THE UNIVERSITY OF MICHIGAN

INDUSTRY PROGRAM OF THE COLLEGE OF ENGINEERING

HUMAN DISCRIMINATION OF AUDITORY DURATION

ouglas
C. D. Creelman

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in the
University of Michigan
1960

December, 1960

IP-483

ACKNOWLEDGEMENTS

The inspiration and encouragement of Dr. W. P. Tanner, Jr., deserve more than can be acknowledged here. To him I owe my interest in auditory research and an orientation to the interrelation of theory and experiment apparent in this thesis. I also wish to express appreciation to Ward Edwards, Gordon E. Peterson, and Paul M. Fitts for their assistance.

Thanks are due for the patience and hard work of the observers whose cooperation made the success of the experiments possible. Clarence Rivette and W. Joe Wiesner assisted in the collection of the data. Carol Creelman gave yeoman service in the calculation of the theoretical functions.

Financial support for the work was provided by the Operational Applications Laboratory of the Air Force Command and Control Development Division, and by the Office of Naval Research.

TABLE OF CONTENTS

ABSTRACT

A series of related experiments was performed which measured human ability to discriminate durations of auditory signals. A two-alternative forced-choice procedure was used: two sine-wave signals of identical amplitude and frequency, differing only in duration, were presented sequentially on each trial. The order of presentation was random, and the observers' task was to state, for each trial, whether the longer signal had occurred first or second. The signals were presented in a background of continuous white masking noise, which was held at a constant level throughout the experiments. Paid observers worked two hours daily for the course of the experiments. Sufficient data were thus available to allow separate treatment for each observer.

The independent variables were the signal voltage, the "base" duration, T, and the increment duration, ΔT. Separate experiments assessed the functional effect of each of these variables on discrimination. These were used to predict the results of two further experiments.

The model used in this prediction was derived from statistical decision theory. Duration measurement was assumed to be accomplished by a "counting mechanism," operating on impulses generated over the relevant duration. The source of these impulses was assumed to be random. Limitations on performance were assumed to come from uncertainty regarding the end-points of the time interval, and from limited memory.

These considerations led to the formula:

$$(d')^2 = \frac{\lambda}{1 + KT} \cdot \frac{\Delta T^2}{2T + \Delta T + \sigma_v^2} ,$$

where d' is the dependent variable, a normal transformation of probability of correct response. The constants λ and K were respectively measures of "counting rate" and memory decay with time, estimated separately for each observer, and σ_v^2 was an inverse power function of signal voltage, a measure of uncertainty regarding starting-time and ending-time of the signals. The decision processes underlying this formula were presented as a general model for discrimination of durations, and shown to agree with the data regarding auditory signals.

LIST OF ILLUSTRATIONS

LIST OF ILLUSTRATIONS-- Continued

1. INTRODUCTION

1.1 Nature of the Problem

The experiments reported in this thesis were an attempt to measure the capacity of human observers to discriminate differences in duration between very short auditory signals. These observations were made for three related reasons. The theoretical problem of how persons utilize temporal information is still not well understood. Performance in a signal detection task depends critically on the ability to attend precisely during the relevant time. A practical problem related to temporal discrimination arises in the field of speech perception, where in some languages cues to linguistic meaning, and in English cues to stress and inflection, seem to depend on the perception of relative durations of speech sounds (cf. Peterson and Lehiste, 1960).

Data on the auditory discrimination of signals differing only in temporal aspects were collected with these three problems in mind. A quantitative theory of temporal discrimination will be developed to account for discrimination in the range ordinarily covered by speech sounds and the signals used in psychoacoustic experiments.

The remainder of this section will present a brief historical summary of the temporal perception problem. In the next section the experimental design and procedure will be presented in detail, and the response measure, d', will be defined from a background of statistical decision theory.

In the following section the results of six experiments will be described which measured different aspects of temporal discrimination. An advantage of the discriminability measure becomes clear in the results of these experiments: predictable functional relations are reported, which carry considerably more information than traditional reports. Ordinarily only one value of the relationship between the stimulus values and responses is presented, namely, the "threshold," defined by some arbitrary level of performance. It has become more and more clear in recent years that useful information is contained in responses, as long as the observer is able to perform at a level above that which we might expect on the basis of chance, whether above or below any arbitrary cutoff. A theoretical model will be developed in section 4 which describes the results of the previous section. The results will be related to previous research on temporal discrimination, and the model will be discussed in relation to other theoretical approaches.

1.2 Historical Perspective

The study of human discriminal abilities in the time domain is, in a sense, a poor stepchild of the growth of psychophysical research. Titchener (1905) characterized the area as a "microcosm, perfect to the last detail," exemplifying in miniature the course of development in psychophysics to that time in methodology, concepts, and empirical knowledge. Titchener went so far as to suggest to the diligent student that he spend a separate year in the sole study of experiments in time perception. We could hardly make the same recommendation now, for with a few notable exceptions the area has been much neglected since the time Titchener wrote, nearly sixty years ago. Study of temporal discrimination, and the estimation of time, remains a microcosm, somewhat isolated from the main stream of empirical research on sensory capacities of men. Seldom has there been an attempt to integrate what was known about how people estimate, react to, or judge elapsed time with what was known about how they react to, estimate, and judge sensory inputs which of necessity are presented in, and are dependent upon, time. Accuracy in psychophysical tasks is necessarily highly dependent on the degree of this ability, and yet it has been only very recently (e.g., Tanner, 1960b) that attempts have been made to account for this as a source of added variance in such studies.

The reasons for this neglect of the area of temporal phenomena do not seem particularly obscure. They can be seen in a scholarly review of the history of the philosophy and psychology of time by Nichols (1891). The views of philosophers from Plato through the British and French empiricists, and the experimental work of the early German psychophysicists were systematically covered. A central problem throughout the development of the subject was the status of the "time sense" as an independent psychic faculty, independent of the content of sensory input. This question does not yet seem to have been resolved, and continues to haunt the researches and theoretical efforts of psychologists.

Reviews of the experimental literature after that of Nichols were presented by Dunlap (1911, 1912, 1914, 1916), Axel (1925), Weber (1933), Gilliland et al. (1946), and most recently by Wallace and Rabin (1960). Chapters on time perception are offered by Titchener (1905), by Boring (1942), and by Woodrow (1951). With all this activity, we still find that there is "...as yet no generally accepted view as to how we perceive or estimate time." (Gilliland et al., 1946), and fourteen years later we find ourselves still on the trail of the "hitherto elusive 'time sense'" (Wallace and Rabin, 1960).

Experimental procedures for the study of time perception have traditionally been limited to four: the methods of estimation, production, reproduction, and comparison. In the estimation and production procedures the subject is required to link a verbal statement in conventional time units, such as seconds, to a time interval marked off by a sensory input. In the former case he attempts to give a verbal estimate of the duration of a signal, and in the latter he attempts to produce a signal of a stated duration on command from the experimenter. These procedures were used in attempts to find the relation between "experienced time" and "real time" long before S. S. Stevens gave such procedures a degree of respectability in his approach to sensory scaling, using them extensively to measure sensory magnitudes. The other two procedures avoid the pitfalls of verbal report. The method of production requires the subject to duplicate an interval provided him by the experimenter, and the comparison procedure requires the subject to state which of two intervals, both provided by the experimenter, was the longer.

The comparison method was extensively used, but early became tied up with the problem of the "time error," the finding that when two stimuli were presented successively for comparison, the same physical value seemed to be judged differently depending on whether it came first or second in the series. Woodrow (1934) and his student Stott (1935) reported extensive investigations of the time-order effect, the most modern and complete of a long line of investigations. Woodrow's data showed a decreasing time error with continued practice, and Stott showed that the time error was greatest among naive subjects. Stott showed the size of the "indifference interval" (that duration at which order of presentation has no effect on comparative judgment) to be a function of the range and size of the stimuli used in a single experiment. Woodrow ascribed these phenomena to a "drifting" of the first stimulus toward a remote standard during the interval between stimuli. This value tended to be the mean of all the durations experienced in the experiment by the subject, plus a standard which the subject brought with him to the experimental situation. This later effect, according to Woodrow, became negligible with extended experience (Woodrow and Stott, 1936). This interpretation received support from experiments by Postman and Miller (1945) and by Phillip (1947), which showed considerable effect in the predicted direction due to "anchor" stimuli in the estimation of durations using a five-point scale, and due to the interpolation of various extraneous signals between two durations which were to be compared. A good deal of experimental effort in the study of temporal perception has been devoted to studies of effects on judgment of order, method of presentation, and practice. Early psychophysics, in attempts to measure the sensory capacities of human observers, regarded these effects as annoyances, to be eliminated by careful experimental design. Later workers saw in these phenomena a fertile field of investigation in its own right. With few exceptions, the circle has failed to close in the area of temporal sensitivity, and there are few modern studies devoted primarily to the limits of human temporal judgment.

Although the experiment by Stott (1934) was primarily concerned with the analysis of time errors, enough data were reported to draw some conclusions about the temporal sensitivity of his observers. The observations were made in a group setting, and the data presented were averaged over subjects, a questionable practice in the light of the degree of individual differences found in such experiments (e.g., Nakajima, 1958). The averaged data will be compared with the data obtained from the present experiments in a later section.

An experiment by Henry (1948) was a later-day attempt to check the validity of Weber's law applied to discrimination of duration in auditory signals. His apparatus presumably afforded greater control of the stimuli than did Stott's in that the signals to the listeners' earphones were electrically switched by means of microswitches activated by rotating cams. Stott's signals were gated by a flag, or shutter, inserted into a tube which carried the signals to the listeners. However, Henry's experimental procedure was probably less reliable than Stott's in that Henry used essentially what has come to be known as the "yes-no" procedure (Blackwell, 1953). A series of signals was presented and the listeners were to state whether the series consisted of alternating durations or of repetitions of the same duration. The dependent variable was the per cent of the time the observer was correct. A complete description with the yes-no procedure requires a breakdown of the data into those correct responses arising from presentation of each possible input sequence (Smith and Wilson, 1953; Tanner and Swets, 1954). Stott used what has come to be known as the two-alternative forced-choice procedure, in which two differing signals are always presented on each trial, in a randomly determined order and the listener is required to identify the order of presentation. The data from Henry's experiment were reported in terms of the Weber ratio, $\Delta T/T$, where ΔT was the increment necessary to give an average performance of 75 per cent correct responses when added to a "base" duration, T. His results are shown in Fig. 1, along with points calculated from Stott's data. The other points shown on Fig. 1 will be discussed below. The results are presented on logarithmic axes, in order to test Henry's assertion that the data follow a modified Weber function, adapted from a proposal by Guilford, namely, $\Delta T/T = KT^{N-1}$, where N is a constant less than unity. If this is in fact the case, the points on Fig. 1 should lie along a straight line with nega-

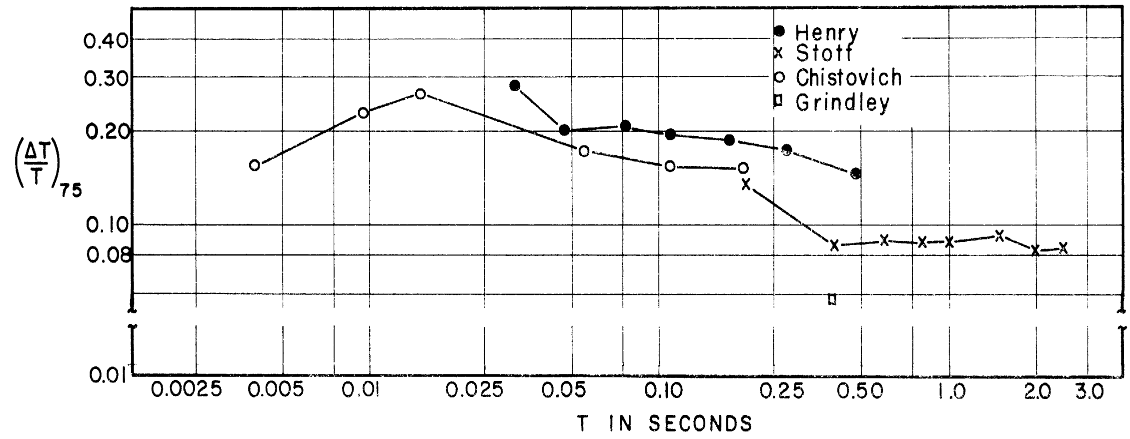

Fig. 1. Weber ratio for added duration necessary to yield 75% correct judgments, as a function of base duration.

tive slope. It is left to the reader to decide whether or not this is an adequate characterization of these data. In the absence of other compelling alternative hypotheses, it does not seem unreasonable. The discrepancy between the levels of the two sets of data can be ascribed to the differences in experimental method and apparatus.

Some question remains as to whether a duration can be judged independently of the sensory events happening during that duration. On the side of a negative answer to the question, we have the considerable weight of the opinion of William James (1890) and, in more modern times, of Fraisse (1957), the author of a text on the psychology of time written mostly from a phenomenalistic point of view. Time seems to pass faster, after all, when we are occupied, and by the same token temporal intervals would be expected to seem to pass faster if they are filled by continuing stimuli than if they are simply marked off by clicks at their beginnings and ends. This is hardly a rigorous proof, and there seems to be no *a priori* reason to expect the two situations to be the same. The available experimental data suggest they are different.

Clausen (1950) has compared estimation of filled and unfilled intervals of 5, 10, and 15 seconds. The results showed no differences in the performance of his subjects between the two procedures. In this study Clausen also showed the methods of production and estimation to be less accurate than the reproduction method. Another deduction from the assumption of dependence of temporal judgment on what is happening during the time judged is that accuracy should depend on the intensity of the signals. Henry found no effect of auditory intensity on the accuracy of judgment. Oleron (1952) and Lifshitz (1935), on the other hand, reported studies which showed dependence of time judgment on the intensity of the signal.

The logical question of the independent status of a sense of duration was considered by Boring (1936). In an argument based on the (then very new) approach of operationism he attempted to show that the old philosophical question disappears as soon as the data are considered in their own right. People were able to give introspective reports involving elapsed durations, and experiments showed subjects able to differentiate sensory events on the basis of duration alone. This was sufficient for the scientist, if he was an operationist, for the existence of the data eliminated any problem. It has become clear since 1936 that the answer is not as simple as it once seemed. More is necessary for understanding processes than the raw data either in the form of introspective reports or behavioral data.

The studies of Chistovich (1959) used short auditory clicks to delimit the stimuli. A duration was marked off by a pair of clicks, and then another duration followed, again bounded by clicks. The task of the observers was to state whether the second was equal to the first or differed from it by ΔT. It was not stated explicitly, but we may safely assume that 50 per cent of the trials actually had the longer duration as the second of the pair. This method again is subject to all the pitfalls of the "yes-no" procedure, and the observers in this experiment were further handicapped by a ten-second wait between presentations of the two durations to be compared. The durations were electronically controlled, and thus probably were the most reliable of any of the studies considered thus far. They would have had to be, to produce stimuli of .0055 sec. and .00635 sec. for comparison (the pair yielded 75 per cent correct responses, according to the report). The data from Chistovich are given along with the data for filled auditory intervals in Fig. 1. For comparison a single point derived from data reported by Grindley (1932) is also shown on the graph. The improvement in performance can most likely be attributed to the fact that Grindley used the Seashore musical abilities test, which gives groups of three clicks which delimit two intervals, one immediately following the other. Thus for these observers there was no wait between the first and the second of the pair of durations to be compared.

Most studies of temporal perception have dealt with the verbal estimation of time intervals. In 1951 two studies departed from tradition and attempted to construct a scale of psychological time from a fractionation procedure. Subjects were given a standard duration by the experimenter and required to produce a duration half as long, by keying an auditory signal. Gregg (1951) used standards from 0.4 to 4.8 seconds and by averaging across subjects obtained a scale of psychological time remarkably linear with elapsed clock time. The scale takes an upward turn toward the longest durations studied, and a calculation based on Stevens' latest (1957) theoretical pronouncement, which predicts a power-function relationship, showed psychological time, t, related to elapsed physical time, T, by the formula, $t = T^{1.022}$. Ross and Katchmar (1951) used the same method to obtain a scale for time intervals from 5.36 to 60 seconds. Unfortunately the results do not fit particularly well with an extrapolation of the Gregg data. In this study only three observers were used, and the results from each were considered separately. The three scales were quite different in shape, but the data from each observer suggested that if a power-law function does hold, the exponent is less, not greater, than unity. That the exponent of such a function should be less than unity agrees with results reported recently in a monograph by Frankenhauser (1959), results which were obtained using the more time-honored method of verbal estimation. A problem recognized early which affects the accuracy of such judgments was pointed out by Urban (1907). This is that people tend to use certain numbers when making reports to the exclusion of others, thus causing at least an increased variability, and perhaps some systematic errors in the data. The criticisms which have been advanced recently regarding the validity of scaling procedures applied to the more usual sensory dimensions apply equally well to time scaling (cf. Garner, 1959).

Some years ago Hirsh (1952) called for a measure of ability to discriminate temporal events analogous to those of visual acuity. He pointed out that lack of such a measure may be in large part the cause of the lack of correspondence found between experimental measures of auditory abilities and the ability to discriminate adequately everyday phenomena. The experiments to be reported in the following pages and the theoretical model which describes the results represent first steps toward such an understanding of these relationships.

2. APPARATUS AND PROCEDURE

2.1 Apparatus

All the experiments followed the same general procedure and were conducted using the N. P. PSYTAR apparatus at the psychophysical laboratory of the Cooley Electronics Laboratory, The University of Michigan. The general nature of the equipment has been described by Bilger (1959). Figure 2 shows a block diagram of the apparatus.

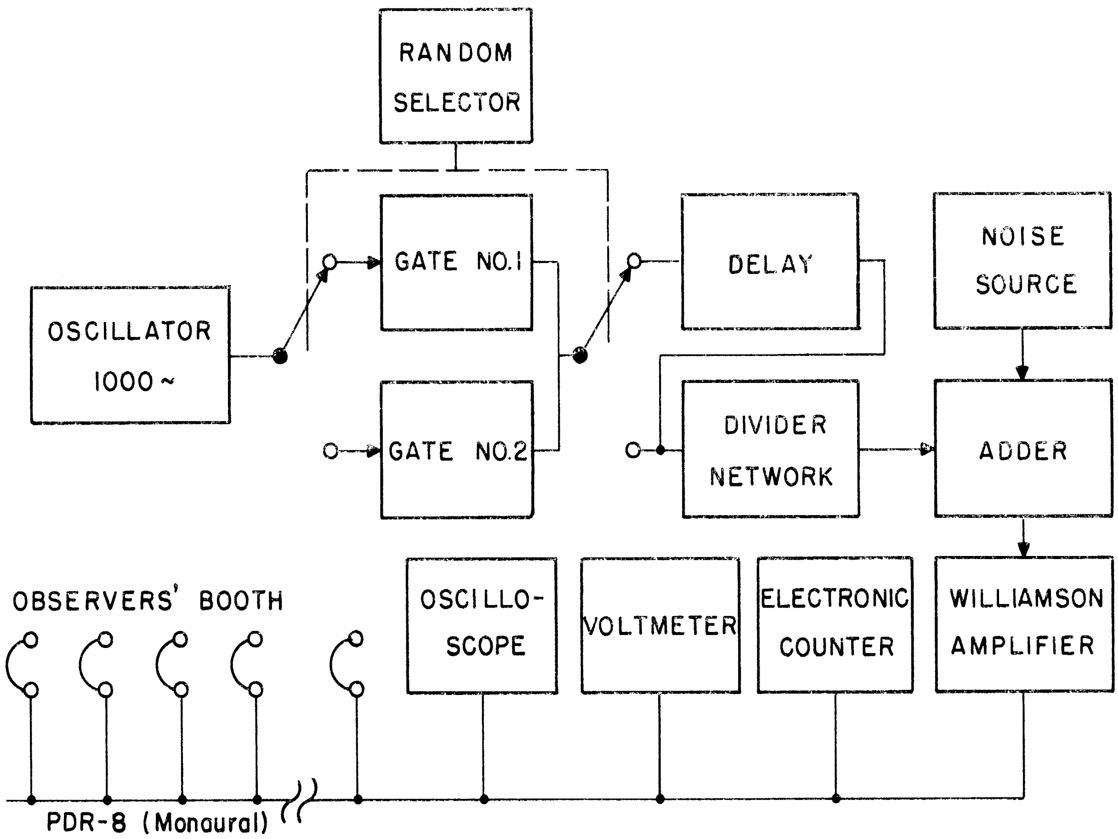

Fig. 2. Block diagram of experimental apparatus.

A modified Hewlett-Packard Model 200AB Oscillator could be connected to one of two gates, which were set to deliver different signal durations. The gates are designed to pass a segment of a sine wave signal which begins at a positive-going zero-crossing of the waveform and continues for an integral number of cycles. The oscillator output was always set at a fixed voltage, and any desired signal voltage could be obtained by means of the divider network, constructed of high-precision fixed resistors. The gated signal was then mixed with continuous wide-band white noise from a General Radio 1390-A noise generator. The observers' PDR-8 earphones were wired in parallel for monaural presentation.

A Hewlett-Packard Model 521 electronic counter was used to set and to check the oscillator frequency and signal durations, as well as the inter-signal times, discussed below. A Ballentine Model 620 "True RMS" meter was used to set the fixed oscillator voltage and the noise voltage.

2.2 Procedure

The events in each experimental trial were automatically programmed electronically. The sequence is listed below in the order of time of occurrence.

1. Make random selection.
2. Observers' ready light.
3. Observers' signal light flash and presentation of first signal, plus .8 sec. delay.
4. Observers' signal light flash and presentation of second signal, plus .5 sec. delay.
5. Observers' response light.
6. Feedback of correct answer information to observers, and record data.
7. Machine clear and recycle.

In interval 1 a radiation-controlled random selector described by Lauder (1959) determined the position of the switch in Fig. 1, and presented the output of the oscillator to the input of one or the other of the two gates. During interval 2 a red warning light on a panel in front of each observer was turned on for about 1/2 second. At the start of interval 3 a brief (.01 sec.) flash of a white neon light signaled the onset of a signal, passed by the gate chosen by the random selector. The end of the fixed third machine interval triggered a delay unit, which was set so that the time elapsed between the end of the first signal and the start of the second was exactly .8 seconds. Since the two gates were set for differing signal durations, the amount of delay was determined by the random selection. Interval 4, triggered by the end of the delay time, contained a signal passed through the gate not selected for interval 3. The start of the second signal presentation was marked by a second flash of the white signal light on the observers' panels. A yellow light during interval 5 signaled the observers to respond with one of two push buttons, to signify whether they believed the first or the second signal interval had contained the longer of the two signals. During interval 6 the observers were informed by means of a separate pair of lights as to the correct answer for the pair of signals just presented; and the responses of each observer, along with the random selections, were automatically recorded on IBM cards and on counters. During the final interval the machine was cleared and the cycle begun again. After 100 trial cycles the machine was automatically stopped, the observers were told their scores, given a brief rest, and another experimental run was begun. The events of a trial cycle are summarized in Fig. 3. The top line represents the times of occurrence of the lights and the bottom line represents the auditory signals. One signal was passed through a gate set to present a signal of duration T, and the other signal was gated for

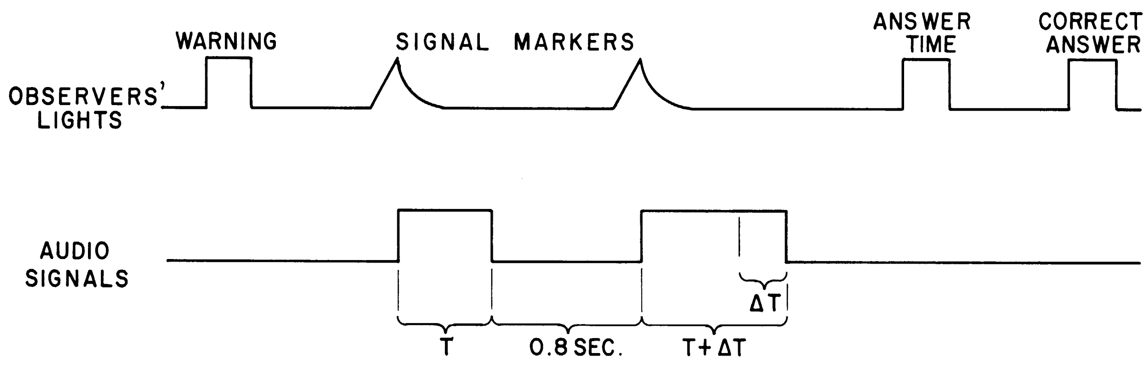

Fig. 3. Sequence of events in an experimental trial.

duration T+ΔT. The order of presentation was decided by the random selector for each trial, with the long or the short signal equally probable as the first of the pair.

The observers were all students, paid an hourly rate. They worked two hours per day, five days a week, for the duration of an academic semester or summer session. A bonus of 0.1 cent was paid for each correct answer, with the same amount subtracted for each incorrect answer.

The experiments attempted to measure ability to differentiate signal durations and each part of the procedure was designed to obtain consistent maximal performance from the observers. The two-alternative forced-choice procedure with equal likelihood of occurrence, the bonus system with equal payoff for correct answers in either of the two intervals, and the instructions given to the observers, all were designed with this in mind. On the first day of experimentation the writer spent a half-hour session with the observers outlining the general nature of the problem, the specific nature of the experimental situation, and the bonus system. It was not unusual for the observers to earn more than sixty cents in the two-hour session in bonus, in addition to their regular hourly rate.

Before any of the data reported below were collected the observers had at least two weeks' practice and their performance had reached a stable level. Before each new experiment was begun, at least a day was spent familiarizing the observers with the new task. In addition, before each experimental run a series of trials was presented with the masking noise turned off, in order to make sure the observers were aware of the nature of the particular signals being used on that run.

The procedure of each experiment was roughly the same. Each value of the independent variable was presented for two or three experimental runs in succession. Then a new value was selected and run, until all had been presented. The order of selection was random. This procedure was repeated until roughly 1000 observations had been obtained for each value, with a new random order used for each repetition.

2.3 Independent Variables

The effects on discrimination of the three variables: "base" duration, T; difference duration, ΔT; and signal voltage, V_S, were considered separately in the experiments. In general while the effects of one variable were estimated the others were held fixed at arbitrary constant values. Throughout all the experiments the background noise was held constant at a wide-band reading of .01 v rms to the observers' earphones. This is proportional to a noise power density of 3.6×10^{-7} watts per cycle per second. The data to be reported are from different sets of observers run at different times, and each group was run under different sets of experimental variables. The results are presented in terms of the detectability measure d' and, in order to make it easier to follow the presentation, an introduction to the concepts underlying this measure will be presented.

2.4 The Dependent Variable, d'

In recent years evidence has accumulated in support of the usefulness of a decision-theoretical model as an interpretive tool for psychophysical data (Creelman, 1960a; Tanner, Birdsall, and Clarke, 1960; Swets, Tanner, And Birdsall, 1960). This approach views the observer as, in essence, a tester of statistical hypotheses. Consider an experimental situation where one of two events can given rise to a sensory input. It can be shown that a single-dimensional number can represent a basis for decision as to which of the two events actually occurred. This number, in the mathematical theory, is likelihood ratio; the ratio of the conditional probabilities that the input signal arose from one or the other of the possible events. A rigorous development of these assertions is presented elegantly by Peterson, Birdsall, and Fox (1954), and discussion of the application to human performance can be found in a chapter by Licklider (1958) and in references cited there. Figure 4 represents the situation, where the occurrence of either event A or event B gives rise

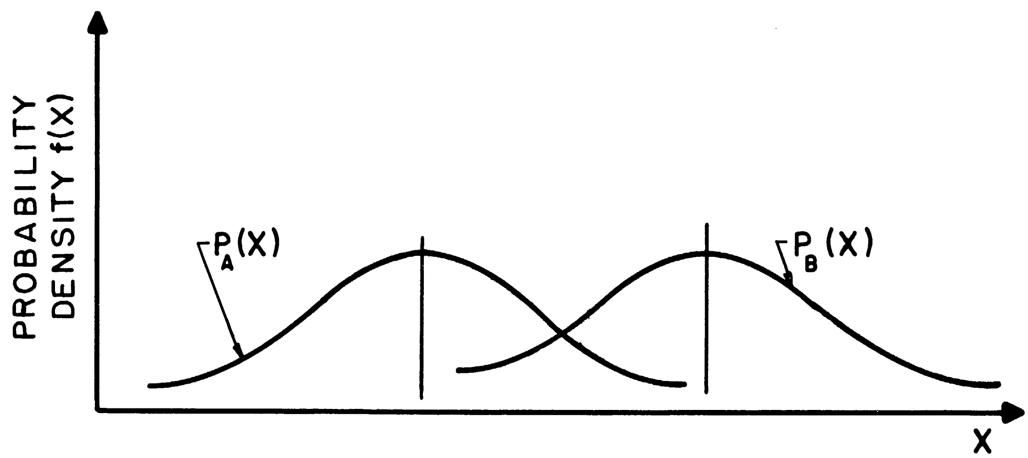

Fig. 4. Probability-density distributions for an input measure X, conditional on whether event A or B led to the observation.

to some value on a decision axis, which must be at least monotonic with likelihood ratio. The conditional probability that the presentation of A or B will lead to a given value on the decision axis is represented by the two probability-density distributions. There is, of course, no *a priori* necessity to assume a particular shape for these distributions, but evidence from human visual discrimination, auditory detection, and recognition of speech signals suggests that the distributions may be assumed to be normal. If the decision axis is carefully specified, the alternatives A and B may be any two signals selected in such a way that some equivocality exists on the basis of the observed input as to which alternative actually gave rise to it. Thus Fig. 4 is, in barest outline form, a general theory of human perceptual discrimination. Any particular application requires that the relevant decision axis be specified and that the shape of the distributions be worked out. However, if the general theory is correct, then any set of discrimination data may be referred back to such a set of hypothetical distributions without the necessity of complete a priori specification of the nature of the underlying decision processes. The measure of performance necessary for such a referral is an estimate of the separation between the two hypothetical distributions. This estimate is the measure d'. The detectability measure can then be used in two ways, both as a way to evaluate performance in any detection or recognition situation and as a framework on which to build theories specifying the processes underlying discrimination. A theory of this type will be developed in a later section; for now we will simply describe the procedure used to go from the obtained experimental data to a value of d'.

The calculation of a value of d' depends on the way in which the data were collected, i.e., on the psychophysical method employed. The two-alternative temporal forced-choice procedure was used throughout the present experiments, and only this procedure will be discussed here. The data were tabulated separately for each experimental condition and for each of the two presentation intervals. The number of correct responses when the longer signal was actually presented in the first interval, and the number of correct responses when the longer signal was actually presented in the second interval were used to obtain estimates of the probability of being correct. These probabilities were then used to enter a table of the normal probability integral. The distance in units of standard deviation from the mean of the distribution necessary to yield the obtained probability was read directly from the table. The procedure is familiar to psychologists as that used with the ordinary t-test. The two deviation values, one representing each presentation interval, were added to give the reported value of d'. If the two obtained per cent correct scores, corresponding to the two temporal intervals, were equal, d' as defined here would be related to per cent correct as shown in Fig. 5. This may be viewed as an operational definition of the dependent variable.

In order to remain consistent with the previously published study by Tanner (1956), the abscissa of Fig. 5 is labeled $d'_{1,2}$. The present case is identical with Tanner's case of recognizing one of two orthogonal signals. Here the signals were orthogonal in time; it is assumed that the occurrence of one input signal has negligible influence on the observation of another signal presented .8 sec. after it. The results of recent experiments by Creelman and Tanner (1960), by Tanner (1959), and by Bilger (1960), all lead to the conclusion that this is a reasonable assumption.

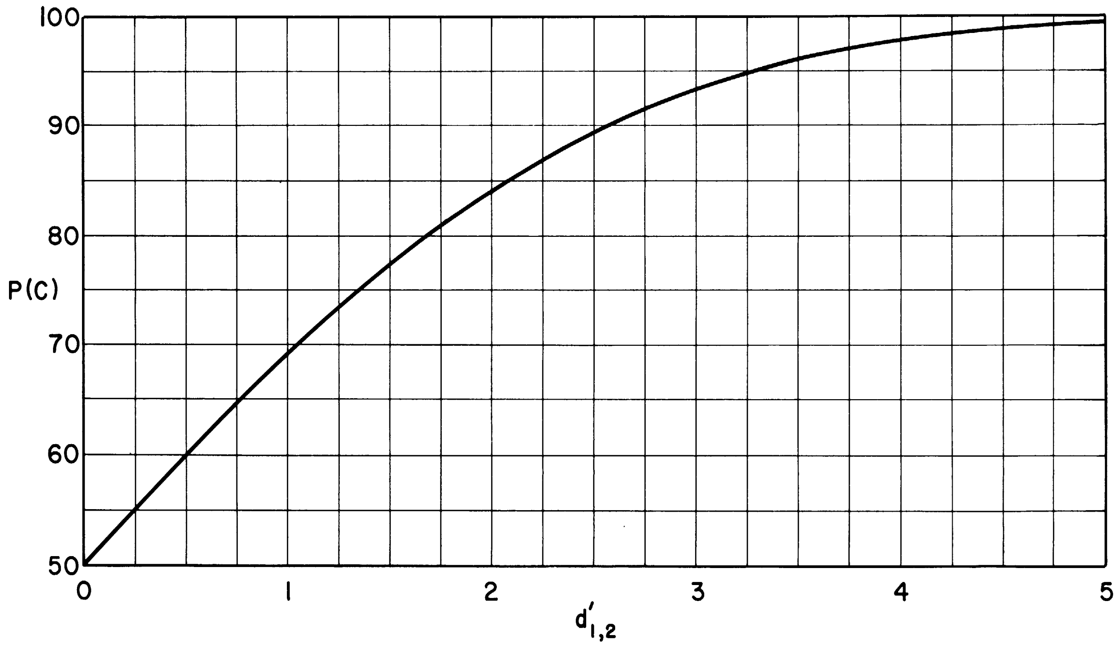

Fig. 5. Per cent correct as a function of $d'_{1,2}$.

These studies showed an effect which might be attributed to nonindependence of the two observation intervals, but such an effect was negligible with delays longer than about a half second.

Figure 4 represents the assumed statistical nature of the underlying problem facing the human observer in a psychophysical experiment. When the two-alternative forced-choice procedure is used, each trial results in the presentation of two values on the decision-axis, X, one drawn from the distribution associated with alternative A, and the other drawn from the distribution associated with alternative B. Under these conditions the observer should respond with the judgment that the input leading to the larger obtained value of X arose from the occurrence of alternative B. He will be correct to the extent that drawings from the distribution under hypothesis B actually do lead to values of X larger than those from hypothesis A. This is equivalent to an obtained difference between the two observations greater than zero. Figure 6 represents the probability of differences of various sizes arising from subtraction of values drawn from the two distributions in Fig. 4. As indicated, the mean of the difference distribution will be the difference between the means, and the variance of the difference distribution will be the sum of the variances of the two parent distributions. This way of looking at the situation is handy when there is reason to suspect that the two underlying distributions are of unequal variance. It will be used in the presentation of the theory for this reason. The obtained $d'_{1,2}$ will be $\sqrt{2}$ times the ratio of the mean to the variance of the distribution of Fig. 6.

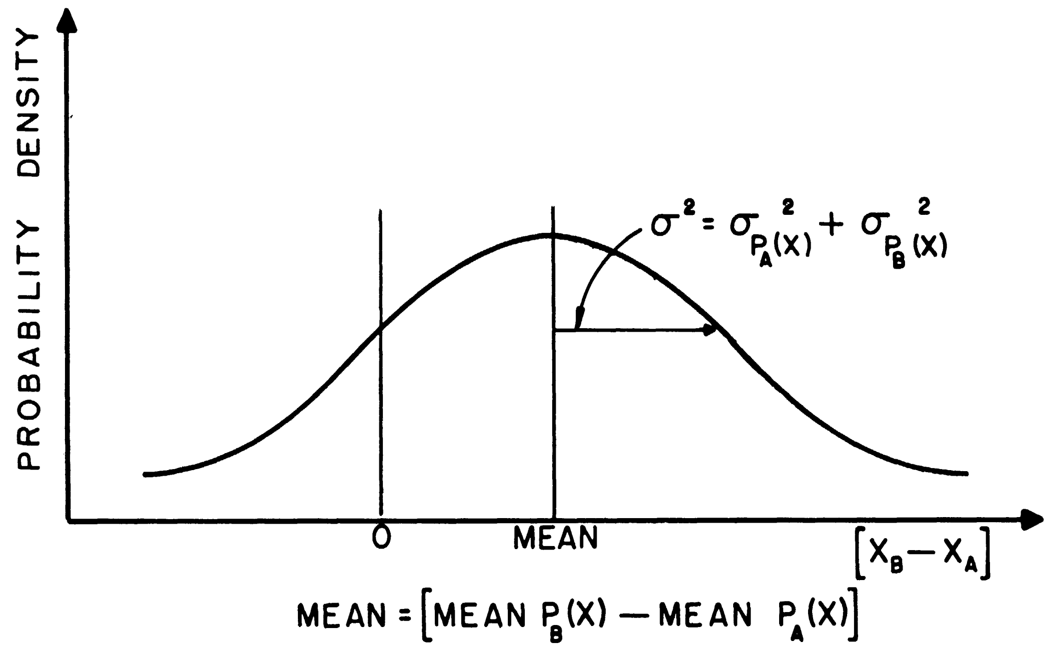

Fig. 6. Probability-density distribution of the difference between values drawn from two overlapping distributions.

When the distributions underlying the detection situation are not of equal variance, the present approach is an alternative to one proposed by Clarke, Birdsall, and Tanner (1959) for the analysis of data collected by the "yes-no" procedure. If the two procedures were used on the same set of conditions, the prediction is that $d'_{1,2}$ will be larger than $\sqrt{d_e}$ by a factor of $\sqrt{2}$. This conclusion is supported by data reported by Swets (1959).

The procedure used to estimate d' in these experiments allows, since the responses to each of the time intervals are analyzed separately, for the possibility of a degree of response bias, or "time error," on the part of the observers. To the degree that such bias is present, calculations which involve an average over the two intervals to obtain an estimated probability of correct responses underestimate the size of the actual d'. As will be shown, there is some evidence for the existence of such biases under some of the conditions of the present study.

3. EXPERIMENTAL RESULTS

3.1 Experiment 1: Effects of Signal Voltage

As was mentioned above, all the experiments were conducted in a continuous background of white masking noise. The problem facing the observers might be characterized as a signal detection task, where the signal to be detected was comprised of the energy present during the interval ΔT. The studies reported in this section compare the performance of observers on a duration discrimination task with performance on an amplitude discrimination task. If qualitative differences are found between the two sets of data, a model for duration discrimination different from that used heretofore in explaining human performance in signal detection tasks will have to be sought.

3.1.1. Effect of signal voltage on duration discrimination. In this experiment, values of T and of ΔT were fixed at .1 sec. and .03 sec., respectively. The signals to be discriminated were thus .10 and .13 sec. in duration. Both signals were presented at the same voltage, mixed with the continuous background noise. Signal voltage values were chosen to cover a wide range of detectability so that a psychophysical function could be obtained. The results are shown in Fig. 7 for three observers. In this figure the obtained $d'_{1,2}$ is plotted as a function of signal voltage on logarithmic axes. On the basis of these data we may say that detection of a duration difference can be expected to increase with signal voltage only at low signal-to-noise ratios, with the dependence becoming negligible as the signals are made "loud and clear" above the noise. The lines fitted to the data are derived from theoretical predictions, described in section 4.2.

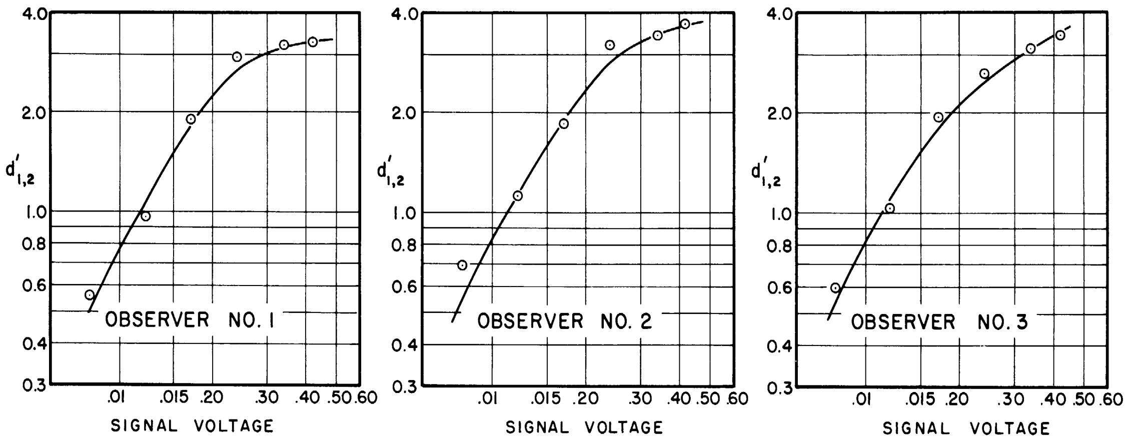

Fig. 7. Duration discrimination as a function of signal voltage.

3.1.2. Effect of signal voltage and base voltage on amplitude discrimination. This experiment is one of the two to be reported in which a procedure was used which deviated from that outlined in the section 2.2. The two-alternative forced-choice procedure was retained, and two signals, both of .1 sec. duration, were presented in the noise background. One of the two was of a fixed voltage, and the other was the fixed base voltage plus an increment. Two levels of base voltage were used, one a relatively low .003 v, and the other the relatively high, 0.42 v. Increments of .002, .004, and .006 volt were used. The results are shown in Fig. 8. In these graphs, detectability is plotted as a function of signal voltage, again on logarithmic axes. The parameter on the curves is the level of the base voltage to which the increments were added. The line marked ideal refers to detection by an ideal receiver, where performance should equal $\sqrt{2E/N_0}$, where E is the energy of the increment signal. The straight lines are a best fit by eye to the data. A recent study by Frank R. Clarke (1960) has considered in some detail

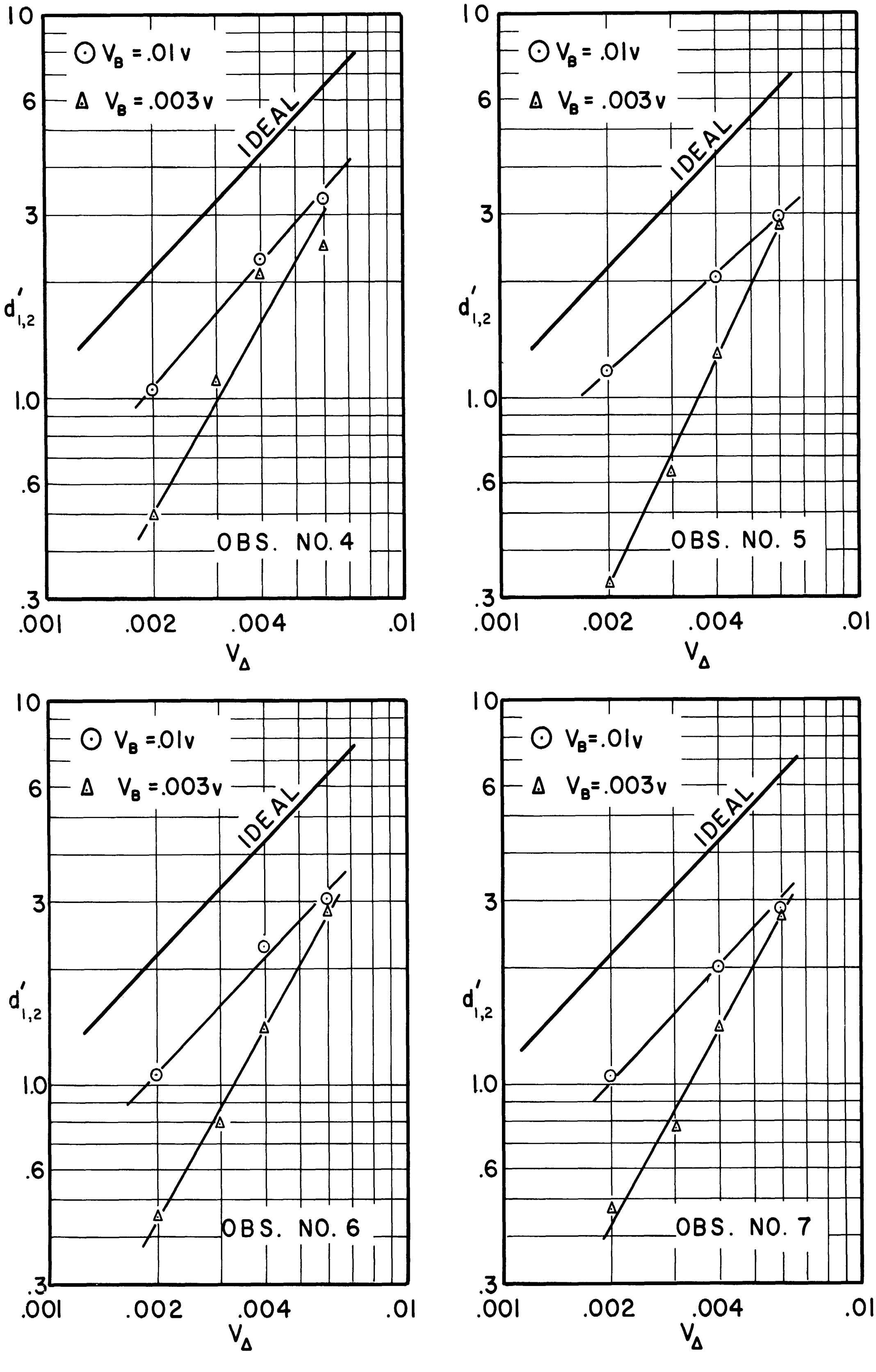

Fig. 8. Amplitude discrimination as a function of voltage increment. Signal voltage is the parameter.

the effects of base voltage and size of increment on the detectability of sine-wave signals. The main purpose of the present data is for comparison to those for duration discrimination, presented above. Amplitude discrimination with a large base varies as if the observers were able to use a process much like that of the ideal detector, with some attenuation, or added noise; while detection of an increment added to a small base is lower for lower-level increments, but increases much more rapidly with increases in amplitude. Such an increase with larger increments has been recently characterized by Tanner (1960b) as due to increased information available to the observer when the signal level is large relative to the noise level. This interpretation is suggested by a recent mathematical development by Birdsall (1960).

The shape of the function in the duration discrimination experiment suggested that the same sort of analysis, in terms of statistical uncertainty, might fruitfully be applied. The lines fitted to the data of the duration discrimination task were computed according to this assumption, by the use of a model which views duration discrimination as a primary process, complicated by lack of complete specification of some of the factors necessary for this discrimination when the signal is masked by noise.

In summary, detection of duration differences seems to be dependent on the amplitude of the signal, but not in the same way as detection of an amplitude increment. As signal voltage is increased, a level of maximal performance is approached in the duration discrimination task. There is no evidence, in the present data or in other data, of such an asymptote for performance in amplitude discrimination. Rather, a line of increasing performance with a constant efficiency is approached.

3.2 Experiment 2: Duration Discrimination as a Function of Base Time

In this section and the ones following the results reported are from a different group of observers. For this experiment a signal voltage of .084 v was used which, on the basis of the above data, can be expected to be well in the range where performance is not influenced by voltage. The effect of uncertainty of the sort discussed above can be assumed to be negligible. The duration of the increment ΔT was constant throughout the experiment at .01 sec. Five durations of T, ranging from .02 sec. to .32 sec., were used in an experimental design like that described for experiment 1. The results for four observers are presented in Fig. 9, where detectability of the duration increment is plotted as a function of base time T. Again, these are presented on logarithmic axes and the lines fitted to the data are derived from the model to be presented later. Detection falls off as the base time is increased. A straight line fit to the data would show this decrease to be not so great as $1/T$, and somewhat faster than $1/T^{1/2}$. That is, the slope of these lines lies somewhere between -1 and -1/2 on the log-log axes. The former is the prediction which could be made for the data from a simplified version of Weber's Law, which assumes that detection would vary inversely with T.

3.3 Experiment 3: Duration Discrimination as a Function of Size of Increment

In this experiment T was held fixed at .16 sec., and ΔT was varied. Two signal voltage values were used, .042 v and .012 v, and the experiment was repeated at each voltage. The results are presented in Fig. 10 in which log $d'_{1,2}$ is plotted as a function of log ΔT. The voltage at which the signals were presented is the parameter. Only three data points were

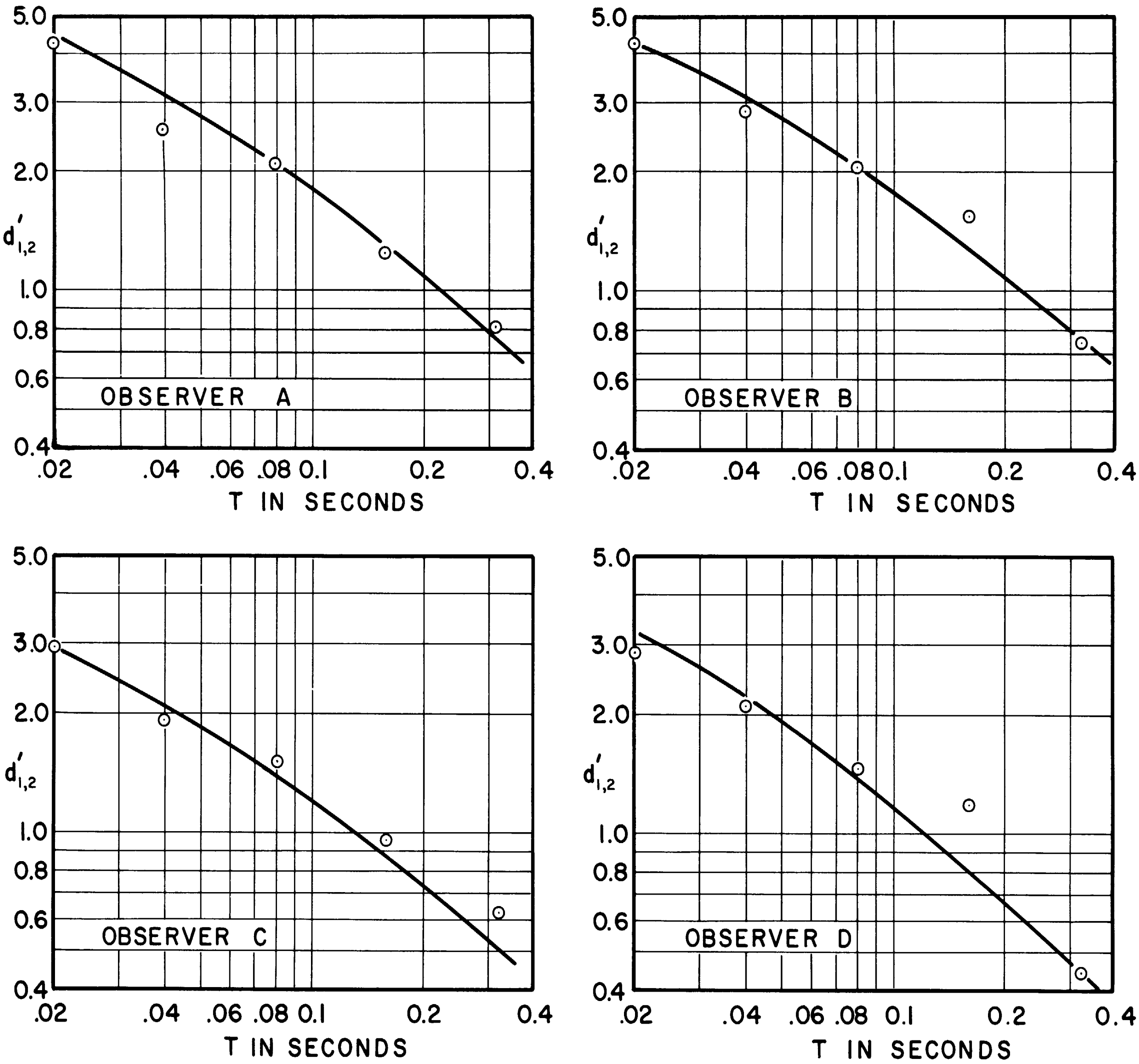

Fig. 9. Duration discrimination as a function of base duration.

obtained for the high-voltage condition, and five points were obtained for the low voltage condition. Again, about 1000 observations were collected to define each point. The lines fitted to the data, derived from the theoretical model, can be seen for all intents and purposes, to show detection to be a linear function of the size of the increment duration. Although signal voltage tends to depress performance, the shape of the two curves is the same, and there is no interaction between the effects of increment duration and signal voltage.

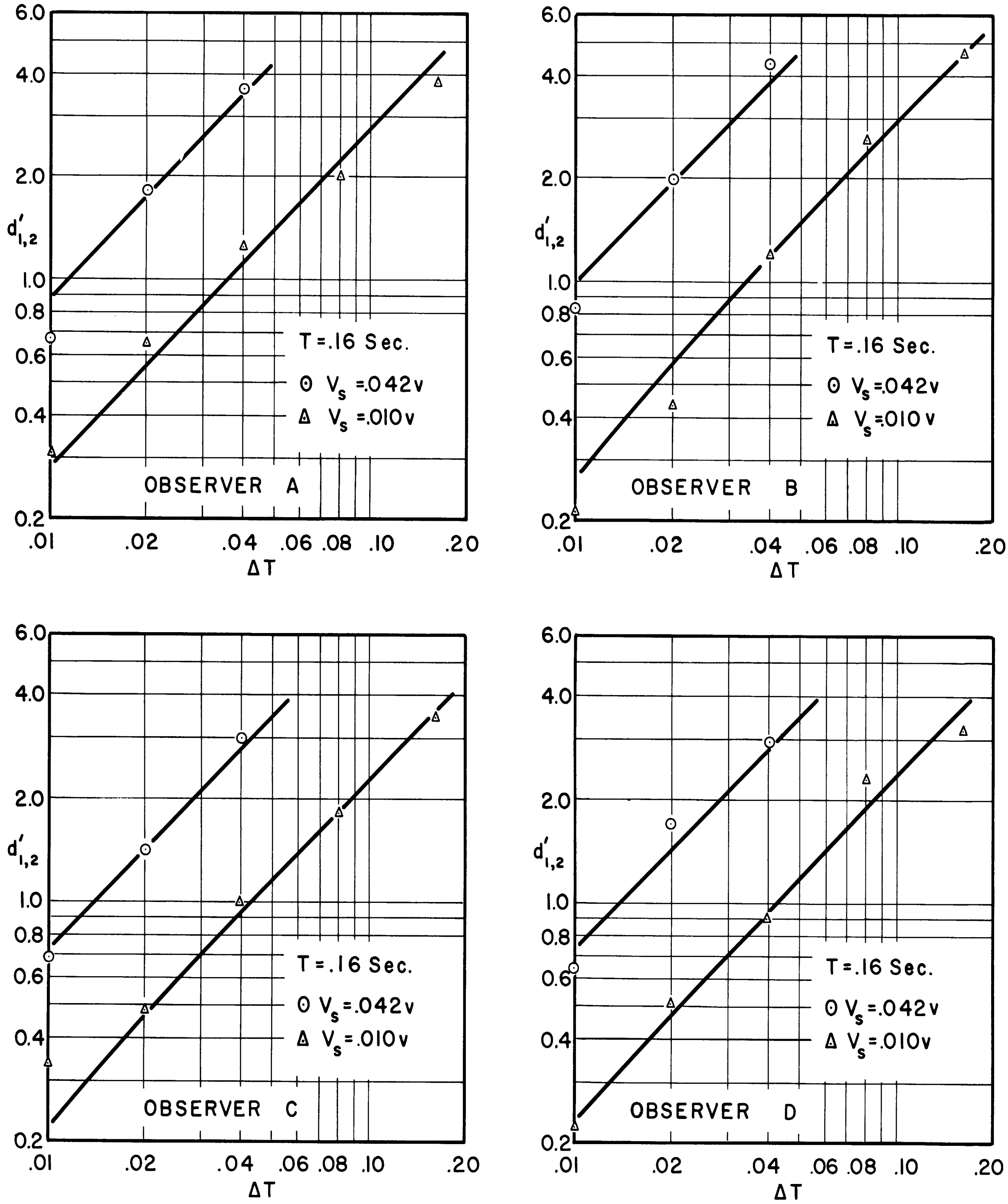

Fig. 10. Duration discrimination as a function of increment duration. Signal voltage is the parameter.

3.4 Experiment 4: Duration Discrimination as a Function of Base Time and Signal Voltage

In a fourth experiment the effect of base time on detection was measured in a situation much like that of experiment 2. In this case ΔT was .04 sec., and the experiment was repeated at two values of signal voltage, .042 v and .012 v, the same as were used in experiment 3. The results are shown in Fig. 11, where detectability is plotted as a function of T on log axes, and signal voltage is the parameter. The lines fitted to these data are predicted from the model on the basis of con-

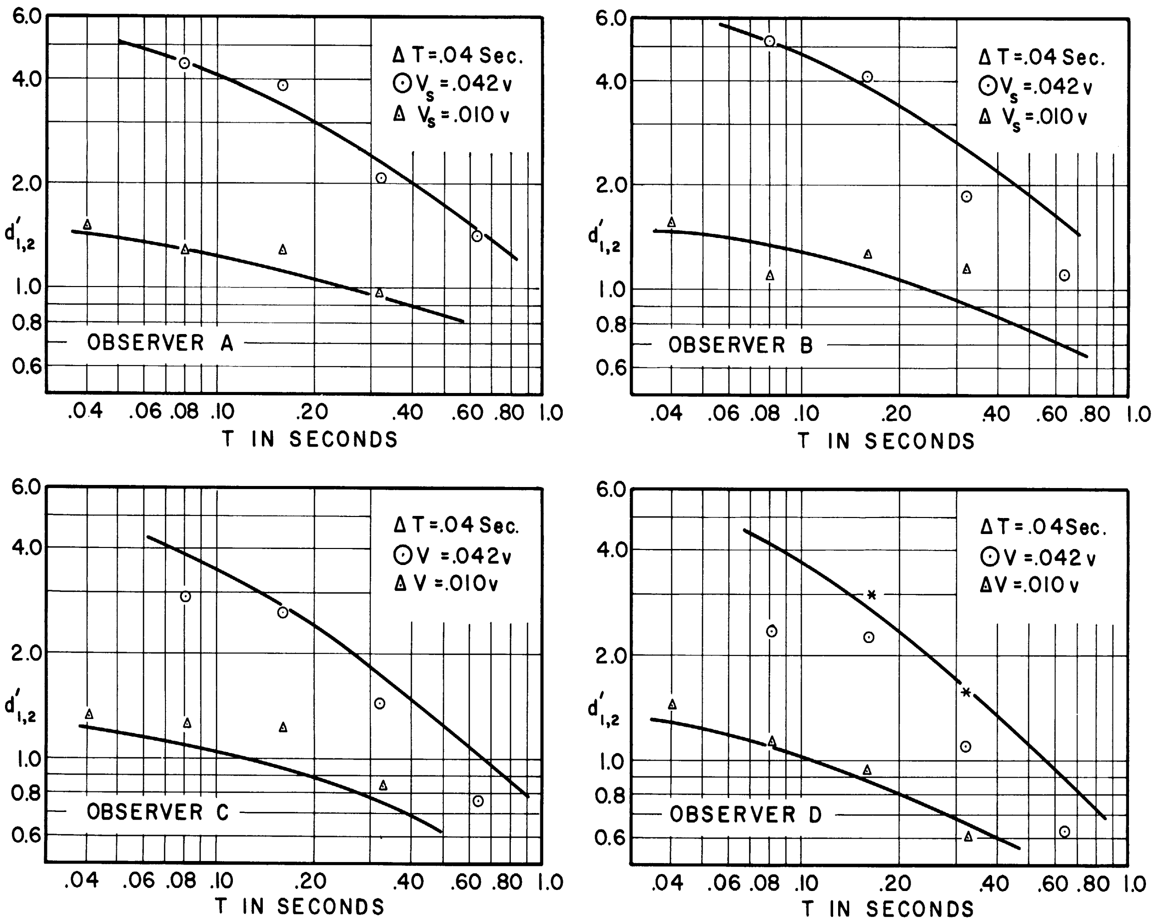

Fig. 11. Duration discrimination as a function of base duration. Signal voltage is the parameter.

stants estimated from the results of the previous two experiments. There seems to be some systematic deviation from the predicted performance, especially in the case of the lower-voltage signals. The data from Observer D for the higher voltage were collected after he had had insufficient practice. He began observing only a few days before this experiment was begun, while the other three observers had been working at this sort of task for over two weeks. The point on the graph marked with a star represents data collected in a later experiment where the same signal values were used. It is probable that the lower performance shown by observer D represents a lack of familiarity with the experimental situation. This is also shown by the excessive number of trials on which he did not respond. In general, performance of observers is remarkably constant over time. Comparisons of data taken at different times during the course of the semester showed no large deviations, except for this one instance.

In this experiment, as opposed to the previous one, there does seem to be interaction between signal voltage and the base duration in their effects on discrimination. Not only does an increase in the voltage of the signals raise the curve of detection as a function of T, but it also seems to make it steeper; or more likely, a decrease in the signal voltage makes the curve less steep.

These same functions were obtained for the observers of experiment 1, with ΔT of .03 sec. With these observers, values of T of .8 sec. were run, considerably longer than the longest T in the above data. The results are shown in Fig. 12. In general, the form of the data is much the same as that for the other observers, with the interaction of T and signal voltage apparent.

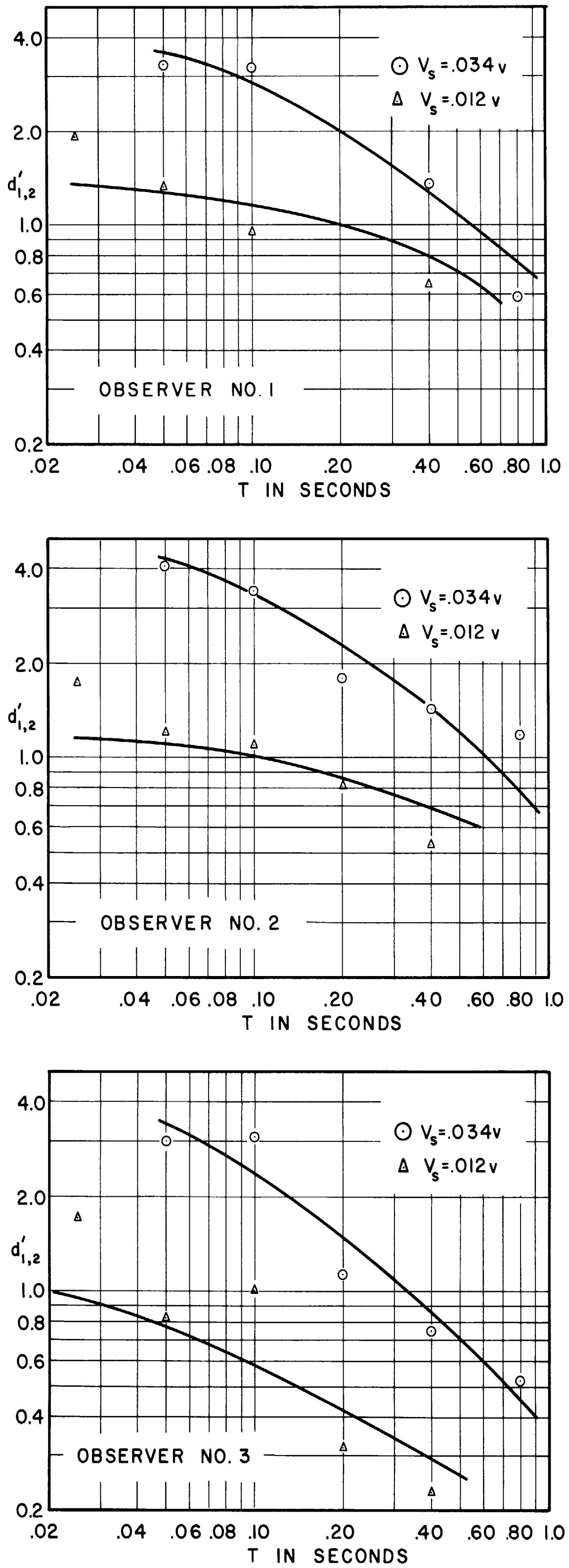

Fig. 12. Duration discrimination as a function of base duration. Data from observers of Experiment 1.

3.5 Experiment 5: Duration Discrimination with $\Delta T/T$ Constant and Difference Energy Constant

In this experiment, both T and ΔT were varied together so that the ratio of the two was constant at 1/8. (When ΔT was .01 sec., T was .08 sec., etc.) At the same time, the energy in the increment ΔT was held constant, so that the product $V_S^2 \times \Delta T$ was constant. Thus when ΔT was doubled, V_S was decreased by a factor of $\sqrt{2}$. In this experiment, at least 600 observations were collected at each point. Figure 13 shows the results. Log d' is plotted as a function of T on the ordinate. In addition, below each value of T is shown the value of ΔT, and below this, the two values of signal voltage used. The two sets of signal voltages led to two different values of constant difference energy, and these two values are the parameters on the graphs, defining two sets of points. We would expect, on the basis of the data reported above, that d' would increase roughly as a function of ΔT, but decrease somewhat as $T^{1/2}$, yielding an overall prediction for these data of increasing performance with a slope of +1/2 on a log-log plot. On the other hand, the signal voltage was decreased from left to right, as ΔT was increased, and this would be expected to have an increasing detrimental effect on performance, These notions are given precise formulation in the theoretical discussion, but for now it is sufficient to note that the overall effect of these influences might be roughly a straight line, i.e., constant detectability for all the experimental conditions. The exact predictions on the basis of the parameters estimated for these observers from previous experiments are a pair of shallow curves, concave downward, as shown on the plots. Probably due to the smaller number of observations collected for this experiment, the data show somewhat more variability than those previously

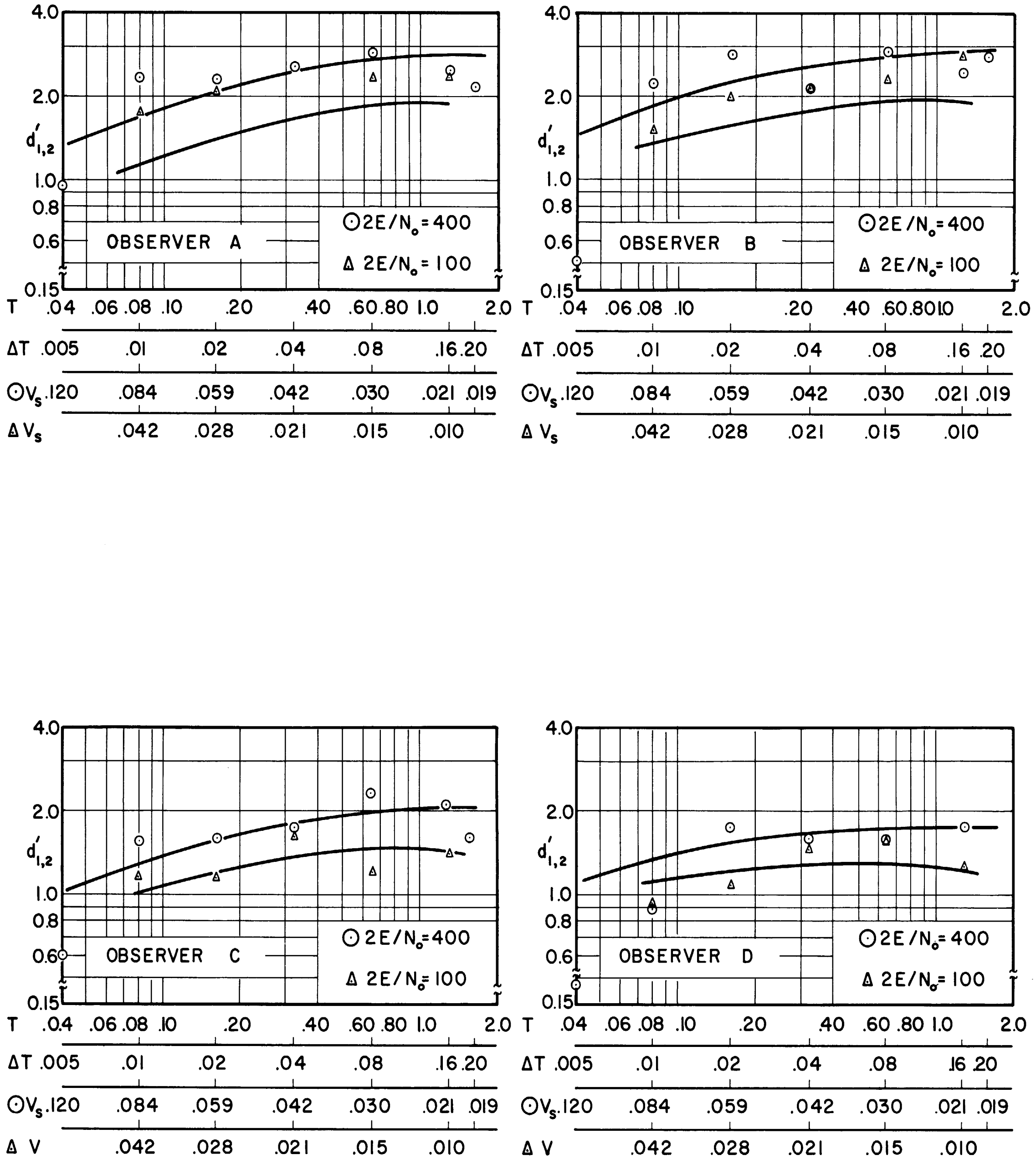

Fig. 13. Duration detection as a joint function of three variables.

presented. However, the trend of the data is clear and follows in the predicted direction. Thus, in a sense, Weber's law has been found to hold approximately for duation discrimination, but only in the case of some very special experimental circumstances.

3.6 Experiment 6: An Experimental Check on Procedure

The critical reader, in considering the procedure used in the experiments, probably will have noticed an unavoidable flaw in the design. There was, at least potentially available, visual information as to the correct response on an experimental trial, which might have been utilized regardless of the auditory signals presented. For reasons having to do with the author's theoretical approach to the empirical problem, two factors were necessary in the presentation of the stimuli. One was that there be a flash of light marking the start of each auditory signal, and the other was that there be a fixed amount of time between the end of the first signal presented in a trial and the beginning of the second signal. If these two restrictions are met, then there must be a difference between the times of occurrence of two light flashes, depending on whether the longer or the shorter of the two signals was chosen by the random selector to be presented in the first interval. In the experiments the observers might have been reacting to an absolute judgment as to whether the time between the two flashes of the signal-marking lights was .8 sec. or .8 + ΔT sec. in length. The only data presented thus far which could negate such an interpretation were those which showed duration discrimination to be a function of the signal voltage level.

In order to test whether visual cues were in fact being utilized in the present experiments, a bit of deception was practiced on the observers, the only break of faith with them in the course of all the experimentation. This was done during the final few days before the end of the observers' employment, so that danger of contamination of data collected after this experiment would be avoided. In fact none seemed to suspect that anything untoward was being done. The observers were those of experiment 1.1.

The procedure was fairly simple: on selected trials, randomly chosen, the oscillator was disconnected from the adder, so no signal was present in the continuous noise, while the trial cycle proceeded as usual. The only information present for the observers to use if they were to perform at better than chance level was the time between the flashes of the signal lights. When the signals were presented, T was .16 sec. and ΔT was also .16 sec., one of the longest values used in any of the experiments. A signal voltage of .004 v was used, one which was almost completely masked by the noise. When this procedure was used, on half the trials, randomly chosen, no signal was presented, and on half the trials two signals were presented in the usual fashion. Since scores under these conditions were very close to 50% or chance level, runs during which this procedure was used were interspersed with runs in which signals of .006 v and .004 v were presented on every trial. A voltage of .004 was sufficient to do better than chance in the experimental conditions, but they apparently could not be sure that the signal was not present when in fact it had been turned off. They assumed on the runs when the signal was intermittently turned off, that an even lower signal level was in fact being presented on all the trials. In Table I the results for four observers are presented. Only 400 observations defined each obtained d' under the no signal condition. The standard deviation figure presented in the table is the theore-

TABLE I

RESULTS OF EXPERIMENT 6

	Obs. 1			Obs. 2		
Signal Voltage	.006	.004	0	.006	.004	0
Average Per Cent Correct	66.91	57.99	57.66	71.37	60.31	49.71
Number of Observations	1000	500	400	1000	500	400
Expected Standard Deviation	1.49	2.20	2.47	1.43	2.18	2.50

	Obs. 3			Obs. 4		
Signal Voltage	.006	.004	0	.006	.004	0
Average Per Cent Correct	54.62	53.65	48.47	65.83	49.67	53.51
Number of Observations	1000	500	400	1000	500	400
Expected Standard Deviation	1.48	2.23	2.52	1.50	2.23	2.44

tical binomial variation around an obtained proportion, and not that actually computed from the data, which is somewhat larger. Only observer 1 shows what might be a significant deviation from 50% correct on trials in which no signal was presented. All except observer 4 showed substantial performance when the signal voltage was always .004, and an improvement when the signal voltage was .006.

There is some evidence that one observer may have extracted information from the signal lights in making his decisions, at least when ΔT was relatively large. In any case the help that the lights gave was negligible, and can be discounted for much smaller values of ΔT. Note that the question the present experiment asked was not "can observers make absolute judgments of empty intervals bounded by flashes of light;" but rather, "did they tend to do so in the experimental conditions as they were run, when as far as they knew the only available cues were auditory." The author is fairly certain, on the basis of his own observation in the experimental situation, as well as the available literature on perception of empty intervals, that, given practice in making such discriminations and instructions as to what were the relevant cues, the observers' performance would have been considerably higher. The primary point is that, with perhaps one exception, the observers did not score above chance with no signal present.

4. A THEORY OF TIME DISCRIMINATION

4.1 Derivation

The measure d' was given an operational definition in a prior section in order that the data could be presented against a meaningful background. It remains to develop a model from which the d' measure may be derived and to fit the model to the data. In the present section the derivation will be presented, and in a following section the method of fitting the data will be described. A final section will contain a brief discussion of some of the implications of the model, in relation to presently available theories and to related research.

It was pointed out in section 2.4 that an observer in a psychophysical experiment can be viewed as a tester of statistical hypotheses, and that Fig. 4 represented an outline of a theory of discrimination. The present task is to fill in the outline for the case of duration discrimination. This will involve specification of the abscissa of Fig. 4, and of the nature of the conditional probability distributions. Without such specifications the model is equivalent to one proposed many years ago by Thurstone (1927). The view of the human observer as rationally testing statistical hypotheses, operating on distributions with precisely specified mathematical properties, characterizes the theory of signal detectability, within the framework of which the present model was derived (cf. Tanner, 1960a, for a general discussion of these points).

The theory pictures human observers as using a separate and independent mechanism to measure short durations. It will be assumed that this mechanism functions by counting during the duration to be judged. It will be sufficient for the present analysis to view the source of pulses for the counter to be a large number of independent elements whose time of firing is randomly distributed. The performance of such a mechanism in a two-alternative duration discrimination task will be derived first, and then two restrictions on the performance of human observers in an actual experiment will be discussed.

The decision axis with which the model will deal has as units the number of pulses which the counting mechanism receives. How many such pulses will arrive at the counter during a duration T? The answer comes from the nature of the assumed source. A large number of elements, each with a fixed probability of firing at any given moment, will produce a total number of pulses over a given time interval whose statistical properties are fairly well understood (cf. Feller, 1957, p. 146 ff.). The probability of _n_ counts occurring in an interval T can be written, $P(n) = \frac{(\lambda T)^n}{n!} e^{-\lambda T}$. The constant λ is a physical parameter reflecting the probability that a given element in the pulse source will be active at a particular time. This is the Poisson distribution, which is closely approximated by the normal distribution when the quantity λT is large. The mean number of counts produced will be λT, and the variance in the number of counts will also equal λT. In the present experiments two durations were presented, T and $T+\Delta T$. When the duration T is presented the number of counts will be distributed as the left-hand distribution of Fig. 14, and when the longer duration is presented the number of counts will be distributed as the right-hand distribution. The means and standard deviations of the distributions are shown on the figure. Note that the size of the standard deviation depends on the mean of the distribution. The longer the time, the greater the variability in the measure of that

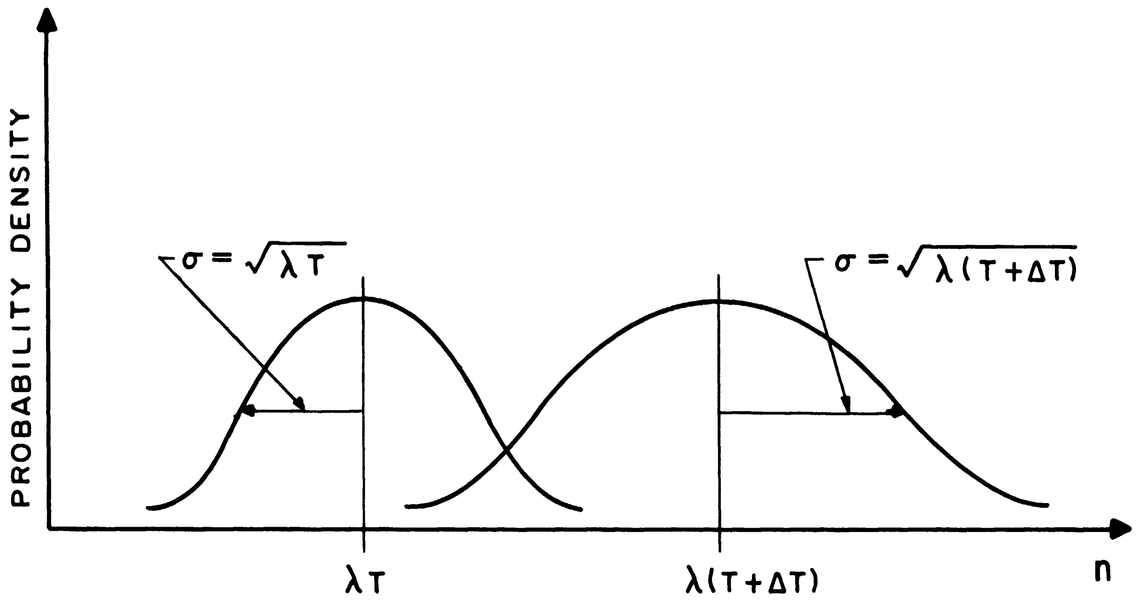

Fig. 14. Probability-density distributions of the number of counts from a random source depending on whether it is active for T or for T+ΔT.

time. In the two-alternative forced-choice experiment, according to this model, the two intervals will produce two numbers of counts, one drawn from each of the distributions of Fig. 14.

In order to receive the best possible score in this situation, the decision rule the observer should follow is to indicate the interval which produced the largest number as having contained the longer signal. This will yield the correct answer as long as the drawing from the distribution associated with the presentation of T+ΔT is larger than the drawing from the distribution associated with T, or when the difference is greater than zero. The probability-density distribution for differences between two such drawings is represented in Fig. 15. Our measure $d'_{1,2}$ is a constant $\sqrt{2}$ times the distance from zero to the mean of this difference distribution, divided by the standard deviation:

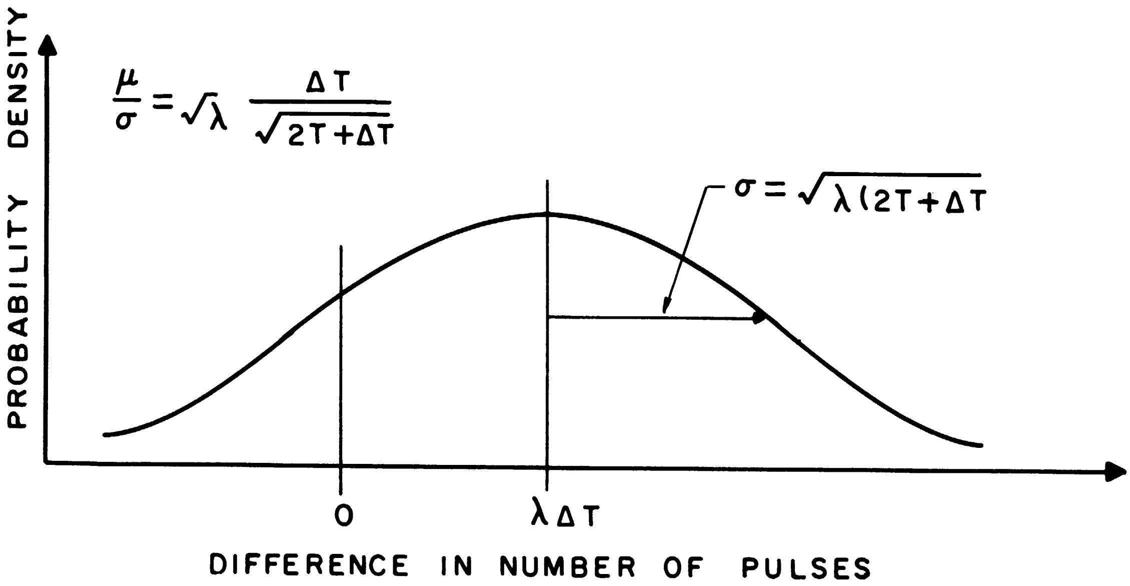

Fig. 15. Probability-density distribution for differences in the number of counts from the distributions of Fig. 14.

$$d'_{1,2} = \sqrt{2\lambda}\ \frac{\Delta T}{\sqrt{2T + \Delta T}}\ . \tag{1}$$

Under the counting model this is the basic formula for the detectability of a difference in duration. For performance to meet this expectation, however, the observer would have to have precise information as to the starting and ending times of the signal. He would also need perfect memory for the number of counts produced by the first signal until it could be compared with the number produced by the second of the pair. It is reasonable to expect neither of these requirements to be perfectly met by human observers.

That memory over the inter-signal interval is not perfect in human observers was shown by Creelman and Tanner (1960) for the case of frequency discrimination and by Tanner (1959) and Bilger (1960) for the case of amplitude discrimination. Bilger varied both the time between presentations and their duration and found that both factors tended to decrease performance. He found that efficiency [proportional to $(d')^2$] varied inversely as $[1 + K(T + \tau)]$, where T was the duration of the signals, τ was the inter-signal interval, and K was a constant characterizing the individual observers. In the present experiments the time between the two stimulus presentations was constant, but the duration of the stimulus was a variable. Longer base times would be expected to cause a decrease in performance, and to be consistent with the prior findings the basic duration detection formula was modified to yield, when squared to eliminate radicals,

$$(d')^2 = \frac{1}{1 + KT} \; \frac{\lambda \Delta T^2}{2T + \Delta T}, \tag{2}$$

where K is again a constant characteristic of the individual observer.

The duration counting mechanism must start precisely when the relevant signal begins, and stop precisely at its end. To the extent that the signals which mark the beginning and the end of the time intervals are masked by a background noise, or are otherwise ambiguous, the observer will be uncertain as to when these occur. This will lead to a larger variance in the number of counts during an observation. This added variance, σ_v^2, should be expected to be an inverse function of the signal power, or signal-to-noise ratio. The nature of this relation will be considered in the discussion under section 4.2. The final expression is then,

$$(d')^2 = \frac{1}{1 + KT} \; \frac{\lambda \Delta T^2}{2T + \Delta T + \sigma_v^2}. \tag{3}$$

The constants λ and K reflect the rate of firing of the pulse source and the ability to hold in mind the number associated with one signal while listening for another. These were fitted for individual observers and the nature of the variance due to signal level was ascertained.

4.2 Empirical Fit to the Data

The data from experiment 1 were fitted by a two-step process. First an asymptote for obtained performance was estimated. For two of the three observers this was not difficult, for they were clearly near it at the largest voltage levels. It was assumed that at the asymptote the variance due to starting-time and duration uncertainty was reduced to zero. Then the data were used to compute values of σ_v^2 necessary to yield the obtained performance at each signal-voltage used. These values were plotted on logarithmic axes, and found to be fitted very well with a straight line. The relationship was fitted by the equation $\sigma_v^2 = A\ v_s^{-b}$, with b somewhat larger than two and varying among observers. The constant A, of course, will depend in any situation on the noise level used in the experiment, as well as the individual observer. The curves drawn to the data in Fig. 7 were derived from the straight line relating σ_v^2 to signal voltage. The interpretation of the effect of lowered signal power, as causing an increase in the variability of a measure of the duration of auditory signals, is supported by this analysis.

Experiments 2, 3, and 4 were a related set performed on the same observers. In experiment 2 the voltage was large enough to assume that the starting-time and duration of the signals were precisely marked. The variance term was thus assumed negligible, and values of K and λ were chosen for each observer to fit the obtained data. In general there was little difficulty in obtaining sets of constants which gave very close fits.

The next experiment used two lower values of signal voltage and allowed an estimate of the added variance due to starting-time and duration uncertainty for each voltage value. The lines shown on Fig. 10 are best fits by eye with the restriction that the slope of the function be nearly unity, as predicted by the theoretical equation.

The estimated values of σ_V^2, together with the estimated constants, were carried over to experiment 4 and used to generate the predicted functions shown in Fig. 11. The two signal voltage levels were retained and a range of values of T was explored with a constant ΔT, different from than used in experiment 2. The data from two of the observers show a suggestion that performance with low voltage signals falls off faster as T is increased than predicted from the mathematical model, and this may be the case. For the present it was decided to retain the present model as a good approximation to the form of the data rather than attempt to complicate it further. In general the fit of the predicted function is close to the form of the obtained data, and the model is given support from them.

Figure 12 represents a similar set of data from the observers of experiment 1. Here the constants were estimated from the data themselves with the restriction that they be consistent with the earlier findings, in particular that the asumptote of Fig. 7 and the obtained values for σ_V^2 be the same. With the exception of Observer 3, who shows the strongest tendency toward a steeper slope than any of the other observers when the signal voltage is low, the agreement is again good. It would seem that for some observers at least a modification of the theory will be needed, but the present data were insufficient to indicate the nature of the necessary modification.

Table II lists the constants used to fit the data for Figs. 11 and 12. Note that the number λ is sufficiently large to support the normal approximation assumption used to find d' from per cent correct and, with the exception of Observer 3 again, the number K is relatively small.

TABLE II

CONSTANTS USED TO FIT THE DATA

Observer	λ	K	σ^2_{034}	σ^2_{01}
1	$.385 \times 10^4$	3.0	.10	1.70
2	$.51 \times 10^4$	3.5	.076	3.17
3	1.40×10^4	20.0	.40	10.0
Observer	λ	K	σ^2_{042}	σ^2_{012}
A	1.13×10^4	6.06	.41	7.00
B	1.105×10^4	5.93	.25	6.50
C	$.49 \times 10^4$	5.94	.15	4.14
D	$.65 \times 10^4$	12.35	.09	4.0

The procedure used to predict the results of experiment 5 was not as precise as those above, in that it involved a considerable extrapolation. The constants in Table II were used, the two values of σ^2_V were plotted on log paper, and a straight line was drawn through them of approximately the right slope on the basis of the data from the other observers. From this line values of σ^2_V were read off and used to compute predicted performance at the various voltage levels. Although not enough data could be obtained to define the curves precisely, they do show fair agreement with the model. A further reason for the variability probably is that in this experiment, where everything was varied at once, the observers did not really have sufficient opportunity to "tune in" and become accustomed to the particular set of experimental conditions in the course of the ten or twenty practice trials given after each change of conditions.

Whatever the inaccuracies involved in the prediction scheme utilized for this experiment and the variability in the data, two points can be made. Over a fairly wide range of conditions the model predicts the form and level of the obtained performance on the duration discrimination task. It is clear in the data of each observer, however, that when ΔT is made as short as 5 milliseconds the model no longer seems applicable. In every case performance was far below the prediction from the model. The upper limit on the range over which the model applied could not be tested, for the equipment used could only reliably produce signals of durations up to the longest used in this experiment, slightly more than two seconds.

Although the constants fitted to the data from each observer were arrived at by a process of trial and error, and the criterion for goodness of fit to the data was informal, it has been demonstrated that the model proposed for the discrimination of duration has considerable predictive ability in moving from one set of conditions to another.

4.3 Discussion

A perceptual theory of the type advanced above must satisfy a number of criteria beyond simply being an adequate fit to the data. The kinds of mechanisms implied by the mathematical statement must be reasonable; suggestive of possible physiological processes. The model must be not inconsistent with data from related areas of experimentation and it must have a foundation in available mathematical techniques.

The model suggests possible neurological processes which could yield the data of these experiments. It does not specify a unique process; a number of possibilities suggest themselves. The "counting mechanism," a simple accumulator, could store neural pulses in reverberatory circuits or, for that matter, store an electrical charge due to a chemical process. The random nature of the source seems feasible in consideration of either chemical or neural processes. It should be pointed out that the mathematical model, although perhaps suggestive of possible mechanisms, makes no commitment to a particular physical model, or for that matter to the existence of a physical entity such as a "counter."

It is difficult to test the model proposed here against alternative formulations, for none could be found in the literature which were precisely enough specified to make comparison possible.

The model does take a stand on the old question of whether the perception of time is an independent process or conditional upon the nature of the sensory input marking the judged time. Fraisse (1957), to take an instance of a modern writer of the opposing camp, states that the experience of an elapsed duration is critically dependent upon the sensory input during it. If this were the case, we might well expect the loudness level of auditory signals to begin to have a determining influence at just those levels where this and other studies show the disappearance of any effect of signal level.

A model proposed by Stroud (1954) is not inconsistent with the data presented or with the mathematical formulation of the counting mechanism model. Stroud proposed that subjective time is quantized and that events happening within any one unit interval of time cannot be differentiated by the human observer. His article brings a great deal of evidence to bear on this hypothesis, but no experimental evidence from the perception of durations _per se_ is cited. The mathematical statements of the two models can be made equivalent if: 1) the length of the "psychological moment" is made very short, in the order of micro-seconds -- much shorter than proposed by Stroud; or 2) successive "moments" are perceptually independent, i.e., the decision during one moment as to the status of the input signal has no effect on the decision with regard to the next "moment." In the former case the practical distinction between the two approaches disappears and the latter case leads to the use of binomial statistics, which lead to generalizations much like those from the Poisson statistics of the present model.

The data reported by Stott (1934) were converted to d' values from the obtained per cent correct scores, and are plotted in Figs. 16 and 17. In Fig. 16 best-fit lines of slope 1 were fitted to the data for each T used. The points of Fig. 17 represent the intersection of each of these lines with the ΔT of .04 sec. For comparison, the data from experiment 5 are also given, averaged across observers. Except for the considerable difference in level of performance, the general form of the curves seems in agreement. Stott used unpracticed observers and, as mentioned earlier, probably did not have as precise control of his auditory signals as present methods afford. These two factors account for the discrepancy between the data.

The theory has had nothing to say about the classical area of investigation in this field, time errors. It predicts an increased variance in judgment with elapsed time, which is an entirely different matter. The

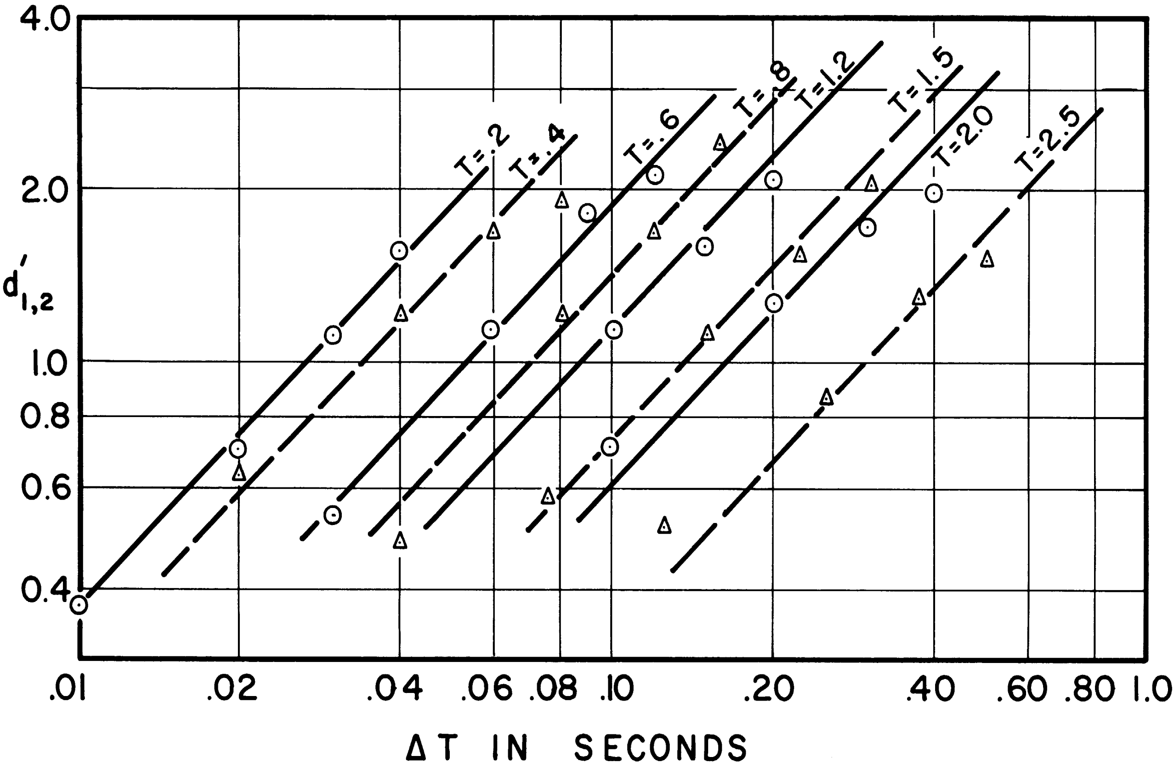

Fig. 16. $d'_{1,2}$ as a function of ΔT. Data recalculated from Stott.

direction of the "drift" with time is not predicted, and in fact should be random. Nothing is said about systematic "fading" or "inhancement" of an image over time. In fact there were in the data some indications of preference on the part of individual observers to respond with one interval or the other under some conditions. This was evidenced by a significantly larger per cent correct score in one interval of the two-alternative situation. However, in agreement with Stott and with Woodrow, no systematic effect could be found which was consistent from experiment to experiment for any observer. In an experimental situation where knowledge of results is given after each trial, and the observers are instructed to be right rather than to "report their experience," there

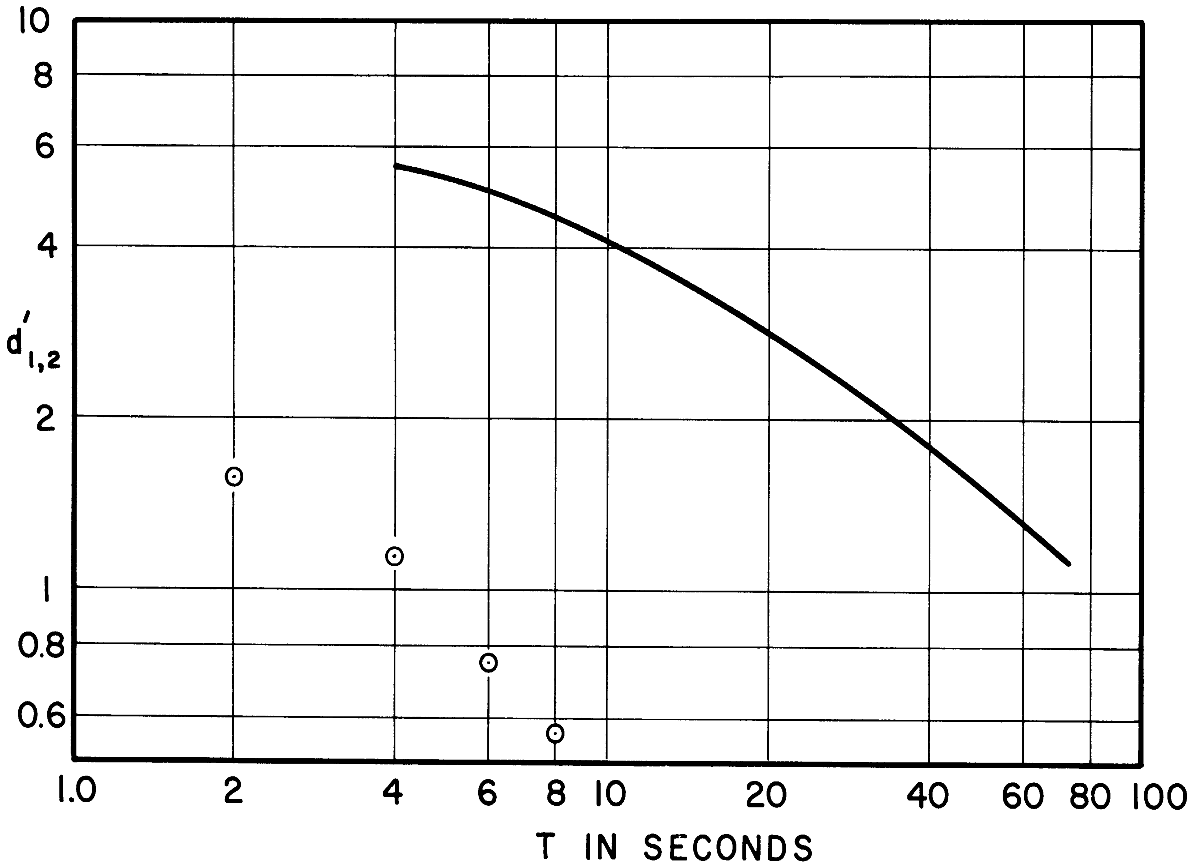

Fig. 17. $d'_{1,2}$ as a function of T, read from Fig. 16. Comparison with average data from Experiment 5.

is no reason to expect such effects. The evidence for response bias tended to occur when the discrimination task was a difficult one, whether the difficulty was due to low signal voltage, long T, or short ΔT. Although the data are insufficient to test it, one speculation as to the cause of "time error" data can be made. This is that the observer enters the situation with some sort of response bias, which is more likely to influence his responses when the discrimination is difficult and the sensory information scanty than when the discrimination is not so difficult and the sensory input has a greater probability of determining the responses of itself, overriding any possible extraneous determiners.

Readers almost certainly will differ in the degree of *a priori* reasonableness they will attach to the hypothetical mechanisms underlying the proposed model. It seems fairly clear that no sort of constantly running "internal clock" will account for the data. The model does not insist that the constant λ retain the same value under all circumstances. Studies of Hoagland (1933) and Hirsh, Bilger, and Deatherage (1956) both showed that external factors, in one case body temperature and in the other case extraneous stimulation, can affect what we refer to as the level of activation of the pulse source. In the terminology of Hirsh, et al., time seems to "run faster" when, for instance, light signals are presented under conditions of fairly loud auditory noise, than under conditions of quiet. These data suggest that λ reflects some generalized level of activation, and the duration counter, if it exists anywhere in the nervous system, can receive counting pulses from many different sources.

Mention should be made here of possible implications of these results for psychophysical theory. A signal-detection task requires that the observer know exactly when the signal is to begin and when it will end and that he consider the auditory input only during that time. Relatively weak signals masked by noise do not carry this information by themselves, and for this we envision a turn-about for the counting mechanism, controlling the input rather than reacting to it. The most recent of the studies on the "time-intensity trade" concept, done by Green, Birdsall, and Tanner (1957), showed a drop in detection for long signals. Keeping energy constant required a decrease in signal power, and thus, perhaps, a greater reliance on the internal clocking mechanism to specify the relevant observation time. Unfortunately, greater time also means, according to the model, greater variance in the specification of duration and thus a lowered detection rate. When an amplitude-discrimination experiment was run on the Green, Birdsall, and Tanner paradigm, the characteristic decrease in detection at long durations was not observed (Creelman, 1960b). Presumably in this case the auditory signals were sufficiently detectable to mark the observation interval quite precisely, eliminating the need to rely on the internal timing mechanism. These are the sorts of interpreations of psychophysical data suggested by the counting-mechanism model.

5. SUMMARY AND CONCLUSIONS

A model has been proposed to account for the way in which human observers discriminate durations of signals. Although the data were all taken using auditory signals and "filled intervals," the model is not restricted with regard to sensory modality, or the way of delimiting the durations compared. Duration discrimination was found to be dependent on the detectability of the signals, and this was taken to be the result of uncertainty as to the starting-time and duration of the relevant signals when signal level was lowered. The effect of signal level on duration discrimination became negligible as soon as the signals were made highly detectable.

The "counting model" uses some simplifying assumptions in order to arrive at a rigorous mathematical formulation which will predict precisely the level of performance for any set of experimental variables. In point of fact, the main finding of the study is not contained in the absolute level of performance, but in the form the data take when the relevant variables are manipulated. The performance level is dependent to a large extent on the individual observers, the equipment, and the experimental procedure. This was shown in the comparison of the present data with those of Stott.

The search for the "lower limits" of discrimination has often proven to be disappointing. On the other hand, much recent research has proven the reasonableness of a search for the factors which influence human performance in specified experimental settings. This is a less ambitious goal, perhaps, but has the advantage of yielding experimental questions which can be rigorously tested, with prediction between related experiments possible. The present experiments were conducted within this framework.

In summary, a set of experiments was conducted in which base time, increment time, and signal level were the independent variables. A two-alternative forced-choice procedure was used throughout, and the results were interpreted within a framework which views discrimination as a statistical decision-making process. The model developed within this framework required the estimation of two constants for each observer, and with these constants accurate prediction could be made over quite a wide range of experimental conditions.

LIST OF REFERENCES

Axel, R. Estimation of time. _Arch. Psychol._, 1925, 12, No. 74.

Bilger, R. C. Laboratory facilities employed in psychophysical memory experiments. Tech. Memo. No. 72, Electronic Defense Group, The University of Michigan, 1959.

_____. Amplitude memory as a function of signal duration. Psychonomic Society, Program of First Meeting, 1960, 17 (abst).

Birdsall, T. G. Detection of a signal specified exactly with a noisy stored reference signal. _J. Acoust. Soc. Amer._, 1960, _32_, 1038-1045.

Blackwell, H. R. Psychophysical thresholds. Engineering Res. Bull. No. 36, Engineering Research Institute, The University of Michigan, 1953.

Boring, E. G. _Physical Dimensions of Consciousness._ V, New York: Century, 1933 .

_____. Temporal perception and operationism. _Amer. J. Psychol._, 1936, _48_, 519-522.

_____. _Sensation and Perception in the History of Experimental Psychology._ New York: Appleton-Century-Crofts, 1942.

Chistovich, L. A. Perception of time interval between two short sound pulses. _Soviet Physics: Acoustics_, 1959, _5_, 480-484.

Clarke, F. R. Amplitude discrimination as a function of three variables. _J. Acoust. Soc. Amer._, 1960, _32_, 931 (abst.).

Clarke, F. R., Birdsall, T. G., and Tanner, W. P., Jr. Two types of ROC curves and definitions of parameters. _J. Acoust. Soc. Amer._, 1959, _31_, 629 (L).

LIST OF REFERENCES-- Continued

Clausen, J. An evaluation of experimental methods of time judgment. J. Exp. Psychol., 1950, 40, 756-761.

Creelman, C. D. Applications of signal detectability to psychophysical research: a bibliography. Tech. Memo. No. 79, Electronic Defense Group, The University of Michigan, 1960a.

_____. Detection of complex signals as a function of signal bandwidth and duration. (Submitted to J. Acoust. Soc. Amer.) 1960b.

Creelman, C. D. and Tanner, W. P., Jr. Frequency discrimination as a function of inter-signal interval. Tech. Report, Cooley Electronics Laboratory, The University of Michigan, 1960 (in preparation).

Dunlap, K. Rhythm and time. Psychol. Bull., 1911, 8, 230-242; 1912, 9, 177-199; 1914, 11, 169-171; 1916, 13, 206-208.

Feller, W. An Introduction to Probability Theory and Some of Its Applications. Vol. I, New York: John Wiley and Sons, 1957, 2nd edition, 146 ff.

Fraisse, P. Psychologie du Temps. Paris: Presses Universitaires de France, 1957.

Frankenhauser, M. Estimation of time. Stockholm: Almquist and Wiksell, 1959.

Garner, W. R. Advantages of the discriminability criterion for a loudness scale. J. Acoust. Soc. Amer., 1958, 30, 1005-1012.

Gilliland, A. R., Hofeld, J., and Eckstrand, G. Studies in time perception. Psychol. Bull., 1946, 43, 162-176.

Green, D. M., Birdsall, T. G. and Tanner, W. P., Jr. Signal detection as a function of signal intensity and duration. J. Acoust. Soc. Amer., 1957, 29, 523-531.

LIST OF REFERENCES-- Continued

Gregg, L. W. Fractionation of temporal intervals. J. Exp. Psychol., 1951, 42, 307-312.

Grindley, P. F. The discrimination of short intervals of time by fingertip and by ear. Amer. J. Psychol., 1932, 44, 18-43.

Henry, F. M. Discrimination of the duration of a sound. J. Exp. Psychol., 1948, 38, 734-742.

Hirsh, I. J. Certain temporal factors in audition. Science, 1952, 116, 523.

Hirsh, I. J., Bilger, R. C. and Deatherage, B. H. The effect of auditory and visual background on apparent duration. Amer. J. Psychol., 1956, 69, 561-574.

Hoagland, H. The physiological control of judgments of duration: evidence for a chemical clock. J. Gen. Psychol., 1933, 9, 267-287.

James, W. Principles of Psychology. Vol. I, Dover, 1950 (1890), 605-642.

Kowalski, W. J. The effect of delay upon the duplication of short temporal intervals. J. Exp. Psychol., 1943, 33, 239-246.

Lauder, J. A. An electronic random number generator. Elect. Engring., 1959, 78, 238-242.

Licklider, J. C. R. Three Auditory theories. in S. Koch (ed.) Psychology: A Study of a Science. Vol. 1, New York: McGraw-Hill Book Company, Inc. 1958.

Lifshitz, S. Apparent duration of sound perception and musical optimum reverberation. J. Acoust. Soc. Amer., 1935-36, 7, 213-216.

Nakajima, S. Time errors in the successive comparison of tonal durations. Jap. J. Psychol., 1958, 29, 18-27.

LIST OF REFERENCES-- Continued

Needham, J. G. The time error as a function of repeated experimentation. Amer. J. Psychol., 1934, 46, 558-567.

Nichols, H. The psychology of time. Amer. J. Psychol., 1891, 3, 453-529.

Oléron, G. Influence of the intensity of a sound on its apparent duration. Année Psychologigue, 1952, 52, 382-392.

Peterson, G. E. and Lehiste, I. Duration of syllable nuclei in English. J. Acoust. Soc. Amer., 1960, 32, 693-702.

Peterson, W. W., Birdsall, T. G. and Fox, W. C. The theory of signal detectability. Inst. Radio Engrs., Trans. Professional Group on Information Theory, 1954, PGIT-4, 171-212.

Phillip, B. R. The effect of interpolated and extrapolated stimuli on the time order error in the comparison of temporal intervals. J. Genet. Psychol., 1947, 36, 173-187.

Postman, L. and Miller, G. A. Anchoring of temporal judgments. Amer. J. Psychol., 1945, 58, 43-53.

Ross, S. and Katchmar, L. The construction of a magnitude function for short time intervals. Amer. J. Psychol., 1951, 64, 397-401.

Smith, M. and Wilson, Edna A. A model for the auditory threshold and its application to the multiple observer. Psychol. Monogr., 1953, 67, No. 9 (Whole No. 359).

Stevens, S. S. On the psychophysical law. Psychol. Rev., 1957, 64, 153-181.

Stott, L. H. Time-order errors in the discrimination of short tonal durations. J. Exp. Psychol., 1935, 18, 741-766.

LIST OF REFERENCES-- Continued

Stroud, J. M. The fine structure of psychological time. in H. Quastler (ed.) Information Theory in Psychology. Glencoe, Ill.: Free Press, 1955, 174-207.

Swets, J. A. Indices of signal detectability obtained with various psychophysical procedures. J. Acoust. Soc. Amer., 1959, 31, 511-513.

Swets, J. A., Tanner, W. P., Jr. and Birdsall, T. G. Decision processes in perception. Psychol. Rev., 1960 (in press).

Tanner, W. P., Jr. Theory of recognition. J. Acoust. Soc. Amer., 1956, 28, 882-888.

_____. Effect of memory for amplitude on amplitude discrimination. J. Acoust. Soc. Amer., 1959, 31, 1575 (abst.).

_____. The theory of signal detectability as an interpretive tool for psychophysical data. Tech. Memo. No. 78, Electronic Defense Group, The University of Michigan, 1960a.

_____. Mathematical models in sensory perception. Proc. N. Y. Acad. Sci., 1960b (in press).

Tanner, W. P., Jr., Birdsall, T. G. and Clarke, F. R. The concept of the ideal observer in psychophysics. Tech. Rept. No. 98, Electronic Defense Group, The University of Michigan, 1960.

Tanner, W. P., Jr. and Swets, J. A. A decision-making theory of visual detection. Psychol. Rev., 1954, 61, 401-409.

Thurstone, L. L. A law of comparative judgment. Psychol. Rev., 1927, 34, 273-286.

Titchener, E. B. Experimental Psychology. Vol. II, New York: MacMillan, 1905.

LIST OF REFERENCES-- Continued

Urban, F. M. Systematic errors in time estimation. Amer. J. Psychol., 1907, 18, 187-193.

Wallace, M. and Rabin, A. I. Temporal experience. Psychol. Bull., 1960, 57, 216-236.

Weber, A. O. Estimation of time. Psychol. Bull., 1933, 30, 233-252.

Woodrow, H. Behavior with respect to short temporal stimulus forms. J. Exp. Psychol., 1928, 11, 167-198.

_____. The effect of practice upon time-order errors in the comparison of temporal intervals. Psychol. Rev., 1934, 42, 127-152.

_____. Time perception. in S. S. Stevens (ed.) Handbook of Experimental Psychology. New York: Wiley and Sons, Inc., 1951.

Woodrow, H. and Stott, L. H. The effect of practice on positive time-order errors. J. Exp. Psychol., 1936, 19, 694-705.