ENGINEERING RESEARCH INSTITUTE
UNIVERSITY OF MICHIGAN
ANN ARBOR

A THEORY OF RECOGNITION

Technical Report No. 50
Electronic Defense Group
Department of Electrical Engineering

By: W. P. Tanner, Jr.

Approved by: _A. B. Macnee_

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

ABSTRACT

The theory of signal detection as applied to the
human observer is reviewed. The theory is then extended to
include the simple case of recognizing a signal as one of a
set of two alternatives, and experiments relating to this case
are reported. The principles upon which the theory can be
extended to cover more complex alternatives are developed.

A THEORY OF RECOGNITION

1. BACKGROUND FOR THE THEORY

The theory of statistical decision has previously been applied to the
oblem of the sensory detection of signals. In this paper a simple recognition
oblem is analyzed in the framework of statistical decision. The extension of
e detection theory to include the problem of recognition is the first phase of
general expansion of the theory to encompass the field of perception. To make
is paper self-contained, a brief review of the detection theory, and its corres-
ndence with data presently available, is presented. This presentation is fol-
wed by a development of the recognition theory for the simple case studied here.
nally, certain principles are outlined for the general extension of the theory.
e paper is presented largely in terms of auditory theory, although it is felt
at the theory is applicable to the entire field of human perception.

1 The Detection Theory
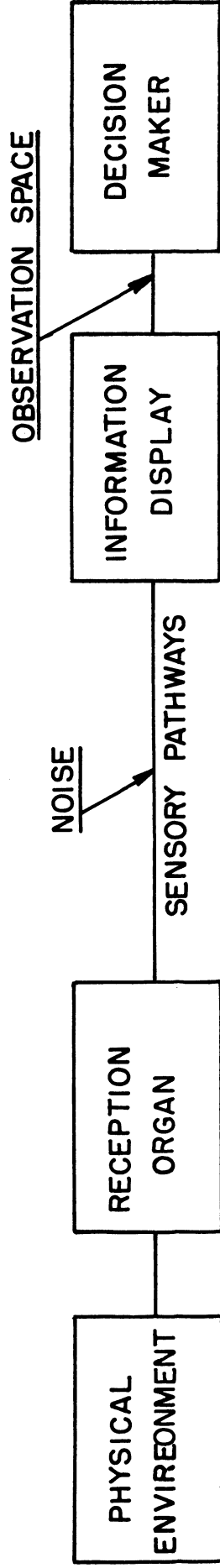
The application of the theory of signal detection, or statistical de-
sion theory, depends on three basic assumptions.

1) Sensory systems function primarily as communication channels.

2) Sensory systems are noisy channels.

3) Central mechanisms, where decisions are made, are capable of
   approximating optimum use of the information gathered by the
   peripheral sensory mechanisms.

The assumptions are illustrated by the block diagram in Figure 1. Th

physical environment, which in the theory discussed below is equivalent to the

signal and noise generators, presents an input to the receptor organ. The func

of this organ is to transform the physical energy into neural activity. The in

mation contained in the neural activity is then transmitted along the sensory

pathways. These pathways are subject to internally generated noise. The infor

tion plus the noise added in transmission, is then presented to, or displayed a

cortical centers. The presentation is here considered as an observation, x, up

which the decision is based. The key points of the theory are that noise is ad

in the transmission of the information, and that the decision making device is

perfect device. A decrement in performance (from that which would be expected

a perfect device were placed at the receptor level) is attributed to the noise

added in sensory transmission.

The fundamental problem in signal detection is the fixed observation

interval problem[1]. That is, the observer is asked to observe the output of a s

sory system, and is then asked to decide on the basis of his observation, wheth

this output arose from noise alone, or from signal plus noise. In this theory

signal is known to be from a certain ensemble of signals. This is the criterio

approach. In other words, the observer chooses a set of observations (the cri-

terion A) which he will say represents signal plus noise; all other observations

in the complement of the criterion, CA, and he will say that these represent no

alone. The notation SN denotes signal plus noise, and N denotes noise alone.

there are only a countable number of possible observations, each observation, x,

the probability $P_{SN}(x)$ of occurrence if signal plus noise is presented and the

---

1. The discussion in Sections 1.1 and 1.2 is based on an unpublished paper by
   T. G. Birdsall of the Electronic Defense Group, University of Michigan.

PHYSICAL
ENVIREONMENT

RECEPTION
ORGAN

NOISE

SENSORY PATHWAYS

INFORMATION
DISPLAY

OBSERVATION SPACE

DECISION
MAKER

BLOCK DIAGRAM ILLUSTRATING BASIC ASSUMPTION.

FIG. I.

probability $P_N(x)$ of occurrence if noise alone is present. The likelihood ratio is defined as $l(x) = \dfrac{P_{SN}(x)}{P_N(x)}$. Usually there are uncountably many observation points (x is a continuous variable), and the probability density functions $f_{SN}($ and $f_N(x)$ must be used; the likelihood ratio is then the ratio of these two quantities.

The evaluation of a criterion is usually in terms of the integrals of the density functions over the criterion A, since the integral $f_{SN}(x)$ over A is the conditional probability of detection, $P_{SN}(A)$, and the integral of $f_N(x)$ over the criterion A is the conditional probability of a false alarm, $P_N(A)$.

1.2 Definitions of Optimum

The theory of signal detectability is essentially this: a class of criteria is defined in terms of likelihood ratio. Six slightly different defini tions of "optimum" are advanced, and under each definition the optimum is found to be in this class of likelihood-ratio criteria. The notation denoting a cri terion in this optimum class is $A(\beta)$, which means that the criterion contains al observations with likelihood ratio greater than β, and contains none of those wi likelihood ratio less than β, (that is, β represents the boundary condition). Whether or not likelihood ratios equal to β fall in or out of the criterion is unimportant.

The six definitions of optimum, and their solutions (the exact values β to be used, called the operating level) are listed below:

1) The Weighted Combination Criterion. This criterion, by definition maximizes $P_{SN}(A) - \beta P_N(A)$. Solution: $A(\beta)$, that is the observer reports that a signal is present is $l(x) \geq (\beta)$, where x is the stimulus magnitude.

2) Seigert's Ideal Observer. The observer employs a criterion that minimizes total error; this is a special case of the Weighted Combination Criterion. Solution: $A(\beta)$ where $\beta = P(N)/P(SN)$, the ratio of <u>a priori</u> probabilities.

3) Expected Value Observer. The observer employs a criterion that maximizes the total expected value, where the individual values are:

$$V_{SN \cdot A} = \text{value of detection}$$

$$V_{N \cdot CA} = \text{value of a correct "no signal present"}$$

$$K_{SN \cdot CA} = \text{cost of a miss}$$

$$K_{N \cdot A} = \text{cost of a false alarm}$$

This is a further refinement beyond Seigert's Ideal Observer. Solution: $A(\beta)$ where

$$\beta = \frac{P(N)\left(V_{N \cdot CA} + K_{N \cdot A}\right)}{P(SN)\left(V_{SN \cdot A} + K_{SN \cdot CA}\right)}$$

4) The Neyman-Pearson Observer. The observer employs a criterion such that $P_N(A) = k$, with $P_{SN}(A)$ a maximum overall criteria.
Solution: $A(\beta)$, where $P_N\left[A(\beta)\right] = k$.

5) <u>A Posteriori</u> Probability Observer. Here the observer does not actually employ a criterion, he makes the best estimate of the probability that signal-plus-noise was the input leading to observation $x = x(t)$

$$P_x(SN) = \frac{\ell(x)P(SN)}{\ell(x)P(SN) + 1 - P(SN)}$$

6) Information Observer. This criterion maximizes the reduction in u

certainty, in the Shannon sense (Ref. 2), as to whether or not a

signal was sent. Solution: $A(\beta)$ where β is the solution to

$$\beta = \frac{P(N)}{P(SN)} \cdot \frac{\log P_{B(\beta)}(N) - \log P_{A(\beta)}(N)}{\log P_{A(\beta)}(SN) - \log P_{B(\beta)}(SN)}$$

1.3 Forced-Choice Optimization

A somewhat different optimization is that involved in the forced-choic

psychophysical experiment. In the forced-choice experiment a signal is presente

in one of n intervals either in time or space, and the observer's task is to sta

in which interval the signal occurred. Optimum behavior requires making an obse

vation, x, in each interval, and choosing the interval for which the likelihood

ratio, $l(x) = \dfrac{f_{SN}(x)}{f_N(x)}$, is greatest. The situation is somewhat different from th

criterion approach, as is the <u>a posteriori</u> observer.

1.4 The Application of the Theory to Human Behavior

A series of papers previously have dealt with experiments designed to

test the applicability of the theory of statistical decision to signal detection

by the human observer. The conclusions drawn from the experiments are the fol-

lowing:

1) A subject can observe well into noise. The observation variable,

x, is indeed continuous (Refs. 4, 5, 6, 7, 8, 9).

2) The observer can act to optimize expected values in the fixed-

observation interval experiment. This is shown to be true for both vision

(Refs. 4, 6) and audition (Ref. 9).

3) The observer can act as a Neyman-Pearson observer (Ref. 4).

4) The observer can act as an <u>a posteriori</u> observer (Ref. 4).

5) The observer can optimize in the forced-choice experiment, as indi-
ted by the predictability from forced-choice to yes-no experiments. The order-
g extends beyond the first choice, as indicated by the second-choice and fourth-
oice experiments (Refs. 4, 9).

6) Within limits the observer can optimize the use of knowledge of sig-
l parameters. He can tune to a narrow band of frequencies in auditory experi-
its, and can, within limits, adjust his auditory bandwidth to knowledge of sig-
l duration.

7) At any instant in time the observer is tuned to exactly one band of
equencies. To act as a wide band receiver is a process requiring time. (Refs.
9).

8) When the observer is listening for a signal known to be at one of
o frequencies, detection performance decreases as a function of the separation
the frequencies (See Section 2.2). This performance suggests a scanning-type
chanism (Refs. 5, 9).

These conclusions furnish the background for the recognition theory.

2. RECOGNITION FOR THE CASE OF TWO SIMPLE ALTERNATIVES

Recognition, by definition, is the process of classifying a signal as
particular one of a set. The fundamental problem treated here is the case of a
t with two members, each a signal of a specified frequency, f_1 or f_2. Through
e experimental design, the a priori probabilities of the two signals $P(S_1N)$ +
S_2N) = 1.00. The observation xy is now associated with two probability density
stributions, $f_{S_1N}(xy)$ and $f_{S_2N}(xy)$. The decision is again based on likelihood
tio, as is shown in Equation 1.

$$\ell(xy) = \frac{f_{S_1N}(xy)}{f_{S_2N}(xy)} \tag{1}$$

If the decision function can be defined along an axis, the problem is similar to the detection problem with one of the probability distribution functions $(f_{S_1N}(x)$ or $f_{S_2N}(x))$ substituted for $f_N(x)$ of the detection problem.

2.1 The Decision Axis for Two Signals

The problem is illustrated in Figure 2^1. The axis, OX, represents the decision axis for the detection case where the signal is known to be at f_1. The axis, OY, represents the decision axis for the detection case where the frequency is known to be at f_2.

The distance, OX, divided by the standard deviation of the noise distribution, $f_N(x)$, is called d_1', the d' for detection of frequency f_1. The distance, O divided by the standard deviation of $f_N(y)$ is d_2', the d' for detection of frequency f_2. The d' for recognition of frequency when the signal is known to be either f_1 or f_2, is designated $d_{1,2}'$. The distributions $f_N(x)$, $f_N(y)$, $f_{S_1N}(x)$ and $f_{S_2N}(y)$ all assumed to have equal variance.

The problem considered is again the fixed observation interval problem. An observation, xy, a function of time for T seconds, is the datum upon which the decision is based. The signal is known to be either f_1 or f_2, but not both. The $f_{S_1N}(xy)$ is the joint probability density function $f_{S_1N}(x)$ and $f_N(y)$, while $f_{S_2N}(xy)$ is the joint probability density functions $f_{S_2N}(y)$ and $f_N(x)$, assuming x and y are independent.

---

1. As the development is suggestive of Thurstone's law of comparative judgment, the similarities and differences between this theory and that law are discus in Appendix II.
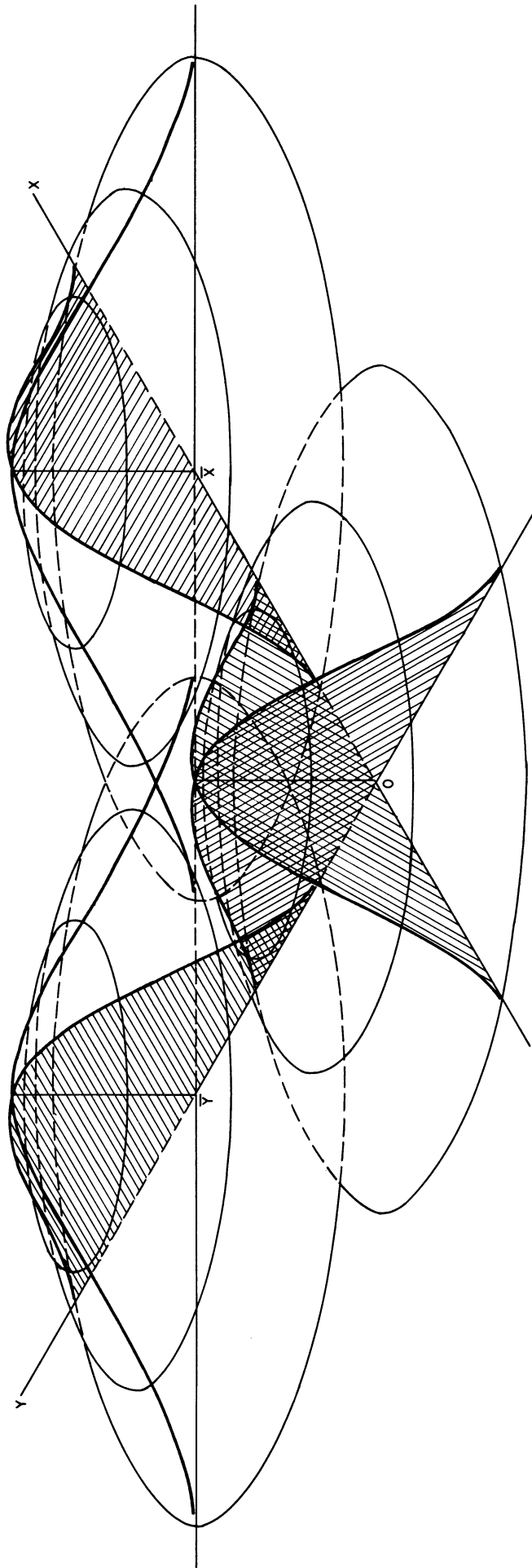
FIG. 2 . THE RECOGNITION SPACE FOR A SIGNAL KNOWN TO BE
ONE OF TWO FREQUENCIES .

$$\text{If } f_{S_1N}(xy) = \frac{1}{2\pi} e^{-\frac{(x - d_1')^2}{2}} e^{-\frac{y^2}{2}} \tag{2}$$

$$\text{and } f_{S_2N}(xy) = \frac{1}{2\pi} e^{-\frac{(y - d_2')^2}{2}} e^{-\frac{x^2}{2}} \tag{3}$$

then, from Eq. 1,

$$\log \ell(x) = \frac{x^2}{2} + xd_1' - \frac{(d_1')^2}{2} - \frac{y^2}{2} + \frac{y^2}{2} - yd_2' + \frac{(d_2')^2}{2} + \frac{x^2}{2}$$

$$= xd_1' - \frac{(d_1')^2}{2} - yd_2' + \frac{(d_2')^2}{2} \tag{4}$$

If $d_1' = d_2'$ and $d' = 0$, then

$$y = x - \frac{\log \ell(x)}{d'} \tag{5}$$

Thus, if $\ell(x)$ is held constant, this is the equation for a straight li:
with slope 1.00, and intercept $\frac{\log \ell(x)}{d'}$. This line passes through the origin
when $\ell(x) = 1.00$.

By equation 5 each value of $\ell(x)$ is represented by a line of slope 1
which intersects the line connecting $d_1' = \bar{x}$ and $d_2' = \bar{y}$ at right angles. From th:
it follows that the decision axis for the recognition problem can be mapped on t!
line, with S_1 normally distributed ($\bar{x}$, 1) and S_2 normally distributed ($\bar{y}$, 1).

Part of Figure 2 has been reproduced as Figure 3(a) to illustrate this
point more simply. The dotted lines on this figure are lines of constant likeli-
hood ratio. The slope of the line $\bar{x}\,\bar{y}$ is -1 while the slope of the lines of con
stant likelihood ratio are +1 (by Eq. 5). Therefore, these lines intersect at
right angles. If the value of y is held constant, say at $y = 0$, then the values
of x are normally distributed along the x axis, indicating the normality of the
mapping along the $\bar{x}\,\bar{y}$ axis.

The assumption of independence implies that the angle $\theta(XOY) = 90°$. erefore

$$(d'_{1,2})^2 = (d'_1)^2 + (d'_2)^2 \tag{6}$$

$d'_1 \neq d'_2$, then Equation 5 becomes

$$y = x \frac{d'_1}{d'_2} - \frac{\log \ell(x) + (d'_1)^2 - (d'_2)^2}{d'_2} \tag{7}$$

i it can be shown that $\ell(x)$ constant is represented by a line which intersects e line $\overline{x}\,\overline{y}$ at right angles, the line $\ell(x) = 1.00$ intersecting at the mid-point. us, equation 6 also applies to this case.

Again, part of Figure 2 has been reproduced as Figure 3(b). The line nnecting $\overline{x}\,\overline{y}$ is at slope $-\dfrac{d'_2}{d'_1}$ while the lines of constant likelihood ratio are slope $+\dfrac{d'_1}{d'_2}$, and again the two lines intersect at right angles. In the figure, e distance $\overline{x}\,\overline{y} = \sqrt{(d'_1)^2 + (d'_2)^2}$. Solving for the intersection of the line x) = 1.00 and the line $\overline{x}\,\overline{y}$:

$$\cos A = \frac{d'_1}{\overline{x}\,\overline{y}}$$

$$\cos A = \frac{a(\overline{x}\,\overline{y})}{(d'_1) - \dfrac{(d'_1)^2 - (d'_2)^2}{2d'_1}} = \frac{2ad'_1(\overline{x}\,\overline{y})}{(d'_1)^2 + (d'_2)^2}$$

uating:

$$\frac{d'_1}{\overline{x}\,\overline{y}} = \frac{2ad'_1\,(\overline{x}\,\overline{y})}{(d'_1)^2 + (d'_2)^2}$$
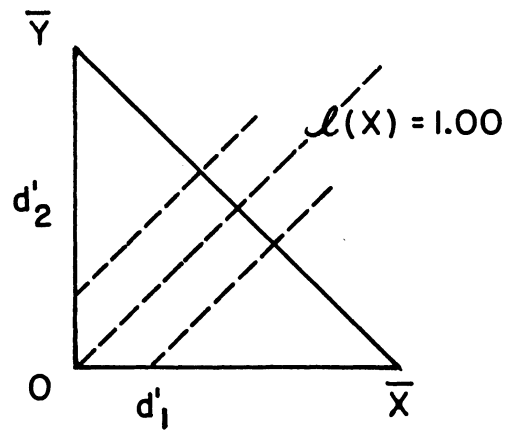
(a)
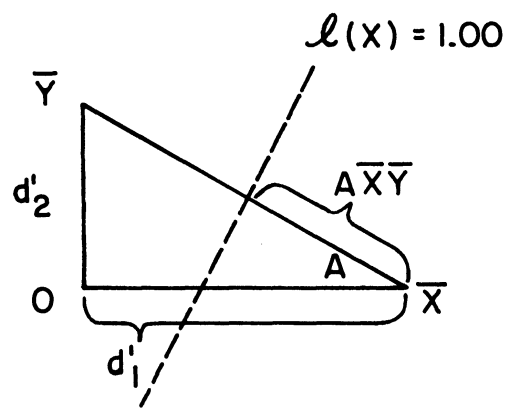
(b)

FIG. 3. SIMPLIFIED DIAGRAMS FOR THE
ORTHOGONAL CASES .

$$1 = \frac{2a(\bar{x}\,\bar{y})^2}{(d_1')^2 + (d_2')^2}$$

ince $(\bar{x}\,\bar{y})^2 = (d_1')^2 + (d_2')^2$, $a = \frac{1}{2}$ and thus, if $l(x) = 1.00$, the line $\bar{x}\,\bar{y}$ is itersected at the midpoint.

If the two signals are not independent, or, in other words, if there is common factor in the observations x and y, then the signal spaces x and y are orrelated. The degree of correlation is defined by the cosine of the angle θ. or this case

$$(d_{1,2}')^2 = (d_1')^2 + (d_2')^2 - 2 \cos \theta \; d_1' \; d_2' \tag{8}$$

Equation 8 is the general form. For the orthogonal or independent case, os $\theta = 0$ and Equation 8 is identical with Equation 6. For the perfectly correla- ed case, such as two signals of the same frequency differing only in d' as a esult of different intensity, $\cos \theta = 1.00$ and

$$d_{1,2}' = \left| d_1' - d_2' \right| \tag{9}$$

Thus, in each case, the decision function has been defined along an axis, lowing that each recognition case is essentially the same as the detection case.

This discussion furnishes the basis for the development of the theory. far it is based on the assumption that the process of observing one frequency bes not interfere with the process of observing at other frequencies. That this s not a valid assumption is indicated by the seventh and eighth conclusions in ection 1.4.

2.2 Modification to Allow for Observation at Both Frequencies

The experiments upon which the 7th and 8th conclusions of Sec. 1.4 are based suggest that for a signal 0.1 second in duration at a frequency of either 900 cycles or 1000 cycles, the detection rate is as if both frequencies can be ob served simultaneously. When the frequencies are separated by more than 100 cycle the detection performance is lower until the separation reaches 300 cycles (f_1 = 700 and f_2 = 1000), at which point the performance is such that only one frequency can be observed during the duration of the signal. The results first reported by Tanner and Norman (Ref. 5) are illustrated in Figures 4 through 7, because they relate closely to the data to be reported below. The curve above the other two is the forced-choice curve for a signal of known frequency. The middle curve is for a signal known to be at one of two frequencies when it is possible to observe at both frequencies simultaneously. The bottom curve is for signal known to be at one of two frequencies, when it is possible to observe at only one frequency. Thus, if the observer happens to be observing at the wrong frequency he is forced to make his choice without relevant information.

The data are placed on the graph as follows. First the d' is determine for the signals when the frequency is known exactly, and then the percentage cor- rect for the experiment in which the signal is known to be at one of two fre- quencies is entered for that d'. Two durations are represented, with the results virtually the same. It should be noted that it is likely that both of the dura- tions are within the range for matching of bandwidth to duration (Section 1.4, Conclusion 6), and the results for the durations should not be markedly different

Thus, for a signal 0.1 second in duration, the signal space is expected to show the angle of correlation, θ, increasing until the frequencies are separat by 100 cycles (900 or 1000 cycles) at which point a maximum of $\theta = 90°$ is expecte

FIG. 4 . DETECTION CASE FOR A SIGNAL KNOWN TO
BE AT ONE OF TWO FREQUENCIES.
OBSERVER I . DURATION 0.1 SEC.

P(C)

1.0

SIMULTANEOUS

0.8

⊙ 900 or 1000 ($\sigma = 0.03$)

800 or 1050 ($\sigma = 0.03$)   SUCCESSIVE
⊙

0.6

⊙ 700 or 1000 ($\sigma = 0.03$)

0.4

OBSERVER 2
t = 0.1 SEC.

0.2

0

0   1   2   3   4   5   6
d'

FIG. 5. DETECTION CASE FOR A SIGNAL KNOWN TO BE AT
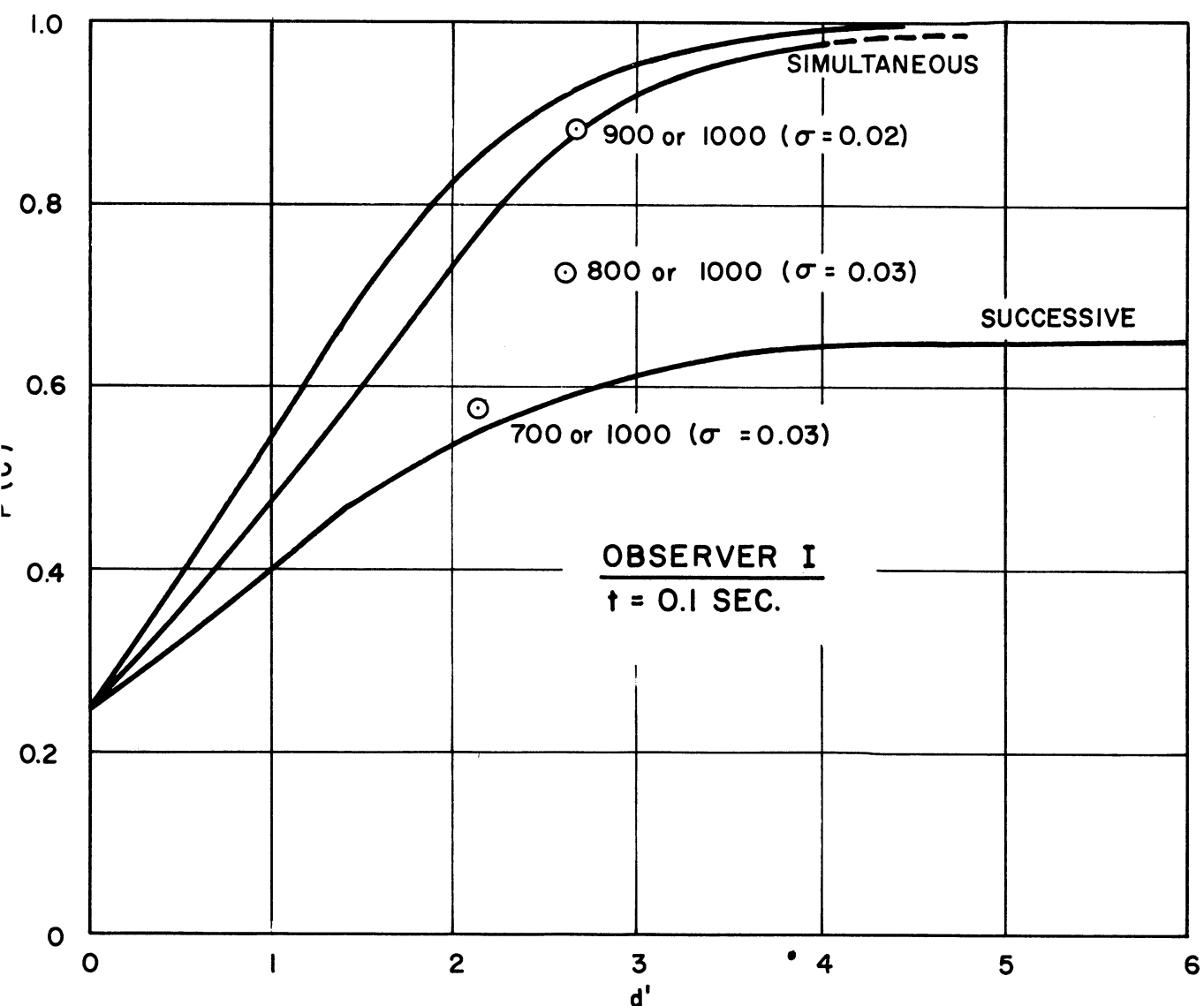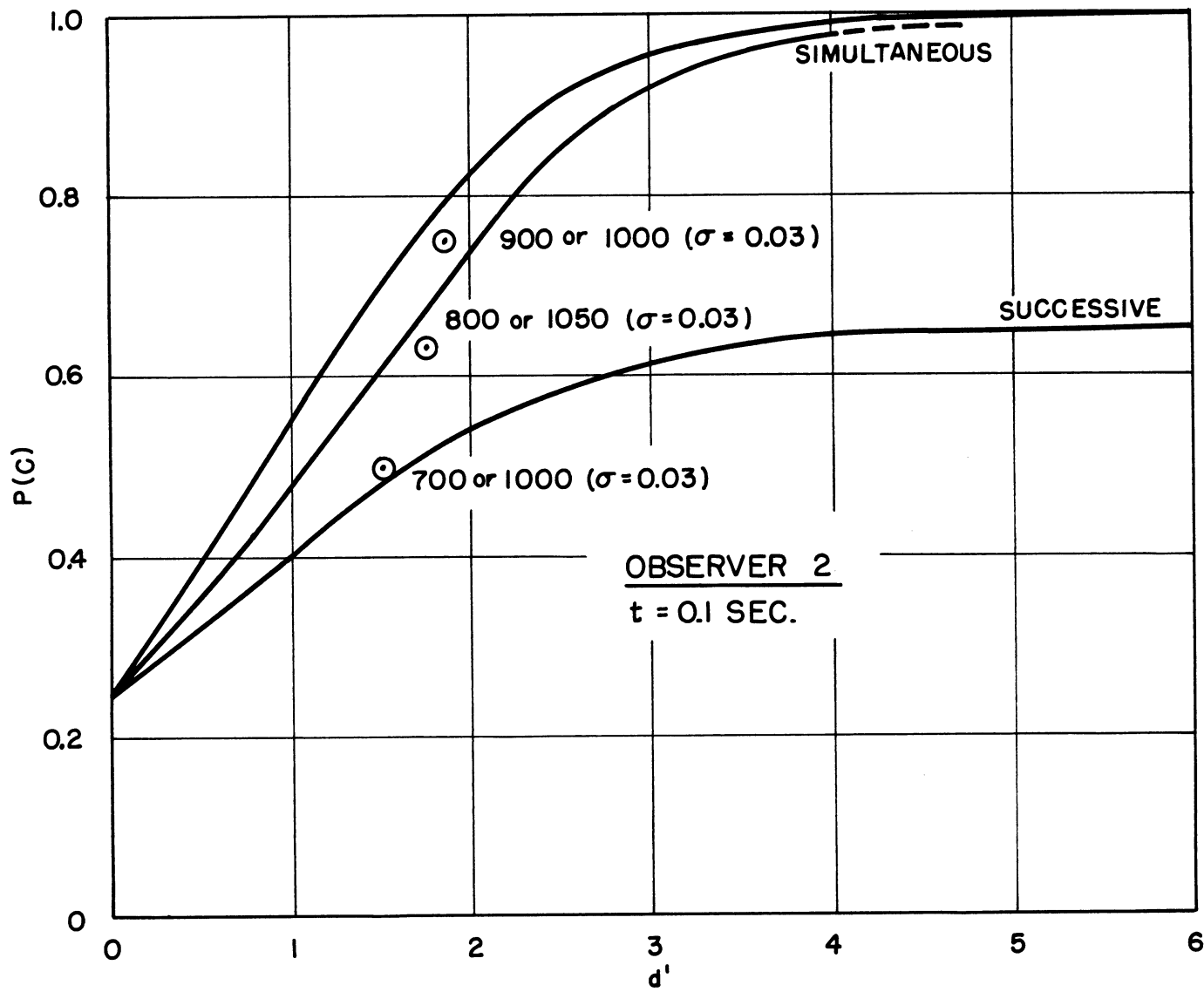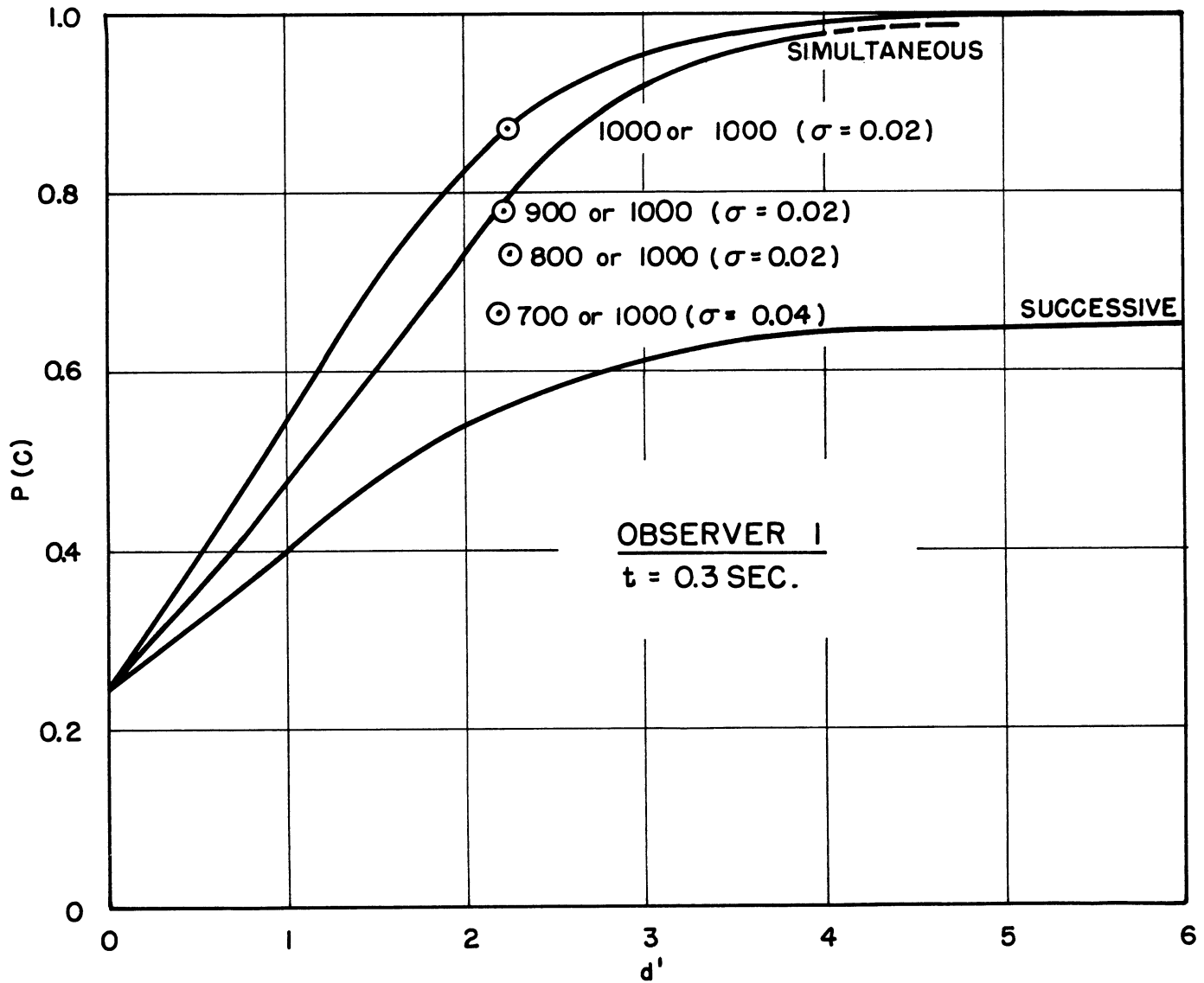ONE OF TWO FREQUENCIES. OBSERVER 2, DURATION 0.1 S

FIG. 6. DECTECTION CASE FOR A SIGNAL KNOWN TO BE AT ONE OF TWO
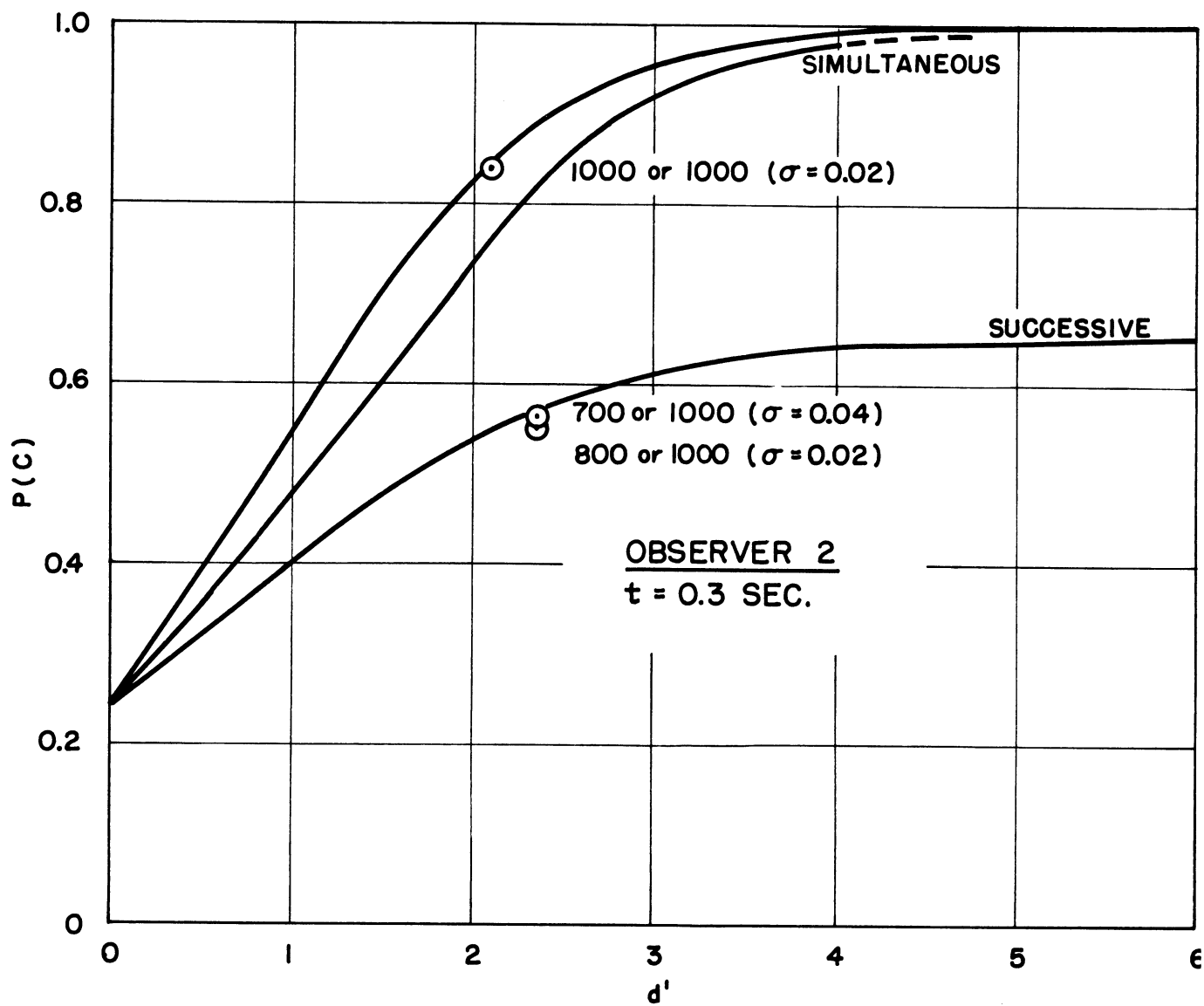FREQUENCIES. OBSERVER I, DURATION 0.3 SEC.

FIG. 7. DETECTION CASE FOR A SIGNAL KNOWN TO BE AT ONE OF TW
FREQUENCIES. OBSERVER 2, DURATION 0.3 SEC.

or frequencies further separated, the calculations in θ (Section 2.3) are expected o yield a decrease in θ until at a separation of 300 cycles θ should appear to be $0°$. An apparent $60°$ can be achieved if the observer attends to one frequency, ffectively performing a yes-no experiment. If he accepts a signal at that fre- uency, he states so. If he does not, then he indicates that the signal was at he other frequency. After the maximum is reached, the decrease in θ does not epresent correlation, but rather a loss due to the observer's inability to ob- erve by x and y. If the signal is sufficiently long in duration the observer hould be able to observe both x and y regardless of the separation, so that once reaches $90°$ it should stay there for frequencies of wider separation.

.3 Experimental Design

The experimental design involved first a two-choice, forced-choice xperiment at each of the frequencies, until approximately equal d's are deter- ined. Then a signal known to be at either one of the two frequencies is presented t a specified time, the observer's task being to state whether the signal is f_1 r f_2. The two-choice, forced-choice experiment is actually the choice of one of wo signals orthogonal in time. The d_1' and d_2' are determined in the following anner. The percentage correct is used as an estimate of the probability of cor- ect. This figure is used to enter normal tables, and the corresponding $\frac{x}{\sigma}$ is etermined. This value is multiplied by $\sqrt{2}$ giving the equivalent yes-no d'. For he recognition experiment, the same procedure is used, except in this case the alue of $\frac{x}{\sigma}$ is multiplied by 2.

From a rearrangement of Equation 8 the d's are then used to find θ.

$$\cos \theta = \frac{(d_1')^2 + (d_2')^2 - (d_{1,2}')^2}{2 d_1' d_2'} \tag{10}$$

ius θ is shown as a function of frequency separation.

TABLE I

RESULTS OF THE EXPERIMENTS[1]

$$N_o = 52.3 \text{ db re } .0002 \text{ dyne/cm}^2$$

Frequency Separation	f_1	f_2	$\sqrt{\dfrac{2E_1}{N_o}}$	$\sqrt{\dfrac{2E_2}{N_o}}$	N_1	N_2	$N_{1,2}$
		Duration .05 seconds					
25 cps	975 cps	1000 cps	3.16	3.16	198	198	198
50	950	1000	3.16	3.16	198	198	197
100	900	1000	3.10	3.16	197	198	198
200	800	1000	3.05	3.16	198	198	198
300	700	1000	3.02	3.16	198	198	198
400	700	1100	3.02	3.25	198	198	198
500	700	1200	3.02	3.42	198	198	197
600	700	1300	3.02	3.42	198	198	198
		Duration .10 seconds					
25	975	1000	3.44	3.44	197	196	196
50	950	1000	3.44	3.44	197	196	197
100	900	1000	3.43	3.44	191	196	292
200	800	1000	3.36	3.44	197	196	198
300	700	1000	3.23	3.44	197	196	194
400	700	1100	3.23	3.60	197	198	198
500	700	1200	3.23	3.67	197	194	196
600	700	1300	3.23	3.67	197	195	195
		Duration .5 seconds					
25	975	1000	5.37	5.37	197	198	197
50	950	1000	5.37	5.37	197	198	197
100	900	1000	5.14	5.37	197	198	296
200	800	1000	4.93	5.37	197	198	197
300	700	1000	4.60	5.37	197	198	
400	700	1100	4.60	5.37	197	197	297
500	700	1200	4.60	5.90	197	196	197
600	700	1300	4.60	5.90	197	197	197
		Duration 1.0 seconds					
25	975	1000	5.81	5.85	198	198	198
50	950	1000	5.76	5.85	197	198	198
100	900	1000	5.63	5.85	197	198	198
200	800	1000	5.06	5.85	198	198	198
300	700	1000	4.93	5.85	198	198	196
400	700	1100	4.93	5.99	198	196	197
500	700	1200	4.93	6.38	198	198	198
600	700	1300	4.93	6.38	198	197	198

---

[1]For an explanation of the term $\sqrt{\dfrac{2E}{N_o}}$, see Appendix I.

d'_1	d'_2	$d'_{1,2}$	θ	d'_1	d'_2	$d'_{1,2}$	θ
1.85	1.85	.89	35	1.15	1.15	.37	25
1.77	1.85	1.51	50	1.00	1.15	.85	51
1.69	1.85	2.33	62	.85	1.15	1.10	51
1.97	1.85	2.18	69	1.12	1.15	1.26	67
2.14	1.85	1.78	53	1.52	1.15	1.00	49
2.14	1.80	1.71	40	1.52	1.23	1.13	28
2.14	1.80	2.08	64	1.52	.80	1.27	89
2.14	1.80	2.05	62	1.52	1.28	1.38	63
2.19	2.19	1.31	35	1.80	1.80	.51	16
2.19	2.19	1.78	48	1.80	1.80	1.37	45
2.06	2.19	3.27	87	1.32	1.80	1.95	76
2.91	2.19	2.80	13	1.46	1.80	1.68	60
2.46	2.19	2.52	72	1.73	1.80	2.20	78
2.46	2.40	2.61	65	1.73	1.69	1.89	69
2.46	2.17	2.38	58	1.73	1.14	1.39	53
2.62	2.62	1.49	33	2.38	2.38	.69	16.6
2.75	2.62	1.56	61	2.10	2.38	1.56	40.1
2.91	2.62	3.76	86	1.84	2.38	2.51	73.6
2.79	2.62	2.80	62	1.68	2.38	2.38	69.2
2.67	2.62	3.38	80	1.76	2.38	2.78	82.9
2.67	2.66	3.04	70	1.76	1.73	2.39	86.5
2.67	2.42	3.74	86	1.76	1.65	2.42	90.7
2.67	2.91	2.86	61	1.76	2.13	2.86	92.5
2.65	2.56	2.15	48	1.52	1.73	.48	16
2.25	2.56	2.74	68	1.73	1.73	1.03	35
2.19	2.56	3.49	93	1.23	1.73	1.86	76
2.61	2.56	3.02	71	1.55	1.73	2.44	93
2.25	2.56	3.38	80	1.40	1.73	2.22	90
2.25	2.25	2.86	79	1.40	1.28	2.72	*
2.25	2.40	2.80	76	1.40	1.73	2.15	86
2.25	2.47	3.28	88	1.40	1.55	2.49	115

2.4 The Experiments

Four durations, .05, .1, .5, and 1.0 seconds were studied for frequency separations of 25, 50, 100, 200, 300, 400, 500, and 600 cycles. Two observers served for the entire set of experiments. The experiments were conducted by the N. P. Psytar programing system, described elsewhere (Refs. 2, 9). Approximately 200 observations are contained in each determination of d'. The results are tabulated in Table I, and are presented graphically in Figures 8 through 15. No effect has been made to fit curves to data, so that the reader can have an unbiased look at the data.

Occasionally the results show θ's greater than $90°$. In all except one case there is the impression that the deviation is within the range of sampling error. The one case of serious deviation can be accounted for largely on the basis of a single run of 100 in which one observer dropped appreciably in the detection experiment from the other run of 100 at that frequency. The result is an indeterminate θ, explaining the absence of a data point for a frequency separation of 400 cycles in Figure 15. Aside from this, the data conform roughly with predictions.

2.5 Generality of the Theory

In the introduction, it is suggested that while the study is in auditio the application extends generally to human information-collecting systems. In order to illustrate the anticipated generality the following discussion is presented on the problem of color vision in terms of experimental design and data interpretation. Suppose that instead of presenting two frequencies, two monochromatic light signals are studied in an experiment. Exactly the same procedure is to be followed, ending with a determination of $\cos \theta$.
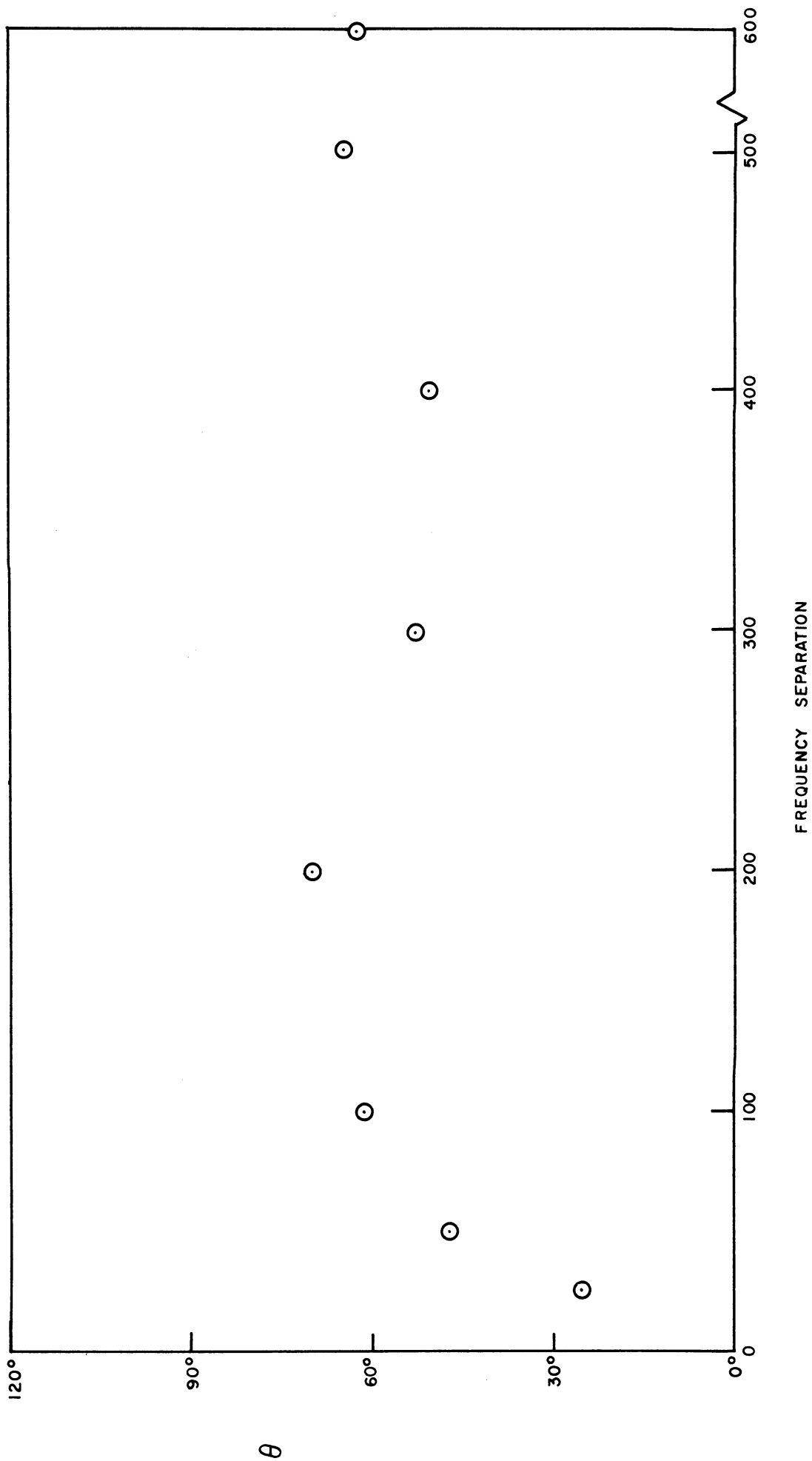
FIG. 8. OBSERVER I. DURATION = 0.05 AROUND 1000~
RECOGNITION OF 1 OF 2 ALTERNATIVES.

FIG. 9. OBSERVER 2. DURATION = 0.05 AROUND 1000 ~
RECOGNITION OF 1 OF 2 ALTERNATIVES.

24

FIG. 10.   OBSERVER I.   DURATION = .1 AROUND 1000~
RECOGNITION OF 1 OF 2 ALTERNATIVES.

FIG. II . OBSERVER 2. DURATION = .I AROUND I000~
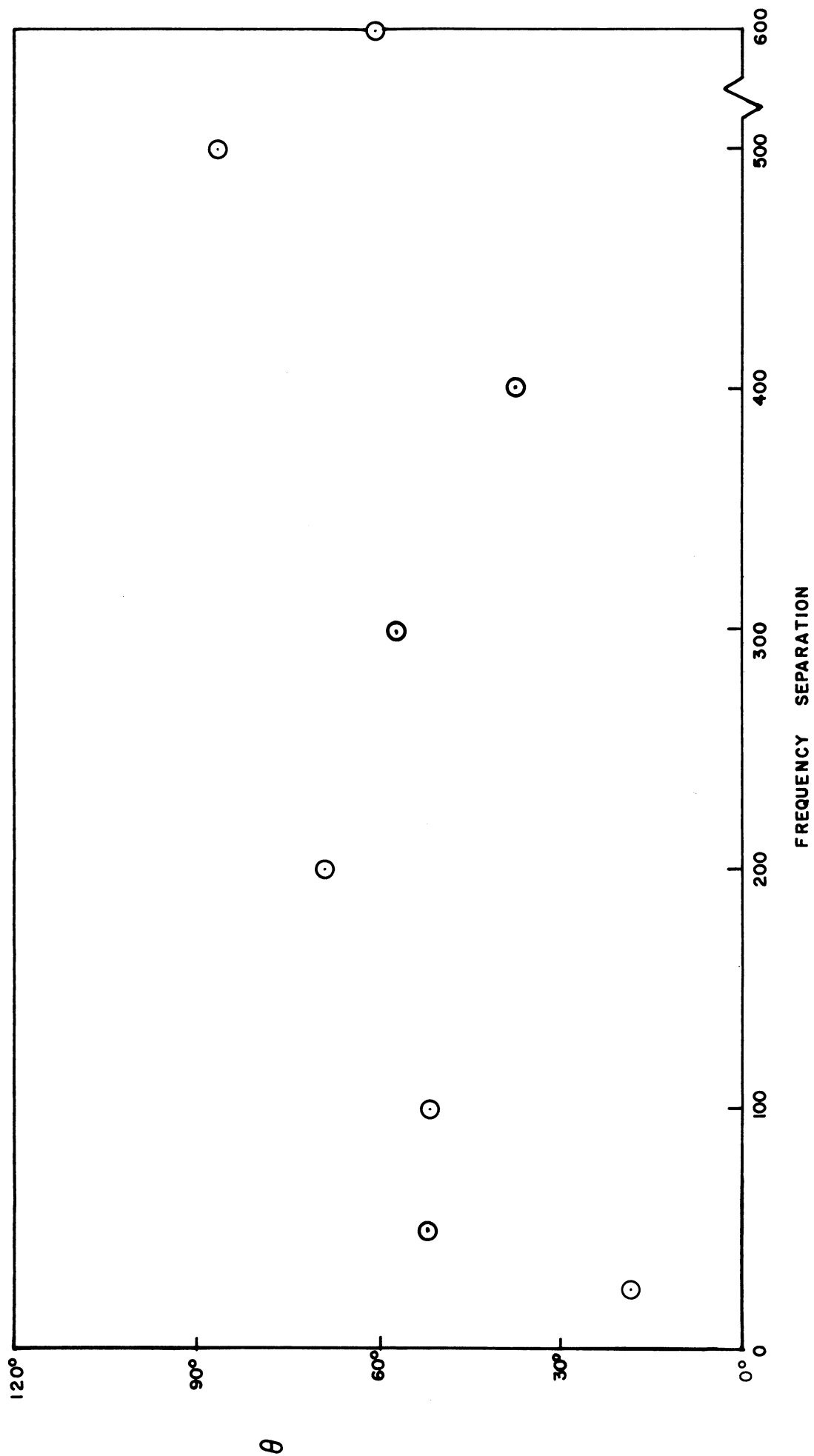RECOGNITION OF I OF 2 ALTERNATIVES.

FIG. 12.   OBSERVER I.   DURATION = 0.5 AROUND   1000~
RECOGNITION OF I OF 2 ALTERNATIVES.

FIG. 13 . OBSERVER 2 . DURATION = 0.5 AROUND 1000 ~
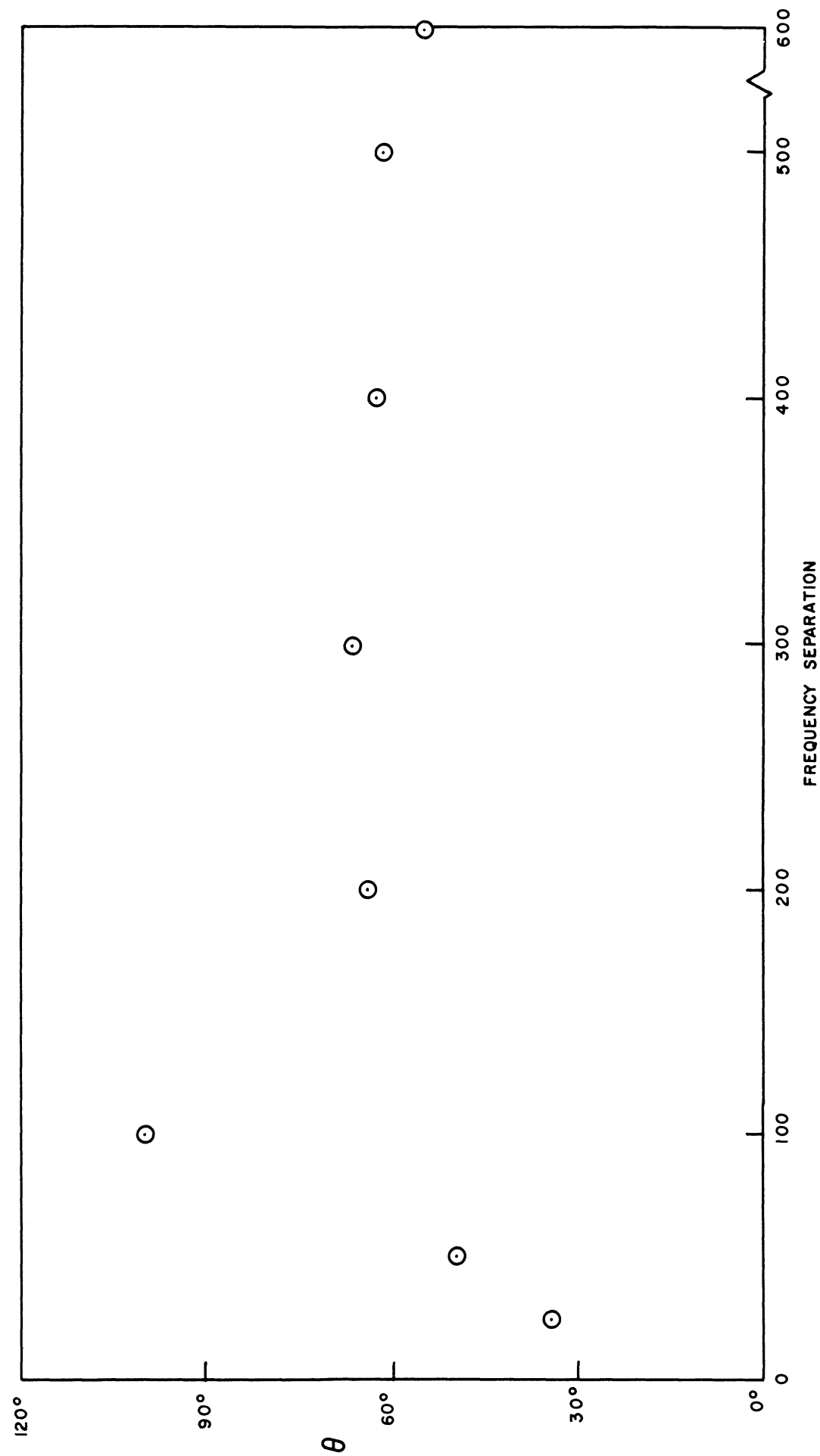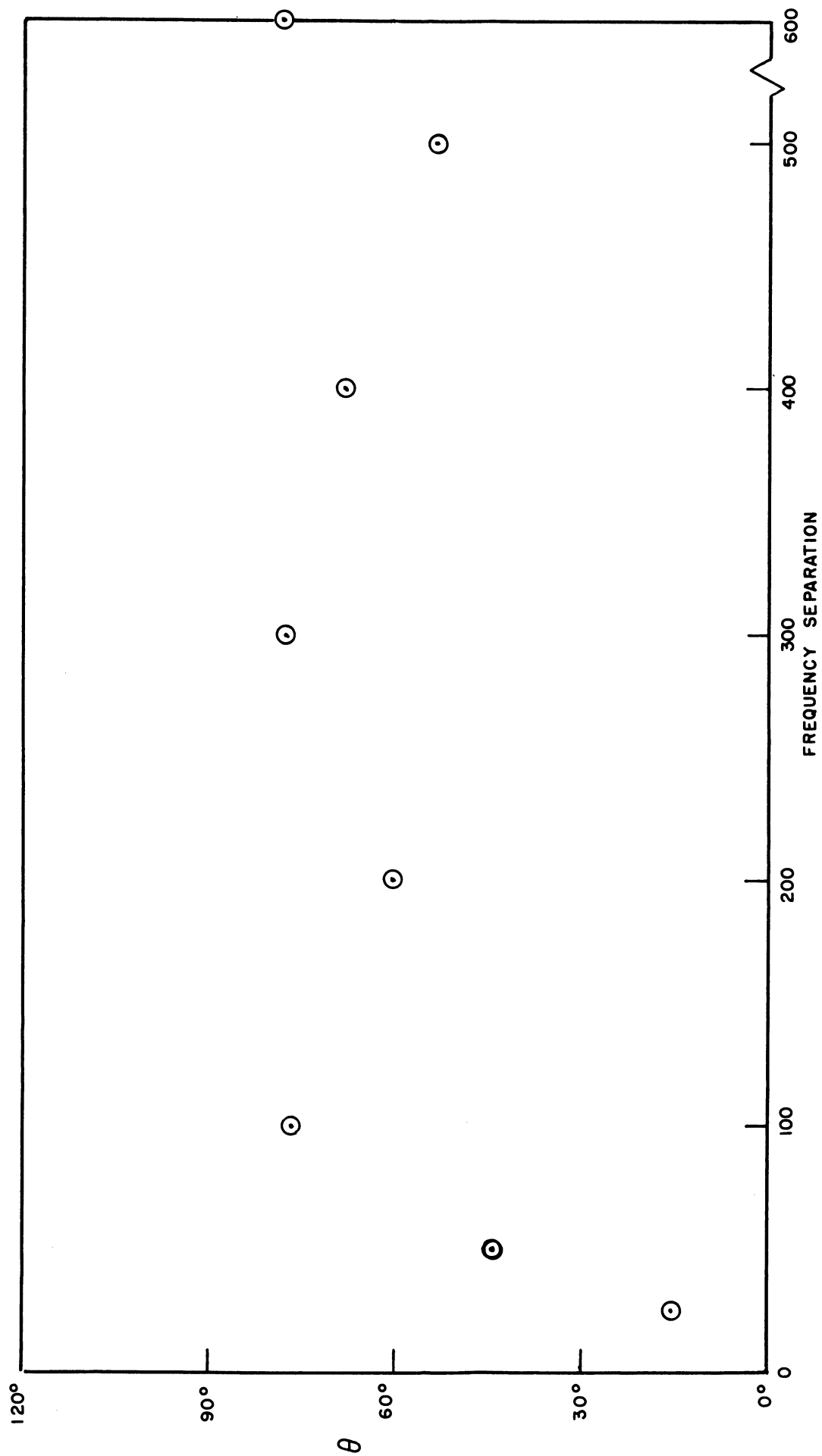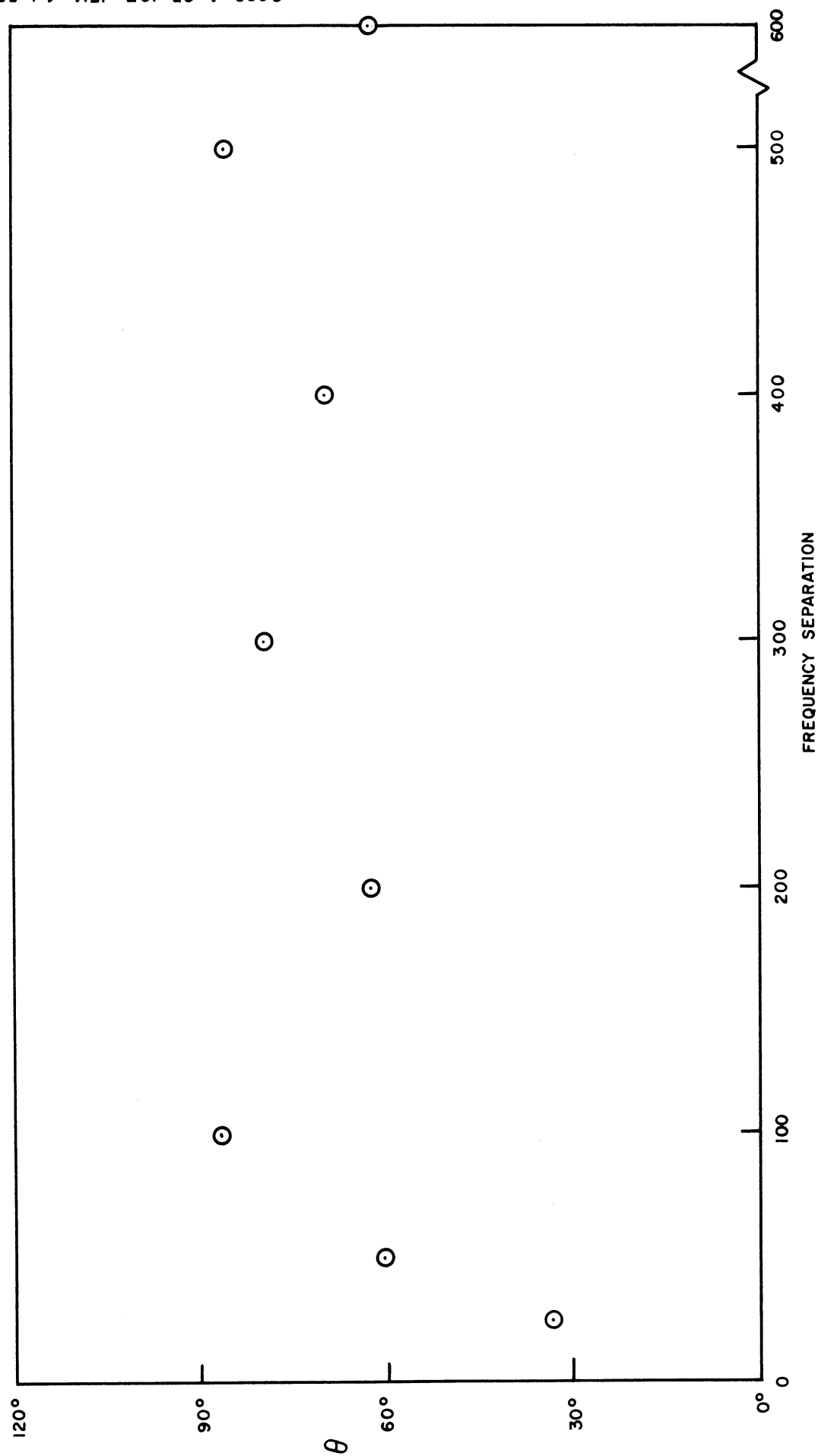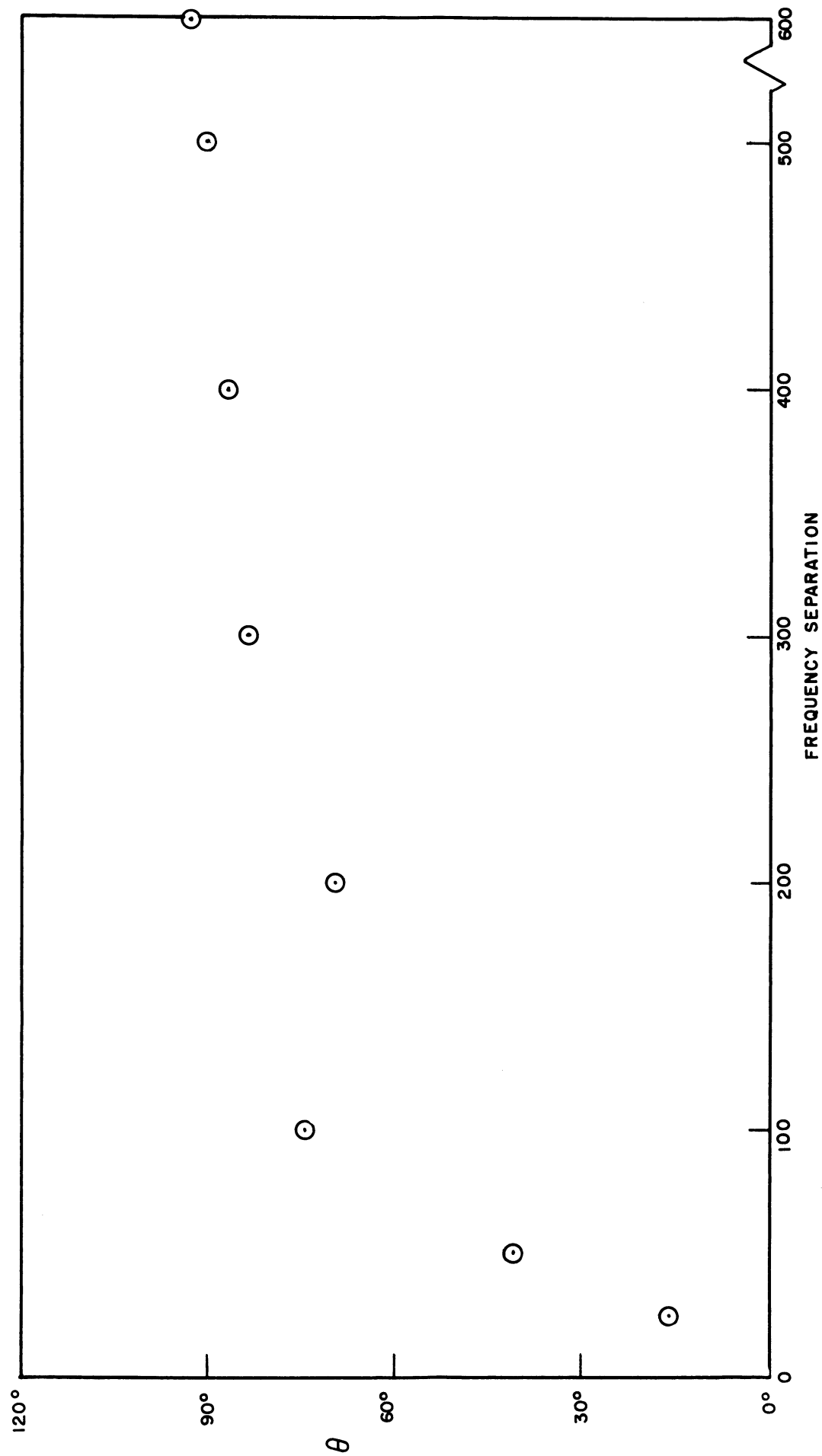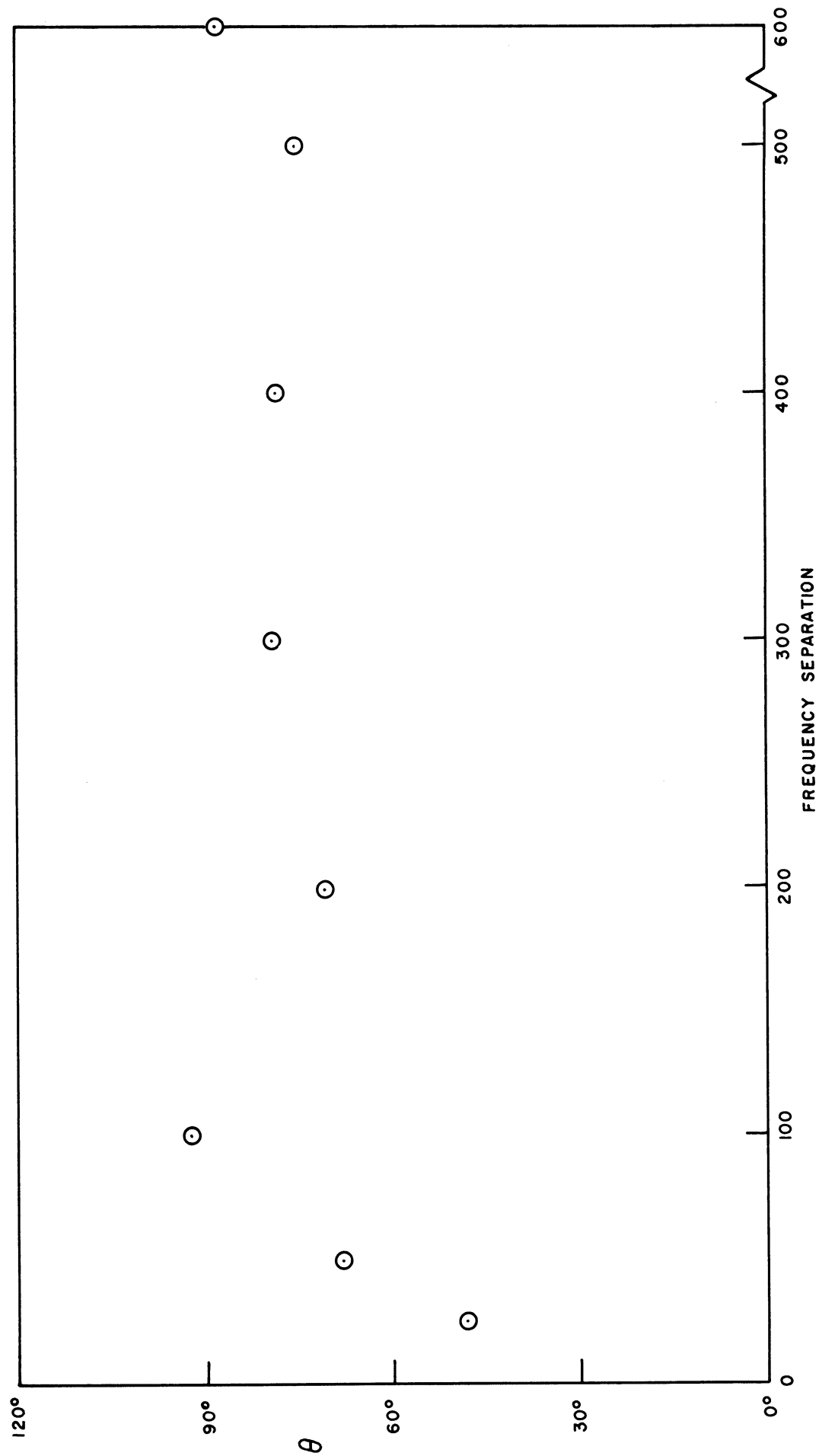RECOGNITION OF I OF 2 ALTERNATIVES.

FIG. 14. OBSERVER I DURATION = I SEC AROUND I000~
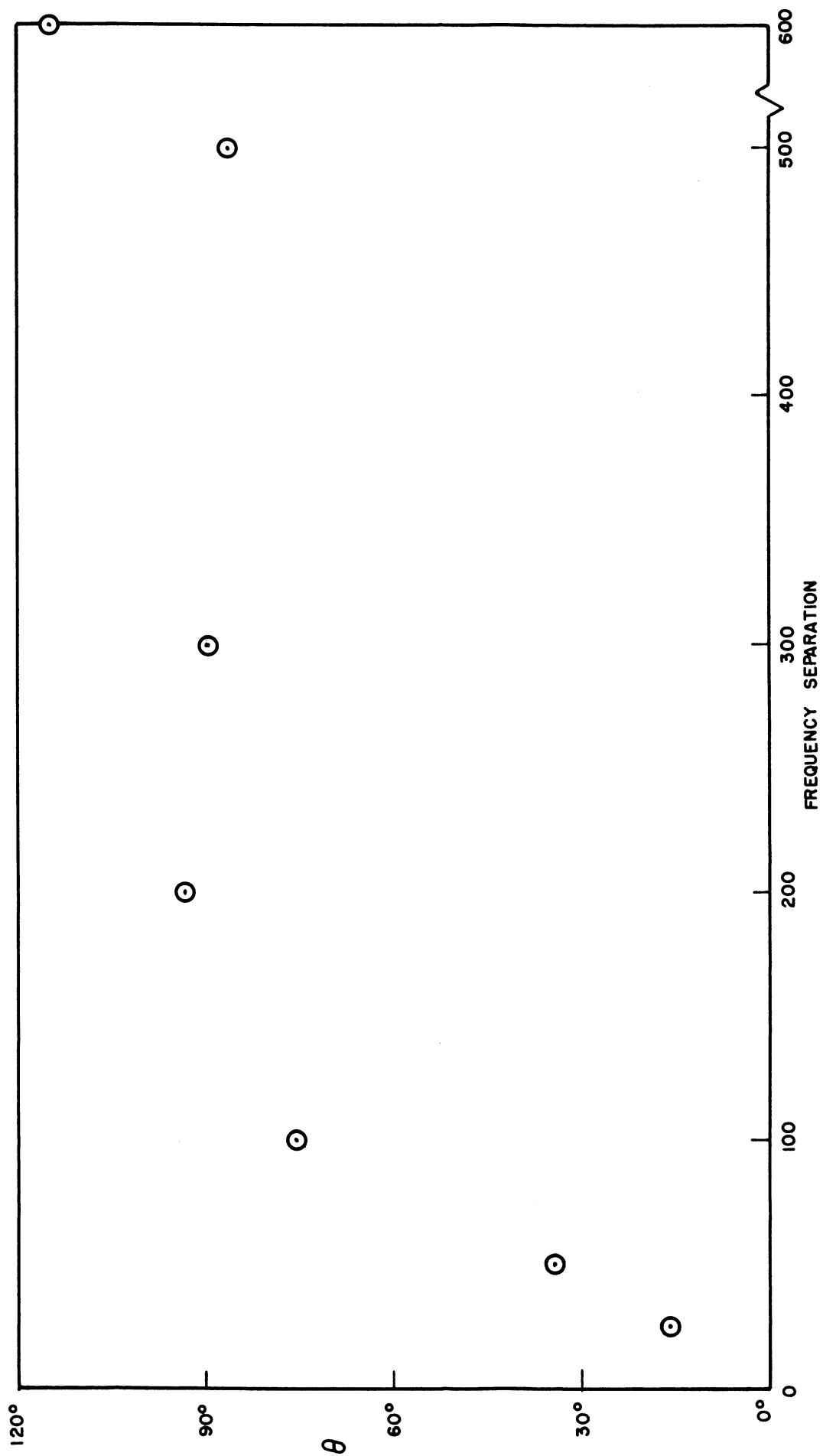RECOGNITION OF I OF 2 ALTERNATIVES.

FIG. 15. OBSERVER 2. DURATION = 1 SEC. AROUND 1000~
RECOGNITION OF 1 OF 2 ALTERNATIVES.

30

In Section 2.1, cos θ is represented as a correlation term. According to the theory, if there is no common variance and the signals are transmitted to the central mechanism over independent channels, then for these monochromatic light signals, θ is 90° (cos θ = 0). This example is used because it does not depend on presumed equivalence between wavelength of light and frequency of sound. This theoretical framework offers a method of psychological determination of the number of independent systems involved in color vision or the number of different types of color receptors. There are many other problems to which such a theoretical framework may be useful.

3. EXPANSION OF THE THEORY

So far the program described in the series of papers dealing with detection and recognition problems, as treated in terms of statistical decision theory, has dealt largely with simple cases readily amenable to study through experimentation. The theory has implications for complex signal structures such that there now appears to be a basis for a more general theory. This section attempts to present a basis for that development.

.1 Requirement of a Set of Alternatives

A decision actually is a choice of one from a set of alternatives. Up to the present time, the theory has dealt with cases involving decisions in favor of one of a set of two alternatives. Implicit in the theory is the fact that for an alternative to have associated with it an a posteriori probability greater than zero, it must also have an a priori probability greater than zero. This is a consequence of Bayes' Theorem. Thus, whenever an observer is placed in a recognition situation his choice is one from a set of alternative signals each of which has an a priori probability greater than zero.

It is further a requirement that the sum of these a priori probabiliti€ be one. This requirement, along with the requirement expressed in the previous paragraph, states that an observer, placed in a recognition situation, assumes ar ensemble of alternative stimuli, A_i (1------- i ------- n) which has the followir properties.

For every i $\qquad 0 < P(i) < 1.00$

$$\sum_{i=1}^{n} p(i) = 1.00 \qquad (11)$$

and

where $P(i)$ is the a priori probability of the ith alternative. It is important t note that all of the experiments so far reported in support of this theory are designed to specify the conditions of Equation (11) for the observer.

The a priori probabilities are not necessarily the true a priori probabilities. It is worth repeating the statement that these are probabilities that the observer assumes. They are, in fact, the observer's beliefs. Before th observer can state an a posteriori probability of an alternative existence, he mu first admit the possibility of its a priori occurrence. Otherwise he would never consider the occurrence. The mere fact that he considers the alternatives implie that the a priori probability is greater than zero.

The a priori probabilities, assigned to the alternatives, depend on the observers past experience, immediate and distant. In an experimental situation t immediate past experience may consist of the experimental instructions and the results of the trials as the experiment progresses, while the more distant past experience may consist of his trust in the experimenter and his idea of the purpo of psychological experiments. The assumed probabilities may or may not approxima the true probabilities. If they fail to approximate the true probabilities there may be adjustments as experience accumulates. The important fact is that the

robabilities assumed by the observer are those which are more likely to determine

ie behavior than are the true probabilities of the form of the signal.

2 The Existence of Hypotheses

By definition, an hypothesis is a probability distribution function. By

ie noise assumption of detection theory, for each signal alternative there exists

ı hypothesis, $f_i(x)$, the probability density that if the ith alternative exists

ie observation x results. Further, by the noise assumption, for every i,

$_i(x) \neq 0$, although for many of the alternatives it may be very close to zero.

3 The Entropy of the Alternative Ensemble

The alternative ensemble has been defined so that it is equivalent to

ie message ensemble of Shannon's communication theory (Ref. 3). Thus, an entropy

ın be assigned to the alternative ensemble, which is:

$$H(x) = - \sum P(i) \log_2 P(i) \tag{12}$$

This entropy is the uncertainty of the set, and defines the amount of

ıformation necessary to resolve the uncertainty. It is necessary to appreciate

ıre that Shannon deals with averages, such that no single trial can describe the

rocess. This means that if the observer is placed in exactly the same situation

large number of times, such that each alternative actually exists according to

ts associated probability, then, on the average, $H(x)$ bits of information are

ıquired to resolve the uncertainty.

Equation 12 is based on the observers assumed probabilities, and ex-

resses the amount of information required by him to resolve an assumed uncer-

ıinty. How a discrepancy between the observers assumed ensemble and the true

ısemble enters will be discussed in the following sections.

3.4 The Equivocation

According to Shannon, the equivocation is the uncertainty remaining after the transmission of information. Here it is the uncertainty remaining afte the observation x. For each alternative in the ensemble there is the associated probability $P_x(i)$ not equal to zero or one, and $\Sigma P_x(i) = 1.00$. The equivocation

$$H_y(x) = \Sigma P(i) \ f_i(x) \ P_x(i) \ \log_2 P_x(i) \tag{13}$$

where $P(i)$ and $f_i(x)$ are not dependent on the observers assumed probabilities, while $P_x(i)$ is dependent on these assumed probabilities. A discrepancy between the observers assumption of entropy increases $H_y(x)$.

3.5 Optimum Behavior Criteria

In Section 1.2 six different definitions of optimum criteria are advanced. The one of particular interest here is the expected value optimum. This interest is based on a simple fact: as far as reducing entropy is concerned it i never optimum to make a decision if one can legitimately avoid making a decision. It is always better to store likelihood ratio, or some monotone function of likel hood ratio such as a posteriori probability; this has been demonstrated by Woodwa and Davies (Ref. 12). It is therefore postulated that, wherever possible, the observer stores the observations, making decisions only when advisable. The decision is for the purpose of determining action, not for maximizing information. If the observer feels that the conditions are such that the expected value of an action based on information available at some time t is greater than the expected value of any action which is based on additional information, then a decision is made at time t. If he feels that the additional information is likely to increas the expected value of the action, the decision is delayed. Thus, at any time

iring the information collecting process the observer is faced with a choice of
ie of n decisions, one of which is to collect further information. To each de-
ision there is attached an expected value. The decision that is associated with
ie greatest expected value is the observers choice.

It is at this point that the observer first suffers from a discrepancy
etween his assumed probabilities and the true probabilities, for it is when he
akes a decision to take action that he becomes aware of errors. The values and
osts are at this point realized, and he finds that he does not realize his ex-
ected values. At this time he may attempt to correct his assumed ensemble.

.6 Complex Alternatives

Up to the present time, only situations where single observations are
equired have been considered. By definition, a complex alternative is defined
s a sequence of simple alternatives. The complex set defines the entropy of the
nsemble. Let A_j represent the jth complex alternative consisting of the sequence
$_{j1}$, a_{j2} ..... a_{ji} ..... a_{jn}, then a set of m complex alternatives has the entropy

$$H(x) = \sum_{j=1}^{m} P(A_j) \log_2 P(A_j) = \sum_{i=1}^{n} \sum_{j=1}^{m} P(a_{ji}) \log_2 Pa_{ji} \quad (14)$$

The complex set may be redundant. Information concerning any simple
lternative in the sequence may furnish sufficient basis for a decision.

.7 Information Basis for a Choice Between Complex Alternatives

In Section 3.5 it is postulated that a decision is made on the basis of
xpected values. Thus, for a set of complex alternatives, each simple alternative
esults in an observation, x_i. The set of observations, x_i, is combined into a
ingle output x, such that for each complex alternative there is the probability
ensity function $f_{Aj}(x)$. This function specifies the probability that, if the jth

complex alternative exists, this particular sequence of observations results. F
each complex alternative a likelihood can now be determined on the basis of the
sequence of observations. The choice is then made on the basis of optimizing
expected value.

The significance of this statement is that it is possible to map the
sequence of outputs into a single output (likelihood) and the problem again re-
solves to the problem of simple alternatives, with the decision again made on the
same basis.

4. CONCLUSIONS

A simple theory of recognition is developed as an extension of the de-
tection theory. Experimental evidence is presented supporting the theory. A
framework is presented for extending the theory to more complex situations, show-
ing how it is possible to map these more complex situations into the same space
that applies to the simple situation.

It remains now to work out cases illustrating the more complex situa-
tions in sufficient detail to permit experimental evaluation. Experimental con-
firmation of the theory developed in Section 3 would provide a basis for the
systematization of recognition data.

APPENDIX I

OBSERVER EFFICIENCY

Table I is self explanatory except for columns headed by$\sqrt{\frac{2E}{N_O}}$.

is column indicates a mathematical upper bound for expected performance. A

rfect detector, operating on the output would be expected to achieve this level

performance. Any detector which achieves this level of performance is using

l of the available information. This quantity represents a standard which can

used for purposes of evaluation of either an operator or a receiver, or a

mbination of an operator and a receiver. The significance of these columns is

us worth some discussion.

$$d'_O = \sqrt{d} = \sqrt{\frac{2E}{N_O}} = \sqrt{\frac{2v^2t}{N_O}} \qquad (A.1)$$

ere d'_O is optimum d', d is the detection index (Ref. 1), E is the signal energy,

is the noise power per unit bandwidth, V is the signal voltage, and t is the

lse duration. The right hand member of the equation is that used for the

lculation of the column head$\sqrt{\frac{2E}{N_O}}$, with the subscript of E refering to the

gnal subscript.

The columns headed d' indicate the value of $\sqrt{\frac{2E}{N_O}}$ which would be required

lead to the same level of performance as that achieved if a perfect device were

aced on the output of the system. The observed d' is thus always equal to or

ss than the calculated value of the$\sqrt{\frac{2E}{N_O}}$. The ratio of the inferred value to

e calculated value can be used as an index of the efficiency of the operating

vice.

The calculations of $\sqrt{\frac{2E}{N_O}}$ are based on measurements made of the output

the earphones used in the experiments. It has thus been possible to calculate

ficiency ratings for the observers performance for the different durations and

the different frequencies studied in the experiments. These are listed in Table II.

TABLE II

OBSERVER EFFICIENCY AS A FUNCTION OF PULSE DURATION AND CENTER FREQUENCY

Pulse duration in secs. Center Frequency	Observer 1				Observer 2			
	.05	.10	.50	1.00	.05	.10	.50	1.00
700	.503	.536	.383	.284	.596	.762	.580	.456
800	.367	.435	.341	.306	.646	.866	.566	.516
900	.274	.385	.358	.218	.545	.601	.566	.389
1000	.364	.523	.443	.296	.585	.637	.488	.438
1100	.378	.469	.322	.213	.554	.667	.495	.376
1200	.234	.310	.280	.271	.526	.591	.410	.376
1300	.376	.420	.361	.243	.526	.695	.493	.387

These tables are not intended to represent a complete study. They are suggestive of a method of study to approach most nearly the optimum use of signal energy in a system involving the human observer. Of the durations studied, the observers most efficient at a duration of 0.1 seconds, and tend to be more efficient at the lower frequencies. These studies involve one particular noise level, and the interpretation of the tables should be made with this in mind.

The discussion in this appendix is presented as a contribution for methodology rather than as a contribution of content.

APPENDIX II

THURSTONE'S LAW OF COMPARATIVE JUDGMENT

In two papers, Thurstone (Ref's. 10,11) presents and develops the Law

Comparative Judgment. Similarities between Thurstone's subject matter and

at of this paper, and in particular, between the form of Thurstone's equations

d those of this paper, justify a discussion of the content of this paper in

rms of Thurstone's earlier work.

By the expression "comparative judgment", Thurstone describes the

perimental design with which he is concerned. It is an experiment of the type

which the observer is presented first with a signal of frequency f_1 and then

th a signal of frequency f_2. He is then asked to state whether f_2 is higher

lower than f_1. Another variation of this experiment is where the observer is

esented first with a signal of energy E_1 and then with an energy E_2. He is then

ked to state whether $E_2 > E_1$ or $E_2 < E_1$. For the case where either E_1 or E_2

zero (or noise alone) this is the two-choice, forced-choice experiment

ployed in determining the detection d' used in this paper.

The definition of d' is $\dfrac{M_{SN} - M_N}{\sigma}$, where M_{SN} is, in Thurstone's language,

e modal discriminal process for signal plus noise, M_N is the modal discriminal

ocess for noise alone, and σ is a measure of the discriminal dispersions σ_N,

d σ_{SN} ($\sigma_N = \sigma_{SN}$). The observations x are assumed in the paper to be a

ntinuous variable corresponding to Thurstone's discriminal processes.

The analysis of forced-choice experiments presented in earlier papers

ef's. 4, 5, 6, 7, 8, 9) can be expressed in terms of comparative judgments.

ppose that a signal of energy $E > 0$ is presented in one of four intervals in

me, while in the other three intervals signals of energy $E = 0$ are presented,

and that the observer is asked to state which interval contained the signal E >0

Observations x are made in each of the four intervals. Three comparative judgmen

are required. First a comparative judgment involving the first and second

intervals is made. Whichever is judged greater is compared to the third interva.

and the greater of this comparison is then compared to the fourth interval.

The greater of the last comparison is then judged to be the signal E > 0. This i:

equivalent to the analysis presented in the previous papers.

The main interest in the theory developed in this series of papers is

not in comparative judgments, however. It is in detection and recognition. The:

are subjects not discussed by Thurstone, although had he recognized the existenc

of a noise distribution such as the one postulated in the theories of detection

and recognition, it seems likely that he would have developed essentially the

same theory as that developed in the current set of papers, only Thurstone would

have been thirty years earlier. It is essentially the noise assumption, along

with the denial of the fixed threshold, which has led to this development.

The detection and recognition theories developed in these papers invol

experiments in which the observer has a single observation, x, and is asked to

state which of a set of alternatives existed to lead to the observation x. It i:

not a comparative judgment in the Thurstone or forced-choice sense. Analysis of

this type of experiment led to the interest in a priori probabilities and risk

functions, variables which are not immediately obvious in Thurstone's discussion

of the law of comparative judgments. Thurstone has assumed that, of two

stimuli (S_1 and S_2) the a priori probabilities $\left[P(S_1 > S_2 \text{ and } P(S_2 > S_1) \right]$ are

equal, and that type I and type II errors are equally costly. Due reflections ar

experimentation should show that these variables (a priori probabilities and risk

functions) also play a part in comparative judgements. The criterion for judgmer

$> S_2$ may not contain all values $S_1-S_2 > 0$ or only values $S_1-S_2 > 0$, but rather
. values $S_1-S_2 > \alpha$ where α is some function of β as defined in Section 1.2.

One further point requires discussion. Thurstone considers a
:relation factor which he considers it safe to assume equal to zero. The
;umption, in view of the noise assumption, is satisfactory for comparative
lgments. If signals of two frequencies are presented successively in time, the
:relation is likely to be zero because of the autocorrelation function of the
lse. However, if a single observation is made, and the choice is between two
;quencies close together, the presence of a signal at one frequency influences
; observation of the components of the other frequency. In these experiments
is necessary to take into account the correlation factor. It is, in fact, this
rrelation factor which determines the "distance" Thurstone discusses. Equation
of Section 2.1 is a general equation for Thurstone's "distance", given the
stance of the signals from the noise, and given the correlation between the
tection axes. It is not the same as Thurstone's general equation which looks
ry much like Equation 9.

REFERENCES

1. Peterson, W. W., and Birdsall, T. G., "The Theory of Signal Detectability," Technical Report No. 13, Electronic Defense Group, University of Michigan, June, 1953.

2. Roberts, G. A., "An Automatic Random Programmer", Electronic Defense Group, University of Michigan, 1955.

3. Shannon, C. E., and Weaver, W., "The Mathematical Theory of Communication", University of Illinois Press, Urbana, Illinois, 1949.

4. Swets, J. A., Tanner, W. P., Jr., and Birdsall, T. G., "The Evidence for a Decision-Making Theory of Visual Detection," Technical Report No. 40, Electronic Defense Group, University of Michigan, Ann Arbor, Michigan, April 1955.

5. Tanner, W. P. Jr., and Norman, R. Z., "The Human Use of Information. II: Signal Detection for the Case of an Unknown Signal Parameter, Transactions of the I. R. E., Professional Group on Information Theory, PGIT-4, September, 1954.

6. Tanner, W. P., Jr., and Swets, J. A., "A New Theory of Visual Detection", Technical Report No. 18, Electronic Defense Group, University of Michigan, Ann Arbor, August 1953.

7. Tanner, W. P., Jr., and Swets, J. A., "The Human Use of Information. I: Signal Detection for the Case of the Signal Known Exactly", Transactions of the I. R. E., Professional Group on Information Theory, PGIT-4, September 1954.

8. Tanner, W. P., Jr., and Swets, J. A., "A Decision Making Theory of Visual Detection", Psychology Review, 1954, 61, 6.

9. Tanner, W. P., Jr., Swets, J. A., and Green, D. M., "The General Properties of the Hearing Mechanism", Technical Report No. 30, Electronic Defense Group, University of Michigan, Ann Arbor, 1955 (In Preparation).

10. Thurstone, L. L., "Psychophysical Analysis", American Journal Psychol. 1927, 38, 368-389.

11. Thurstone, L. L., "A Law of Comparative Judgment", Psychol. Review, 1927, 34, 273-286.

12. Woodward, P. M., and Davies, I. L., "Information Theory and Inverse Probability in Telecommunication", Proc. I.E.E.(London), v. 99, p. 37, March, 1952.

DISTRIBUTION LIST

1 Copy	Director, Electronic Research Laboratory Stanford University Stanford, California Attn: Dean Fred Terman
1 Copy	Commanding General Army Electronic Proving Ground Fort Huachuca, Arizona Attn: Director, Electronic Warfare Department
1 Copy	Chief, Research and Development Division Office of the Chief Signal Officer Department of the Army Washington 25, D. C. Attn: SIGEB
1 Copy	Chief, Plans and Operations Division Office of the Chief Signal Officer Washington 25, D. C. Attn: SIGEW
1 Copy	Countermeasures Laboratory Gilfillan Brothers, Inc. 1815 Venice Blvd. Los Angeles 6, California
1 Copy	Commanding Officer White Sands Signal Corps Agency White Sands Proving Ground Las Cruces, New Mexico Attn: SIGWS-CM
1 Copy	Commanding Officer Signal Corps Electronics Research Unit 9560th TSU Mountain View, California
75 Copies	Transportation Officer, SCEL Evans Signal Laboratory Building No. 42, Belmar, New Jersey FOR - SCEL Accountable Officer Inspect at Destination File No. 22824-PH-54-91(1701)
1 Copy	H. W. Welch, Jr. Engineering Research Institute University of Michigan Ann Arbor, Michigan

1 Copy	Document Room Willow Run Research Center University of Michigan Willow Run, Michigan
11 Copies	Electronic Defense Group Project File University of Michigan Ann Arbor, Michigan
1 Copy	Engineering Research Institute Project File University of Michigan Ann Arbor, Michigan