister, A., & Ellis, N.R. Delayed response performance of retardates. *Amer. J. ment. Defic.*, 1963, 67, 714-722.

Fifteen retardates performed on a m/s task with delays up to 120 sec between the removal of ST and presentation of COs. The stimuli were visual patterns. Five COs were present on each trial. As the delay increased, there was a decrease in per cent correct responses and an increase in the latency of the response. When exposed to a visual "distractor" Ss improved their accuracy and responded more quickly.

Berryman, R., Cumming, W., & Nevin, J. A. Acquisition of delayed matching in the pigeon. J. exp. Anal. Behav., 1963, 6, 101-107.

Several different procedures were employed to teach a delayed m/s key-pecking task involving colored stimuli. The methods are discussed in terms of chaining and mediating behavior.

Berryman, R., Cumming, W., Nevin, J.A., & Jarvik, M.E. Effects of sodium pentobarbital on complex operant discriminations. Psychopharmacologia 1964, 6, 388-398.

The effects of sodium pentobarbital on pigeons are reported for (a) zero and variable delay m/s and (b) simultaneous oddity, using colored lights as stimuli. There were decrements in accuracy in all cases with increasing dose levels and delays. The oddity performance was much less sensitive to drug effects than zero-delay matching. In variable-delay matching, position preference increased with an increase in delay and drug desage.

Berryman, R., Jarvik, M. E., & Nevin, J. A. Effects of pentobarbital, lysergic acid diethylamide and chlorpromazine on matching behavior in the pigeon. Psychopharmacologia, 1962, 3, 60-65.

The effects of chlorpromazine on m/s behavior were found to be variable. LSD produced initial inactivity, followed shortly by relatively normal performance. Pentobarbital produced initial decrements in accuracy, but recovery to normal levels was rapid.

Bitterman, M.E., & Wodinsky, J. The solution of oddity and non-oddity problems by the rat. Amer. Psychologist, 1953, 8, 458. (Abstract)

Two groups of rats were trained on a three-window jumping apparatus. The black and white cards were arranged BBW, WBB, WWB, and BWW. One group was trained to respond to the odd card and the other group to the side card that matched the center card. The oddity group learned significantly faster.

Blough, D. S. Delayed matching in the pigeon. J. exp. Anal. Behav., 1959 2, 151-160.

Four pigeons were trained in m/s with a flickering or steady light, with ST-CO delays from 0 to 10 sec. For two birds the accuracy of matching was a stable decreasing function of delay, whereas the function was unstable for the other two birds. Stereotyped responding to each display, as mediating behavior, is discussed.

Blough, D. S. Effects of drugs on visually controlled behavior in pigeons. In S. Garattini, & V. Ghetti (Eds.) Psychotropic drugs. Amsterdam: Elsevier, 1957. Pp. 110-118.

The effects of three drugs, LSD, chlorpromazine, and pentobarbital on matching, delayed matching and visual threshold, are discussed. Steady and flickering lights were stimuli. Three techniques of studying drug effects on pigeons are also described.

Cumming, W. W., Berryman, R., & Nevin, J. A. Search for an effect of satiation on delayed matching-to-sample performance. *Psychol. Rep.*, 1965, 16, 645-652.

The effects of satiation on a well-established delayed m/s performance were studied in three pigeons. In order to manipulate drive, *Ss* were (1) allowed to run to satiation or (2) were prefed 10% of their working weight. At low-drive levels there was no reduction in per cent correct, and performance was slightly better than with high-drive levels at long delays.

Cumming, W. W., & Berryman, R. Some data on matching behavior in the pigeon. *J. exp. Anal. Behav.*, 1961, 4, 281-284.

M/s behavior was investigated using colored stimuli.

Cumming, W. W., & Berryman, R. The complex discriminated operant: Studies of matching-to-sample and related problems. In D. Mostofsky (Ed.), *Stimulus generalization*. Stanford, Calif.: Stanford Univ. Press, 1965. Pp. 284-330.

Reviewing experimental animal work in m/s discrimination learning, the authors discuss interrelations between *ST* and *CO* sets, the temporal relation of the *ST* and *COs* (simultaneous, zero delay, and delayed presentation of the *COs*), and the reinforcement contingencies: matching, "amatching," and oddity. Also discussed are motivation and delayed matching, ratio reinforcement of matching behavior, the effects of drugs, a technique called "titrating delayed matching," and the relationship of "hypotheses" to matching acquisition. The authors conclude their article with an argument that "the *ST* is exercising an instructional function and acts as a selector of S^Ds rather than of S^Δs."

Eckerman, D. A. An example of the control of matching-to-sample accuracy by a chained response to the sample stimulus. Paper read at Eastern Psychol. Ass., Philadelphia, March, 1965.

Three pigeons, with a long history of m/s training, served as Ss in an experiment designed to determine the importance of responding to the ST. It was concluded that "both the proximity (physical) to the standard stimulus and active responding to it were correlated with high matching accuracy."

Ellis, N. R. The stimulus trace and behavioral inadequacy (delayed response). In N. R. Ellis (Ed.), Handbook of mental deficiency. New York: McGraw-Hill, 1963. Pp. 134-158.

Six female retardates were trained to march colors in a delayed m/s experiment. The number of correct responses was a decreasing function of the delay intervals, which ranged from one to 68 seconds. The chapter includes a general discussion of the Stimulus Trace Theory.

Ferster, C.B. Intermittent reinforcement of matching-to-sample in the pigeon. J. exp. Anal. Behav., 1960, 3, 259-272.

M/s in pigeons was reinforced on fixed-ratio, fixed-interval, variable-interval, and multiple schedules. Stimuli were colored lights. Each schedule produced characteristic patterns of responding and accuracy levels.

Ferster, C. B. Arithmetic behavior in chimpanzees. Scient. American, 1964, 210, 98-106.

Two chimpanzees were trained to assign the appropriate binary number to a group of geometric forms, ranging from one to seven figures in each group.

The training was accomplished by starting with simple color matching and gradually increasing the complexity of the matching task. Approximately 800,000 trials were required to form the discriminations.

Ferster, C. B., & DeMyer, M. K. Increased performances of an autistic child with prochlorperazine administration. J. exp. Anal. Behav., 1961, 4, 84.

The number of m/s sequences emitted was used to assess the effects of prochlorperazine on the activity level of an autistic child. In every session m/s activity increased then the drug was introduced.

Ferster, C. B., Levitt, E. E., Zimmerman, J., & Brady, J.P. The measurement of hypnotic effects by operant-reinforcement techniques. Psychol. Rec., 1961, 11, 427-430.

Oddity-matching was used to determine the influence of hypnotic suggestion on one human S. Varying hypnotic instructions produced different levels of accuracy and rates of responding. M/s is proposed as a useful tool for studying many variables relating to hypnosis.

Ferster, C. B., & Appel, J. Punishment of S^{Δ} responding in matching-to-sample by time out from positive reinforcement. J. exp. Anal. Behav., 1961, 4, 45-46.

Time out was used to punish incorrect responses of two pigeons in an m/s task. Effects on performance were examined as a function of the duration of the time out. Longer time outs produced more accurate performance, except at extreme values, where accuracy broke down.

Finch, G. Delayed matching-from-sample and non-spatial delayed response in chimpanzees. J. comp. Psychol., 1942, 34, 315-319.

Two chimpanzees were tested on non-spatial and spatial-delayed m/s tasks. Results indicated that the spatial matching was easier than the non-spatial delayed response.

Geis, G. L. Matching-to-sample by retarded children: the effects of delayed response and number of alternatives. Unpublished doctoral dissertation, Columbia University, 1965.

A total of 27 of 40 retarded children learned accurate m/s discriminations involving colored stimuli. The nonlearners showed stereotyped place and color responses. The number of COs, which varied from 2-5, was not a significant variable in the study but, as the delay increased up to 30 sec, performance deteriorated.

Ginsburg, N. Matching in pigeons. J. comp. physiol. Psychol., 1957, 50, 261-263.

Eighteen pigeons were tested on matching, non-matching and "amatching" tasks. Non-matching proved to be easiest. When tested under high- and low-drive levels, results shower that increased motivation hindered learning in the more difficult tasks but facilitated mastery of the easier problems.

Harlow, H. F. Responses by rhesus monkeys to stimuli having multiple sign values. In Q. McNemar, & M. A. Merrill (Eds.), Studies in personality. New York: McGraw-Hill, 1947. Pp. 105-123.

Four monkeys solved m/s, non-m/s, sign-discriminated-antagonistist-position habits, and reversed S. D. A. P. H. problems. In the final stage the four separate tasks were presented in random sequence. The solution of the final response set was possible only if the animals attended to

four variables: (a) nature of the sample object, (b) nature of the food sign, (c) the positions of the choice objects, and (d) the identity or non-identity of the sample-object with either or both choice objects. All four monkeys learned the discriminations.

Harlow, H. F. Alternation of the matching-from-sample and non-matching-from-sample problems accompanied by appropriate cues. In R.H. Seashore (Ed.), Fields of psychology. New York: Holt, 1942. Pp. 191-196.

Two monkeys solved alternating matching and non-matching problems by using differential cues furnished by the color of the object tray and the reward or lack of reward beneath the sample object.

Harlow, H.F. Solution by rhesus monkeys of a problem involving the Weigl Principle using the matching-from-sample method. J. comp. Psychol., 1943, 36, 217-227.

Four rhesus monkeys were tested, using a m/s technique and a step-wise training procedure on a problem requiring differential responses to each of two visual attributes (color of form) of a single stimulus. Three of the four monkeys achieved criterion in 2400 to 4000 trials.

Hively, W. A framework for the analysis of early reading behavior. Amer. educ. Res. J., in press.

An operant analysis of reading with examples of m/s in reading tasks.

Hively, W. Programming stimuli in matching-to-sample. J. exp. Anal. Behav., 1962, 5, 279-298.

Pre-school and first-grade children were taught a series of progressively difficult m/s discriminations, using a fading technique and a "step-by-

step" program. The stimuli were geometric patterns of various sizes. The size of the steps and the amount of training on each discrimination affected the error rate. Theoretical and practical implications of these findings are discussed.

Itard, J. The wild boy of Aveyron. New York: Century, 1932.

An anecdotal report (dated 1799) about a feral boy. In order "to induce him to employ the simplest mental operations," he was trained first to match primary colors and simple geometric figures; then, he was taught to match printed letters of the alphabet with metal cutouts. Finally, he learned by m/s procedures "not only to use class nouns appropriately, but adjectival and verbal concepts as well."

Ladygina-Kots, Nadezhda N. Untersuchungen Über die Erkenntnissfähigkeiten des Schimpansen. Moscow: Zoopsychologischen Laboratorium des Museum Darwinianum, 1923, 453.

(See resumé below by Yerkes, 1925.)

Ladygina-Kots, Nadezhda N. Research on the intelligence of the chimpanzee using the choice by sample method. J. de Psychol., 1928, 25, 255-275.

A four-year-old chimpanzee was trained to match samples varying in color and form. In one phase of the experiment, S was required to discriminate by touch alone, the COs having been placed in a bag. The task was successfully learned.

Lanson, R. M. Effects of sodium pentobarbital on matching behavior in the pigeon. Paper read at Eastern Psychol. Ass., Philadelphia, March, 1965.

Nine pigeons were used to assess the effects of sodium pentobarbital on matching behavior. Decrements in accuracy, with increasing dose level, were shown to be correlated with the development of stimulus preferences and position preferences. "Color preference" seemed to be related more to the hue of the CO, then to the hue of the ST.

Nevin, J. A., Cumming, W. W., & Berryman, R. Ratio reinforcement of matching behavior. J. exp. Anal. Behav., 1963, 6, 149-154.

Three pigeons were exposed to various schedules of reinforcement in an m/s experiment. Accuracy was lower with fixed ratio (FR) than with continuous reinforcement (CRF). There was a high rate of errors immediately after reinforcement, with an increase in accuracy as the ratio progressed. Accuracy was high throughout the ratio on a variable ratio (VR) schedule of reinforcement.

Nissen, H. W., Blum, Josephine, & Blum, R. A. Analysis of matching behavior in chimpanzees. J. comp. physiol. Psychol., 1948, 41, 62-74.

Within 177 to 547 trials each of seven chimpanzees was taught a m/s discrimination using a variety of junk stimuli. When the animals were tested for generalization with 77 new object combinations, they responded with an accuracy ranging from 75 per cent to 95 per cent. Three possible mechanisms involved in m/s discrimination are discussed.

Riesen, A. H., & Nissen, H. W. Non-spatial delayed response by the matching technique. J. comp. Psychol., 1942, 34, 307-313.

Two slightly differing forms of non-spatial delayed response by the matching technique are described. After training a different "mediating" response to red and to green, gradually increasing delays were interposed. Chimpanzees gave performances significantly above chance, with delays up to 60 seconds.

Scott, Keith G. A comparison of similarity and oddity. J. exp. Child Psychol., 1964, 1, 123-134.

Similarity and oddity tasks were found to be equally difficult for retarded children.

Skinner, B. F. Are theories of learning necessary? Psychol. Rev., 1950, 57, 193-216. (Also in B. F. Skinner, Cumulative record. New York: Appleton-Century-Crofts, 1959.)

The article includes a brief mention of matching studies and the comment "the discriminative response of striking-red-after-being-stimulated-by-red is apparently no easier to establish than striking-red-after-being-stimulated-by-green."

Smith, D. E. P. Michigan successive discrimination reading program. Ann Arbor: Ann Arbor Publishers, 1964.

This comprehensive self-instructional program is an outstanding example of the m/s procedure applied to teaching.

Spaet, T., & Harlow, H. F. Problem solution by monkeys following bilateral removal of the prefrontal areas. J. exp. Psychol., 1943, 32, 424-434.

Two rhesus monkeys, following bilateral removal of the prefrontal areas, were tested on spatial and non-spatial delayed reaction problems, using the m/s technique. Four methods of presentation were used. The maximum delay intervals successfully bridged were 30 and 10 seconds for each The data indicate that monkeys, following bilateral removal of the prefrontal areas, can respond successfully to delayed reaction problems.

Tellier, M. The choice according to model in the macaque. Bull. Soc. roy. scient. de Liège, 1933, 3, 41-45.

A macaque monkey, four and one-half years old, learned a m/s discrimination very easily.

Verlain, L. De la connaissance chez le macaque. La substance. Bull. Soc. roy. scient. de Liège, 1935, 4, 239-242.

Verlain, L. Histoire naturelle de la connaissance chez le singe inférieur. Actualités Sci., 1936, 360, 1-58.

Verlain, L. La vision des formes chez le macaque. Mem. Acad. de Belgique Cl. des Sci., 1935, 14, 1-85.

This m/s study uses the textures of various substances (e.g., iron, zinc) as cues.

Weinstein, B. Matching-from-sample by rhesus monkeys and by children. J. comp. Psychol., 1941, 31, 195-213.

Two young children and two rhesus monkeys responded with relative success on m/s tasks involving non-spatial delayed responses, generalization to new stimulus objects, and the alternation of approach and avoidance behavior to the same choice object in successive trials.

Weinstein, B. Stanford-Binet intelligence test type performance by a rhesus monkey. Psychol. Bull., 1942, 39, 471-472.

The paper deals with a m/s color-form categorizing experiment similar to the one discussed below.

Weinstein, B. The evolution of symbolic behavior in rhesus monkeys: color categorizing. Unpublished doctoral dissertation, Univer. of Wisc., 1942.

Rhesus monkeys were tested for ability to categorize colors, using m/s and sorting problem techniques. An attempt was made to analyze experimentally the stimulus-response factors in behavior that can be characterized as "intelligent." It is concluded that rhesus monkeys can be trained in problem solving performances analogous to those in human intelligence tests.

Weinstein, B. The evolution of intelligent behavior in rhesus monkeys. Genet. Psychol. Monogr., 1945, 31, 3-48.

Monkeys, trained by the m/s technique, learned to sort out red objects when presented with a white triangle and blue objects when presented with an uncolored ellipse.

Weinstein, B. Delayed discriminative matching-from-sample by a rhesus monkey. Psychol. Bull., 1942, 39, 591. (Abstract)

In a study using the m/s method, Ss examined two STs, one movable (correct) and one fixed (incorrect). The Ss were first required to choose the "correct" ST. After a 15-60 second delay, the Ss matched the movable ST from five COs, and refrained from matching the fixed sample. Performance deteriorated with increasing delay.

Weinstein, B. Simple matching-from-sample. In R. H. Seashore (Ed.), Fields of psychology. New York: Holt, 1942. Pp. 179-184.

Rhesus monkeys were trained to perform a m/s task. Once the task was mastered, Ss succeeded in matching a large number of new sets of stimulus objects.

Weinstein, B. Discriminative matching-from-sample. In R. H. Seashore (Ed.), Fields of psychology. New York: Holt, 1942. Pp. 184-188.

A rhesus monkey solved a series of m/s problems involving five COs.

Weinstein, B. Discriminative delayed matching-from-sample. In R. H. Seashore (Ed.), Fields of psychology, New York: Holt, 1942. Pp. 188-191.

(See: Weinstein, B. Psychol. Bull., p. 591, above.)

Yerkes, A. W. Experiments with an infant chimpanzee. J. genet. Psychol., 1935, 46, 171-181.

A chimpanzee learned matching and non-matching tasks.

Yerkes, R. M., & Petrunkevitch, A. Studies of vision by Ladygina-Kots. J. comp. Psychol., 1925, 5, 98-108.

The article reviews the 1921 Russian Government Report of the work of Ladygina-Kots which was carried out in Moscow from 1914-1917. A five-year-old chimpanzee was trained to perform m/s and delayed matching tasks, using two- and three-dimensional colored plates and various geometric figures as stimuli. Reinforcement consisted of allowing S to play games and wrestle. The animal was able to match the sample with delays up to 15 secs.

Zimmerman, J., & Ferster, C. B. Intermittent punishment of S^{Δ} responding in matching-to-sample. J. exp. Anal. Behav., 1963, 6, 349-356.

Two pigeons were punished for incorrect m/s responses by a time out (TO) on FR and CRF schedules. Accuracy was examined as a function of TO duration and frequency. With TO durations of 10 sec to 1 min, accuracy increased as the frequency of TO presentation increased. With extremely long (10 min), or extremely short (1 sec) delays, accuracy was poor over the entire range of frequencies.

Zimmerman, J., & Baydan, N. F. Punishment of S^{Δ} responding of humans in conditional matching-to-sample by time out. J. exp. Anal. Behav., 1963, 6, 589-597.

In a m/s situation, human Ss were intermittently reinforced with money for accurate performance and punished for S^{Δ} responding by a TO on CRF and FR schedules. Accuracy increased with longer TOs and decreased when TOs were presented less frequently.

The Effects of "Step Size" in Shaping

Harlan L. Lane

James L. Kopp

Thomas Anderson

David M. Carlson

Recent studies show that the topography of human vocal responses, like that of other operant behavior, may be altered by selective reinforcement (Lane, 1964; Lane, Kopp & Sheppard, 1965). In order to manipulate vocal duration, these studies determine the initial frequency distribution of response durations (which is characteristically normal), then select a value of duration that is n standard deviations above the mean, and finally reinforce all responses with durations longer than this criterion value. As a consequence, the mean and variance of the distribution of response durations increase. Complementary methods and findings are associated with the shaping of shorter durations. In both cases, the alteration in vocal topography is still evident twenty-four hours after the experiment.

Perhaps the most important parameter of the technique of shaping is the initial probability of those responses that have been selected for exclusive reinforcement. Keller and Schoenfeld (1950) advise that we "pick out, in advance, a variation that has been of fairly frequent occurrence"-- but how frequent is "fairly"? If the initial probability of the responses selected for reinforcement is too low, there will be no shift in the distribution of response topographies. If the initial probability is quite high, on the other hand, the shaping process is secure, but large changes in topography can be accomplished only slowly and inefficiently. In quantitative terms, and taking vocal duration as a case in point, we want to know how many standard deviations (σ) away from the mean duration (M) of current responding should the reinforcement criterion (C) be set for effective and efficient shaping. The present study explores the effects of this parameter of shaping ($C-M/\sigma$), called the shaping index (Lane, 1964), on the control of response topography. In addition we are concerned with finding a simple way to predict the topography of responding at any moment in the shaping process, given the topography of the preceding responses.

Method

Each of 39 college students served as subjects from one to two hours while seated in front of a microphone, a small light, and a penny dispenser in a sound attenuating room. These instructions were read to each S:

> "In this experiment you can earn money simply by saying [u] whenever the light in front of you flashes. You will know that you have earned money when a penny falls into the tray in front of you. Occasionally the light will remain off for a period of time. When the light is off please do not respond. Please make no other sounds whatever besides [u] throughout the experiment."

The stimulus light was flashed automatically once every five sec. The penny dispenser was controlled by the experimenter, located in another room. The apparatus for response measurement, which duplicated that employed in the studies of vocal duration cited above, included, basically, a device for detecting the amplitude envelope of the speech waveform and an electronic timer (the timer measured the interval between the onset and offset of the waveform, and this value was displayed by a printer).

In the first phase of the experiment, reinforcement was delivered after each of 20 responses (CRF). The mean and standard deviation of those response durations were then computed and the criterion for the first shaping phase was set at a prescribed number of standard deviations above the mean. During shaping, those responses with duration greater than criterion were reinforced exclusively, until 10 consecutive reinforcements had been received. Then the phase was terminated and a new criterion for the next shaping phase was computed, using the mean duration and standard deviation of the last ten responses. Shaping continued in this fashion until three shaping phases or two hours had elapsed.

The number of standard deviations above the mean, at which the criteria for reinforcement were set, varied across <u>Ss</u> and across successive phases. The values employed were:

$$C-M/\sigma = .5,\ 1.0,\ 1.5,\ 2.0,\ 3.0.$$

The orders in which they were used are shown in Table 1. These values of the

---

Insert Table 1 about here

---

shaping index correspond respectively to the following probabilities of reinforcement (assuming a normal distribution of response durations):

.310, .160, .070, .020, and .005.

At the end of each session, a questionnaire was administered to <u>S</u>. The response to each question was completed before the next was presented:

1. What do you think was the point of this experiment?
2. Did you know what it was about your responding that earned pennies for you?
3. Did you know that you received pennies depending on the length or duration of your responses?

If the words "duration" and "length" never appeared in S's answers, and if the reply to questions 2. and 3. was negative, the questionnaire was scored negative.

Results and Discussion

Control of Response Topography

Approximately half of the Ss completed at least one shaping phase and did not tact the contingencies of reinforcement on the questionnaire. The mean duration and coefficient of variation (σ/M) in the phases completed by these 18 Ss are given in Table 2. These results are comparable to those

---

Insert Table 2 about here

---

found in the earlier experiments cited above, with the mean duration increasing almost twofold in three shaping phases, and σ/M constant at about .10. (The value of σ/M in the last shaping phase is a little higher than usual.)

The effect of step size on the efficiency of shaping is shown in Fig. 1

Insert Fig. 1 about here

---

in terms of the length of the first shaping phase when it was completed, and also the number of failures to complete it at all: the mean number of unreinforced responses and the percent of Ss who failed to shape are plotted as a function of the size of the shaping index. It is apparent that the lower the probability of occurrence of those responses selected for reinforcement, the longer the shaping process and the greater the danger of its unsuccessful outcome. The same relations were obtained, despite confounding with order effects, when the data from all Ss were pooled over all phases.

The redistribution of response topography brought about by shaping is shown in detail in Fig. 2 for a representative S. Responses with duration

Insert Fig. 2 about here

less than criterion were sorted into two classes of approximately equal size and their frequency of occurrence in each class was plotted cumulatively as a function of consecutive responses. The same procedure was followed for reinforced responses. The cumulative record for responses with short durations, well below criterion, is negatively accelerated, whereas that for responses with durations well above criterion is positively accelerated. Typically, the intermediate functions, for class intervals bounded by the reinforcement criterion, have two points of inflection, indicating that the probability of the corresponding topography first increases at the expense of responses with shorter duration, and then decreases, in favor of responses with longer duration.

Prediction of Response Topography

Several families of mathematical functions were examined as a means of predicting the topography of successive responses during shaping. In one quite general method employed for predicting the n+1st response duration from the preceding n response durations, all of the preceding durations were assigned weights and their mean computed (1); then the differences in duration between all pairs of consecutive responses were assigned weights and the weighted mean difference computed (2); finally, the sum of the weighted mean duration and the weighted mean was taken as the predicted duration (3):

$$(1)\quad \bar{X}_n = \sum_{j=1}^{n} (j+\gamma)^{\alpha} X_j \Big/ \sum_{j=1}^{n} (j+\gamma)^{\alpha}$$

(2) $\overline{\overline{\Delta X}}_n = \sum_{j=1}^{n} (j+\gamma)^{\alpha}(X_j - X_{j-1}) / \sum_{j=1}^{n} (j+\gamma)^{\alpha}$

(3) $\hat{X}_{n+1} = \overline{X}_n + \overline{\overline{\Delta X}}_n$

In this way, the topography of each response by a S (except his first) was predicted from those preceding, and a product-moment correlation between the predicted and obtained response durations was computed for each pair of test values of the parameters γ and α.

The initial computations, performed on an IBM 7090 computer, yielded the highest predictive accuracy with weighting functions that had strong positive acceleration; in effect, very early responses in the shaping series were not entering the prediction of later responses at all. Consequently, a third parameter was added, a "lookback" variable, L, which specified the number of responses preceding the n+1st entering in its prediction. When the computations were carried out with three parameters for all Ss and phases, it was found that the predictive accuracy was not appreciably better when γ and α were permitted to vary than when they were set equal to zero. Under the latter condition, each predicted response duration is simply the mean of the L preceding response durations plus the mean of the L-1 preceding increments in duration.

The accuracy of predictions using the mean duration plus mean difference is plotted in Fig. 3 as a function of the number of preceding responses (L) entering the prediction. Correlations between predicted and obtained response durations, averaged over Ss, are shown separately for shaping phases that

---

Insert Fig. 3 about here

---

were completed, for those that were not, and for CRF phases (where there was no progressive increase in response duration). It appears that, for those phases in which response topography was under experimental control, the prior four responses suffice to permit a prediction of the topography of the following response with moderate accuracy (r=.55). For those phases that did not reveal an orderly redistribution of response topography, however, it may actually detract from predictive accuracy to give weight to remote responses; this is especially true of the prediction of response topography during continuous reinforcement, where it appears that sequential effects in response topography are exclusively short term.

When the accuracy of the simple predictive formula described above was compared with the results of a linear multiple regression analysis of the same sets of data, it was found that linear regression accounted, on the average, for about the same amount of variance (16% to 25%). The estimated partial correlations between each n+1st predicted duration and its predecessors X_n, X_{n-1}, ... X_{n-9} considered singly were all close to $r = .10$, with the exception that the correlations between the duration of a response and that of its immediate predecessors were somewhat higher.

References

Lane, H. L. Differential reinforcement of vocal duration. J. exp. Anal. Behav., 1964, 7, 107-115.

Lane, H. L., Kopp, J. L., & Sheppard, W. C. Retention of vocal topography. J. exp. Anal. Behav., in press.

Keller, F. S., & Schoenfeld, W. N. Principles of psychology. New York: Appleton-Century-Crofts, 1950.

Table 1.

The values of the shaping index, $C-M/\sigma$, used in an experiment on step size and the control of vocal duration.

Group	Shaping Phase 1st	2nd	3rd
A (N=6)	.5	1.0	1.5
B (N=10)	1.0	1.5	2.0
C (N=5)	1.0	2.0	1.5
D (N=5)	1.5	1.0	2.0
E (N=6)	1.5	1.0	2.0
F (N=7)	2.0	3.0	4.0

Table 2.

The effects of CRF and shaping on the mean duration and variability of vocal responses. The data from phases completed first, second and third were pooled across Ss.

	Phase			
	CRF N=18	1st shaping N=18	2nd shaping N=12	3rd shaping N=8
Mean Duration	236	264	309	430
Percent increase from duration in CRF		19	31	82
Coefficient of variation	.10	.11	.11	.19

Figure Captions

Fig. 1. The effect of step size on the efficiency of shaping. The mean number of unreinforced responses during the first shaping phase and the percent of Ss who failed to complete that phase are shown as a function of the size of the shaping index: .5, 1.0, 1.5 and 2.0. The number of Ss in each case are, respectively, 4, 9, 6, and 4.

Fig. 2. Redistribution of response durations during shaping. Response durations above and below criterion were sorted into two approximately equal classes and their frequency of occurrence in each class was plotted as a function of consecutive responses. The records have been displaced from their common origin to make them more legible. These data are representative of those obtained from 18 Ss.

Fig. 3. Accuracy in predicting response topography during successful and unsuccessful shaping phases and CRF. The average correlation between obtained response durations and those predicted from a mean-duration-plus-mean-difference formula is plotted as a function of the number of preceding responses entering the prediction of the next.

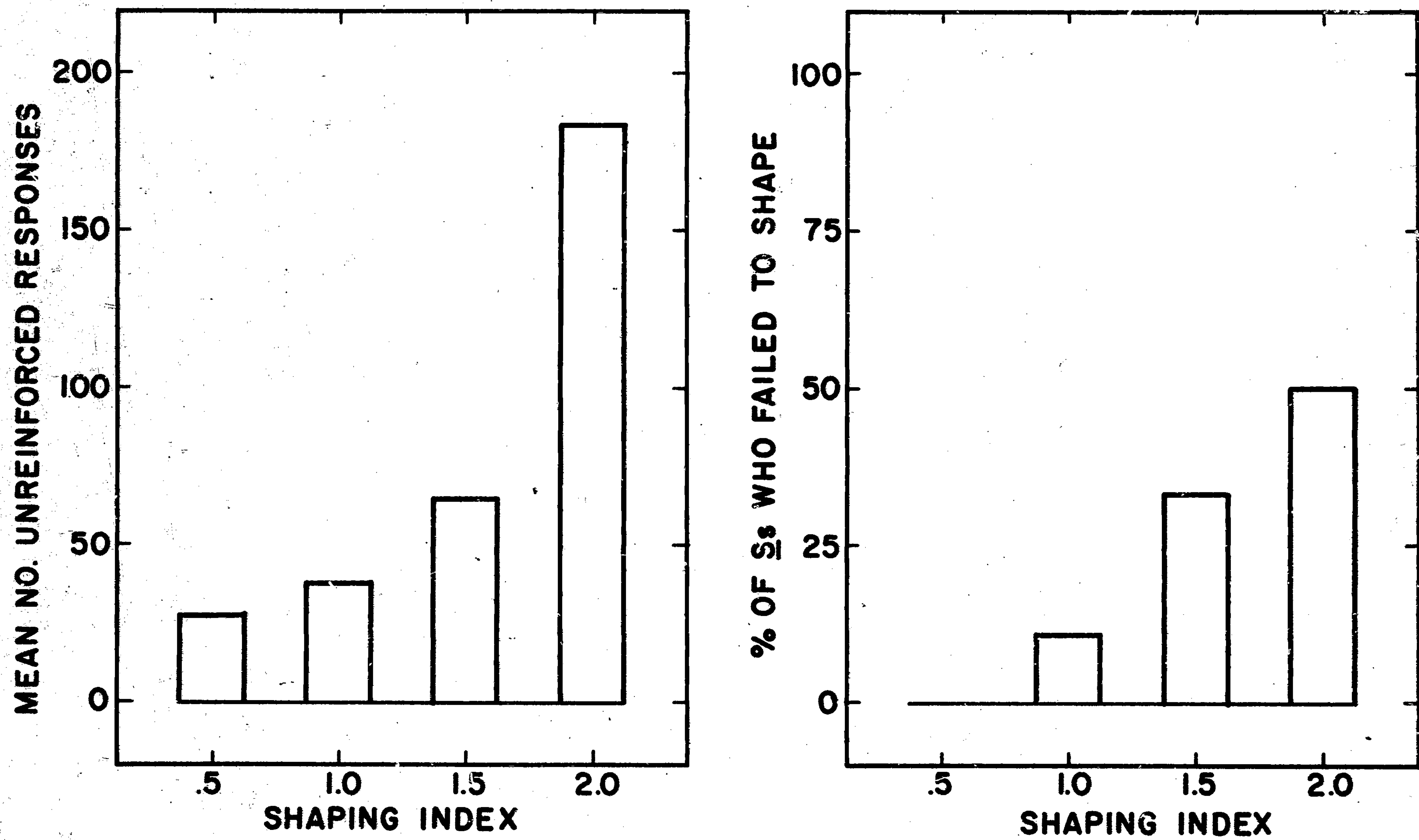

MEAN NO. UNREINFORCED RESPONSES
200
150
100
50
0
.5
1.0
1.5
2.0
SHAPING INDEX
% OF Ss WHO FAILED TO SHAPE
100
75
50
25
0
.5
1.0
1.5
2.0
SHAPING INDEX

Fig. 1

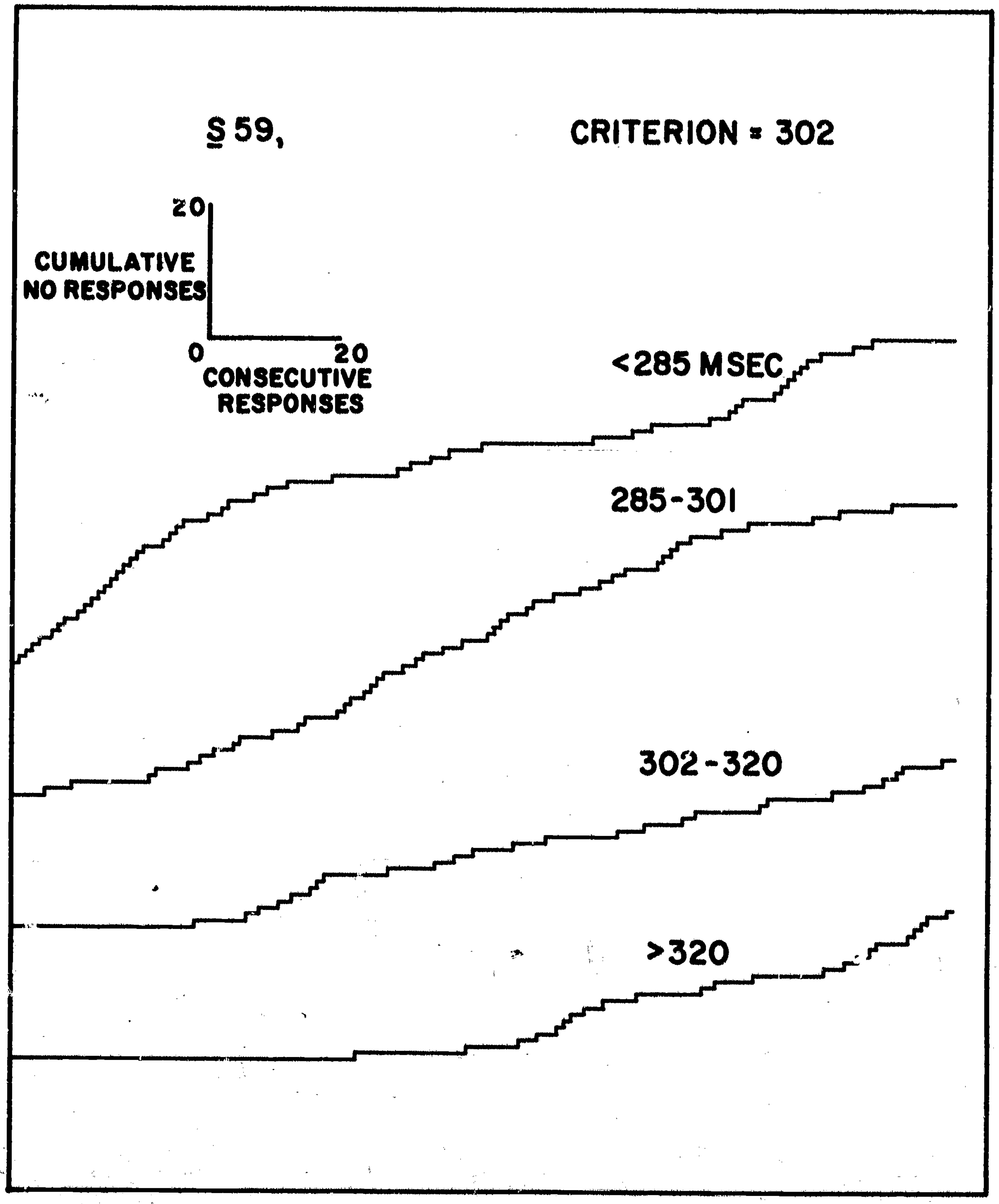
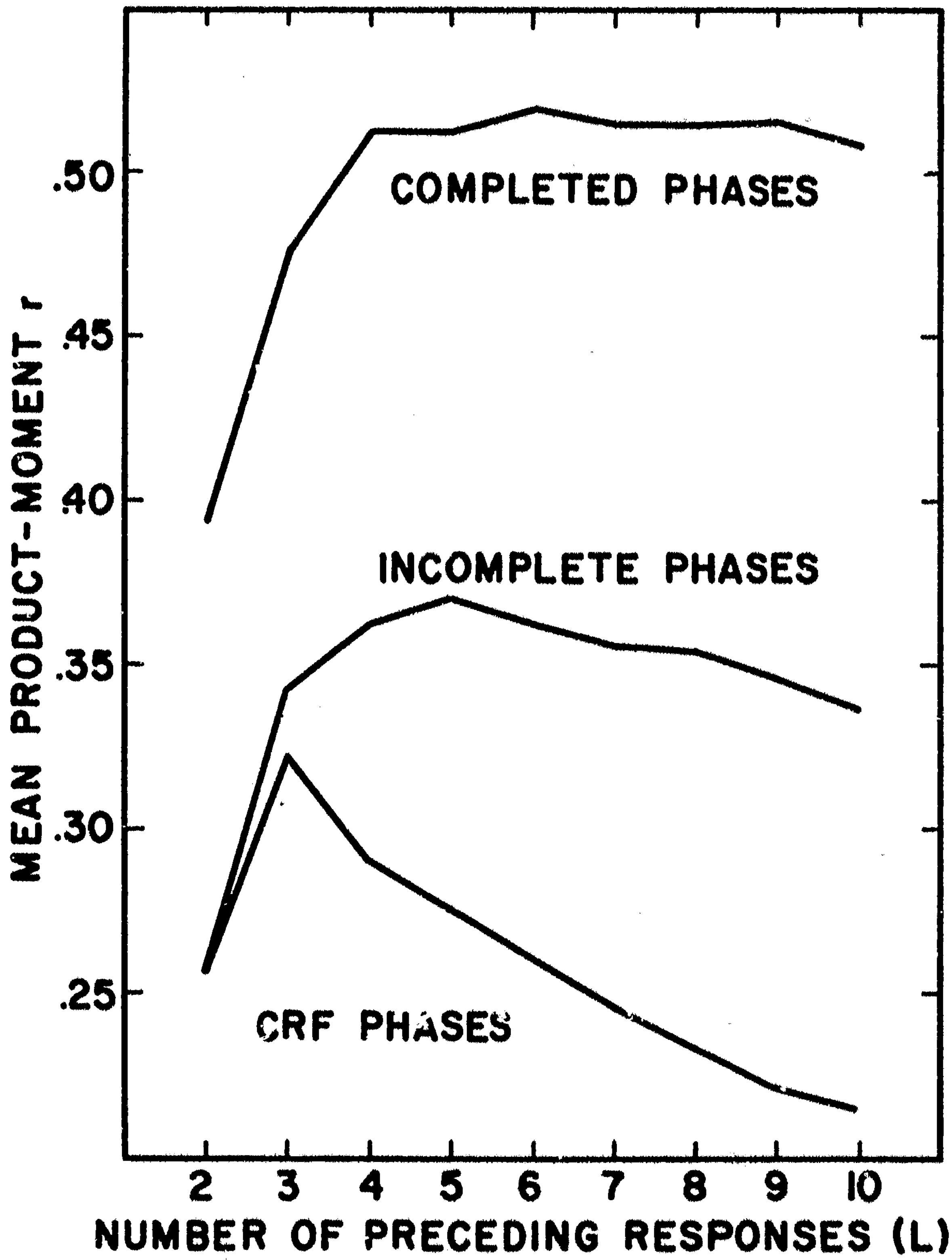

S 59,
CRITERION = 302
20
CUMULATIVE NO RESPONSES
0
20
CONSECUTIVE RESPONSES
<285 MSEC
285-301
302-320
>320

Fig. 2



.50
.45
.40
.35
.30
.25
MEAN PRODUCT-MOMENT r
COMPLETED PHASES
INCOMPLETE PHASES
CRF PHASES
2 3 4 5 6 7 8 9 10
NUMBER OF PRECEDING RESPONSES (L)

Fig. 3

Grammatical Form-Class in Word Associations of Educable Mentally-Retarded and Normal Children

Melvyn I. Semmel, Loren S. Barritt

Stanley W. Bennett, and Charles A. Perfetti

Studies of retarded language-functioning may permit greater understanding of normal language development (Carroll, 1963, 1964; Wepman, 1964). It is argued that the language changes which occur in normal children are often so rapid that intensive study of them is extremely difficult. Retarded children are thought to go through the same early stages in developing language abilities as normal children, but the stages are somewhat protracted. Hence, Carroll suggests that, "The study of the slow-motion language development of mental defectives might throw much light on how grammar develops and how various grammatical phenomena rank themselves in conceptual difficulty" (Carroll, 1964, p. 70).

Recent reviews of the research literature on linguistic problems of the mentally-retarded (McCarthy, 1964; Smith, 1962; Spradlin, 1963) suggest the rapidly-growing interest in the subject. Research efforts have primarily concentrated on the acquisition of semantic elements in retarded-language functioning. Little attention has been given to the processes by which educable mentally-retarded (EMR) children acquire the syntactic features of the language. The paucity of information on the grammatical habits of EMR children is dramatically demonstrated by the absence of a single reference to the topic in comprehensive reviews (McCarthy, 1964; Smith, 1962; Spradlin, 1963), and by the absence of references in a bibliography of world literature on mental retardation (Heber, Simpson, Gibson, & Milligan, 1963) which covers the 23-year period from 1940 through 1963.

Brown and Berko (1960) contend that results of word-association tests yield important information pertinent to the development of syntax in normal children. Brown (1957) found that children are able, by age three to five years, to decipher the grammatical class of novel stimuli. A subsequent stage of development occurs in which the child progressively moves from a purely sequential approach to the relationship betwsen verbal stimuli to a stage in which associations are made between words with similar "privileges of occurrence" (grammatical form-class). It is hypothesized that as children acquire the syntactic elements of language, syntactic equivalences of words emerge as progressively important determiners of responses in word-association tasks. The developmental process of shifting from heterogeneous (sequential) responses toward homogeneous (same form-class) responses "is a manifestation of this great step forward into syntactic operations" (Brown & Berko, 1960).

Ervin (1961), Brown and Berko (1960), and Entwisle, Forsyth, and Muuss (1964) demonstrated a relationship between chronological age and the grammatical form-class of free-associate responses. If the response word and the stimulus word do not belong to the same form-class, and/or if the response word can be used sequentially with the stimulus word, then the response is generally categorized as syntagmatic (Deese, 1962; Ervin, 1961). A response of the same form-class as the stimulus word is assigned to the paradigmatic category. Items in the syntagmatic category have been referred to by some authors as heterogeneous responses; while paradigmatic items are designated homogeneous (see Brown & Berko, 1960). Ervin (1961), Brown and Berko (1960), and Entwisle et al. (1964) revealed that young children (ages 5 to 7) tend to give many more syntagmatic than paradigmatic responses in word-association tasks. As the chronological age (CA) of Ss increased, there was a progressive trend toward increased paradigmatic responding with a corresponding decrease in syntagmatic associates.

The frequency of usage of a particular stimulus word has been related to paradigmatic responding in word-association tasks. High-frequency adjectives produce a greater number of paradigmatic utterances than low-frequency adjectives (Deese, 1962). The grammatical form-class of the stimulus word appears to be another important variable in such tasks (see Glanzer, 1962). Deese (1962) reported that noun stimuli were generally followed by paradigmatic associates, whereas adjectives and verbs tended to be followed by an equal number of responses from both categories; and adverb stimuli were predominantly followed by syntagmatic responses. Entwisle et al. (1964) reported a convergence toward an equal number of paradigmatic responses to nouns, verbs, and adjectives, as the chronological ages of Ss increased. It should be noted that Deese (1962) systematically varied the frequency of usage of stimuli presented, whereas Entsiwle et al. (1964) used only high-frequency noun, verb, and adjective stimuli.

The purpose of this study was to compare the word associations of educable mentally-retarded children, both institutionalized and non-institutionalized, with normal children having comparable mental age or chronological age. Responses were analyzed using syntactic, rather than semantic, criteria for the purpose of exploring the relationsbetween mental retardation and the level of syntactic operations in school-age children.

Method

Subjects. Four samples of 20 Ss each were selected for inclusion in this study. There were two EMR subgroups: Institutionalized (I-R) and Public School (PS-R). All retarded Ss were drawn by stratified random sampling on the sex and race variables. The retarded Ss were randomly selected from a population with CA range of 10 through 14 years, and an IQ range of 60 through 80.

Two samples of normal children were selected according to the same procedures used with retarded subgroups, except that one subgroup of normal Ss was selected from a population having a mental age (MA) range comparable to the EMR samples (MA range 6-3 to 10-7)--equal MA normal group (MA-N). A second sample of normal Ss (CA-N) was drawn from the same CA range as the two retarded subgroups. The normal Ss and the PS-R samples were drawn from the Ypsilanti and Wayne Public Schools. The socio-economic status of Ss was not specifically controlled; the Ss in all four subgroups were assumed to come from relatively low socio-economic families. Table 1 is a summary of the characteristics of the subgroups. There were 12 Caucasian and 8 Negro Ss in each sample. The

---

Insert Table 1 about here

I-R sample contained 9 males and 11 females; the remaining three subgroups contained 10 males and 10 females.

Selection of stimuli. Forty stimuli from six grammatical form-classes were selected from the Mein and O'Connor (1960) list of words uttered most commonly by severely retarded children. All the stimuli that were selected appeared among the 500 words used most commonly by the retarded population and they predominantly fall within one of six grammatical form-classes (noun, verb, adjective, adverb, pronoun, and preposition). All the stimuli also appear in the word-association norms reported by Palermo and Jenkins (1963); 38 stimulus words have AA frequency on the Thorndike-Lorge general word count, the remaining two stimuli have A frequency. Table 2 presents a list of the stimuli arranged by grammatical form-class. Webster's New Collegiate Dictionary was used to

---

Insert Table 2 about here

---

corroborate the assignment of each stimulus word to an appropriate form-class.

Procedure. Stimulus words were typed on 5" x 8" unlined index cards (24-point type). The forty index cards were randomized by hand shuffling prior to presentation to S. Each S was introduced to the task as follows:

"We are going to play some word games today. Now if you're ready, I'll tell you the rules for the game.

"In this game I'll show you a word on each one of these cards. I'll say the word to you. The idea of the game is for you to say the first word you think of when I say the word to you. You should say just one word--not more than one."

The S was then presented three sample stimuli to assure comprehension of the task. E presented the first sample word by saying:

"Now let's see if you understand the rules of the game--What's the first word you think of when I say 'cow'?"

Any single-word response was reinforced by E with "Good!" or "Fine!" This general procedure was followed in presenting the remaining two sample words (draw and cold). The test stimuli immediately followed S's response to the last sample word.

Coding procedures. Figure 1 is a flow diagram of the model developed to make judgments for categorizing the free associate responses of Ss. Each response word was evaluated independently by three judges following the same procedure. Each judge first determined whether or not the response word was a repetition of the stimulus word. If the response was a repetition of the stimulus word it was placed in the category labelled repetition of stimulus (Fig. 1). If the response was judged not to be a repetition of the stimulus

---

Insert Figure 1 about here

---

word, the judge then decided whether the response could appear in an immediately contiguous relationship with the stimulus word (sequential order). A response judged sequential was next judged homogeneous (of the same form-class as the stimulus word) or heterogeneous (of different form-class as the stimulus word). Responses relegated to the sequential-homogeneous category met these criteria: they could occur contiguous to the stimulus word, and had the same form-class as that word (e.g., stimulus: was; response: gone). A sequential response that differed from the stimulus word in form-class was assigned to the sequential-heterogeneous category (e.g., stimulus: yellow; response: tree). It should be noted that the sequential-heterogeneous and sequential-homogeneous categories are two of the elements of the syntagmatic category used in previous research (Deese, 1962; Ervin, 1961).

A response judged non-sequential (i.e., rarely occuring contiguous to the stimulus word) was next judged with regard to form-class. Thus, non-sequential responses of the same form-class as the stimulus word were classified as non-sequential-homogeneous (e.g., stimulus: table; response: chair). Non-sequential responses of different form-class than their respective stimulus words were placed in the non-sequential-heterogeneous category (e.g., stimulus: red; response: sit). The former category is equivalent to the paradigmatic category used by other researchers (Ervin, 1961; Deese, 1962), whereas the latter is the third element of the syntagmatic category (Ervin, Deese).

Clang response (rhyming response words) and minor modifications of the stimulus word were included in the coding process and identified within the five categories for separate analysis. Minor stimulus modifications were defined as responses in which there was either a substitution, deletion, or addition of one letter to the stimulus (e.g., bed--bad), or the addition of a suffix to the to the stimulus (e.g., go--going). Combinations of sounds judged not to be English words were eliminated from the analysis.

Two psychology graduate students (males) and one graduate student majoring in speech pathology (female) served as judges for coding responses. Approximately four hours of training were provided, using pilot data for coding. An analysis of the variance between judges in their assignments to each of the response categories appears in Table 3 (repetitions of the stimulus were omitted due to the relative infrequency of responses in this category.) Since the F ratios

---

Insert Table 3 about here

---

obtained from comparing judges across the four categories did not achieve significance, it was assumed that inter-judge variance was not a significant variable in interpreting the pooled subgroup data. Each S's score for each category was the mean number of judgments made for that category by the three judges. Hence, if two responses were considered sequential-heterogeneous by one judge, and three responses assigned to the same category by the second and third judges, the S's score for the sequential-heterogeneous category would be 2+3+3/3=2.67.

Preliminary Results

Table 4 presents means and standard deviations of the four major response

---

Insert Table 4 about here

---

categories across the four subgroups. Figure 2 is a graphic representation of these data.

Insert Figure 2 about here

The NS-Hm (paradigmatic) responses were arranged for a two-way analysis of variance in which differences as a function of the form-class of the stimulus words and subgroups were analyzed. Table 5 is a summary of this analysis.

Insert Table 5 about here

The results indicate the absence of a reliable interaction between subgroups and the form-classes of stimulus words presented to Ss. However, both main effects of form-classes and subgroups are reliable ($p < .01$).

The analysis of form-class differences appears in Table 6. The Table shows critical differences between respective mean per cent responses across subgroups for each of the six form-classes. A significantly greater percentage of NS-Hm responses was associated with nouns when compared with adjectives, adverbs, prepositions ($p < .01$) and verbs ($p < .05$). NS-Hm responses to prepositions were proportionately fewer than to verbs, adjectives, adverbs ($p < .01$).

Insert Table 6 about here

Table 7 presents a comparison of the mean percentage NS-Hm responses for the four subgroups. The CA-N Ss produced a significantly greater mean per cent

Insert Table 7 about here

NS-Hm associations to stimulus words than the other three subgroups ($p < .01$). The I-R subgroup mean was significantly lower than that for the PS-R and MA-N subgroups ($p < .01$). It should be noted that the mean difference between the public school retarded Ss and the equal MA normal Ss were not reliable.

Since the NS-Ht, S-Hm, S-Ht responses in the categories have been scored in previous research as syntagmatic (Deese, 1962), the responses falling in these categories were pooled for the purpose of obtaining a comparable dichotomy between paradigmatic and syntagmatic associations. Figure 3 presents a histogram showing the mean percentage of paradigmatic and syntagmatic associations produced by each of the four subgroups. The greatest discrepancy between the two

Insert Figure 3 about here

categories is revealed in the I-R group, while the smallest difference is reflected in the CA-N group. The MA-N and PS-R groups are approximately equal in the discrepancy between paradigmatic and syntagmatic responding.

Rhyming responses (Clangs) were analyzed independently across the four major response categories (viz., NS-Hm, NS-Ht, S-Hm, S-Ht). Table 8 summarizes the results of a simple analysis of variance in which the significance of sub-

Insert Table 8 about here

group mean differences for clang associations was determined. The main effects of subgroups were significant ($p < .01$). Table 9 summarizes the analysis of

---

Insert Table 9 about here

---

simple effects. The MA-N Ss produced significantly more clang responses than the I-R or CA-N subgroups. The PS-R subgroup mean was significantly greater than the mean for the CA-N subgroup ($p < .01$).

Analysis of variance for the stimulus-modification category showed no significant differences between subgroups ($F=1.95$/df 3, 156/ $p > .05$).

A simple analysis of variance was conducted to determine the reliability of differences between Negro and Caucasian Ss on mean percent NS-Hm responses. The resulting F ratio was not significant ($F=.91$/df 1,78/ $p > .05$).

A one-way analysis of variance was used to compare males and females on mean percentage of NS-Hm responses. Female Ss responded with significantly more NS-Hm responses than male Ss ($F=9.44$/df 1,78/$p < .01$). The mean percentage of NS-Hm responses for the male Ss was 25.64 (SD=19.44) whereas the mean for females was 38.61 (SD=36.28). Figure 4 presents these comparisons graphically for the four subgroups. While the statistical significance of the interaction has not been tested, inspection of the function suggests that the effects of differences in sex are qualified by the interaction of the subgroup and sex variables.

Insert Figure 4 about here

Correlations between the percentage of NS-Hm responses with chronological age and mental age for the combined retarded subgroups (I-R and PS-R) and the normal groups (CA-N and MA-N) were computed separately. The retarded group showed a product moment correlation of .31 between CA and NS-Hm responses, and r= .19 between NS-Hm and MA. The normal group shower a correlation of .55 for CA and NS-Hm and .47 between MA and NS-Hm. The correlation of CA with NS-Hm responses for all Ss (pooled across subgroups) was .21, while the correlation with MA was .45. The relations between CA and MA, respectively, with NS-Hm responses of retarded Ss were not significant (p > .05). The remaining correlations (Normal and Total groups) were significant at < .05 level.

A Type-Token Ratio (TTR) was calculated for each S by dividing the number of different word responses by the total number of responses to the 40 stimuli. Table 10 presents the mean TTR's for the four subgroups and also the correlations between the percentage of NS-Hm responses and TTR for each subgroup.

Insert Table 10 about here

A variance analysis did not confirm significant differences between subgroup mean TTR's (F= .22/df 3, 76/ p > .05). All of the correlations between NS-Hm and TTR's are significant for all the subgroups at < .05 level.

Preliminary Conclusions

1. Category by subgroup interaction was not tested because of assumptions of analysis of variance, but is shown in Fig. 2. It is obvious that S-Hm and NS-Ht responses are relatively equal in number across subgroups, but an interaction becomes apparent in the analysis of the remaining two categories (NS-Hm and S-Ht).

2. Although the two retarded groups were matched on MA, the I-R group functioned significantly below the PS-R group in NS-Hm responding. The similarity between the PS-R and MA-N groups in their responses to stimuli of

the six form classes lends support to Carroll's suggestion (1964) that retarded children's language development may be retarded quantitatively but not qualitatively. Therefore, it may be possible to study language development in normal children by analysing the protracted verbal skills of retardates.

3. The difference found in mean percentage of NS-Hm (paradigmatic) responses between the PS-R and I-R groups might be due to a selection factor in the I-R population (e.g., more brain-injured children would be anticipated).

4. Since there were unequal numbers of stimuli for the six form classes used, the differences in means must be interpreted with some caution. The ranking of highest to lowest number of NS-Hm responses is: nouns, pronouns, verbs, adjectives, adverbs, and prepositions. Glanzer (1962), in a free-association task, measured latency of response to seven form classes and found the following ordering (from smallest to largest latency): nouns, adjectives, pronouns, verbs, prepositions, adverbs, and conjunctions. Although this order is not precisely the same as the mean percentages of NS-Hm responses, the obtained order from least to most NS-Hm responses is similar to Glanzer's latency measures.

5. In regard to clang responses, it is difficult to explain why the I-R group is not different from the CA-N group. Possibly qualitative differences explain this result (i.e., the I-R subjects apparently are not functioning like the PS-R Ss).

6. When Ss were matched on CA, MA, and IQ, Negro Ss were not inferior in the incidence of NS-Hm responses.

7. The sex variable is apparently complex. Our data suggest that in very young children sex does not appear to be an important variable in NS-Hm responding, but that by the sixth grade in school, females give significantly more of these responses. In the retarded groups, both of which are equivalent

to the CA-N group in chronological age, females consistently score above the males. Therefore, this study strongly suggests that chronological age is an important variable in the obtained differences between males and females in paradigmatic responding. When comparing across subgroups, it can be noted that PS-R females give about as many NS-Hm responses (to the high frequency stimuli employed in this study) as do the CA-N males.

8. Although the mean TTR's for individuals do not differ significantly between subgroups, the correlation of TTR's with NS-Hm responding is moderately high in each of the subgroups. Thus it would appear that the variability of response to the stimuli is related to paradigmatic responding to about an equal degree in <u>all</u> the subgroups, but that this does not account for the differences obtained between the subgroups in the incidence of NS-Hm responses.

9. Correlations obtained between incidence of paradigmatic (NS-Hm) responding and mental and chronological age, both within the retarded and normal samples and in all the groups combined, indicate that, besides chronological age, mental age may be an important predictor of NS-Hm responses.

References

Brown, R. Linguistics determinism and the part of speech. J. abnorm. soc. Psychol., 1957, 55, 1-5.

Brown, R., & Berko, Jean. Word association and the acquisition of grammar. Child Develpm., 1960, 31, 1-14.

Carroll, J.B. Psycholinguistics in the study of mental retardation. In Schiefelbusch, R.L., & Smith, J.O. (Eds.) Research in speech and hearing for mentally retarded children. Lawrence, Kans.: Univ.of Kansas, 1963.

Carroll, J.B. Language and thought. Englewood Cliffs, N.J.: Prentice-Hall, 1964.

Deese, J. Form class and the determinants of association. J. verb. Learn. verb. Behav., 1962, 1, 79-84

Entwisle, Doris, Forsyth, D., & Muuss, R. The syntactic-paradigmatic shift in children's word associations. J. verb. Learn. verb. Behav.. 1964, 3, 19-29.

Ervin, Susan. Changes with age in the verbal determinants of word-association. Amer. J. Psychol., 1961, 74, 361-372.

Glanzer, M. Grammatical category: A rote learning and word association analysis. J. verb. Learn. verb. Behav., 1962, 1, 31-41.

Heber, R., Simpson, Nancy, Gibson, Ann, & Milligan, G. Bibliography of world literature on mental retardation. Wash., D.C.: Public Health Service, 1963.

McCarthy, J.J. Linguistic problems of the retarded. In Scheerenberger, R.C. (Ed.) Mental retardation abstracts. Wash., D.C.: NIMH, 1964.

Mein, R., & O'Connor, N. A study of the oral vocabularies of severely sub-normal patients. J. ment. Defic. Res., 1960, 4, 130-143.

Palermo, D.S., & Jenkins, J.J. Word association norms, grade school through college. Minneapolis: Univ. of Minnesota Press, 1964.

Spradlin, J.E. Language and communication of mental defectives. In Ellis, N.R. (Ed.) Handbook of mental deficiency. New York: McGraw-Hill, 1963.

Smith, J.O. Speech and language of the retarded. Train. Sch. Bull., 1962, 58, 111-123.

Thorndike, E.L., & Lorge, I. The teachers' word book of 30,000 words. New York: Bur. Pub., Teachers College, Columbia University, 1944.

Webster's New Collegiate Dictionary. Cambridge, Mass.: Riverside Press, 1961.

Wepman, J.M. Conference on planning research on the mentally retarded child. In Schiefelbusch, R.L., & Smith, J.O. (Eds.) Research in speech and hearing for mentally retarded children. Lawrence, Kans.: Univ. of Kansas, 1963.

Table 1

Characteristics of the Four Subgroups

(N=20 each subgroup)

Variable		Subgroup			
		I-R	PS-R	MA-N	CA-N
CA*	Mean	141.15	141.65	97.00	140.30
	SD	14.01	13.03	12.96	11.95
	Range	120-164	120-160	76-123	123-165
MA*	Mean	98.95	98.00	98.55	143.15
	SD	12.80	10.34	13.30	12.65
	Range	82-128	77-117	75-127	119-161
IQ	Mean	70.15	69.60	102.25	102.30
	SD	5.34	5.79	6.34	7.56
	Range	61-78	60-80	90-118	92-126

*In months

Table 2

Free-Associate Word Stimuli Arranged by Grammatical Form-Class

Nouns (n=10)	Verbs (n=10)	Adjectives (n=6)	Adverbs (n=5)	Pro-nouns (n=5)	Prepo-sitions (n=4)
bed	tell	sweet	there	you	for
window	was	read	here	it	with
bread (A)[1]	see	a	always	we	of
children	sit	the	where	my	at
table	get	yellow	very	he	
gun (A)	have	black			
door	is				
boy	come				
dog	take				
car	go				

[1]Word stimuli classified as A frequency on the Thorndike-Lorge word list. All other word stimuli are classified as AA frequency on the T-L list.

Table 3

Summary of Analysis of Variance

Between Judges for Scoring Categories

Category	SS Between	df	MS	F
Sequential-Homogeneous	22.82	2	11.41	0.7973*
Sequential-Heterogeneous	196.54	2	98.27	0.5102*
Non-Sequential-Homogeneous	30.81	2	15.45	0.0716*
Non-Sequential-Heterogeneous	238.65	2	119.33	1.6986*

*All F ratios are not significant ($p > .05$ level)

Table 4

Mean Per cent Responses within Categories

Across the Four Subgroups

(N=20 each subgroup)

Subgroups		Paradigmatic	Syntagmatic		
		NS-Hm	NS-Ht	S-Hm	S-Ht
CA-N	Mean	48.14	7.19	2.21	42.29
	SD	15.60	4.33	1.65	16.19
MA-N	Mean	30.91	13.30	1.84	53.82
	SD	18.60	6.74	1.67	20.44
PS-R	Mean	28.54	13.39	1.49	56.43
	SD	16.97	5.18	1.64	17.59
I-R	Mean	21.57	12.51	1.65	64.02
	SD	17.77	8.82	1.47	19.02

Category Abbreviations:

NS-Hm -- Non-Sequential Homogeneous

NS-Ht -- Non-Sequential Heterogeneous

S-Hm -- Sequential Homogeneous

S-Ht -- Sequential Heterogeneous

Subgroup Abbreviations:

CA-N -- Equal CA normal Ss

MA-N -- Equal MA normal Ss

PS-R -- Public School retarded Ss

I-R -- Institutionalized retarded Ss

Table 5

Summary of Analysis of Variance for Subgroups

and Form-Class of Stimuli

Source	SS	df	MS	F
Form-Classes (A)	56320.35	5	11264.07	22.17**
Subgroups (B)	47209.32	3	15736.44	30.98**
A x B (Interaction)	1[illegible]764.77	15	784.32	1.54
Error	231648.52	456	508.00	
Total	346942.96	479		

** Significant at the .01 level.

Table 6

Critical Differences between Mean Percentage

NS-Hm Responses for Form-Classes of

Stimulus Words

	Nouns	Pronouns	Verbs	Adj.	Adv.	Prepositions
Mean %	42.62	34.29	33.17	29.61	28.96	7.24
Pronouns	8.33					
Verbs	9.45*	1.12				
Adj.	12.01**	4.68	3.56			
Adv.	13.66**	5.33	4.21	.65		
Prep.	35.38**	27.05**	25.93**	22.37**	21.72**	

*Critical difference 8.92 significant at .05 level

**Critical difference 9.75 significant at .01 level

Table 7

Critical Differences between Mean Percentage

NS-Hm Responses for Subgroups

	CA Normal (CA-N)	MA Normal (MA-N)	Pub. Sch. Ret. (PS-R)	Inst. Ret. (I-R)
Mean %	48.14	30.91	28.54	21.57
MA-N	17.23**			
PS-R	19.60**	2.37		
I-R	26.57**	9.34**	6.97**	

** Critical difference 6.93 significant at .01 level

Table 8

Summary of Analysis of Variance of Clang

Responses for Subgroups

Source	SS	df	MS	F
Between (Subgroups)	167.80	3	55.93	12.65**
Within (Subgroups)	711.30	156	4.56	
Total	879.10	159		

**Significant at the .01 level.

Table 9

Critical Differences between Mean Number of

Clang Responses per Stimulus Word

	MA-N	PS-R	I-R	CA-N
Mean #	3.08	2.48	1.32	.42
PS-R	.60			
I-R	1.76**	1.16		
CA-N	2.66**	2.06**	.90	

** Critical difference $\geq$1.62 significant at < .01 level

Table 10

Mean TTR's and Correlations for Subgroups

	CA-N	MA-N	PS-R	I-R	Total
Mean	.86	.85	.83	.85	
SD	.11	.11	.01	.13	
r (NS-Hm and TTR)	.63	.46	.49	.44	.52

Differences between subgroup _r's_ not significant ($p > .05$)

Figure Captions

Fig. 1. Stages in the prescribed routine for assigning word associations by the Ss to formal categories.

Fig. 2. The percent of all word associations, emitted by the 20 Ss in each of the four groups, that were assigned to each of the following categories: Sequential-homogeneous (S-Hm), Sequential-Heterogeneous (S-Ht), Non-sequential Homogeneous (NS-Hm), Non-sequential Heterogeneous (NS-Ht). Retarded Ss in public schools (PS-R) and in institutions (I-R), as well as normal Ss with comparable chronological age (CA-N) or mental age (MA-N) were employed.

Fig. 3. The percent of syntagmatic and paradigmatic word associations given by the 20 Ss in each of the four subgroups.

Fig. 4. The difference between male and female Ss in each of the four subgroups in the relative frequency of paradigmatic free associates.

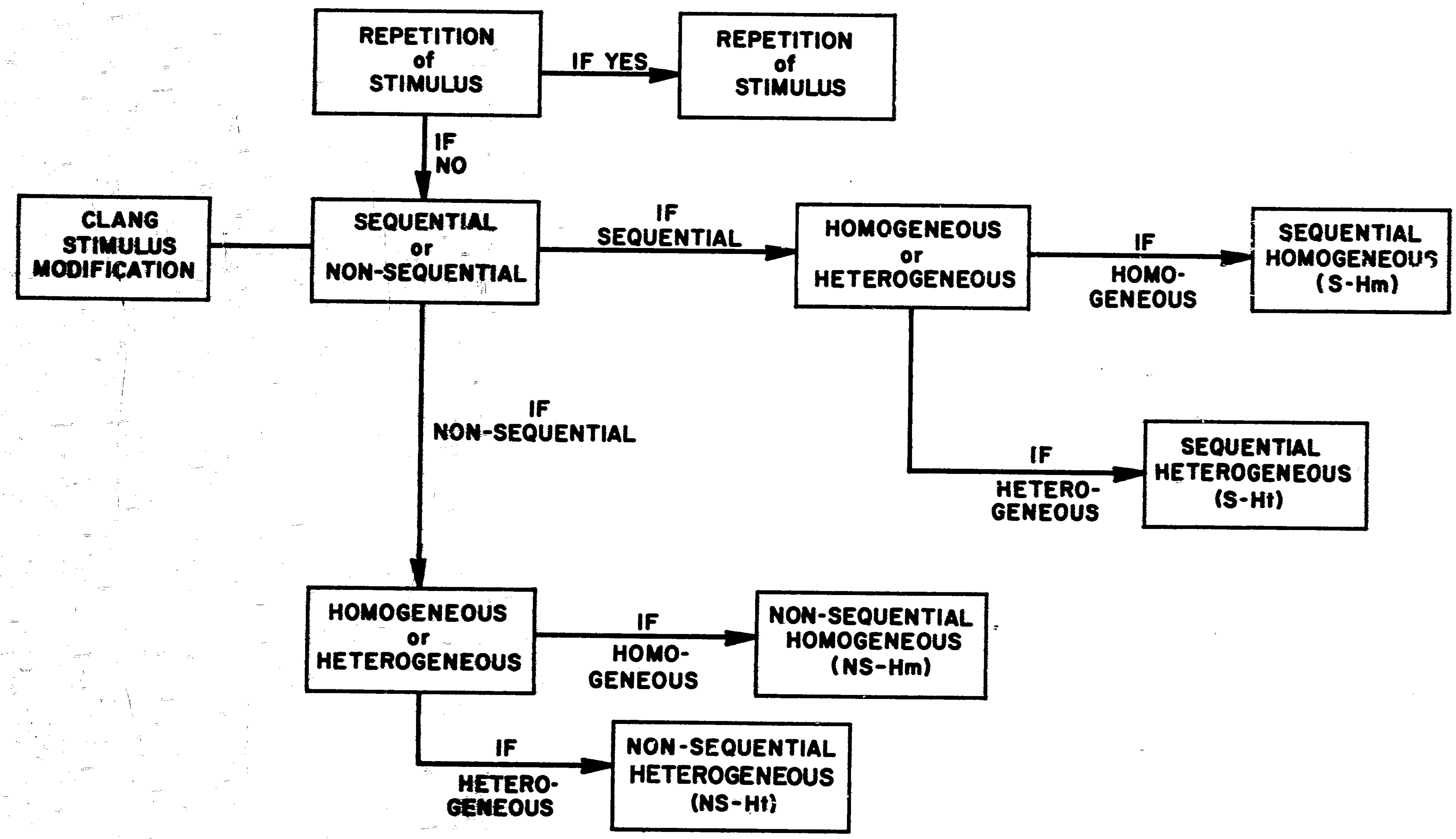

REPETITION of STIMULUS
IF YES
REPETITION of STIMULUS
IF NO
CLANG STIMULUS MODIFICATION
SEQUENTIAL or NON-SEQUENTIAL
IF SEQUENTIAL
HOMOGENEOUS or HETEROGENEOUS
IF HOMO-GENEOUS
SEQUENTIAL HOMOGENEOUS (S-Hm)
IF HETERO-GENEOUS
SEQUENTIAL HETEROGENEOUS (S-Ht)
IF NON-SEQUENTIAL
HOMOGENEOUS or HETEROGENEOUS
IF HOMO-GENEOUS
NON-SEQUENTIAL HOMOGENEOUS (NS-Hm)
IF HETERO-GENEOUS
NON-SEQUENTIAL HETEROGENEOUS (NS-Ht)

Fig. 1

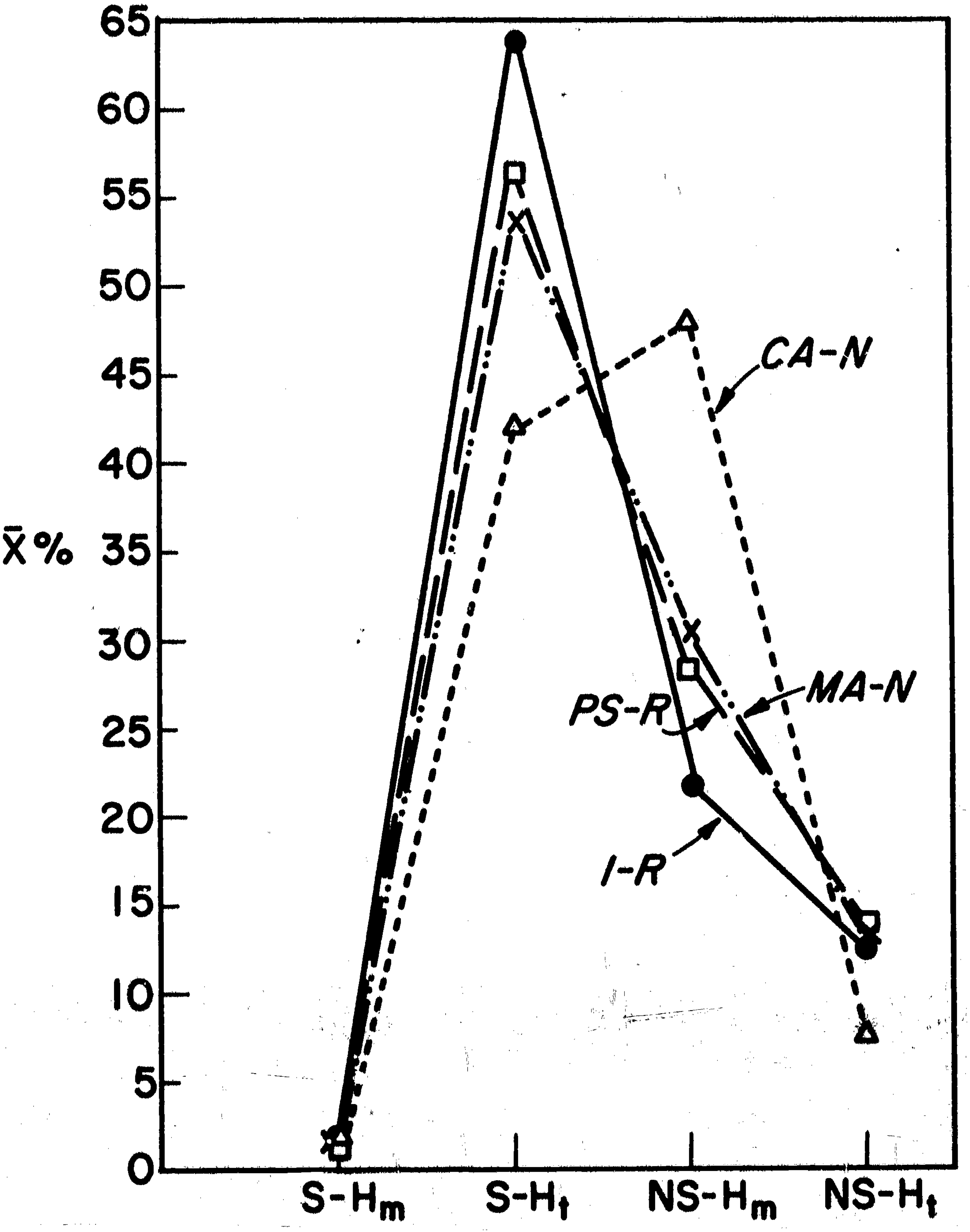

65
60
55
50
45
40
X̄%
35
30
25
20
15
10
5
0
S-Hm
S-Ht
NS-Hm
NS-Ht
CA-N
MA-N
PS-R
I-R

Fig. 2

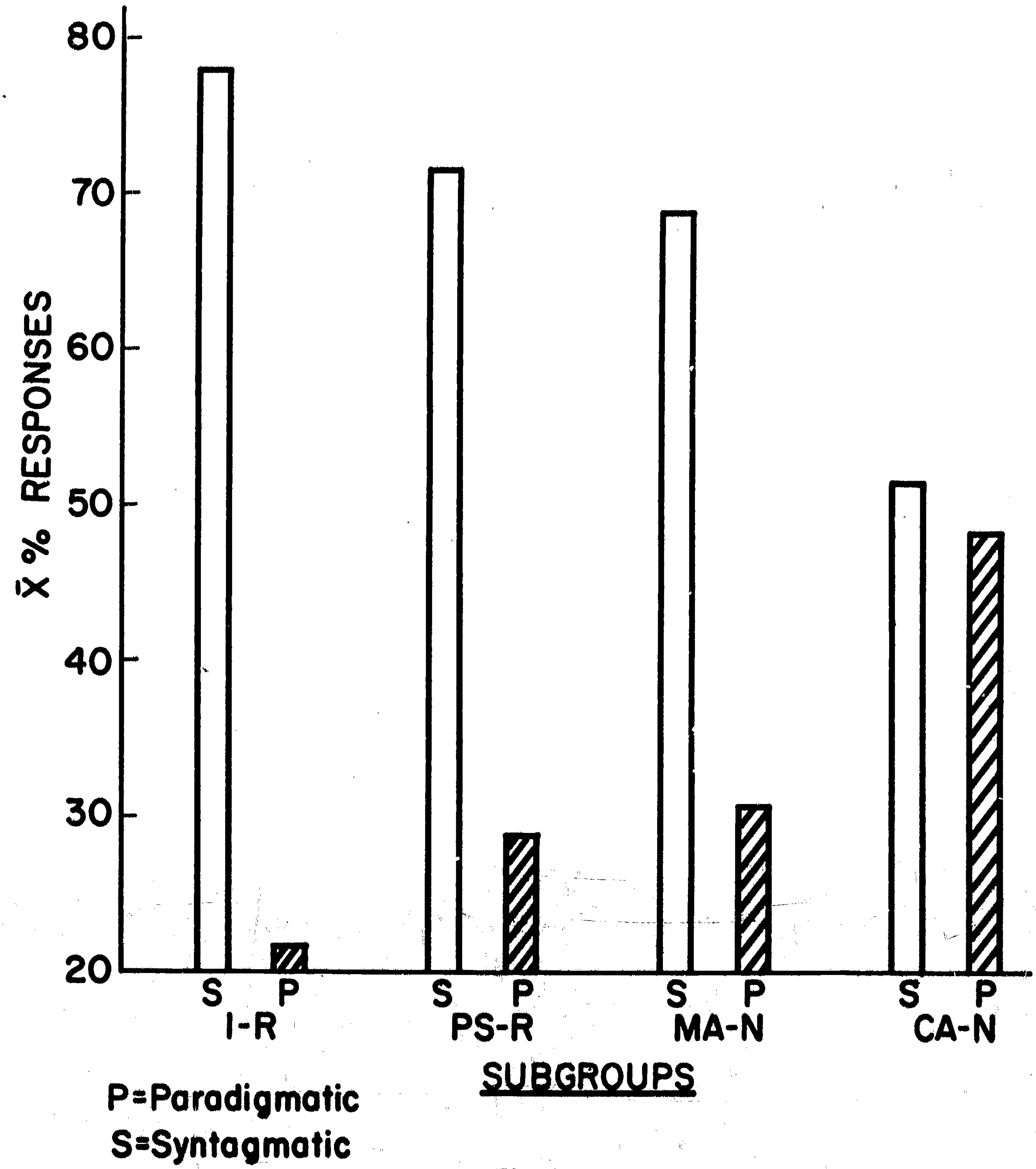
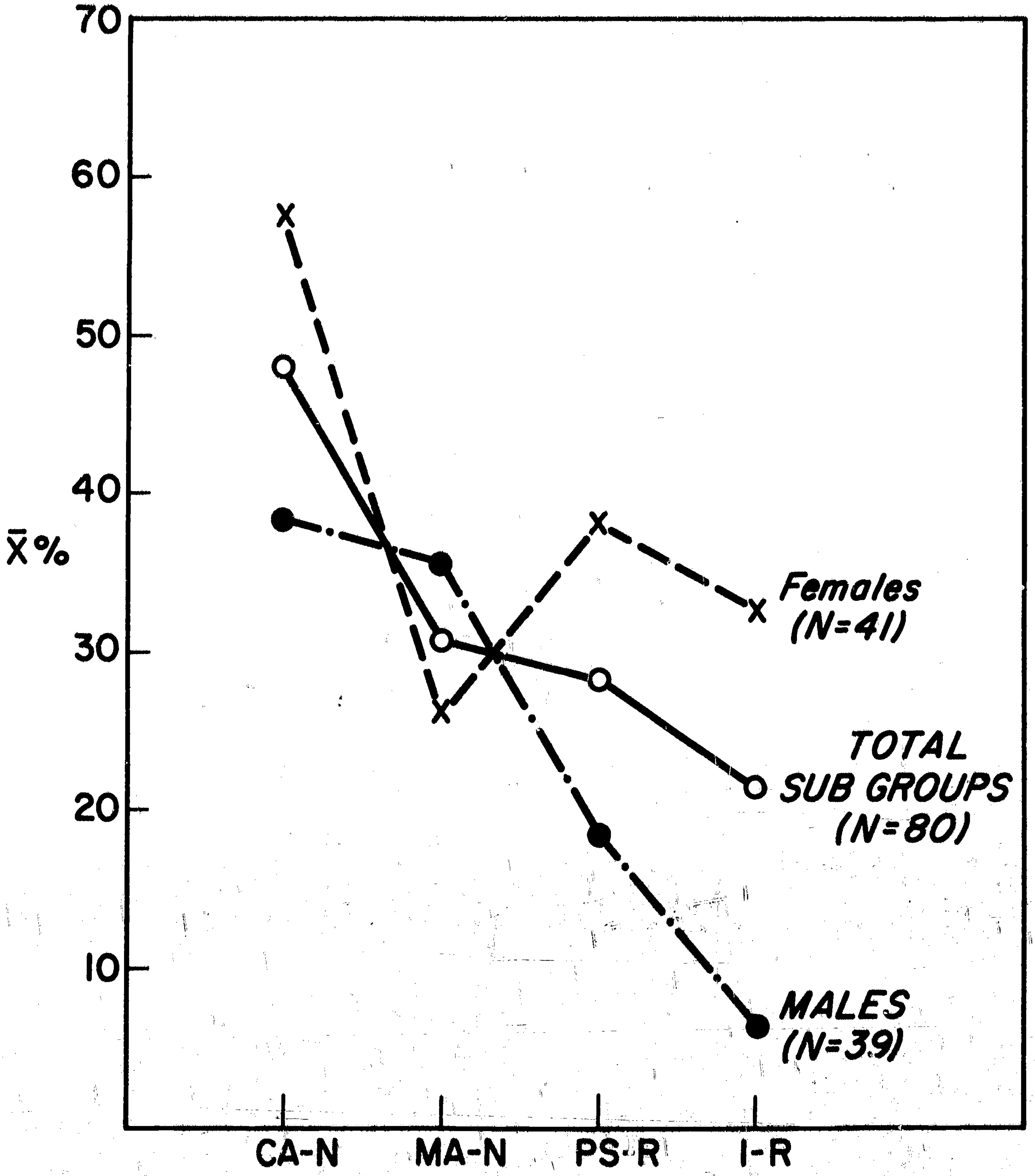

80
70
60
50
40
30
20
X̄ % RESPONSES
S
P
I-R
PS-R
MA-N
CA-N
SUBGROUPS
P=Paradigmatic
S=Syntagmatic

Fig. 3



70
60
50
40
X̄ %
30
20
10
Females
(N=41)
TOTAL
SUB GROUPS
(N=80)
MALES
(N=39)
CA-N
MA-N
PS-R
I-R

Fig. 4

A Pilot Study of the Verbal Interaction of Regular Elementary Classes and Special Classes for the Educable Mentally Retarded

Melvyn I. Semmel

Beverly Herzog

Fredericka Jorgenson

The purpose of this study was to compare verbal interaction in regular elementary school classes with that in special classes for educable mentally-retarded children.

The investigators used the method of Classroom Interaction Analysis developed Dr. Ned Flanders (1964) of the University of Michigan School of Education. A trained observer visited each classroom for a three-hour period in the morning. At three-second intervals the observer recorded a code number which indicated the category of interaction which best represented the communication events that were just completed. Approximately 3600 numbers were recorded in each observation.

Flanders' system has ten categories. Seven of these are assigned to teacher talk and two to student talk. Another category is used for periods of silence or confusion. These categories are designed to include all verbal interaction in the classroom and yet not overlap. Briefly the ten categories are:

<u>Teacher talk</u>

1. Accepts feelings
2. Praises or encourages

3. Accepts or uses ideas of student
4. Asks questions
5. Lectures
6. Gives directions
7. Criticizes

Student talk

8. Responds to teacher
9. Initiates own ideas
10. Silence or confusion

The observers participated in a six-week training period prior to beginning the study. Following training, a pilot study was undertaken during the fall semester, 1964, involving nine regular and special classes. Several hypotheses resulted from this pilot study:

1. There is more teacher talk in the special than in the regular class. (Categories 1-7)
2. The special teacher uses more praise and encouragement than the regular teacher. (Category 2)
3. There is more student-initiated talk in the special than in the regular class. (Category 9)

Method

The present study included ten regular elementary classrooms and ten special (Type A) classrooms for educable mentally-retarded children. There were five primary and five intermediate special classes. Each class was matched with a regular class on the basis of chronological age.

A primary special was matched with a first, second, or third regular class, and an intermediate special with a regular fourth, fifth, or sixth grade class.

Schools in two school districts were involved. Six regular and six special classes were located in the Wayne Community School District, Wayne, Michigan, and four regular and four special classes were in the Ann Arbor Public Schools, Ann Arbor, Michigan.

Preliminary Findings[1]

1. Category 2, "Praise and encouragement," received greater loading in the special class than in the regular (6.4% to 3.9%). This agrees with Hypothesis 2. (These percentages are based on total teacher talk [Categories 1-7].)

2. Category 3, "Accepting and using ideas," received greater loading in the special class than in the regular (14.8% to 13.7%). (These percentages are based on total teacher talk [Categories 1-7].)

3. Category 6, "Giving directions," received greater loading in the regular class than in the special (22.6% to 20.7%). (These percentages are based on total teacher talk [Categories 1-7].)

4. Category 9, "Student initiated response," received greater loading in the special class than in the regular (13.2% to 7.5%). This agrees with Hypothesis 4. (These percentages based on all ten categories.)

5. The percentage of total teacher talk for the regular class was 62.91%; in the special class the percentage was 63.79%.

6. The Chi Square test of significance was performed on all findings reported and all were found significant at the .01 level.

Discussion

Most studies have found little difference between the regular elementary class and the special class for the educable mentally retarded. Past studies have concentrated on achievement data in the academic and social areas. The present study has selected one aspect of the classroom <u>process</u> - verbal interaction - and has demonstrated that differences do exist between groups. The qualitative differences between the patterns of interaction in the two placement groups suggest that there are much more important distinctions to be made between regular and special education than the rather superficial differences presently emphasized--e.g., differences in homogeneity of grouping or in number of pupils per class.

Appendix

Categories for Interaction Analysis[2]

Teacher Talk

1. <u>Accepts feeling</u>: accepts and clarifies the feeling tone of the students in a non-threatening manner. Feelings may be positive or negative. Predicting or recalling feelings are included.
2. <u>Praises or encourages</u>: praises or encourages student action or behavior. Jokes that release tension, but not at the expense of another individual; nodding head, or saying "um hm?" or "go on" are included.
3. <u>Accepts or uses ideas of students</u>: clarifying, building, or developing ideas suggested by a student. As teacher brings more of his own ideas into play, shift to category five.
4. <u>Asks questions</u>: asking a question about content or procedure with the intent that a student answer.
5. <u>Lecturing</u>: giving facts or opinions about content or procedures; expressing his own ideas; asking rhetorical questions.
6. <u>Giving directions</u>: directions, commands, or orders to which a student is expected to comply.
7. <u>Criticizing or justifying authority</u>: statements intended to change student behavior from nonacceptable to acceptable pattern; bawling someone out; stating why the teacher is doing what he is doing; extreme self-reference.

Student Talk

8. <u>Student Talk — Response</u>: talk by students in response to teacher. Teacher initiates the contact or solicits student statement.

Appendix (Cont.)

Student Talk	9. Student Talk - Initiation: talk by students which they initiate. If "calling on" student is only to indicate who may talk next, observer must decide whether student wanted to talk. If he did, use this category.
	10. Silence or confusion: pauses, short periods of silence and periods of confusion in which communication cannot be understood by the observer.

Reference

Flanders, N. A. Interaction analysis in the classroom--a manual for observers. Ann Arbor: University of Michigan Press, 1964.

Footnotes

1. There is no scale implied by these numbers. Each number is classificatory; it designates a particular kind of communication event. To write these numbers down during observation is to enumerate, not to judge a position on a scale. See Appendix for explanation of Categories.

2. From Flanders (1964).

Some Differences in the Verbal Behavior of Preschool Children from Contrasting Social Environments

Paul Weener[1]

Recent research by Bernstein (1960, 1962a, 1962b, 1964) and Lawton (1963) has shown that distinct forms of spoken language are associated with different sociological strata.

> "It is proposed that the two distinct forms of language use arise because the organization of the two social strata is such that different emphases are placed on language potential. Once this emphasis or stress is placed, then the resulting forms of language use progressively orient the speakers to distinct and different types of relationships to objects and persons" (1960, p. 271).

The "two forms of language" Bernstein (1962a) calls "codes," and he distinguishes between a <u>restricted</u> code, characteristic of the lower classes, and an <u>elaborated</u> code, characteristic of the middle classes.

Distinguishing features of the restricted code include high structural redundancy, limited range of syntactic alternatives, and speech restricted by its social context. It is a status-oriented code, manifested in its extreme in military and ritualistic settings, where the setting determines the use and structure of speech. The status orientation of the lower class society, with its clearly defined roles, fulfills the conditions necessary for the development of a restricted language code. Children brought up in this environment learn a speech form "which discourages the speaker from verbally elaborating subjective intent and progressively orients the user to descriptive, rather than abstract, concepts" (Bernstein, 1960, p. 271).

The elaborated code has less structural redundancy with a wide range of syntactic alternatives to choose from. The meaning of the verbal interaction in this code does not depend primarily on the roles of the people involved, but is expressed by elaboration and expansion.

The middle class, with its flexible roles and extensive interaction among status positions, encourages the development of an elaborated code. The middle-class child uses language to express his subjective intent.

> "Speech becomes an object of special perceptual activity and a theoretical attitude is developed towards the structural possibilities of sentence organization. This speech mode facilitates... sensitivity to the implications of separateness and difference, and points to the possibilities inherent in a complex, conceptual hierarchy for the organization of experience" (Bernstein, 1960, p. 271).

The research reported here concerns the differences between samples of language of two preschool populations--a group of children from a remedial preschool program, and a group of middle- or upper-middle-class preschool children. In the context of the theory outlined above, it was hypothesized that the former group would reveal a restricted language code, the latter an elaborated code. In particular, H_1: children from upper-middle-class homes exhibit more complex grammatical patterns than children from lower-class homes. H_2: children from upper-middle-class homes exhibit a larger vocabulary than children from lower-class homes.

Method

Subjects

Two groups of 12 preschool children were selected from disparate social environments. The "high" social group was selected from first-year pupils in the preschool program at The University of Michigan Elementary School. All of the parents of this group had at least a college degree and the fathers were professionally employed. The "low" social group consisted of 12 pupils from the Perry Preschool Nursery Project in Ypsilanti (Michigan). Pupils in this project came from culturally-deprived homes, and tested below 85 on the Stanford-Binet Intelligence Scale, Form L-M. At the time of the experiment both groups had been in the preschool programs for about six months. The "low" social group had a mean age of 45.9 months; the "high" social group had a mean age of 46.4 months.

Procedure

The Ss were taken in groups of two into a room with a slide projector, a screen, and a tape recorder. There was a five-minute warm-up session in which the Ss were told about the equipment and were acquainted with the procedure. Two sample slides were shown, and, to encourage spontaneous verbalization from Ss, the experimenter asked several questions about each slide.

The experimental session consisted of showing 20 colored slides for 20 secs each. The slides depicted people and animals, usually engaged in some activity. The experimenter's participation in the session was limited to a few standard questions which were intended to promote Ss' participation. Wireless microphones and FM transmitters were used to facilitate transcription. This equipment made it possible to place the tape recorder out of sight in another part of the room. There was no evidence that any of the Ss knew that their talk was being recorded.

Data Analysis

The recorded tapes were played back and a record of each session was written out completely. The rules for classification of words and sentences were the same as those used by several investigators, notably Templin (1957,

Appendix 3). The following features were scored:

1. Percentage of words in each form class based on total number of intelligible words uttered;
2. Percentage of words in each form class based on number of different words uttered;
3. Number of different words per 100 words uttered (type-token ratio);
4. Sentence length;
5. Sentence complexity;
6. Total words used (tokens);
7. Total different words used (types).

The number of unintelligible words for each S was estimated and included in the total word-count, but only intelligible words were included in the totals for form-class percentages. The data on sentence length included unintelligible words but ratings of sentence complexity excluded sentences containing unintelligible words.

Two Ss from the low social group did not respond to any of the pictures. This behavior is reflected in the scores for total words and total different words but not in the remaining measures. The recordings of two Ss from the high social group were faulty and were not included in the analysis.

Results and Discussion

Four indices of grammatical structure and three measures of vocabulary were considered relevant to the two hypotheses, respectively. There were significant differences between groups on all four measures of grammatical structure: length of sentence, sentence complexity, percentage of nouns used, and percentage of different nouns used (Table 1). Subjects in the high social

---

Insert Table 1 about here

---

group uttered longer and more complex sentences with relatively fewer nouns than <u>S</u>s in the low social group. These results support the first hypothesis.

The low level of grammatical complexity exhibited by the low social group reflects a very limited use of the possible syntactic structures. Only about 16 per cent of the utterances by <u>S</u>s in this group could be classified as containing a complete simple sentence. There was only one compound sentence used and there were no examples of subordinate clauses in complex sentences. In the middle-class sample, on the other hand, about 57 per cent of the utterances were complete sentences and all but two <u>S</u>s used compound or complex sentences.

The two groups differed significantly in the total vocabulary used and in the number of different words used. These two indices of vocabulary support the second hypothesis.

A further comparison of the grammar of the two groups can be made from the distributions of frequency of occurrence of the remaining parts of speech. Inspection of Table 2 indicates that the two groups differed most in the relative use of the noun and verb classes.

Insert Table 2 about here

The differences in the verbal behavior of the high and low social groups observed in this study accord with the contrasts between the elaborated and restricted language codes described by Bernstein. It may not be correct, however, to infer that these differences are the result of social-class, since the two groups also differed in mean IQ. To the degree that the

IQ scores are related to the language measures used in this study, decreasing the mean difference in IQ scores between the groups would diminish the differences between the groups on the language measures. For the purpose of certain applications, including classroom instruction, the quantitative evidence presented here for the use of a more restricted code by the low social group may be important in its own right, apart from the search for its controlling variables.

Summary

Seven measures of the verbal behavior of two groups of preschool children (one of middle-class background; the other of lower-class background) were compared. The language samples were spontaneous responses to 20 colored slides. The middle-class group uttered longer and more complex sentences with relatively fewer nouns and greater type-token ratios. Thus, some evidence was provided for the theory previously put forth which attributes a more restricted language code to lower-class children and a more elaborated code to middle-class children.

References

Bernstein, B. Language and social class. Brit. J. Sociol., 1960, 2, 271-278.

Bernstein, B. Linguistic codes, hesitation phenomena, and intelligence. Lang. & Speech, 1962, 5, 31-48. (a)

Bernstein, B. Social class, linguistic codes, and grammatical elements. Lang. & Speech, 1962, 5, 221-242. (b)

Bernstein, B. Elaborated and restricted codes: their social origins and some consequences. Amer. Anthro., 1964, 66, 55-69.

Lawton, D. Social class differences in language development. Lang. & Speech, 1963, 6, 120-143.

Templin, M. C. Certain language skills in children. Minneapolis: University of Minnesota, 1957.

Footnote

[1]The author is indebted to Carol Huth for her assistance and active interest at every stage of the research.

Table 1

An Analysis of the Verbal Behavior of Children from Lower-Class and Upper Middle-Class Homes

Tests	Low Social Group			High Social Group			t (one-tailed)
	n	Mean	Sigma	n	Mean	Sigma	
1. Length of sentence	10	3.47	1.14	10	4.83	.96	2.89*
2. Complexity	10	.17	.10	10	.63	.23	
3. Total words (tokens)	12	149.1	85.6	10	215.3	93.0	1.70
4. Total different words (types)	12	45.9	27.3	10	86.1	26.1	3.42*
5. Number of different words per 100 words used (type/token ratio)	10	37.4	6.0	10	46.0	8.8	2.58*
6. Percentage nouns out of total words	10	30.8	7.1	10	22.1	3.9	2.79*
7. Percentage different nouns out of total different words used	10	38.5	7.1	10	30.4	5.9	3.06*

*Significant at the .01 level

Table 2

Form Classes in Children's Speech

(Relative frequency of use [in per cent] of form classes in children's speech of various parts of speech based on the total number of words uttered and the number of different words uttered by two groups)

		Noun	Verb	Adjective	Adverb	Pronoun	Conjunction	Preposition	Article	Miscellaneous
Percentage of total	High Social Group	22.1	20.8	8.2	5.8	15.6	5.7	5.9	12.6	3.3
	Low Social Group	30.8	13.2	7.4	5.6	11.5	4.7	14.7	17.9	4.2
Percentage of different words	High Social Group	30.4	24.2	10.5	7.2	11.2	3.5	5.1	2.7	5.2
	Low Social Group	38.5	19.8	7.5	8.6	10.8	1.7	6.3	3.9	2.9

A Critical Evaluation of the Illinois Test of Psycholinguistic Abilities

Paul Weener

Loren S. Barritt

Melvyn I. Semmel

This report precedes a series of papers describing the language performance of 189 kindergarten and first-grade pupils from the Ann Arbor Public Schools. The Illinois Test of Psycholinguistic Abilities (ITPA) and several other assessment techniques were used as part of a testing program designed to measure the effects of the reassignment of pupils from the Jones School in Ann Arbor.

The ITPA was used because it is the only standardized test available which purports to assess comprehensively the psycholinguistic abilities of children. Because this test is relatively new and there is very little published information available by which the test can be evaluated, it seems desirable to precede the study of performance on the ITPA with a technical report which describes and evaluates this test.

The emphasis here is on measures of the reliability and validity of the test rather than its theoretical foundations.

Introduction to the ITPA

General information

The Illinois Test of Psycholinguistic Abilities, published by the Institute for Research on Exceptional Children, University of Illinois, is designed to identify "psycholinguistic abilities and disabilities in children between the ages of two and one-half and nine" (McCarthy & Kirk, 1963). In the original description of the test's development (Kirk & McCarthy, 1961), the authors emphasized that the instrument was a diagnostic rather than a classificatory test, such as the Wechsler or Binet. A profile of nine different language skills and a total language age are provided.

The theoretical model and subtest description

The theoretical structure of the test is based on a communication model adapted from Osgood (1954) which is, in turn, an extension of Hull's mediation hypothesis. Three dimensions of psycholinguistic abilities are postulated--level of organization, channel of communication, and process. The two levels, four channels and three processes, are organized in Table 1.

Insert Table 1 about here

The six tests at the representational level require comprehending and giving meaning to symbols used in interpersonal communication, i.e., spoken words, written words, and pictures. Decoding (1, 2), which is the ability to understand the meaning of symbols, and encoding (8, 9), which is the ability to express this meaning in words or gestures, can be tested by utilizing the input and output side of the channel, respectively. The association process involves both an auditory or visual input and a vocal or motor output. Four complete channels could be tested but only two, the auditory-vocal and visual-motor (3, 6), are included in the ITPA battery.

Two types of tests evaluate decoding at the automatic-sequential level. The first (4) provides a measure of (automatic) habits which produce syntactical and inflectional aspects of language. The second (5, 7) measures the ability to remember a sequential structure; performance on these tests depends on auditory and visual memory. These tests are defined in terms of channel, but the model does not clearly specify processes at the automatic-sequential level.

The model is weakest in its theoretical description of the levels of organization in language functioning. The relationship between the theoretical description of the levels and their operational manifestation in the subscales is not very clear. For example, the representational level is nebulously defined as "sufficiently organized to mediate activities requiring the meaning or significance of linguistic symbols" (McCarthy & Kirk, 1961). At the automatic-sequential level, the processes involved are not specified "because of the lack of theoretical clarity at this level." The apparent loose relationship between the theoretical model and the test makes theoretical inferences from test results difficult.

The description of the actual behavior assessed in the tests listed in Table 1 raises the question of what kinds of abilities should be called psycholinguistic. Some of the tests seem to be measuring general cognitive factors (e.g., ordering geometric shapes in the Visual-Motor Sequential subtest) which are only tangentially related to psycholinguistic abilities.

Standardization Procedures and Statistical Characteristics

A recent publication by the authors of the ITPA (McCarthy & Kirk, 1963) provides information about the standardization procedures and the statistical characteristics of the test. It should be considered as a necessary supplement to the examiner's manual. Another publication (McCarthy & Olson, 1964) summarizes the validity information available from the standardization data and an independent study.

Standardization procedures

The standardization sample was obtained by testing 1,100 children between the ages of 2-0 and 9-0 from the Decatur, Illinois public schools. This sample was randomly selected from the school population with some noteworthy exclusions. Subjects with an IQ less than 80 and greater than 120 were excluded, as were children with serious sensory or physical handicaps, Negroes and parochial school students. The preschool children were primarily siblings of the already-selected school sample.

These initial steps in the standardization procedure reduce the generalizability of the normative data. Restricting the IQ range to exclude the upper and lower ten per cent gives a spuriously low estimate of variability. These restrictions on the normative sample were imposed by the authors of the test because "the main purpose of the standardization is to provide a reference group of relatively normal children..." (McCarthy & Kirk, 1963). It can be argued, however, that a standardization sample should represent the population with which the instrument will be used. If a test-user assumes this to be the case with the ITPA, he might come to unwarranted conclusions about the performance of a group not represented in the normative sample. For example, a score at the 15th percentile on the ITPA does not place a child in the same position relative to the general population as a score at the 15th percentile on the Binet or WISC. In the latter cases the normative samples are broader than in the former case, since only the children scoring above the 10th percentile on tests like the Binet and Wechsler have been included.

Seven hundred subjects were used to derive standard scores separately for fourteen age-groups at half-year intervals. There were consequently 50 subjects at each age level from 2-1/2 to 9 years. Each subject within an age group was within two months of his full-year or half-year birthday. Because of this restriction on the age range, the reported variance for each age-group is lower than would be expected for groups with a full age span.

Statistical characteristics of the ITPA

Fourteen different half-year age-levels, from 2-6 to 9-0, with fifty subjects in each group comprise the final sample used for developing norms. Means and standard deviations for males and females were calculated separately for each age-group. No systematic differences were found between the sexes, so the two groups were combined.

Average scores on each subtest showed a regular increase with age. The differences between means of successive age-groups for each of the nine subscales were tested by t-tests. About one-half (56 out of 117) of these differences were significant ($p < .05$). Only four group comparisons showed a decrease in mean score from one age-group to the next older group, and these differences were not significant. Mean comparisons indicated that the poorest discrimination among age-groups occurred at the older levels. Consequently, the 7-6 and 8-0, as well as the 8-6 and 9-0, groups were combined. This left twelve age-groupings for which standard scores and grade equivalents are provided for each subtest.

Reliability. An over-all measure of internal consistency was calculated for each subtest and for each age-group within each subtest by using coefficient alpha (Cronbach, 1951). Coefficient alpha is theoretically the mean of all the possible split-half coefficients and an upper bound measure of first factor concentration, or item homogeneity, of a test.

Table 2 summarizes the ranges and medians of the internal consistency coefficients for the twelve age-levels for each subtest, as well as for all age-levels combined.

Insert Table 2 about here

Odd-even split-half reliabilities were also calculated for each age-level for respective subtests yielding 108 coefficients with a median of .75 and a range of .39-.91. Table 3 summarizes these coefficients. The values of split-half reliability are similar to those for coefficient *alpha*, as would be expected since the measures sample the same general variance.

Insert Table 3 about here

Seven of the subtests have ceilings which introduce a spurious factor into the calculations of the split-half reliabilities. Ceilings based on the performance of subjects can spuriously raise the split-half reliabilities of a test by forcing a homogeneity of item performance. For example, the most extreme use of a ceiling would be to stop testing with the first error, in which case the split-half reliability would be almost perfect. The appropriate method of obtaining split-half reliabilities for tests that have ceilings is to establish equal but independent ceilings on each half of the test.

A measure of test-retest stability was obtained by retesting 69 of the 6-0 and 6-6 age group three months after the original testing. The stability coefficients for the nine subscales range from .18 to .86; six of the coefficients are less than or equal to .50. Estimates of the stability coefficients for the full age range (from 2-6 to 9-0) were calculated and found to range from .73 to .96 for the nine subscales. Since judgments made with the ITPA will generally involve comparisons of scores within similar age-groupings, the reliabilities at each age-level, rather than for the total group, would seem to be most important. Reliabilities based upon the total group have little meaning or value for users of the test.

The standard errors of measurement reflect the relatively low reliabilities and high variances for the subscales. Fifty-three of the 108 standard errors for standard scores are greater than or equal to .50. This means that for about half of the subscales a 2 Sem band about a subject's score would include more than ±1 standard deviation.

Summary of Reliability. Internal consistency, test-retest stability, and standard errors of measurement are the three types of reliability measures available. Internal consistency measures for a restricted age range are moderately high for each subscale indicating homogeneity or unifactorness for the subscales. The test-retest stabilities for subscales are quite low; standard errors of measurement for subscales are large.

In general, the reliability of subtests is too low for adequate prediction. The over-all score, which is used to derive the subject's language age, does demonstrate reliability measures in an acceptable range. But if the test is to accomplish its stated purpose of differential diagnosis, the subtest reliabilities must be increased. In their present form the subtests can be used only for gross discriminations, as was illustrated by the confidence intervals.

Validity. Validity studies were carried out in Middleton, Wisconsin, using a sample of 86 subjects from 88-110 months old (McCarthy & Olson, 1964). The validation sample was similar to the original standardization sample with respect to social class, sex, and mental age. Concurrent, predictive, content, construct, and diagnostic validity are reported.

To obtain measures of concurrent and predictive validity, criterion tests were administered at approximately the same time as the ITPA, and the criterion tests were readministered about three months later. The criterion measures in the validity study were "existing language tests and the linguistic portions (reading, spelling) of achievement tests" (McCarthy & Olson, 1964). A criterion test was pre-selected for each of the nine subtests on the basis of its apparent similarity to the respective subtests. Correlations between the ITPA and the first administration of the criterion scores provided concurrent validity estimates; correlations between the ITPA and retest criterion scores provided predictive validities.

The test-retest interval is in several cases so short that no real distinction can be made between concurrent and predictive validity. The average time in months between administration of the ITPA and the criterion test for the concurrent measure of validity was 1.48 months for two of the tests; the average difference in months for the predictive validity measures on the same two tests was 1.94 months. Such small differences in the test-retest interval do not warrant classifying these measures as two different kinds of validity.

The concurrent validity coefficients for the nine subtests ranged from .03 to .65 with a median of .15. The measures of predictive validity ranged from -.19 to .53 with a median of .23. Several explanations can be given to account for these low validity coefficients. (1) Good criterion tests which assess the same abilities as the ITPA subscales may not be available. (2) Low reliabilities for both criterion tests and the ITPA subscales restrict the amount of common variance available. (3) The restricted nature of the tests and the relative homogeneity of the sample would tend to depress the validity coefficients. The criterion tests were subscales and parts of tests designed to measure a single aspect of language development and the age range of the sample was from 7-0 to 8-6.

Two different aspects of <u>content validity</u> were analyzed. A logical analysis of the items in each subscale and a description of the item selection procedures were used to demonstrate the "<u>face</u>" validity of item selection. Internal consistency coefficients indicate to what extent test items within a subscale are measuring the same ability. The internal consistency coefficients of the twelve age-groups for each subtest were given in Table 2.

The "face" validity or the rationale for defining a universe of items and the method of sampling from this universe are discussed at length by the authors. This is helpful information which tells what the test purports to measure. But the "face" validity of an instrument is not a necessary or sufficient condition to establish its actual validity; that is, how well it actually measures what it purports to measure. For this purpose, only statistical data will suffice.

In assessing construct validity, predictions from the theory or model are tested. Most of this section consists of an analysis of intercorrelation data. Gross factors known to affect language development, such as mental age, social class, birth order, etc. are correlated with total battery scores and each subtest. Mental age correlated positively with ITPA performance; smaller, but significant, negative correlations were found with social class, birth order, and number of siblings.

The theoretical meaning of each subscale was made clearer by predicting which factors should or should not influence subscale performance. Eight of the nine subtests were matched with tests measuring behavior which the authors predicted would not be related to performance on the respective subtests, although they appeared to be measuring similar factors. Seven out of the eight tests showed no significant correlation with subtest scores, as was predicted. Such correlations, which define a subtest by specifying measures which are unrelated, are a necessary part of discriminant validation methods described by Campbell & Fiske (1960).

These findings, however, should be compared with the measures of concurrent and predictive validity described earlier. Five of the nine tests selected to match the subscales as measures of concurrent validity showed no significant relationship, nor were three of the eight comparisons significant in assessing predictive validity.

The analysis of the statistical relationship of levels, channels, and processes provides a test of the theoretical model. Although a degree of heterogeneity is desired for the subscales, tests that are of the same linguistic level, or are in the same channel, or that assess the same process should correlate more highly with each other than with tests from other levels, channels, and processes. The intercorrelation data from the standardization sample cannot be consistently explained in terms of the theoretical development of the test.

This "muddiness" in the internal structure is also evidenced in factor analyses of the test. The over-all ages intercorrelation matrix of standardization data was factor-analyzed, as was each of the matrices for the 14 age-levels. The Principal Axis analysis was followed by the orthogonal rotation of the first six factors resulting from each analysis. In the analysis of over-all results, a general linguistic factor accounts for about 80 per cent of the common variance; one factor (general expressive ability) contributes about 4 per cent and four other factors contribute about 3 per cent each for a total of about 95 per cent of the total variance. The factor analysis of the over-all ages intercorrelations is quite meaningless because the between-ages covariance is mixed with interscale covariance. This confounding of two large sources of variance tends to yield too large a first factor and too few other factors.

In the factor analysis of the different age levels, the general factor accounts for less than 40 per cent of the variance. Each of six tests loads from .50 to .80 on the general factor with little evidence of regularity from one age-level to the next in the distribution of the loadings on the remaining five factors. Five of the subtests do give singularly high loadings at five or more of the fourteen age-levels and contribute an average of 10% to 12% of the variance when present.

The factor tables indicate that unities were used in the diagonal of the intercorrelation matrices for the Principal Axis extraction. The factors extracted, therefore, account for the error variance as well as common and specific variances (Fruchter, 1954). Because the error variance is large for the subscales, a better choice of factoring techniques would have been to insert an estimate of the communalities in the diagonals of the correlation matrices. In view of the low reliabilities for subscale scores, error variance influenced the factor structure considerably and accounts for some of the inconsistencies across age groups.

Summary of Validity: Concurrent and predictive validity measures were obtained from a sample of 86 subjects, ranging in age from 7-0 to 8-6. The median of the nine concurrent validity coefficients was .15; the median of the nine predictive coefficients was .23. Content validity was demonstrated by a description of the item selection procedures and calculation of subtest homogeneity using the normative data. The internal consistency coefficients for the subtests were about .75 for each age group. Construct validity was assessed by demonstrating which relevant factors were or were not related to subtest performance. Mental age, social class, birth order, and number of siblings were found to be significantly related to ITPA performance.

The factor-analytic results indicate at least a partial failure of the attempt to construct "single ability" tests. A general linguistic factor was by far the largest after rotation; individual subscales came out as singular factors with varying degrees of consistency. The appropriateness of the factoring technique selected is questionable. The effects of error variance partially account for the overlapping factor structure.

Testing the theoretical model is difficult because the model does not clearly specify behavioral correlates. The theoretical terms are not well-defined and therefore the specification of operational definitions is quite arbitrary. The theory must further develop a nomological network of theoretical variables before meaningful assessment of construct validity can be made.

Summary

This evaluation of the statistical characteristics of the ITPA indicates that the reliability and validity are not yet adequately established. The standardization sample is small and of a very restricted nature, the stability of subtest scores is low, and the validity of subtests is not proven. The model does not adequately integrate the nine subscales nor does it specify the nature of the relationships among subscales or between subscale performance and other relevant behavior.

The available information does indicate that a range of abilities is assessed by this instrument. Subtest performance can be used to supplement or corroborate existing data on a subject or a group. If possible, research projects using the ITPA should use the instrument in such a way that information about the reliability and validity of the test is provided.

Papers following this report will describe the performance of 189 Ann Arbor school children on the ITPA. In addition, the ITPA itself will be examined by comparing a subject's performance on it with other cognitive and social variables. A factor analysis of ITPA results and other cognitive variables may point to a more parsimonious description of the psycholinguistic skills measured by the ITPA.

References

Campbell, D. T., & Fiske, D. W. Convergent and discriminant validation by the multitrait-multimethod matrix. Amer. Psychol., 1960, 15, 546-553.

Cronbach, L. J. Coefficient alpha and the internal structure of tests. Psychometrika, 1951, 16, 297-334.

Fruchter, B. Introduction to factor analysis. New York: Van Nostrand, 1954.

Loevinger, J. Objective tests as instruments of psychological theory. Psychol. Rep., 1957, 3, 635-694, Monogr. Suppl. 9.

Kirk, S. A., & McCarthy, J. J. The Illinois Test of Psycholinguistic Abilities--an approach to differential diagnosis. Amer. J. ment. Def., 1961, 66, 399-412.

McCarthy, J. J., & Kirk, S. A. Examiner's manual--Illinois Test of Psycholinguistic Abilities. Urbana, Illinois: Institute for Research on Exceptional Children, 1961.

McCarthy, J. J., & Kirk, S. A. The construction, standardization and statistical characteristics of the Illinois Test of Psycholinguistic Abilities. Madison, Wisc.: Photo Press, 1963.

McCarthy, J. J., & Olson, J. L. Validity studies on the Illinois Test of Psycholinguistic Abilities. Madison, Wisc.: Photo Press, 1964.

Osgood, C. E. (Ed.) A behavioristic analysis. In Contemporary approaches to cognition. Cambridge, Mass.: Harvard University Press, 1957, 75-118.

Table 1

Derivation and Description of ITPA Subtests

Processes	Channels	Levels	Tests
Decoding	Auditory	Representational	1*
		Automatic-Sequential	
	Visual	Representational	2
		Automatic-Sequential	
Association	Auditory-Vocal	Representational	3
		Automatic-Sequential	4 5**
	Visual-Motor	Representational	6
		Automatic-Sequential	7
	Auditory-Motor	Representational	
		Automatic-Sequential	
	Visual-Vocal	Representational	
		Automatic-Sequential	
Encoding	Motor	Representational	8
		Automatic-Sequential	
	Vocal	Representational	9
		Automatic-Sequential	

*Number refers to subtest descriptions presented below.

**There is an auditory-vocal test at both an automatic and sequential level.

Table 1 (Cont.)

1. Auditory Decoding--Vocabulary test requiring only "yes" or "no" answer, e.g., Do females slumber?
2. Visual Decoding--Matching a stimulus picture to its perceptual counterpart, e.g., Office table and coffee table.
3. Auditory Vocal Association--A verbal analogies test, e.g., Soup is hot. Ice cream is _____.
4. Auditory Vocal Automatic--Correct grammatical form must be provided in sentences, e.g., Here is an apple. There are two _______.
5. Auditory Vocal Sequencing--Digit repetition as in Binet.
6. Visual Motor Association--Relate pictures on some conceptual basis, e.g., sock with shoe.
7. Visual-Motor Sequencing--Sequence of geometric shapes must be reproduced from memory.
8. Motor Encoding--Expressing one's ideas in terms of meaningful gesture, e.g., "Show me what you should do with this." (hammer)
9. Vocal Encoding--Describe a simple object verbally, e.g., block, nail.

Table 2

Internal Consistency Coefficients for Subscales of the ITPA

	Aud.-Voc. Auto.	Vis. Dec.	Mot. Enc.	Aud.-Voc. Assoc.	Vis. Mot. Seq.	Voc. Enc.	Aud.-Voc. Seq.	Vis. Mot. Assoc.	Aud. Dec.
Range for age groups	.61-.83	.50-.84	.64-.86	.51-.76	.58-.76	.54-.82	.53-.83	.70-.89	.75-.84
Median of age groups	.75	.72	.75	.71	.69	.75	.74	.84	.78
Over-all age group	.93	.90	.89	.95	.91	.92	.92	.95	.92

Table 3

Split-Half Reliability Coefficients for Subscales of the ITPA

	Aud.-Voc. Auto.	Vis. Dec.	Mot. Enc.	Aud.-Voc. Assoc.	Vis. Mot. Seq.	Voc. Enc.	Aud.-Voc. Seq.	Vis. Mot. Assoc.	Aud. Dec.
Range for age groups	.62-.90	.45-.90	.60-.86	.51-.83	.39-.84	.48-.84	.72-.89	.55-.89	.58-.91
Median of age groups	.78	.76	.76	.72	.70	.72	.80	.76	.80
Over-all age groups	.95	.91	.90	.96	.93	.91	.94	.93	.96

The Role of the Printed Answer in Programmed Instruction

George L. Geis, William Jacobs, and Dwight Spenser

Some Assumptions in Programmed Instruction

From the operant conditioning laboratories of psychology have come principles, procedures, and personnel which in large measure define the field of programmed instruction. If ever there was a direct, systematic application of psychological research to "the classroom" it is programmed instruction. Yet the applied instructional situation is so complex that it deserves more than a mere transfer of principles. Further, it well may be that such a bald application of general principles simply will not prove to be of much help to those interested in education. The efficient production of learning in a human student may be a more complicated procedure than producing learning in a human subject.

One concept transferred from the laboratory and readily used by programmers is "reinforcement." This is the term to be examined in the present studies.

It is often said that programmed instruction presents not only a tool for teaching but one for research in learning as well-research in both basic and applied learning problems. Programmed instruction may provide the opportunity for such research, but it seems not to have provided the inspiration, judging from the shocking poverty of the literature in the area. The present set of studies is designed on the one hand to investigate the variables controlling a particular operant response, namely, answer-observing in a program. On the other hand, the findings of this research should provide support for, or a challenge to, what has come to be a leading commandment in the programmer's bible.

Almost every published program, as well as innumerable articles and texts concerning programmed instruction, has a statement similar to this one gleaned almost randomly from the literature. "It is the immediate checking of the right answer that serves to reinforce it. The more often it occurs and the sooner it follows the writing of a response, the better the learning" (Brown, 1964, Preface).

That students learn from programmed materials cannot be disputed. And, while there are differing opinions among psychologists concerning the status of the concept of reinforcement, many would agree that it is a crucial concept in any discussion of learning. Nevertheless, putting the case that a student learns by doing a programmed text and, further, that in order to have learned he must have been reinforced, does not constrain one to assume the presentation of a correct answer is the reinforcer at work. A previous paper (Geis & Knapp, 1963) has demonstrated at least the possibility that the control of learning in a program rests with aversive rather than positively reinforcing stimuli.

If one does assume that the answer in a program is the reinforcer, certain other basic assumptions might be made and tested. For example, one would assume that the student and the answer are in contact. The answer cannot have an effect on the student if he does not see it. (Programmers have been accused of being overly artistic but no one has yet suggested that they are metaphysical.) Yet our own observations of students have made us suspect that they often do not look at the answers in a program.

Suppose we do assume that the student looks at an answer. Still, a wide variety of conditions can obtain. The reinforcing effect of the answer, one might expect, depends on the student's discovering the correctness of his answer. In many programs it is the student who makes this discovery by recognizing whether his answer matches a printed one. In most programs, matching may not present a problem but in some programs (of special interest to us, in second-language programs) such matching behavior cannot be assumed as given in the naive student.

Suppose a student looks at and matches the printed answer with his own only to find that his answer is incorrect. Ugelow (1962) suggests that "while working with knowledge of results may be preferable to working without such knowledge, knowing that one is consistently doing poorly cannot be said to be rewarding" (p. 9). The student who repeatedly discovers, by observing answers in a program, that his own answers are erroneous and who, nevertheless, continues to observe answers might be suspected of masochism. On the other hand, it cannot be assumed that the student is reinforced solely by seeing he is correct. It has been argued elsewhere (Geis, 1961) that if an answer is redundant, either because of overcuing or because other frames may confirm the answer, seeing the correct answer may not be reinforcing. This sequential set of logical assumptions seems to us to indicate that the matter of reinforcement in programmed instruction deserves attention and that the concept of confirmation and its role in programmed instruction deserve clarification.

Knowledge of Results

The view that knowledge of results (KR) is reinforcing draws strength not only by extrapolation from the animal experimental laboratory but also from many studies of human learning. (See, for example, Bilodeau & Bilodeau, 1961, where feedback is reviewed in the context of motor skills learning.) The need for clarification of the definition of KR was recognized long before the advent of programmed instruction. Brown, in 1940, distinguished among three functions of KR: as a reinforcer, as a discriminative stimulus for later behavior, and as an incentive. To summarize the literature on the role of KR in human learning, instruction often appears more effective when the student can obtain information about the correctness of his own responses. As in the general literature on KR so in the context of programmed instruction, a wide variety of terms has been used interchangeably (e.g., confirmation, feedback, knowledge of results) despite the need for distinctions among them. A student can be told whether he is right or wrong without the program's actually supplying a matching answer. Similarly, the student may see a printed answer without being able to determine whether it confirms his own answer or not, as has been pointed out previously. There have been attempts to clarify the various terms that refer to the role of the printed answer in the program. Porter (1957), for example, considers confirmation a descriptive concept referring to a particular sort of sensory feedback and not to the reinforcing value of that feedback. Carr (1959) and others think it is necessary to distinguish between confirmation, which provides information regardless of the student's motivation, and reinforcement, which does depend on the student's motivation.[1]

If learning is affected by the reinforcement variable and if the answer in the program is considered to be reinforcing, then it follows directly that performance on post-program criterion tests should vary with availability of answers within the program. Several studies have compared criterion test scores of a group of students who had KR presented in the program with those of students who had not. (See, for example, Krumboltz & Weisman, 1960; Moore & Smith, 1964; Rosenstock, 1965.) "Feedback" was presented in various ways: the answer in some cases was revealed in a teaching machine; the answer was presented in an accompanying answer book for some text programs; answers were at the back of the program, or on the following page in the program, or imbedded in the text itself, or in a column adjacent to the program frame.

In almost all cases there was no significant difference found between criterion test performance of students who had the answer present and those who had the answer absent in the programs. Partial feedback with answers supplied on a given percentage of frames similarly produced little effect on performance on criterion tests. These findings conflict somewhat with most of the KR studies using other than programmed material. And it should be mentioned that a few programming studies (e.g., Meyer, 1960) do show some effect when KR is removed. At best a search of the programming literature reveals inconsistent findings; at worst it reveals evidence which challenges the view that confirmation is reinforcing and also challenges the generality of findings in the other literature on KR (see Bilodeau & Bilodeau, 1961).

In good part, the whole issue may be thoroughly contaminated by the differences in materials used among programmed instruction studies and between programmed instruction studies and the more traditional knowledge-of-result studies. It has become almost reflexive for researchers on programmed instruction to begin to answer any research question with "It depends on the program being used." Certainly it is naive to compare the effect of KR in a program which is carefully sequenced and designed to produce correct answers with, say, the effect of KR when the material consists of nonsense syllables carefully designed to provide difficulty in learning.

Some Studies of Confirmation

Earlier it was pointed out that a distinction between confirmation and reinforcement might well be made. The studies reported below do not start with such a distinction a priori; instead they are aimed at clarifying what is reinforcing in a program by investigation of the following questions:

(1) Is a general description of the control of answer observing or checking possible when a variety of students, a variety of programs, and a variety of formats are used?

(2) Is checking or not checking peculiar to treatment, format, or students?

(3) Is checking or not checking peculiar to certain variables which, broadly speaking, might be designated interactions between students and frame characteristics--for example, the difficulty of the frame?

(4) Is answer checking related to the subject's previously emitted answer in that frame? Specifically, do subjects tend to check those frames on which they have made errors more than those on which they have not made errors?

(5) Is there a relation between the number of errors that the student makes in a program, and/or the number of checks he is observed to make, and the score he obtains on a posttest?

(6) Does finer analysis, subject by subject and frame by frame, reveal certain frames which are generally checked by most subjects and certain frames which are never checked by any subjects?

(7) Is it possible to find a relationship between some measure of a student's confidence of being correct and his answer-observing behavior?

Study 1

Since all of the studies are of a similar nature, one will be described in detail; for purposes of this report, the others will be summarized.

Method

Materials. One self-instructional program was prepared in two formats.

1. Card-flip program. The first 113 frames (3 sets) of The Analysis of Behavior (Holland & Skinner, 1961) were cut from the book and each was mounted on a 3" x 5" white index card. The correct answer was inverted and pasted on the reverse side of the card. Thus, by flipping the card, the answer could be

revealed. Two slotted boxes marked CHECKED and UNCHECKED were provided and S was told to drop each frame, when he was finished with it, into one of the boxes.

2. Tab-Pull Program. The first 19 pages (113 frames) of the same program were mounted intact on both sides of 8-1/2" x 11" acetate sheet protectors. The correct answer to each frame, which appeared to the right of the question, was concealed by a gummed-back paper tab. The S could remove the tab to reveal an answer. In this program a cardboard mask was provided which S was asked to slide down the page as he progressed, thereby covering previously revealed answers and frames.

The one-page exhibit, or panel, accompanying Set 3 of the program was provided in a loose sheet protector with both versions of the program. A four-page blank answer booklet was provided each S. Two 11-question criterion tests were constructed based on material in the first three sets of the program. The Test-Anxiety Questionnaire (Mandler & Sarason, 1951; Sarason & Mandler, 1951) was administered to each S.

Subjects. Twenty-four university students were paid Ss in the experiment. No one had previously used the Holland-Skinner program.

Procedure. Each S was given the TAQ before beginning the experiment. The Ss were divided into four groups. Group C (n=8) used the card-flip program. Group T (n=8) used the tab-pull program. Group C/T (n=4) used the card-flip program for the first 54 frames and the tab-pull program for frames 55 through 113. Group T/C (n=4) used the tab-pull program for frames 1 through 54 and the card-flip program for frames 55 through 113.

All Ss were familiarized with the conventions of the program (see pages vii-viii of the text).

Group C Ss were read these instructions:

This is an experiment in programmed instruction. Answer each question in the answer booklet. The correct answer to each question is provided on the back of the question card. If you wish to check your answer, turn the card over, and after having checked your answer, place the card in the box marked CHECKED. If you do not wish to check your answer, place the card in the box marked UNCHECKED. You may check as many answers as you want to, but you don't have to check any. Once you have answered a question to your satisfaction, do not change it. Do not go back to any question once you have left it. Are there any questions?

Group T Ss were read the following instructions:

This is an experiment in programmed instruction. Answer each question in the answer booklet. The correct answer to each question is provided under the paper tab beside the question. If you wish to check your answer, carefully remove the tab and place it in the box. You may check as many answers as you want to, but you don't have to check any. Once you have answered a question to your satisfaction, do not change it. Do not go back to any question once you have left it. In order to help you avoid looking back at previous questions, a cardboard mask is provided. Please slide this mask down the page so as to cover each question after you have answered it. Are there any questions?

The Ss in Groups C/T and T/C were given the instructions appropriate to the format just prior to its presentation. If Ss asked questions, they were instructed to reread the instructions. The Ss were allowed to arrange the apparatus to suit themselves. While E was present at all times, he did not observe Ss closely.

After completing the 113 frames of the program S was immediately given the posttest. Sixteen Ss were given the first form of the test and eight Ss were given the second.

Results and Discussion

In Figure 1, the 113 frames in the program were ranked for error with all Ss' responses included in these data. Thus, the frame with the highest error

---

Insert Figure 1 about here

---

was ranked number one, and is plotted against the total number of errors occurring on that frame. The negative acceleration of the cumulative curve shows that a few frames contribute heavily to the total number of errors the Ss made on the program. Twenty-one, or approximately 17 per cent of the frames, account for 50 per cent of the total number of errors committed. No error was recorded on almost 25 per cent of the frames in the program. Figure 2 demonstrates another important feature of the program and the population inter-

---

Insert Figure 2 about here

---

acting. In this figure the 24 Ss were ranked for total number of errors committed in the 113 frames. Once again, the cumulative plot shows a high error contribution by a small number of Ss. (The error total in the two figures is different because failure to answer at all was counted as an error when plotting Fig. 1 but these blanks were not included as errors when plotting Fig. 2.)

Figure 3 shows the distribution of checking behavior on the program, disregarding format. One interesting finding of the study confirmed by other

---

Insert Figure 3 about here

---

studies, to be summarized later, is that the population seems to fall into three categories: Ss who never or almost never check an answer; Ss who always, or almost always, check an answer; and an intermediate group of Ss, constituting the majority of the population, who check on occasion. The wide variability among Ss in their answer-checking behavior has both theoretical and practical implications. The contention that the answer is automatically a reinforcer in a program is severely questioned when one observes that at most six Ss, or one quarter of the population, could be said to be taking advantage of the availability of the answer most of the time. The answer in a program may be reinforcing but, as has been suggested earlier, probably it is reinforcing under a specific set of circumstances. Two practical implications from the data might be stated. The first is that the programmer can assume, at the very least, that other kinds of reinforcers are at work in a program. Four Ss, or one sixth of this population, almost never "took advantage of" what might have been designated a priori as a reinforcer. A large number of Ss exposed the answer only infrequently. The programmer should be aware, then, that other reinforcers may be at work within the program or may have to be built into the instructional system. A second practical implication concerns the use of the answer space for additional information, correction, and the like. It would seem dangerous to place in this space any additional material for the students, since most students consult the answer space only infrequently. There is no guarantee, in other words, that any information placed in that area will be read by the student.

Figure 4 compares the answer-checking behavior of Ss in the different groups. The investigators were inquiring whether mere manipulation of the

Insert Figure 4 about here

format of the answer space would produce marked differences in answer observing. Although there are such differences indicated in the data, they are not statistically significant. Furthermore, the answer checking of the T/C and C/T groups on the tab-pull format is not markedly different from their behavior on the card-flip format. The differences present appear to be a function of the location of the frames (first or second half of the program) rather than of the format.

Table 1 indicates that the posttest scores among the four groups are not different. Although the per cent wrong answers seems to be higher for the

---

Insert Table 1 about here

Group T, once again the differences among the data are not statistically significant. Table 2 presents data grouped along the lines indicated in Fig. 3. The

---

Insert Table 2 about here

Ss who very rarely checked answers are included in Group Lo, Ss who very frequently checked answers are identified as Group Hi, and the majority of Ss who checked answers periodically are included in Group Inter. Scores on the Scholastic Aptitude Test, taken from the college records of these Ss, show almost no difference between members of the extreme groups. Similarly, the differences among groups in posttest scores and in mean percentage correct responses in the program were nonsignificant. Although the mean TAQ differences, as represented in this table, prove not to be statistically significant, this seems to be an artifact of the grouping since a Pearson r of .41 ($p < .05$) was found between TAQ scores and the number of frames checked.

Figure 5 presents the probability of checking given that a response is incorrect, P(C/W), plotted against the per cent of frames checked for each S

Insert Figure 5 about here

---

in the intermediate group. The solid line represents the chance probability for checking a response given that it was incorrect. All points above that line indicate P(C/W) is greater than P(C). Since a majority of Ss were above the chance level, a sign rank test of difference was employed. This test indicated that the Ss in the intermediate group had a conditional probability of checking when wrong that was significantly greater than the probability of checking ($p < .02$). Thus, the null hypothesis was rejected and it can be concluded that, at least on this program, Ss in the intermediate group were checking partly on the basis of the correctness or incorrectness of their own answers. That is, the likelihood of checking is greater for an S in this group when he has made an incorrect response on the program than when he has made a correct response.

Summary and Conclusions

The results of this study indicate that one must be wary of characterizing the answer in a program or the act of observing an answer as a reinforcer. If the answer were reinforcing at all times one would expect that all subjects would, of course, look at the answer when they are allowed to do so. Rather, as these data have indicated, Ss vary a great deal in their answer-checking behavior. Furthermore, it seems that answer-checking is, at least in part, under the control of S's own answering behavior. Roughly, S tends to check when he knows or suspects that he is wrong. What is being reinforced under such circumstances is not clear. If seeing the answer at such a time reinforces previously-emitted behavior then for some Ss only incorrect answers are being reinforced! A more complicated but more acceptable hypothesis might be that when Ss are uncertain as to the correctness of their answer, exposing the answer makes possible a correction on their part and increases the likelihood that they will emit a correct answer on the next occasion that such an answer is called for.

Summaries of Other Studies

Study 2

Another study, quite similar to the first in many respects, was conducted using another program (Basic Systems, 1963).[2] Again it was found that there was wide variation in the answer-confirming behavior of Ss, with most Ss confirming some, but not all, of their answers. Also, it was found that there was no correlation between confirming behavior and performance on the posttest. The Ss who used the card-flip version of this program consistently checked more answers than the Ss with the tab-pull version. The Ss in the half and half groups each checked more under the card-flip treatment than under the tab-pull treatment. Once again, it was found that Ss in the intermediate group tended to check more frequently when they had previously emitted an erroneous answer than when they had previously emitted a correct answer, confirming the study recounted at length above.

Study 3

Sections of two programs were utilized in this study (Charles Pfizer, 1964; McGraw Hill, 1963). As in the previous studies there was wide variation in the number of observing responses made by Ss. A trimodal distribution developed with a few Ss always looking at the answer, a few Ss rarely looking at the answer, and most Ss looking occasionally. Confirming the previous studies, almost without exception the Ss tended to observe answers more often after having emitted an incorrect answer than after having emitted a correct answer. An attempt was made with one of the programs to manipulate the level of difficulty, since it was suspected that difficulty level (or, probability-of-error level) was, in part, responsible for answer observation. When Ss were instructed to skip from the first part to the middle part of the program, there was a sharp increase in errors and a sharp increase in answer observing, as hypothesized.

Study 4

In a final study in this series, Ss were asked to rate their confidence in their answer after answering each frame in a program (the one employed in Study 2). Judging from the observations made in previous studies and preliminary analysis of the present findings, there is evidence that S tends to check the answer to a frame when he is not confident of the correctness of his own answer.

Conclusions

These studies have examined answer observing in diverse programs that vary in programming style and in difficulty level. The findings in the first study detailed earlier have been repeatedly confirmed: (1) there is great variation in answer observing among Ss working through a program; (2) answer observing is, at least in part, a function of the correctness or incorrectness of S's own emitted response; (3) criterion performance, as indicated by post-test scores, does not seem to be affected by frequency of answer observing within [illegible] program.

References

Basic Systems. Allergy and hypersensitivity. (2nd ed.) New York: Pfizer, 1964.

Basic systems. The diagnosis of myocardial infarction by electrocardiogram. New York: Basic Systems, 1963.

Bilodeau, E. A., & Bilodeau, I. McD. Motor-skills learning. Annu. Rev. Psychol., 1961.

Brown, J. J. Programmed vocabulary. New York: Appleton-Century-Crofts, 1964.

Brown, J. S. A proposed program of research on psychological feedback (knowledge of results) in the performance of psychomotor tasks. In Research planning conference on perceptual and motor skills. Unpublished report, USAF HRDC, San Antonio, 1949.

Carr, W. J. Self-instructional devices: a review of current concepts. Tech. rept. 59-503, Wright Patterson Air Devel. Cent., Ohio, 1959.

Geis, G. L. Possible reinforcements in a program. In F. R. Morton (Ed.), Programming of audio-lingual skills for self-instructional presentation. Ann Arbor: Publications of the University of Michigan Language Laboratory, 6, 1961.

Geis, G. L., & Knapp, S. A note on nonfunctional branching in a linear program. J. programed Instruc., 1963, 2, 15-17.

Holland, J., & Skinner, B. F. The analysis of behavior. New York: McGraw-Hill, 1961.

Krumboltz, J. D., & Weisman, R. G. The effect of immediate confirmation in programed instruction. J. educ. Psychol., 1962, 53, 250-253.

McGraw-Hill. A programmed introduction to the game of chess. New York: McGraw Hill, 1963.

Mandler, G., & Sarason, S. B. A study of anxiety and learning. J. abnorm. soc. Psychol., 1952, 47, 166-173.

Meyer, Susan R. Report on the initial test of a junior high school vocabulary program. In A. A. Lumsdaine, & R. Glaser (Eds.), Teaching machines and programed learning. Wash., D. C.: N.E.A., Dept. of Audiovisual Instruction., 1960.

Moore, J. W., & Smith, W. I. Role of knowledge of results in programed instruction. Psychol. Rep., 1964, 14, 407-423.

Porter, D. A critical review of a portion of the literature on teaching devices. Harvard educ. Rev., 1957, 27, 126-147.

Rosenstock, E. H. The effects of several schedules of knowledge of results on mathematical achievement in a programmed instructional setting. Unpublished report AF-AFOSR-61-54, Bucknell Univ., Lewisburg, Pa., 1965.

Sarason, S. B., & Mandler, G. Some correlates of test anxiety. J. abnorm. soc. Psychol., 1952, 47, 810-817.

Ugelow, A. Motivation and the automation of training: a literature review. Unpublished Rep. (MRL-TDR-62-15) Wright Patterson AFB, Ohio, 1962.

Footnotes

1. We thank Miss Charlotte Cook who aided in preparing this review.

2. We are grateful to Mr. Robert Pachella who assisted in this study.

Table 1

	% Checked	% Wrong	Mean Post Test
Card	62.9%	7.6%	9.8
Card-Tab	40.7%	5.0%	9.5
Tab	45.5%	15.8%	9.0
Tab-Card	44.9%	6.3%	9.7

Results, in terms of per cent of total frames checked, per cent of incorrect answers, and mean posttest scores are presented for the four groups which differed in format of presentation.

Table 2

	Mean Percentage of Frames Checked	Mean T.A.Q.	Mean S.A.T.	Mean Percentage of Frames Checked	Mean Posttest Score
Group Hi	97.7%	67	118	88.3%	9.3
Group Inter.	37.7%	62	117	91.9%	9.9
Group Low	1.5%	51	116	84.6%	8.3

The Ss were sorted into three groups: high, low and intermediate frequency of answer-observing. Results are given for each group for mean score on the Test Anxiety Questionnaire, the Scholastic Aptitude Test, per cent of correct answers in the program and mean posttest score.

Figure Captions

Fig. 1. Cumulative errors in the program are plotted as a function of frames in the program, the frames having been ranked from high-to-low error frames.

Fig. 2. Cumulative errors committed in the program are plotted as a function of subjects ranked in order from high-to-low error Ss.

Fig. 3. The number of Ss falling into each category is shown as a function of the number of frames checked in the program. Categories represent 10 frame intervals with the extreme classes of ALL, NONE, < 10 and, > 10 indicated.

Fig. 4. Per cent of the total frames in the program which were checked by Ss in each of the four format groups is shown as a function of the format group. Separate histograms are plotted (for the card and the tab format) for the T/C and C/T groups.

Fig. 5. For the intermediate group of answer observers (n=14) the probability of checking an answer after the emission of an incorrect response P(C/W) is shown as a function of the per cent of all frames checked.

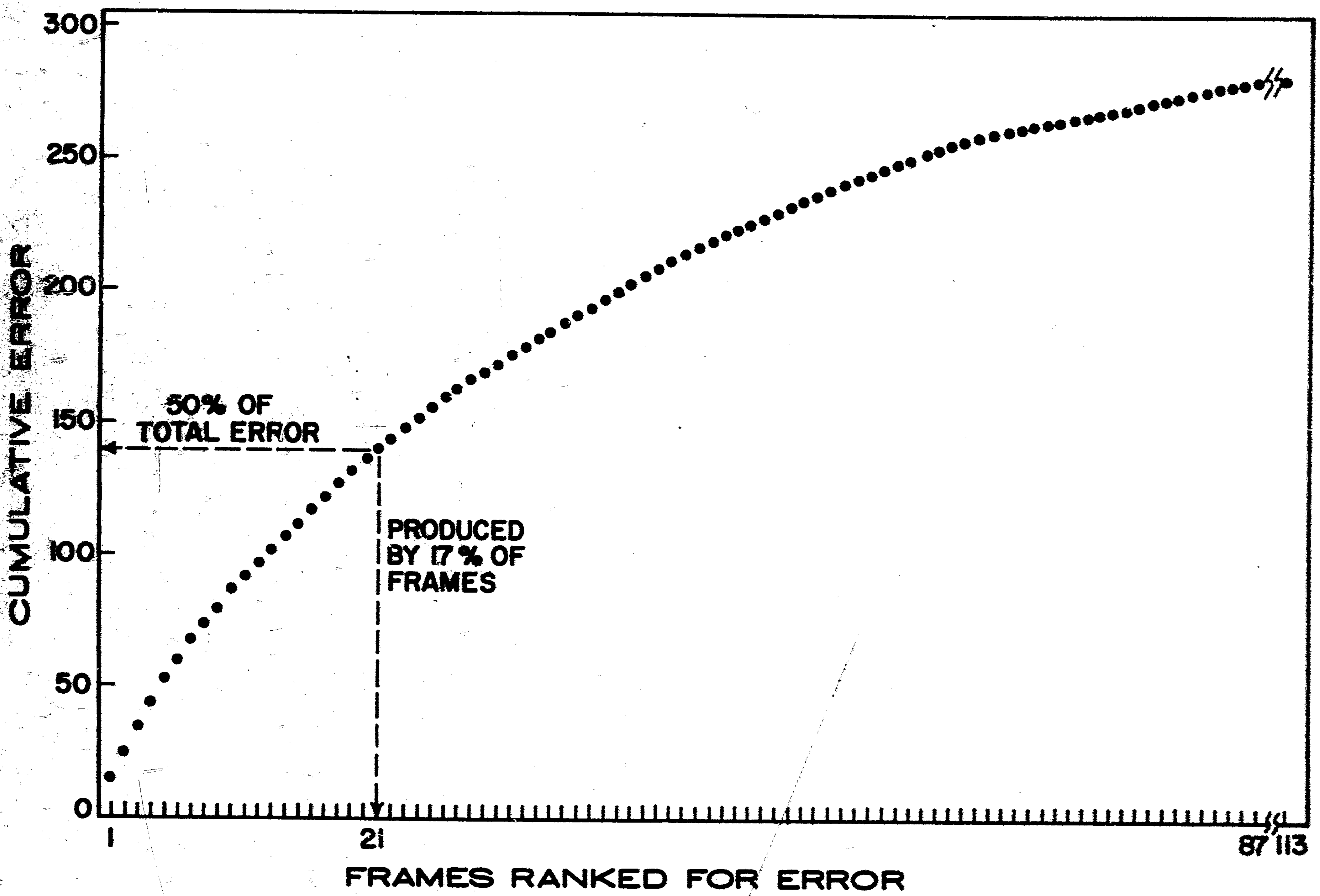
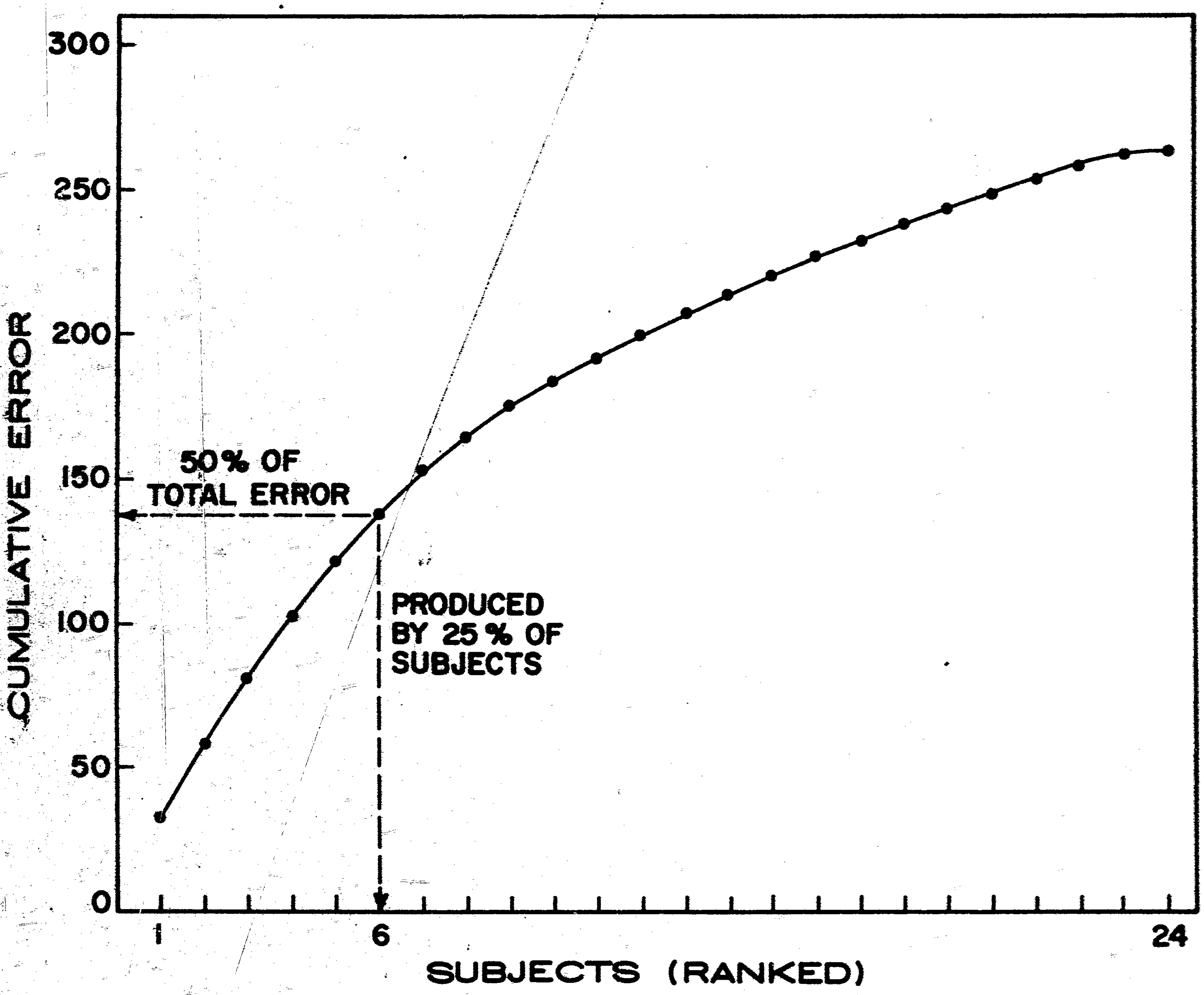

300
250
200
150
100
50
0
CUMULATIVE ERROR
50% OF
TOTAL ERROR
PRODUCED
BY 17% OF
FRAMES
1
21
87
113
FRAMES RANKED FOR ERROR

Fig. 1



CUMULATIVE ERROR
300
250
200
150
100
50
0
50% OF
TOTAL ERROR
PRODUCED
BY 25% OF
SUBJECTS
1
6
24
SUBJECTS (RANKED)

Fig. 2

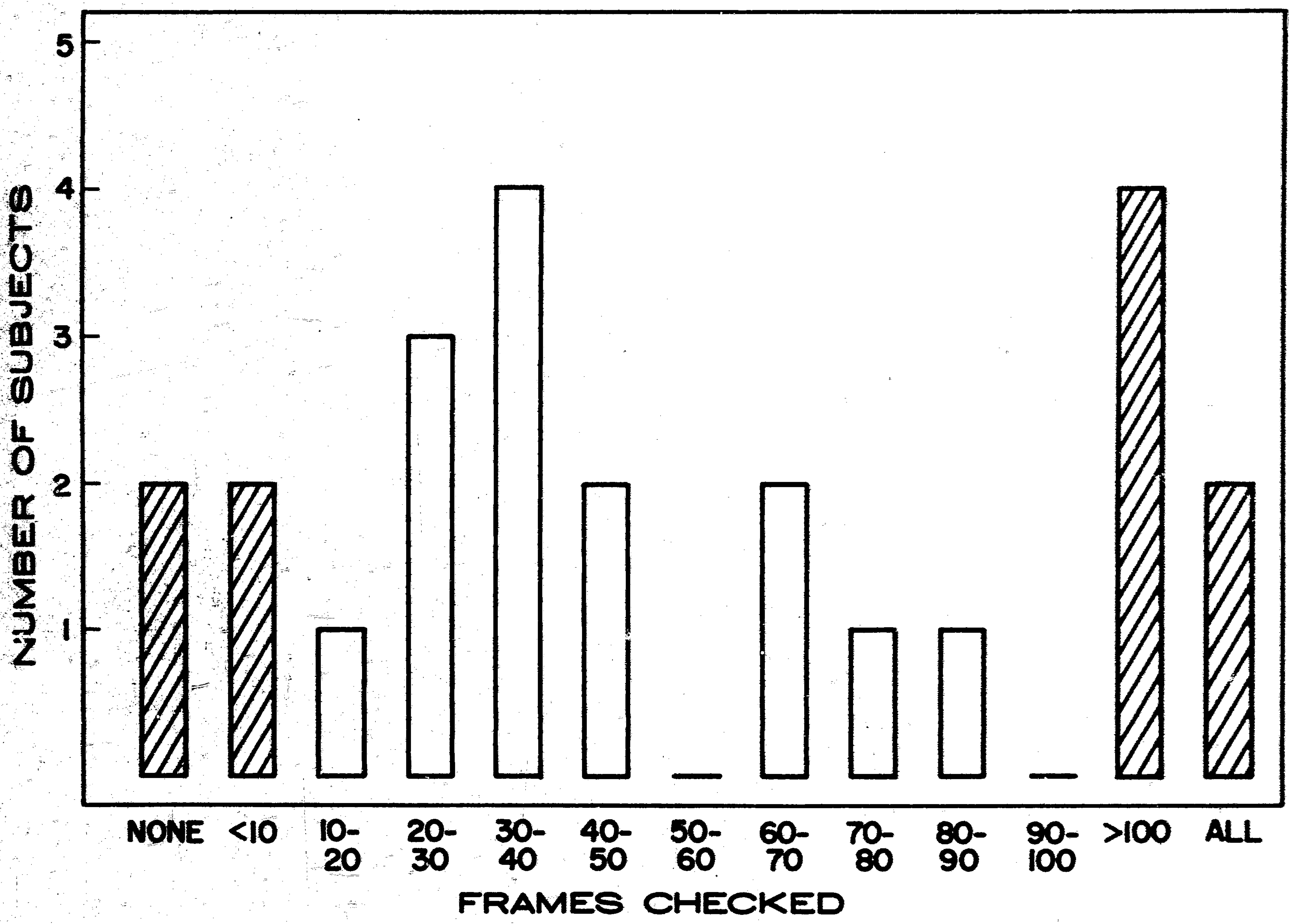
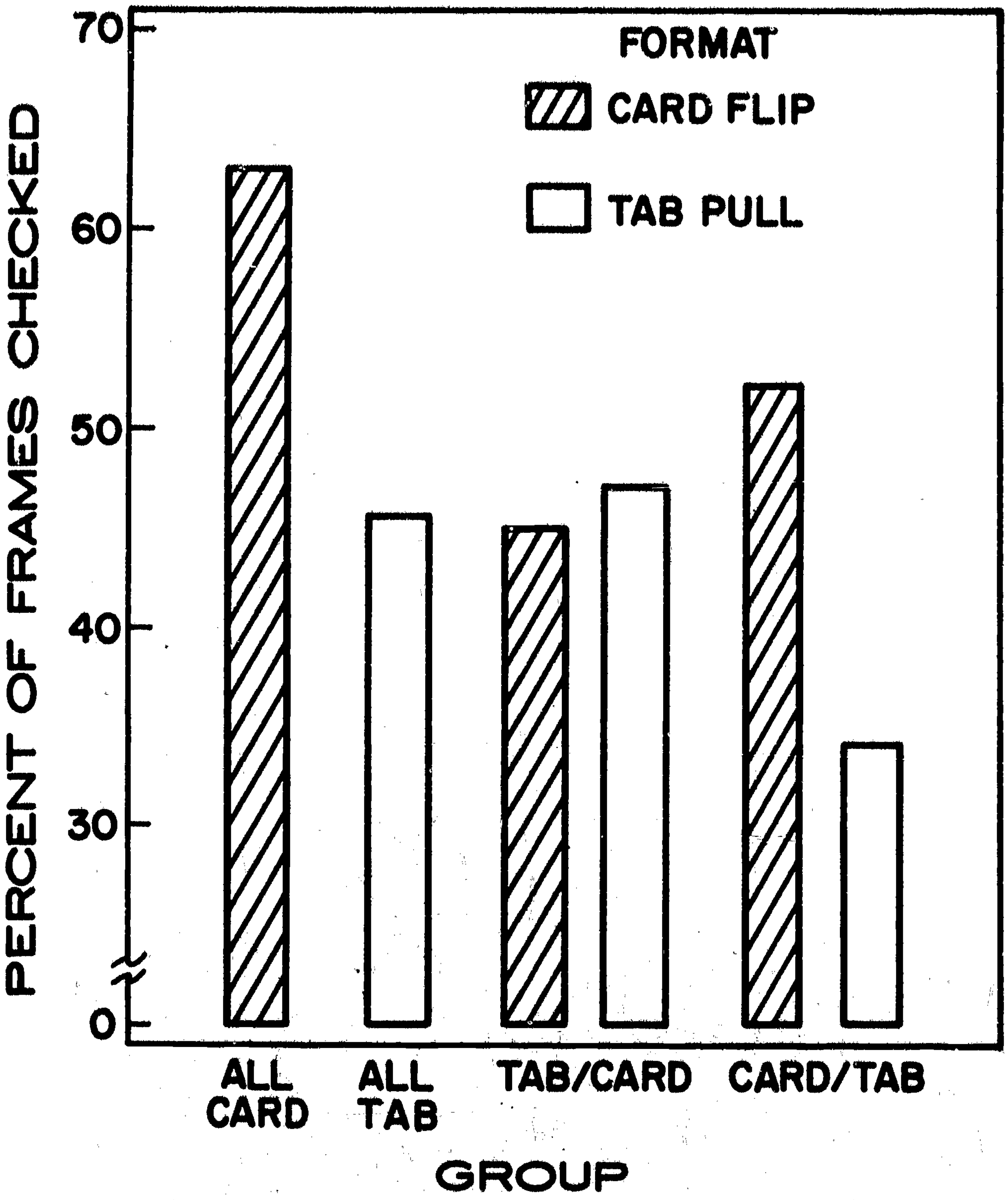
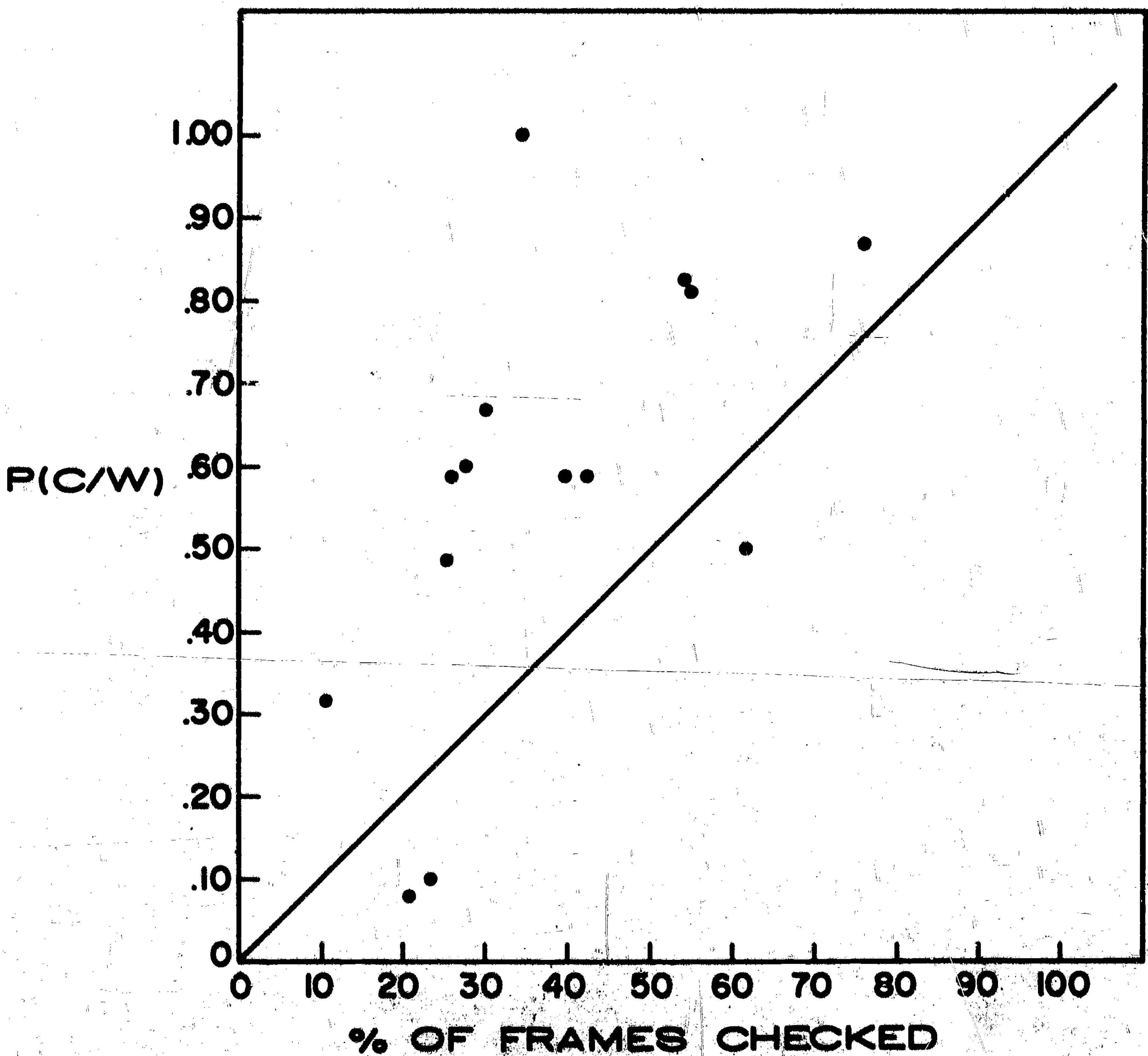

NUMBER OF SUBJECTS
5
4
3
2
1
NONE
<10
10-20
20-30
30-40
40-50
50-60
60-70
70-80
80-90
90-100
>100
ALL
FRAMES CHECKED

Fig. 3



FORMAT
CARD FLIP
TAB PULL
PERCENT OF FRAMES CHECKED
70
60
50
40
30
0
ALL CARD
ALL TAB
TAB/CARD
CARD/TAB
GROUP

Fig. 4



P(C/W)
1.00
.90
.80
.70
.60
.50
.40
.30
.20
.10
0
0
10
20
30
40
50
60
70
80
90
100
% OF FRAMES CHECKED

Fig. 5

Classroom Management:

A Statement of Principles[1]

Dale Brethower

Three basic principles of classroom management are:

I -- People continue doing things for which they are rewarded: the process of reinforcement.

II -- People stop doing things for which they are punished: the process of punishment.

III -- People stop doing things for which they are not rewarded: the process of extinction.

There is nothing new about these principles, but people often overlook the crucial role of the teacher in applying them in the classroom. If the teacher is the primary source of the reinforcers, reinforced behavior undergoes extinction every time the teacher leaves the room, and sometimes when the teacher just has a headache.

Teacher-as-punisher has the same problem. The behavior that is punished also has been reinforced elsewhere; when punishment stops, reinforcement can occur again and thus the behavior recurs. If the teacher is the primary source of the reinforcers and punishers, the child soon learns a very important discrimination: the presence or absence of the teacher.

Teachers recognize this problem in the area of deportment, but they often fail to notice that precisely the same problem exists for the behaviors involved in learning school subjects. When the teacher leaves the children, or when the children leave the teacher, the behaviors involved in learning extinguish quickly--as is evidenced, for example, by the statistics concerning the number of books read by college graduates.

Teachers have always used the principles of reinforcement, punishment, and, to a much lesser extent, extinction. The problem is that they tend to make themselves the source of reinforcement and punishment; it is difficult not to. The teacher's challenge is to find sources of reinforcement that do not depend upon her presence.

In the classes at the University of Michigan Reading Service the challenge is met by providing a source of reinforcement in the materials themselves, with which the student can learn entirely on his own. In addition, an attempt is made to eliminate other sources of reinforcement by teaching in an uncluttered environment, using procedures designed to extinguish irrelevant behavior. Typically, the teacher does not react to irrelevant behavior, thus allowing it to extinguish. If the behavior distracts other children, or if it breaks one of the very few rules, however, the teacher reacts in a predictable way. (The rules are set up in the first place to limit distractions.)

For example, if a child pounds his desk noisily enough to distract other children, he is asked to go into another room where he may pound and he is told that he may return when he finishes. At first, this procedure is usually punishing or reinforcing to the child. (Some people, as you know, are reinforced by disciplinary action.) Because it is carried out calmly, routinely, consistently, and predictably, however, the effectiveness of this treatment as a reinforcer soon diminishes and the behavior then extinguishes. The teacher's action is designed as an extinction procedure and it becomes one as soon as its initial punishing or reinforcing properties are eliminated.

The teacher makes no other use of punishment or reinforcement for irrelevant behavior; instead, it is allowed to extinguish.

This is another way of describing the same procedure: the child is allowed, within the limits of the very few rules, to do whatever he wants. The teacher arranges the environment so that the child wants increasingly to work on the instructional materials.

There is, typically, an habituation period during which a child explores the environment, tests the rules, etc. The first few days of any class are devoted largely to settling into a routine. In many classes every Monday morning is another habituation period.

For some children the habituation occurs very gradually. A child who has experienced a great deal of failure usually habituates slowly. Rather than testing out and getting acquainted with the teacher, the child at first avoids the teacher. If these avoidance reactions are allowed to extinguish, the child can then begin to seek out the teacher and to turn his undivided attention to the instructional material.

In managing a classroom, the teacher seeks to extinguish irrelevant student behavior (such as getting the teacher to solve the problem), to arrange the classroom environment, and to provide instructional materials so that task-relevant behavior will be reinforced. As better instructional material becomes available, the teacher's job becomes easier.

But the techniques of classroom management are always important.

[1]A version of this article appeared in the Southeastern Michigan Reading Association [SEMRA] Journal, 1965, 3, 2-5.

Classroom Management of a Reading Program: The Teacher's Role

Dale M. Brethower[1]

One result of programmed instruction is to stabilize some parts of the educational system so that we can look more closely at other parts. It is giving us a sensitive measuring instrument. The use of the Michigan Successive Discrimination Language Program (Smith & Kelingos, 1964) provides an example. As the children's reactions to the program became more and more consistent, we became aware of what was happening in the rest of the classroom and could see more clearly the many possible influences of the teacher on learning in the classroom.

The first draft of the program was based on the authors' experiences in public school teaching and in teaching remedial reading at the University of Michigan's Reading Improvement Service. The draft was tried out in a public school classroom taught by a competent and experienced teacher. The purpose of the tryout was to find out what the program could and could not do, so that it could be improved. At the same time, we wanted to provide adequate reading instruction for the children. The following pages present some data, some informal observations, and some guidelines concerning classroom management which are based on these tryouts of programmed reading materials.

The Teacher's Role

Among the several results of the cooperative efforts of the programmers, the teachers, and the children, one of the most important was a clearly specified role for the teacher.[2] One part of the role involves presenting the auditory portion of the programmed material. Another involves arranging the classroom environment so that the children can make optimal use of the materials.

In his role as presenter of material, the teacher follows a script. The teacher's behavior is so well prescribed that many teachers elect to relegate this function to a tape recorder. The teacher's second role, that of arranging the environment, presents more challenges and rewards. The program insures that behavior relevant to it has positive consequences. It is designed so that the children can succeed. The teacher's job is to insure that other behavior, behavior not relevant to the task, has no positive consequences. This involves the enforcement of rules, the arrangement of space, and the observation of behavior. Before considering each of these, it will prove useful to examine a case history in classroom learning in some detail.

One Child's Behavior

The effects of the program and environment on some aspects of one child's behavior are shown in Fig. 1. The numbers along the x-axis represent successive class sessions. The class met for one hour each day, Monday to Thursday, at the University of Michigan Reading Improvement Service. The numbers along the

---

Insert Figure 1 about here

---

y-axis represent the frequency of occurence of behaviors, tallied by an observer sitting in the classroom. The solid line represents task-relevant behavior: the number of minutes that the child, Jimmie, elected to spent working on the program.

The other lines represent task-irrelevant behavior: the broken line shows the number of times the child interacted with one of the other eight children in the class and the broken line shows interactions with the teacher. All three kinds of behavior are frequent at the outset, then decline rapidly. Beginning with the fourth session, the task-relevant behavior increases steadily; interactions with the teacher stay low, and interactions with other children fluctuate considerably.

Concurrent observation of the classroom environment indicated that these changes in the child's behavior were brought about primarily by certain features of the teacher's behavior, and secondarily by that of other students. For example, a minimum amount of time is spent on the program perforce since each class session starts with the teacher saying "Sit quietly and fold your hands," and then, when all children have complied "Find your places [in the program]...begin.". However, the child was able to stop work at any time without coercion to resume; the behaviors entailing stopping work were neither punished nor reinforced. This extinction procedure was applied also to interactions with the teacher. Flattery, hand-raising, and even certain requests for assistance all went without response.

Interactions with peers also declined initially and for functionally similar reasons. The other children were busy working and did not respond to Jimmie's actions. On day 3 he spent almost all hour playing with modeling clay. Subsequently, he was more successful in obtaining attention from his peers. (On day 15, the teacher arranged for Jimmie to work in a three-sided booth which isolated him from these social reinforcers.) In the absence of reinforcement for task-irrelevant behaviors, Jimmie increasingly came under the control of the reinforcing properties of the reading program.

Classroom Management in an Experimental Classroom

Enforcement of Rules. Consistency of rules in classroom management seems to be much more important than precisely what the rules are. Whenever a child breaks a rule, the teacher enforces the rule by saying the child's name followed by the question "What's the rule?" (in an emotionless voice). She waits for the answer, and then continues whatever she was doing. That seems to be all that is required for minimizing rule-breaking in our experimental classroom. For a while we also used what Dr. Ogden Lindsley of the University of Kansas calls an "inconsequation room"--a room outside the class where behavior has fewer consequences. The teacher would take the child to the room and say, "You may continue in here where it won't disturb anyone. Come back to the classroom when you finish." This treatment may resemble a non-punishing trip to the principal's office. Like stealing watermelons from your own garden, it quickly loses whatever original appeal it may have had for the child. The use of an inconsequation room has not proven necessary in our classroom, however; asking the child to say the rule seems to be all that is required.

Arrangement of Space. The environment in our experimental classroom has freedom within structure. The children do what they please as long as they do not disturb other children. For example, they may sit under their desks and work; although this may disturb the teacher at first, if she does not show it the rest of the class does not become disturbed.

The teacher arranges seating, takes down distracting bulletin board displays, uses the coat rack to provide a protected corner for a fearful child, and does whatever is necessary so that the children will work on their own. She observes the children carefully to see what each one needs to make him safe and secure enough to try a new reading task. When a child completes particular segments of the program, the child asks the teacher to listen while he reads. If the child fails to read correctly, the teacher hands the book back to the child without comment, signaling that more work needs to be done. If the child reads it correctly, the teacher presents without comment the next segment of the program.

It is unnecessary for the teacher to criticize the child. The program can fail and occasionally does. The child is not blamed for the program's failure or for the teacher's failure to provide an environment where the child can attend to the program. The child can and does go back over the program, finding what was missed the first time.

It is unnecessary for the teacher to praise the child for success. Success in reading is sweet enough. There are always things to read but there are not always teachers around to pat the child on the head. One goal is independent reading.

The thirty children work independently, giving the teacher time to observe them carefully and to record how many pages each of them completes each day. By looking at these records, she can see problems beginning to arise. She can take corrective action and then look at the records to see if it was, in fact, successful.

The teacher as expert. The teacher's role in the classroom that we have described requires that she observe each child carefully and systematically, monitoring his progress in quantitative form. This becomes possible when much of the teaching is accomplished by self-instructional programmed materials.

It has been our experience that competent classroom teachers can fulfill this and the other facets of their role described above without extensive and specialized training. Initially they need help in learning to observe and to interpret their quantitative observations. Once they have gotten started they are capable of being experts. They are capable of collecting the data other experts in educational research and classroom management have to obtain research grants to get.

The departure from more traditional patterns of teacher behavior is difficult at first, but it is soon sustained by ample and appropriate rewards--teaching effectiveness.

References

Smith, D. E. P., & Kelingos, J. Michigan successive discrimination language program. Ann Arbor, Michigan: Ann Arbor Publishers, 1964.

Smith, D. E. P., & Kelingos, J. Michigan successive discrimination language program manual: a program for teachers. Ann Arbor, Michigan: Ann Arbor Publishers and Operant Associates, 1964. (a)

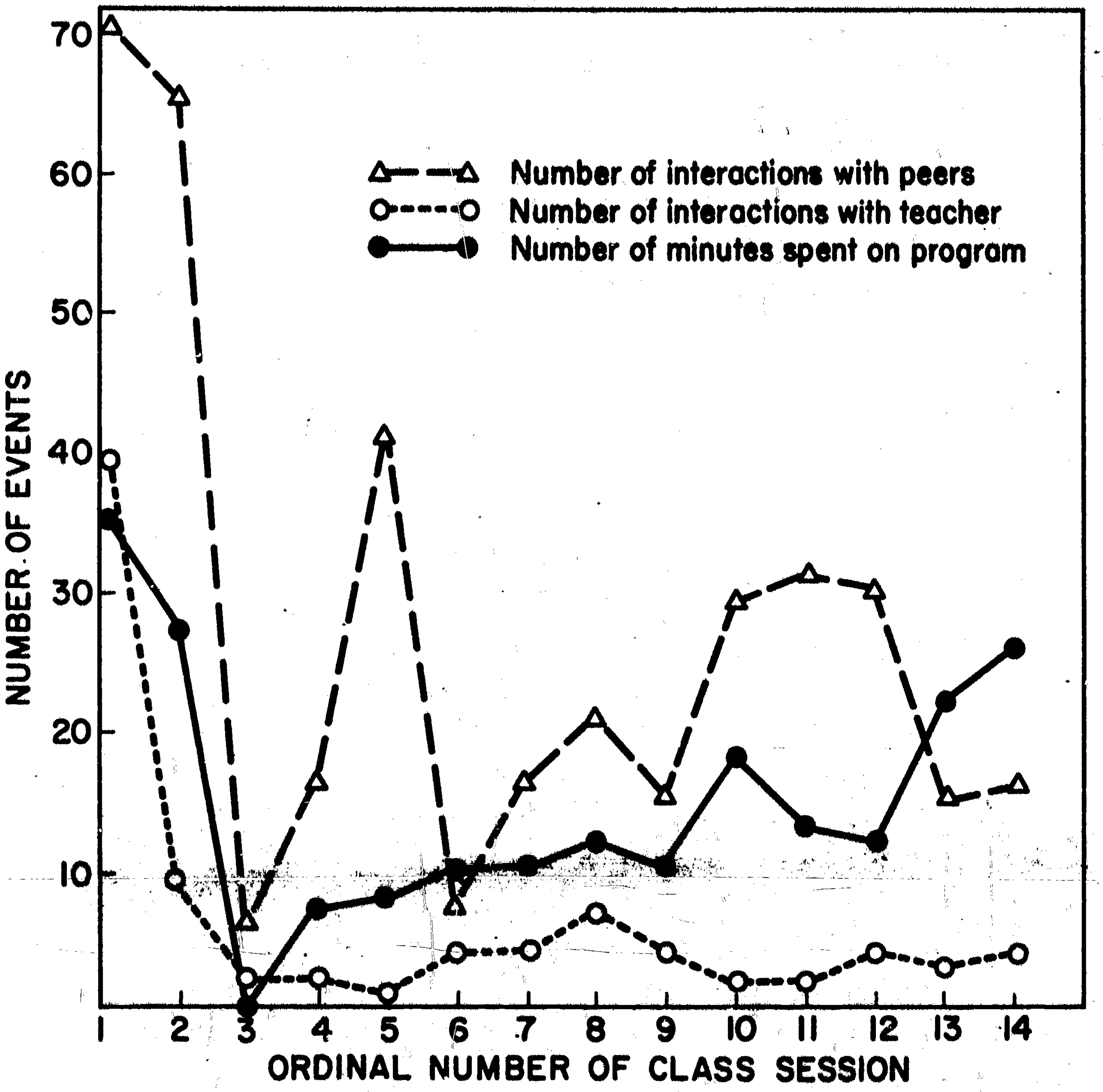
Footnotes

1. This research was carried out in cooperation with the Reading Improvement Service, Bureau of Psychological Services, The University of Michigan. The present report is adapted from a paper presented at the 1965 convention of the National Society for Programmed Instruction in Philadelphia, Pa.

2. See Smith & Kelingos, 1964. (a)

Figure Caption

Fig. 1. The number of interactions with peers and with teacher and the number of minutes spent on programmed self-instruction in reading by one child in consecutive hour-long class sessions.



Number of interactions with peers
Number of interactions with teacher
Number of minutes spent on program
NUMBER OF EVENTS
70
60
50
40
30
20
10
1
2
3
4
5
6
7
8
9
10
11
12
13
14
ORDINAL NUMBER OF CLASS SESSION

Fig. 1

Toward A Modern Theory of Rhetoric:*
A Tagmemic Contribution

Richard E. Young

Alton L. Becker

> Our discussion will be adequate if it has as much clearness as the subject-matter admits of, for precision is not to be sought for alike in all discussions.... We must be content...in speaking of such subjects and with such premises to indicate the truth roughly and in outline, and in speaking about things which are only for the most part true and with premises of the same kind to reach conclusions that are no better. In the same spirit, therefore, should each type of statement be received; for it is the mark of an educated man to look for precision in each class of things just so far as the nature of the subject admits; it is evidently equally foolish to accept probable reasoning from a mathematician and to demand from a rhetorician scientific proofs.
>
> Aristotle, Nichomachean Ethics i.e.1094b 1228.
> trans. W. D. Ross

Traditional Rhetoric: A Brief Survey

Years ago the heart of a liberal education was the trivium of grammar, logic, and rhetoric. Modern linguistics has come to encompass more and more of this trivium, and has itself become transformed in the process. Traditional grammar is no longer anathema to the linguist, and linguistic description has adopted many of the techniques of logical analysis. Furthermore, linguistics is becoming increasingly interested in the analysis and description of verbal structures beyond the sentence, traditionally a rhetorical concern. It seems fitting, therefore, to explore the relationships of linguistics and rhetoric, discovering, hopefully, just what contributions a theory of language can make

*See Harvard Educ. Rev., 35, 4, Fall, 1965

to a modern theory of rhetoric.

As Aristotle states in the quotation given above, the nature of the subject matter imposes some constraints on the statements we make about it. It is our intention, therefore, to define the subject matter of rhetoric as it has been understood traditionally and then illustrate how aspects of one modern linguistic theory, tagmemics, can form the basis for a new approach to rhetorical problems. The field is broad and there are many points of contact between linguistics and rhetoric which will be passed over here. Nor can we hope to consider all linguistic points of view, each with important contributions to make. We will limit ourselves to a description of the three traditional stages in the rhetorical process--invention, arrangement, and style--and then approach the problems of each stage via tagmemic theory.

There are four rhetorical traditions, which taken together constitute the history of rhetoric. There is sophistic rhetoric, which has as its goal the effective manipulation of language without regard to truth and logic. The tradition continues in modern propaganda and advertising techniques. There is Platonic anti-rhetoric, which stresses not the art of writing but the quality of the writer in his adherence to truth and virtue: a good writer is a good man writing. There is the rhetoric of literary criticism, which applies the categories and techniques of rhetoric to the analysis and evaluation of poetry, drama, and narration. And finally, there is Aristotelian rhetoric, which had its origins in the law courts of early Greece and which was expanded, systematized and given a philosophic foundation by Aristotle. After being brought to perfection by Cicero and Quintillian, it was a basic, at times the basic, discipline in Western education for fifteen hundred years. It survives today but with greatly diminished influence. Because this is still the most complete rhetoric ever developed and because it best defines what traditionally has been the scope of rhetoric, we shall focus our attention almost exclusively on the Aristotelian tradition.

For Aristotle, rhetoric was "the faculty of observing in any given case the available means of persuasion."[1] Its immediate end was to persuade a popular audience of what is true and just; its ultimate end was to secure the cooperation necessary for a civilized society. The classical art of rhetoric consists of five separate arts, which taken together embrace the entire process of developing and presenting a persuasive discourse: invention, arrangement, style, memory, and delivery. As the last two concern speaking rather than writing, which has become the principal concern of modern rhetoric, we shall consider only the first three: invention, arrangement, and style, which apply to writing as well as speaking.

"Invention," wrote Cicero, "is the discovery of valid or seemingly valid arguments to render one's cause plausible."[2] Rhetoricians distinguished two kinds of arguments: extrinsic arguments which came ready-made to the writer (e.g., eye-witness testimony, documents, confessions) and intrinsic arguments. The latter were of special interest to rhetoricians because they were subject to discovery by a system of topics. These topics were a kind of checklist of mental acts one could use when investigating and collecting arguments on a subject (e.g., definition by genus and differentia, comparison and contrast, cause and effect). Certain of these topics, the "common" topics, were appropriate to all types of speeches; others, the "special" topics, were appropriate to only one of the three types of speeches studied in the classical system, i.e., to forensic, or political, or ceremonial speech.

Use of the topics presupposed wide learning since they were primarily a method for putting the writer in contact with knowledge which already existed. Edward Corbett has remarked that Mortimer Adler's Syntopicon of great ideas of the Western World would have been an ideal reference work for the ancient rhetorician.[3] It was the art of invention which made rhetoric the core of humanistic education until the late Renaissance.

In the Renaissance under the influence of Bacon and Descartes, logic increasingly came to be seen not as the art of learned discourse, as it had been since Greek times, but as an instrument of inquiry. Rhetoric gradually enlarged its boundaries to include both the arts of learned and popular discourse. The process was finally completed in the nineteenth century in the work of John Stuart Mill, commenting on the proper domains of logic and rhetoric, Mill remarked that

> "...the sole object of Logic is the guidance of one's thoughts: the communication of those thoughts to others falls under the consideration of Rhetoric, in the large sense in which that art was conceived by the ancients...."[4]

That spirit of modern science which was modifying the nature of logic and the scope of rhetoric also had its effect on the art of invention. For since the seventeenth century we have increasingly regarded facts and experimental evidence as the basis for sound argument and not, as our ancestors did, the wisdom of the past. That is, we have increasingly put our faith in extrinsic arguments. We have become much more interested in techniques for discovering what is unknown than in techniques for bringing old beliefs to bear on new problems. Thus the classical art of invention has diminished in importance while the modern art of experimental inquiry has expanded immensely. But this art of inquiry is no longer a part of modern rhetoric, each academic discipline having developed its own discovery procedures. The strength and worth of rhetoric seem, however, to be tied to the art of invention; rhetoric tends to become a superficial and marginal concern when it is separated from systematic methods of inquiry and problems of content.

The second art in classical rhetoric was arrangement. Rhetoricians developed persuasive patterns for organizing their materials, flexible systems of slots into which appropriate categories of subject matter were fitted. One common arrangement, the Ciceronian, had six slots: the exordium, the narrative or exposition of the problem's history, the proposition, the demonstration, the refutation of alternative propositions, and the peroration. The functions and structures of each of these slots were systematically developed and described. Arrangement was the art of distributing in this pattern the subject matter gathered in the process of invention; it also involved modifying the pattern by expanding, omitting, or rearranging the various steps to meet the needs of the audience, speaker, and subject matter. The pattern was employed in all three types of speech, i.e., forensic, political, and ceremonial.

Since rhetoric was the art of persuasion, patterns of other modes of discourse (e.g., description, exposition) were given little attention. In the seventeenth century, however, developments in science led to an increasing interest in expository prose, a movement which parallels the shift from intrinsic to extrinsic argument. Other developments, such as the decline in the power of the aristocracy and the growing importance of evangelical religion, led to a rejection of elaborate patterning and the development of simpler, more manageable rhetorical forms, though none were described in the same detail as the classical patterns.

Implicit in classical theory was a dualistic conception of discourse. Form was treated as independent of both subject matter and the writer. Since the Renaissance there has been a tendency to see form as the product of a particular mind or as discoverable in the subject matter itself. In the latter case the form of a discourse is not separable from the content—the discourse is seen as having an organic unity. In either case the form of a work is not predictable. If form is a personal matter or implicit in the subject matter, the rhetorician can make fewer generalizations about arrangement. Classical rhetoric was a rhetoric of precept; in modern times it has become, for the most part, a rhetoric of practice.

Style, the third of the rhetorical arts in classical rhetoric, was largely the technique of framing effective sentences. Its function was to give clarity, force, and beauty to ideas. Although grammar was its foundation, it was clearly a separate art, concerned with the effective use of language while grammar was concerned with the correct use. Both, however, were concerned with language at the sentence level.

Aristotle justified the study of style on practical grounds: ideally rational argument alone should be sufficient to persuade. Since experience suggests that this is often not sufficient, the art of style must be employed if wisdom is to be persuasive. The art of style tended, however, to become an end in itself, at times preempting the entire field of rhetoric, possible--as in the classical conception of arrangement--because of a dualistic view in which content and style were separable.

In the classical tradition, good style was a deliberate departure from the speech of everyday life. Renaissance classicists ingeniously isolated and systematized figures of speech. Henry Peacham's Garden of Eloquence (1577), for exa le, lists 184 schemes and tropes, artistic departures from ordinary syntax and meanings of words. Clarity and appropriateness became less frequent constraints than elegance and ingenuity. As a result, "rhetoric" gained its pejorative connotation: elegant but empty verbosity.

As with the other rhetorical arts, there was a reaction against this concept of style, and rhetoricians sought a norm closer to the speech of every-day life. In the eighteenth century the dualistic conception of style and content began to compete with monistic conceptions. Style came to mean either the characteristic expression of a particular personality (Le style c'est l'homme même) or the mode of expression organically a part of the subject matter itself.

Since the eighteenth century the analysis of style has become almost exclusively the concern of literary criticism. In rhetoric courses today style is still seen, by and large, as the art of framing effective sentences, but the art is much simpler, less systematic, and considerably more intuitive than it was in classical rhetoric.

The classical art of rhetoric has a number of weaknesses which make it inadequate for our time. Without involving ourselves directly in a criticism of the philosophical assumptions upon which classical rhetoric is based, we can note, in general, four major problems: the classical art of invention stresses authoritative confirmation of present beliefs while modern modes of inquiry stress imaginative discovery of new facts and relationships; the art of arrangement includes only patterns of persuasion, neglecting considerations of form in other important rhetorical modes such as description, narration, and exposition. Both the art of arrangement and the art of style divorce form from content, failing to consider the importance of the act of discovery in shaping form. And, finally, the art of style is concerned primarily with embellishing, clarifying, and giving point to sentences, neglecting both the deeper personal roots of style and the ways in which style is manifest in patterns beyond the sentences.

In recent years numerous rhetoricians have been seeking a new rhetoric as effective practically and as stimulating and coherent theoretically as classical rhetoric. As Daniel Fogarty puts it, there are numerous "roots for a new rhetoric."[5] While other members of the trivium have changed greatly from their earlier forms—witness the revolution in Mill's System of Logic and later in symbolic logic and the recent revolution in grammatical theory—there has as yet been no comparable change in rhetoric. That is, there has been no change which includes both a complete theory and an explicit practical method. Rhetoric is still in the midst of a chaotic transition period. I. A. Richards is right, unfortunately, when he describes the general state of rhetoric today as

> the dreariest and least profitable part of the waste that the unfortunate travel through is Freshman English! So low has Rhetoric sunk that we would do better just to dismiss it to Limbo than to trouble ourselves with it—unless we can find reason for believing that it can become a study that will minister successfully to important needs.[6]

A Tagmemic Approach to Invention, Arrangement, and Style

The tagmemic approach to language analysis and description, developed primarily by Kenneth L. Pike and his associates in the Summer Institute of Linguistics,[7] has for many years been concerned with problems which have traditionally been within the scope of rhetoric. This concern results, in part, from the strong motivation the model gives for moving beyond the specification of well-made sentences. In tagmemic theory any linguistic unit is assumed to be well defined only when three aspects of the unit are specified: its contrastive features, its range of variation, and its distribution in sequence and ordered classes. This constraint on grammatical description (necessary and sufficient to include all relevant aspects of any linguistic unit) has meant that a complete description of sentences, for example, should include a specification of their distribution in paragraphs and other larger units of discourse.

This concern with problems which traditionally have been a part of rhetoric also results from the desire of many who use the tagmemic model to provide a means for producing extended discourse, primarily Bible translation. Translators frequently encounter instances of grammatical constraints extending beyond the sentence. In some Philippine languages, for example, there is a system of focus, somewhat like active and passive voice in English though vastly more complex. To produce understandable discourse in these languages requires a consistent marked correlation between situational roles (actor, goal, instrument, setting, etc.) and grammatical roles (subject, predicate, object, locative, etc.) in a sequence of sentences.[8]

As the linguist moves beyond the sentence, he finds himself asking questions which have long concerned rhetoricians. The description of the structure of a sentence and the description of the structure of an expository paragraph, an extended argument, or a novel are not sharply different kinds of activity, for all involve selecting and ordering language in a significant way. The traditional separations between grammar, logic, rhetoric, and poetics begin to break down.

Selecting and ordering language, however, has two aspects. One sort of inquiry into the selection and ordering of language leads us deeply into the mental activity of the writer and to questions which are difficult, perhaps impossible to answer except intuitively. Can we specify the reasons why a writer choses to write "John loves Mary" rather than "John is in love with Mary"? Probably not; we can only describe the choices he does make, the characteristic features of his style. Another sort of inquiry, however, leads us to the conventions which constrain the writer. We can specify the reasons why "Love John is Mary in" does not make sense to us, except in rather far-fetched ways. In the same way we believe we can also specify why the following sequence of sentences doesn't make sense either:

The trees are budding. Coal is a form of
carbon. He has been singing for three hours now.
The world used to be round. It seems enough.

If we were to prod the reader, insisting that the above "sentence" and "paragraph" do have meaning, he could probably find some sense in them, as many have in Chomsky's "Colorless green ideas sleep furiously." In each case one "discovers" meaning by imposing conventional formal patterns on the deviant sequences.

Both the process of imposing (or discovering) pattern on apparently meaningless utterances and the process of describing the conventions of language are important to the linguist. To do the former, he uses some sort of discovery procedure; to do the latter, he employs a descriptive model which specifies the structures of conventional utterances. Although the act of discovery is in part intuitive, the model does provide both a method for finding significant linguistic patterns and a taxonomy of the sort of patterns the analyst is likely to find, the still tentative universals of language. Discovery procedures are not mechanical; there is as yet no completely systematic way of analyzing a language, just as there is no algorithm for planning an effective literary composition. But there are important guides to the processes: one can learn to analyze a language and he can learn a great deal about how to write an essay or a novel.

We believe that the procedures the linguist uses in analyzing and describing a language are in some important ways like the procedures a writer uses in planning and writing a composition, and hence that tagmemic theory can provide the basis for a new approach to rhetoric. Tagmemic discovery procedures can provide a heuristic, comparable to the Aristotelian system of invention; and the tagmemic descriptive model can give us a vehicle for describing conventional rhetorical patterns. If our beliefs are sound, this approach will provide a bridge between the traditionally separate disciplines of grammar and rhetoric.

A heuristic is a method of solving problems, a series of steps or questions which are likely to lead an intelligent analyst to a reasonable solution of a problem. There are two different though related kinds of heuristic: a taxonomy of the sorts of solutions that have been found in the past and an epistemological heuristic, a method of inquiry based on assumptions about how we come to know something. Bacon's statement of the distinction is worth quoting:

> The invention of speech or argument is not properly an invention: for to invent is to discover that we know not, and not to recover or resummon that which we already know; and the use of this invention is no other but out of the knowledge whereof our mind is already possessed, to draw forth or call before us that which may be pertinent to the purpose which we take into our consideration. So as, to speak truly, it is no Invention, but a Remembrance or Suggestion, with an application; which is the cause why the schools do place it after judgment, as subsequent and not precedent. Nevertheless, because we do account it a Chase as well of deer in an inclosed park as in a forest at large, and that it hath already obtained the name, let it be called invention: so as it be perceived and discerned, that the scope and end of this invention is readiness and present use of our knowledge, and not addition or amplification thereof.[9]

Aristotelian rhetoric provides a taxonomy of effective rhetorical arguments which a speaker can use to attain specific ends with specific audiences. Tagmemic theory, on the other hand, provides an epistemological heuristic.

Tagmemic epistemology is based largely on two principles, though other principles are necessary for a complete statement of the theory. These two principles emphasize the active role of the observer in discovering pattern, and hence meaning, in the world around him. The first principle contrasts external and internal views of human behavior, in tagmemic jargon, etic and emic views. This distinction can be seen in the differences between phonetic and phonemic contrasts in linguistic phonology. A phonetic inventory provides a systematic statement of the phonological distinctions which occur in various human languages, while a phonemic description provides a systematic statement of the significant phonological distinctions in a particular language. A distinction is judged significant, and hence phonemic, if it signals a difference in the lexical meaning of linguistic units. Though there is much controversy about how phonological signals are to be described, the basic distinction remains valid: the contrast, for example, between aspirated and unaspirated consonants is lexically significant for a native speaker of Hindi or Burmese but not for a native speaker of English, who has difficulty in learning to hear this contrast.

The distinction is especially important when two emic systems come in contact, as when the speaker of English is learning to speak Hindi and is forced to recognize that his native distinctions are emic and not etic only (i.e., not universal). Likewise one who finds himself in a different culture must learn to distinguish between universals of human behavior and particular customs and mores which taken together comprise the emic distinctions of a culture. The ways of treating time and space, for example, vary throughout the world and one must learn them if he wishes to communicate and cooperate outside his own culture.

Though it is interesting to envision a universal etics of rhetoric, an orderly classification of the rhetorical forms found throughout the world, for the present our concern must be with the writer of English writing for readers of English. Even with this restriction, one confronts frequent clashes of emic systems, for if a writer has anything new to say, his image of the world is in some way different from his reader's. It is at this point of difference that his message lies. He may seek to expand or clarify some feature of the reader's image, thus making it more nearly like his own, or he may seek to replace some feature of the reader's image. In the first instance he would be informing; in the second, persuading.

Before developing this discussion of rhetorical intention further, we must introduce a second major principle in tagmemic epistemology. This principle asserts that a complete analysis of a problem necessitates a trimodal perspective. After the principle had been worked out in tagmemic theory and the modes had been labelled feature mode, manifestation mode, and distribution mode, Pike noted a striking similarity between these modes and the trimodal perspective of modern physics, the complementary views of physical phenomena as particles, waves, and fields, and he adopted these terms for his behavioral model.[10]

Language phenomena, and presumably all human behavior, can be viewed as particles (discrete contrastive bits), waves (unsegmentable physical continua), or fields (orderly systems of relationships). For example, a sentence can be viewed as a sequence of separate words or morphemes, as a physical continuum with definite peaks but indeterminate boundaries between the peaks (as seen on a sound spectrogram), and as a system of interrelationships manifesting the grammatical, lexical, and phonological rules of English. Tagmemic theory asserts that only by this complementarity of perspectives is a complete analysis of language structure possible.

The principle of trimodalism gives the analyst both a procedure for approaching new problems and a corrective for a limited view of the data. Only when he has described his data from all three perspectives can he be reasonably sure that his analysis is complete. Likewise, the writer can use the principle as an aid in discovering a wide range of perspectives on his topic. Though a writer often emphasizes one mode in a particular work, he should be aware of the other possibilities, particularly if his readers customarily emphasize a different mode. Let us consider a simple example. A particle description of a flower emphasizes those features which make it distinctive from other flowers. A wave description emphasizes the flower as a moment in a process from seed to final decay (even this is only a peak in a larger wave) or as merging into a scene. A field description may partition a flower into its functional parts or classify it in a taxonomical system. The flower may also be seen metaphorically or symbolically, in which case it is conceived as part of a new conceptual field, religious, say, or geometric, where certain of its features--its beauty or its shape--are hypostasized, allowing it to manifest a category in a new field. One can view any topic trimodally and soon discover a wide range of significant perspectives.[11] The process is broad, flexible, and intuitive, though the intuition is guided by what has proved to be a very fruitful principle. It is especially useful since it is not limited to a particular subject matter. In this sense it is similar to the topics of Aristotelian logic or the "common" topics of classical rhetoric. A generally applicable approach helps to free us from the built-in limitations of a conventional, specialized approach. The discovery procedure, then, also has a corrective function.

This heuristic procedure, based on the emic-etic distinction and trimodal perspective, in addition to helping the writer explore his topic generates a set of questions which he can use to analyze his reader's preconceptions, his emic system:

1) What are the distinctive features of the reader's conception of the topic? What characteristics does it have that lead him to contrast it with similar things? (Particle view).

2) How are the reader's views on this topic part of a mental process, a phase in the continual development of his system of values and assumptions? (Wave view)

3) How does the reader partition the topic? What are its functional elements for him? How does he classify it? (Field view)

The answers to these questions provide criteria for selecting and ordering the writer's subject matter as he develops his discourse.

The missionary linguist in the field seeks to translate his message into the language and cultural conventions of the people, not to teach them English and his own emic conventions. He does not seek to replace their emic system with his own, but to modify their image after finding within it the motivations for receiving his message.[12] For he realizes that change is most effective and enduring when it occurs within the emic system of those he is trying to convince. Unlike traditional rhetoric, which sought to persuade people principally by confirming authoritative attitudes , modern rhetoric, we believe, must seek identification. That is, the writer must seek to have his readers identify his message with their emic system.

Because it seeks identification rather than persuasion and because this assumption often leads the writer to modify his own position, modern rhetoric, still in the process of development, is characterized by Kenneth Burke and others as "discussion rhetoric." The basis for a rhetoric of this sort has been developed by Anatol Rapoport in his book Fights, Games, and Debates, where it is called Rogerian debate, its assumptions having been derived from the methods of the psycho-therapist Carl Rogers.[13] This principle of identification of the writer with his audience points toward a rhetoric not of opposition but of mutual respect.

A comparison of emic systems, different systems of selecting and grouping followed by writer and reader, leads the writer to find what he shares with his reader in his conception of the topic and what he does not share. One of the assumptions of tagmemics is that change can occur only over the bridge of a shared element. There can be no action at a distance. The key to understanding language change, for example, is the identification of the shared features of the initial state and the subsequent altered state. The writer's message is an unshared item in the comparison, while the shared items, in so far as they are relevant to the message, provide the means by which the reader can identify--and identify with--the message. Shared items are the potential bridges over which change can take place. These bridges may be broad cultural conventions or more specific things like common social roles, problems, philosophical assumptions. Among the most important of these shared items is a common language, a common set of patterns and rules governing selection and grouping of words or morphemes within a sentence and of sentences and paragraphs within still larger units of discourse. It is here that the linguist can make his unique contribution to a new theory of rhetoric, especially as he broadens his focus to include larger units than the sentence.

So far we have dealt chiefly with what might be called prewriting problems, problems of discovery. We believe, as did Aristotle and Cicero, that a complete theory of rhetoric must include the entire sequence of acts which result in the finished discourse, beginning with the initial act of mental exploration. We have offered two principles of tagmemic heuristic as an indication, hardly an exhaustive one, of how linguistics can contribute to this aspect of rhetorical theory. We now turn to a description of rhetorical patterns beyond the sentence, extending techniques which have been used in the past for the description of lower-level patterns.

Tagmemics is essentially but not entirely a slot-and-substitution grammar, describing linguistic patterns as sequences of functional slots which can be filled, or manifested, by a class of fillers. These functional slots can be seen as functional parts of a pattern and stated in a formula like the following simplified one for an English transitive sentence:

+ Subject + Verb + Object ± Manner ± Locative ± Temporal

(E.g., He walked the dog slowly around the block yesterday.)
Some of these slots are obligatory (+), some optional (±). Each may be manifested by one of a set of filler constructions; thus the subject slot can be filled by a noun phrase, a pronoun, an adjective phrase, a verbal phrase, a clause, etc. More fully represented, the subject slot in the formula above would be:

+ Subject: np,p,ap,vp...c

Tagmemics assumes that language is composed of interlocking lexical, phonological, and grammatical hierarchies. Here the internal surface structure of the fillers of the subject slot of the sentence are described at the clause, phrase, word, and morpheme levels of the grammatical hierarchy.

In at least two important ways, however, tagmemic grammar goes beyond the surface level descriptions of other slot and substitution grammars.[14] First, tagmemic grammars go on to represent the filler class of a functional slot as an ordered set, or, in tagmemic jargon, a matrix. The categories of these ordered sets indicate concord relationships of one tagmeme with another; thus the filler class of the subject tagmeme is ordered into categories such as singular-plural and human-nonhuman in concord with these same categories in the predicate, so that, for example, a singular, nonhuman subject specifies the selection of a singular, nonhuman verb, preventing such collocations as "the tree jump fences."

Second and more important for our present discussion, tagmemic grammars specify in addition to the surface structure of patterns an ordered set of operations to be carried out on the patterns. These include ordered reading rules by which all possible readings of a formula are generated. Then each reading is reordered according to permutation rules. Finally, in each reading and its permuted variants the tagmeme symbols are replaced by each of the possible filler constructions according to a set of exponence rules. These operations are carried out repeatedly until only morphemes or symbols for morpheme classes manifest the formulas, which are then terminal grammatical strings, not yet sentences until phonological and lexical specifications have been met.

Though a description of English will not specify sentences like the one mentioned earlier, "Love John is Mary in," it so far contains no constraints to prevent it from accepting a sequence of sentences like:

> The trees are budding. Coal is a form of
> carbon. He has been singing for three hours now.
> The world used to be round. It seems enough.

This is not a paragraph because there is no formal connection between the sentences. We can discern no conventional pattern relating them, as we can, for example, in this pair of sentences:

> What is John doing?
> He's washing his face.

This sequence manifests a conventional rhetorical pattern, Question and Answer. The question is marked by three formal features: the word order, the question word what, and (in writing) the punctuation. The second sentence is recognized as an answer to the question by the pronoun reference (he has to be a substitute for John here), by the parallel grammatical structure in which the functional slots of the question words in the first sentence (What...doing) are filled in the second (washing his face), by the parallelism of verb form (is --ing), by the fact that washing is a possible lexical equivalent for doing, and (in writing) by the period. Question-Answer is a formal pattern illustrating a number of formal constraints which extend beyond the sentence.

The relationship of these two sentences can be described in numerous ways (probably most simply by seeing the first as a permutation of the second), but they can only be described as a sequence by positing the larger pattern, Question-Answer, and specifying the formal ways the two functional slots in this larger pattern are related, just as we specify the relationship between subject and predicate in a sentence. A number of these relatively simple two-part patterns can be described, including greetings, cause and result (hypothesis), topic and illustration, topic and partition, disjunction, and so forth. These patterns can be manifested by a single sentence or by two or more sentences. A large number of higher-level units of discourse can be described as chains of these simple two-part patterns.

As we move on to larger rhetorical patterns, the complexity increases. Formal signals become redundant: we identified the Answer in the Question-Answer pattern above by five of its contrastive features. Furthermore, lexical and semantic features become increasingly important in recognizing patterns: in the example above we recognize washing as a lexical equivalent of doing. Lexical equivalence chains are probably the most important markers of higher-level patterns.[15] We can illustrate some of this complexity by attempting to describe the paragraph as a formal structure, limiting ourselves here to only one, rather simple pattern.

We believe that written paragraphs are emically definable units--not just groups of sentences isolated by rather arbitrary indentations--and that this can be demonstrated. We are presently carrying out controlled testing of the recognition of these units in collaboration with psychologists at the Center for Research on Language and Language Behavior, The University of Michigan. Informal investigation has shown that, given a text in which all paragraph indentations have been removed, readers can mark paragraph breaks, with only limited indeterminacy at predictable points. Furthermore, readers can recognize a number of recurring paragraph patterns and partition these patterns in predictable ways.

One of the most common of these patterns is the one we have labelled TRI (Topic, Restriction, Illustration), or more formally,

$$+ T^2 \pm R + I^n$$

(The raised numbers indicate that in reading the formula, T may be read twice, R once, and I n number of times recursively.) This is the Topic-Illustration pattern with an optional intermediary slot in which the topic is restricted in some way (e.g., by definition, classification, partition). The following paragraph illustrates this pattern:

> (T) The English Constitution--that indescribable entity--is a living thing, growing with the growth of men, and assuming ever-varying forms in accordance with the subtle and complex laws of human character. (R) It is the child of wisdom and chance. (I) The wise men of 1688 moulded it into the shape we know, but the chance that George I could not speak English gave it one of its essential peculiarities--the system of a Cabinet independent of the Crown and subordinate to the Prime Minister. The wisdom of Lord Grey saved it from petrification and set it upon the path of democracy. Then chance intervened once more. A female sovereign happened to marry an able and pertinacious man, and it seemed likely that an element which had been quiescent within it for years--the element of irresponsible administrative power--was about to become

> its predominant characteristic and change completely the direction of its growth. But what chance gave, chance took away. The Consort perished in his prime, and the English Constitution, dropping the dead limb with hardly a tremor, continued its mysterious life as if he had never been.[16]

The slots in this tripartite pattern are marked by lexical equivalence classes, two of which have extended domains: 1) English Constitution, indescribable entity, living thing, It, child,...English Constitution; 2) men, human character, wise men of 1688, George I, Lord Grey,...Consort. Note that the domain of the first chain is the entire paragraph, while that of the second chain is the I slot. Chains can thereby be ranked as head and attribute chains, each paragraph including a head chain and one or more attribute chains.

The slots are also marked by grammatical parallelism (e.g., the first and second sentences, the third and fourth sentences), verb shift (e.g., shift to past in the I slot), pronoun domains, determiners, and transitional function words (e.g., then, but).

The TRI pattern has a number of variant forms which can be specified by the reading, permutation, and exponence rules. Only a few of these variants will be illustrated. As R is optional the pattern can be read + T + I. For example, this paragraph by Marchette Chute:

> (T) The only safe way to study contemporary testimony is to bear constantly in mind this possibility of prejudice and to put almost as much attention on the writer himself as on what he has written. (I) For instance, Sir Anthony Weldon's description of the Court of King James is lively enough and often used as a source material; but a note from the publisher admits that the pamphlet was issued as a warning to anyone who wished to "side with this bloody house" of Stuart. The publisher, at any rate, did not consider Weldon an impartial witness. At about the same time Arthur Wilson published his history of Great Britain, which contained an irresistibly vivid

account of the agonized death of the Countess of Somerset. Wilson sounds reasonably impartial; but his patron was the Earl of Essex, who had good reason to hate that particular countess, and there is evidence that he invented the whole scene to gratify his patron.[17]

If I is read a number of times, the pattern may be broken by indentation into more than one paragraph, although it remains a single emic unit. Indentation, like line ends in poetry, can either correspond to formal junctures or, for various reasons, interrupt the structure in a way somewhat similar to poetic enjambment.

The TRI pattern can be permuted to IRT, producing the so-called funnel effect or inductive structure. This is comparable to such permutations at the sentence level as "Home is the sailor" from "The sailor is home." Another illustration by Marchette Chute:

> (I) The reason Alice had so much trouble with her flamingo is that the average flamingo does not wish to be used as a croquet mallet. It has other purposes in view. The same thing is true of a fact, which can be just as self-willed as a flamingo and has its own kind of stubborn integrity. (R) To try to force a series of facts into a previously desired arrangement is a form of misuse to which no self-respecting fact will willingly submit itself. (T) The best and only way to treat it is to leave it alone and be willing to follow where it leads, rather than to press your own wishes upon it.[18]

This permutation is frequently used to begin or end discourse, probably because it imparts a greater sense of closure than the more open-ended TRI order.

Other permutations include TIRI, ITR, and TRIT, to list only the most common. Following exponence rules, slots in paragraph patterns may be filled by other rhetorical patterns. In the following example by Bernard Iddings Bell, the Answer slot in the Question-Answer pattern, which we discussed earlier, is filled by a TRI pattern, producing a compound paragraph structure:

(Q) Is the United States a nation composed chiefly of people who have not grown up, who think and act with the impulsiveness of adolescents? (A-T) Many shrewd observers of the American scene, both abroad and here at home, are saying that this is indeed the case. (R) They intentionally disturb our patriotic complacency. (I) They bid us view with alarm cultural immaturity revealed by current trends in journalism, by the radio, by the motion picture, by magazines and best-selling books, by mass response to emotionalized propaganda--political and otherwise; by a patent decay of good manners, by the spread of divorce and by other manifestations of parental irresponsibility; by all the various aspects of behavior which indicate to a student of human affairs the health or sickness of a civilization.[19]

Tagmemic matrix theory provides further insight into another traditional problem of rhetoric. We said earlier that form and idea are seen by many as organically unified, a view that we share. The literary statement contains within itself its own dimensions of development. It constitutes a semantic field which is clearly perceived when we try to extend it. The relevant categories of the Strachey paragraph discussed above can be displayed in the rows and columns of an emic paragraph matrix: (see Table 1.)
If we extend the paragraph we are obliged to supply a still more recent illustration of the effect of wisdom on the Constitution. It should be possible from a study of a large number of paragraph matrices to generalize further about various types of paragraph development. The investigation of paragraphs as semantic fields is as yet only beginning.

A writer's style, we believe, is the characteristic route he takes through all the choices presented in both the writing and prewriting stages. It is the manifestation of his conception of the topic, modified by his audience, situation and intention--what we might call his "universe of discourse." These variables directly affect selecting and grouping in all three linguistic hierarchies: grammatical, phonological, and lexical. An analysis and description of style involves the specification of the writer's characteristic choices at all points in the writing process, although usually only the final choices are directly accessible to the analyst.

The classical conception of style has a number of limitations. To see style as an addition to the message, an affective layer imposed on conventional language, ignores the close connection between language and idea. Seeing it as essentially a matter of sentences ignores stylistic patterns beyond the sentence. Furthermore, the theory grew out of a very specialized sort of practice, formal public speaking in the courts and legislatures and at ceremonial occasions. As a result it has a limited range of applicability. Seventeenth century critics were right in saying that its generalizations were inappropriate to a wide range of important topics, audiences, and situations. Finally, the highly normative approach of classical rhetoric tends to ignore the individuality of the writer, describing a style rather than style itself.

To see style, as some modern rhetoricians do, as the expression of a particular personality lays too much stress on one variable in the universe of discourse and too little on the others. Some stylistic features of a work inevitably remain unexplained if one commits himself to this definition strictly. To see style as a vision of the topic also has limitations; it ignores the influence on choice of situation and audience. It assumes that the act of writing is essentially expressive, not communicative. Neither of these views of style have been very useful in the teaching of rhetoric, for they make pedagogic generalization difficult.

To see style, in the way many linguists do today, as deviation from conventional language leads to the difficulty of defining conventional language. Somehow the deviations must be separated from the corpus, perhaps by measuring the frequencies of patterns. However it is done, it leaves conventional language as a styleless language. This view, like the classical view, tends to conceive of style as an embellishment, an added affective layer. Though very unconventional styles can be identified as linguistic deviations, there are "conventional" styles which this approach does not explain. These include the different styles we all use in different situations, with various audiences and in writing on assorted topics.

It seems to us that a full discussion of style must include the prewriting process in order to interpret the formal manifestations on the written page, the purely linguistic choices that the writer has made. Without the context of a linguistic unit, the universe of discourse, we are able only to describe stylistic features in a fairly trivial way. With the context there is the possibility of explaining the writer's choices. In a complete theory, then, a particular style is a characteristic series of choices throughout the entire process of writing, including both discovery (invention) and linguistic selection and grouping (arrangement).

We have presented what we believe to be the traditional problems of rhetoric and have suggested how a linguistic model, which includes both a discovery procedure and a descriptive technique, may provide the base for a new approach to rhetoric, a bridge between the humanities and sciences. A tagmemic rhetoric stands somewhere between the rigorous theories of science and the almost purely intuitive theories of the humanities. We see no reason to reject the insights of either the former or the latter, believing that all new knowledge, like the process of writing itself, involves both intuitive analogy and formal precision.

Footnotes

1. Aristotle, Rhetoric, i.2.1355b 26-27., trans. W. D. Ross, in The Basic Works of Aristotle, ed. Richard McKeon (New York: Random House, 1941) p. 1329.
2. Quoted in Wilbur S. Howell, Logic and Rhetoric in England, 1500-1700 (Princeton, N.J.: Princeton University Press, 1956) p. 66.
3. Edward P. J. Corbett, Classical Rhetoric for the Modern Student (New York: Oxford University Press, 1965) p. 171.
4. Quoted in Howell, p. 350.
5. Daniel Fogarty, Roots for a New Rhetoric (New York: Bureau of Publications, Teachers College, Columbia University, 1959).
6. I. A. Richards, The Philosophy of Rhetoric (New York: Oxford University Press, 1936) p. 3. Kenneth Burke and S. I. Hayakawa have both developed extremely interesting theories of rhetoric and must be mentioned, along with Richards, as having made notable contributions to the development of a new rhetoric.
7. The basic source of tagmemic theory is Kenneth L. Pike, Language (in Relation to a Unified Theory of the Structure of Human Behavior) (Glendale: Summer Institute of Linguistics, Part I, 1951; Part II, 1955; Part III, 1960). A new edition, to be published by Mouton, is in preparation. Pike applies tagmemic theory to problems of rhetoric in "Beyond the Sentence," College Composition and Communication 15.3 (1964) and "Discourse Analysis and Tagmeme Matrices," Oceanic Linguistics (April, 1965).
8. Kenneth L. Pike, "A Syntactic Paradigm," Language 39.216-30. (1963).
9. Quoted in Howell, p. 367.

10. Kenneth L. Pike, "Language as Particle, Wave, and Field," *The Texas Quarterly*, 2.2.37-54 (1959).

11. For further illustrations of the use of tagmemic discovery procedures in rhetorical invention, see Hubert English, "Linguistics as an Aid to Invention," *College Composition and Communication* 15.3 (1964).

12. Our conception of the image here is drawn in large part from Kenneth Boulding, *The Image* (Ann Arbor: The University of Michigan Press, 1956) and from William Angus Sinclair, *Conditions of Knowing* (London: Routledge and Kegan Paul, 1951).

13. Anatol Rapaport, *Fights, Games, and Debates* (Ann Arbor: The University of Michigan Press, 1961).

14. A full description of tagmemic grammatical theory can be found in Robert Longacre, *Grammar Discovery Procedures* (The Hague: Mouton, 1964). Tagmemics is contrasted with transformational and other models in Longacre, "Some Fundamental Insights of Tagmemics," *Language* 41.1.65-76 (1965).

15. The concept of lexical equivalence chains is derived in large part from Zellig S. Harris, *Discourse Analysis Reprints* (The Hague: Mouton, 1963) pp. 7-10.

16. Lytton Strachey, *Queen Victoria* (New York: Harcourt, Brace, 1921), pp. 300-301.

17. Marquette Chute, "Getting at the Truth," *The Saturday Review*, Sept. 19, 1953, p. 11.

18. *Ibid.*, p.12.

19. Bernard Iddings Bell, "We are Indicted for Immaturity," *New York Times Magazine*, July 20, 1947, p. 8.

Table 1

The emic structure of the paragraph by Strachey that appears on p. 20.

Forces shaping the English Constitution / historical manifestations	wisdom	chance
1688	The wise men ... molded it into the shape we know	
(1714)		George I ... gave it ... the system of a Cabinet independent of the Crown and subordinate to the Prime Minister
(1832)	Lord Grey saved it from petrification, and set it upon the path of democracy	
(1840)		(Victoria's marriage made it seem likely that a quiescent element) was about to become its predominant characteristic and change ... the direction of its growth.
(1861)		[With the death of the Consort] the English Constitution ... continued its mysterious life as if he had never been.

II. STUDIES IN PROGRESS

Extension of the Speech Auto-Instructional Device to teaching segmental features of language (H. L. Lane, J. C. Catford)

Functions of speaker and listener vowel-spaces in the imitation and identification of spoken vowels (S. Ross)

Investigation of the rhythm of spoken American English (G. D. Allen)

Acoustic analysis of the development of the prosodic features of infants' vocalizing (W. C. Sheppard, H. L. Lane)

Perceptual and cognitive processes in speech recognition (J. F. Hemdal)

Syntactic generalization and perception of grammaticalness by aphasic adults (R. S. Tikofsky)

Acoustic analysis of the prosody of Jordanian Colloquial Arabic (J. C. Catford, R. M. Rammuny)

A comparison of the psycholinguistic functioning of "educationally-deprived" and "educationally-advantaged" children (L. S. Barritt)

The use of the Cloze technique in the study of the grammatical skills of retarded and normal subjects (M. I. Semmel)

An examination of the conditions under which passive and declarative sentences are produced (L. S. Barritt)

Latency and accuracy of responding by retarded and normal subjects to sentence structures (L. S. Barritt, M. I. Semmel, J. Prentice, S. Bennett)

Differential reinforcement of a vocal operant along three parameters (S. E. Knapp)

Some effects of the behavior of the teacher on discrimination learning by the student (D. M. Brethower)

Objectives and techniques for training teachers in classroom control (D. E. P. Smith)

The relation between arousal and the recall of verbal material in connected discourse (F. M. Koen)

The role of distinctive features in phonology in the perception and recall of speech (F. M. Koen)

II. STUDIES IN PROGRESS (cont.)

The acquisition of Japanese by children (D. McNeill)

Tagmemic and matrix linguistics applied to selected African languages (K. L. Pike)

Identification, description and testing of the psychological reality of the paragraph (A. L. Becker, R. E. Young, F. M. Koen)

III.

AFFILIATED PERSONNEL

CENTER FOR RESEARCH ON LANGUAGE AND LANGUAGE BEHAVIOR

ALLEN, George. M.A. Graduate student, Communication Sciences. Research assistant in Group A.

BARRITT, Loren S. Ph.D. Indiana. Assistant Professor of Education, School of Education. Chairman, Group B.

BECKER, Alton L. M.A. Graduate student, Department of Linguistics. Instructor in English and Linguistics. Research assistant in Group E.

BENNETT, Stanley. Graduate student, School of Education. Research assistant in Group B.

BREND, Ruth. Ph.D. Michigan. Research associate in Group E.

BRETHOWER, Dale M. Graduate student, School of Education. Research assistant in Group C.

CAROLINE, Lawrence. M.A. Graduate student, Department of Philosophy. Research assistant in Group A.

CATFORD, John (Ian) C. University of Edinburgh. Professor of Linguistics. Member of Center's Executive Committee. Member of Group A.

ELLIOTT, Dale. Graduate student, Department of Linguistics. Research assistant in Group D.

GEIS, George L. Ph.D. Columbia. Chairman, Group C.

GINET, Sally. M.Sc. Computer programmer in Group A.

HANNA, Susan J. Undergraduate student in Psychology. Research assistant in Group F.

HAYS, William L. Ph.D. Michigan. Professor of Psychology and Associate Dean of the College of Literature, Science and the Arts. Member of Center's Executive Committee.

HEMDAL, John F. Ph.D. Purdue. Research associate in Group A.

HODGSON, Helena. Undergraduate student in Sociology. Research assistant in Group B.

HORSWELL, Roxanne, B.S. Michigan. Research assistant in Group C.

JAREMA, Louise. Secretary-typist.

KEILER, Allan R. Ph.D. Harvard. Assistant Professor of Linguistics.

KNAPP, Stephen. B.A. Hamilton. Research assistant in Group C.

KOEN, Frank M. Ph.D. Vanderbilt. Assistant Professor of Psychology. Chairman, Group D.

KOPP, James L. Graduate student, Department of Psychology. Research assistant in Group A.

LANE, Harlan L. Ph.D. Harvard. Associate Professor of Psychology. Director of the Center and Chairman, Executive Committee. Chairman, Group A.

LEHMAN, Robert. Undergraduate student in Psychology. Research assistant in Group A.

LEHMANN, Charles F. Ph.D. Michigan. Associate Dean, School of Education. Member of Center's Executive Committee.

LI, Ying Chê. M.A. Graduate student, Department of Linguistics. Research assistant in Group D.

MAEDA, Yasuko. B.A. Doshisha. Research assistant in Group D.

McNEILL, David. Ph.D. Harvard. Assistant Professor of Psychology. Member of Group D.

McNEILL, Nobuko. B.A. UC Berkeley. Research assistant in Group D.

NEYER, Edna. Administrative secretary.

OSTER, Ruth. Graduate student, Department of Linguistics. Research assistant in Group A.

PAPER, Herbert H. Ph.D. Chicago. Chairman, Department of Linguistics. Member of Center's Executive Committee.

PETERSON, John R. B.A. Graduate student, Department of Far Eastern Languages and Literatures. Research assistant in Group D.

PETERSON, Lydia G. Dissemination secretary.

PHILIP, Franklin J. M.A. Graduate student, Department of Philosophy. Research assistant in Group F.

PIKE, Kenneth L. Ph.D. Michigan. Professor of Linguistics. Chairman, Group E.

RAMMUNY, Raji. Graduate student, Department of Near Eastern Languages and Literatures. Research assistant in Group A.

RAND, Timothy. Assistant research engineer.

ROSS, Strange. Mag.Art. University of Copenhagen. Research associate in Group A.

ROTHKRUG, Françoise. B.A. University of Paris. Research assistant in Group F.

SEMMEL, Melvyn I. Ed.D. George Peabody Institute. Assistant Professor, School of Education. Research associate in Group B.

SEMMELROTH, Carl C. Graduate student, School of Education. Research assistant in Group C.

SHEPPARD, William C. Graduate student, Department of Psychology. Research assistant in Group A.

SMITH, Donald E. P. Ph.D. Cornell. Associate Professor of Education. Research associate in Group C.

SUTCLIFFE, Rita. Undergraduate student in Zoology. Research assistant in Group B.

TARTE, Robert D. Ph.D. Michigan. Associate Director of the Center (Administration).

TIKOFSKY, Rita. M.S. Pittsburgh. Research assistant in Group C.

TIKOFSKY, Ronald S. Ph.D. Utah. Associate Professor of Speech. Member of Center's Executive Committee. Member of Group A.

VANDER YACHT, David. B.S. Graduate student, Department of Electrical Engineering. Computer programmer.

WALKER, E. L. Ph.D. Stanford. Professor of Psychology. Member of Center's Executive Committee.

WEENER, Paul. Graduate student, School of Education. Research assistant in Group B.

YOUNG, Richard E. Ph.D. Michigan. Assistant Professor of English. Member of Group E.

ZALE, Eric M. Ph.D. Michigan. Associate Professor of English, Eastern Michigan University. Associate Director of the Center (Dissemination).