NASA CONTRACTOR REPORT

NASA CR-434

AN INVESTIGATION OF THE VISUAL SAMPLING BEHAVIOUR OF HUMAN OBSERVERS

by J. W. Senders, J. I. Elkind, M. C. Grignetti, and R. Smallwood

Prepared under Contract No. NAS 1-3860 *by*
BOLT BERANEK AND NEWMAN, INC.
Cambridge, Mass.
for Langley Research Center

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION • WASHINGTON, D. C. • APRIL 1966

NASA CR-434

AN INVESTIGATION OF THE VISUAL SAMPLING BEHAVIOUR OF HUMAN OBSERVERS

By J. W. Senders, J. I. Elkind, M. C. Grignetti,
and R. Smallwood

Distribution of this report is provided in the interest of information exchange. Responsibility for the contents resides in the author or organization that prepared it.

Prepared under Contract No. NAS 1-3860 by
BOLT BERANEK AND NEWMAN, INC.
Cambridge, Mass.

for Langley Research Center

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

For sale by the Clearinghouse for Federal Scientific and Technical Information
Springfield, Virginia 22151 – Price $2.25

TABLE OF CONTENTS

ACKNOWLEDGMENTS

I wish to acknowledge the invaluable assistance of Jane Ward of Bolt Beranek and Newman Inc. in obtaining the subjects and training them, designing the apparatus and getting it built, running the experiments, and proof-reading and editing the reports and for making suggestions which proved useful in the design of the experiments.

ABSTRACT

We have undertaken an investigation of human visual sampling behaviour. The experimental portion of this study has been aimed at verifying and extending results obtained in a single study done in 1954. Those results suggested that the theoretical notions which had been advanced in Reference 14 would make a valuable contribution to the solution of practical design problems of aerospace vehicles. For that reason it was felt desirable to verify the results for at least one other condition as well as to investigate a number of other quasi-operational situations for which no analytical solution was then available. In addition, further theoretical investigations were carried on. These have led to a much more comprehensive theory about human visual sampling behaviour in particular and about attention in general. The results of the experiment, taken in the aggregate, strongly support the simple theories about frequency and duration of visual fixation, and about transition from one point of fixation to another.

The extended theory incorporates ideas about conditional sampling behaviour, in which the observer's inter-sample interval is a function of the value of the signal read on the previous sample. In addition, the idea is advanced that as a consequence of the single-channelness of attention, queueing theory provides a general method of analysis of the switching of attention, of the attentional demand of a stimulus source, of the probability of simultaneous demand from two or more sources of stimuli, and of the notion of overload. The conditional sampling models provide the probability distributions which enter into the queueing model.

The results of the study suggest that some parts of the model proposed can be used in the analysis of real systems. Appendix II shows the results of applying a transition probability equation to data taken from the literature. The whole set of models could be applied with profit to preliminary analysis of manned systems, if the analyst is careful to take into account the limitations of the simple theories which were tested, and to use with caution the ideas, as yet untested, which are presented in the theoretical discussions. The more complex theoretical model should ultimately make possible the analytical solution of some of the human factors design problems which have been treated only empirically in the past.

PART I--INTRODUCTION

Human beings in their normal living activity must receive and organize information taken from the environment both to manipulate that environment and to satisfy intensive needs for stimulation. With the exception of highly restricted and quite artificial laboratory situations which, being totally contrived, are knowable, most natural environments have in them a large number of different kinds of stimuli. Most natural situations are difficult to describe quantitatively but this is usually due more to our inability to analyze and characterize natural stimuli and to the inadequacy of our understanding of human information processing, than to the number of stimuli, however large. For some small set of situations which are more characteristic of a mechanical aspect of modern life, and therefore more nearly approach the laboratory situation, the stimuli which present information for reception and organization are more clearly defined and are more susceptible to measurement, analysis, and scientific understanding. Examples of this are operator positions of a chemical plant, an airplane or a space vehicle.

It is obvious that when sources of useful information (places where stimuli occur), are sufficiently separated in space or in time, some kind of overt sampling behaviour must take place if such a dispersed multiplicity of information sources is to be observed. Thus the eyes cannot look in two places at the same time if the places are more than 180 degrees apart. The distance need not, of course, be as large as this since for the observation of fine detail foveal vision is required. The eyes will be seen by an external observer to fixate on one or the other location. Similarly the hand of a "tactual observer" cannot be in more than one place at a time; and if the things to be felt are more than the span of the fingers apart, the hand must touch first one and then the other. If human observers had sufficiently mobile ears, like those of donkeys, then overt auditory sampling behaviour would be observed. As it is only shifts of head position indicate that spatially separated sounds are being attended to.

Less obvious is the notion that information sources which exist in the same place at the same time, or are presented at the same time to different sense modalities, must also be sequentially sampled by the human observer. Under these conditions, the sampling is, of necessity, covert, and not directly observable. In the case where sampling is overt, i.e., the sources are separated and involve vision, there is no question as to whether an observer can deal with many sources absolutely simultaneously. He can only look in one place at a time. For the covert case the question remains unanswered, although for some situations (and some experiments) data have been accumulated which bear on the question. Historically, the problem of multi-sensory sampling arose with the use by astronomers of the so-called "eye and ear method". This technique involved simultaneous watching of a star and listening to the tick of a clock. The task was to estimate that portion of the inter-tick interval which had passed when the star crossed a reference line in the field of view. The method was precise but different observers generated different constant errors.

The observation was made that different observers tend to favor different sense modalities. For example, if two objectively simultaneous signals occur, one to the eye and one to the ear, some observers would observe the sound before the sight; others the sight before the sound. Bessel, in 1822 (1) stated: "If it is assumed that impressions on the eye or the ear cannot be compared with each other in an instant and that two observers use different times for carrying over the one impression upon the other, a difference originates; and there is a still greater difference if one goes over from seeing to hearing and the other from hearing to seeing. That different kinds of observation are able to alter this difference (between observers) need not seem strange, if one assumes as probable that an impression of one of the two senses alone will be perceived either quite or nearly in the same instant that it happens, and that only the entrance of a second impression produces a disturbance which varies according to the differing nature of the latter." In other words, one of the things will be perceived when it happens; the other will "come in" later. Bessel's statement clearly suggests the notion that simultaneous observation through two sensory modalities of two objectively simultaneous events is impossible; and that it is instead the case that sequential observation must occur. Boring says also referring to Helmholtz's work of 1850: "Half a century later psychologists were ready to accept the principle that the latent times for perception vary so greatly that attentive predisposition may cause an impulse to mill around in the brain waiting for the attention to be ready to receive it." (1, p. 147)

The hypothesis that this kind of simultaneity is impossible is further supported by the evidence on response latencies or reaction times. To examine the evidence in detail is unnecessary in light of the large number of general discussions already in existence. Reference (2) is a good example. A variety of studies have shown that in general the larger the number of different kinds of possible events there are, the longer the time required to respond to any one of them. Merkel found a regular increase in reaction time with an increase in number of alternative stimulus-response pairs. His data (from Woodworth (3)) showed an increase from 187 ms for a simple reaction to a single stimulus to 622 ms for a reaction to any of 10 stimuli. The notion of the psychological refractory period* lends additional support to this idea.

Further, much of the results of the early research on perceptual-motor skills has supported the idea that there is a discontinuous functioning in the central nervous system. Such discontinuous functioning also supplies sequential mechanisms in attention. Welford (4) hypothesized that the psychological refractory period was the result of the inability of the central nervous system to permit an overlap of the times and functions required to organize the two or more responses involved. In other words, the operator behaves like a single channel system.

The work which has been done in measuring human information transmission where additional sensory channels of information have been used has led to the conclusion that such additional channels do not markedly increase the total amount of information that can be processed by human observers. On the other hand, it has also been shown that the addition of extra channels of information reduce the probability of missed signals in a vigilence task. This result suggests that there may be involuntary alternation of attention among sensory modes. More recently, Kristofferson (6) has postulated an involuntary internal switching mechanism and presented data which support the hypothesis. Thus, for a variety of situations covert sampling appears to be an genuine a function as does overt sampling even though it is less easily measured and not directly observable. Whether the difference between these two kinds of sampling behaviour is more than one of observability alone is not clear.

* The experiments of Telford (5) showed that if stimuli in a reaction time experiment followed one another too closely the second response was delayed by as much as 150 ms. Hence, the Psychological Refractory period.

Broadbent, (7), has proposed a general model of attention, which attempts to deal with both kinds of attentional or observing behaviour. There are always difficulties associated with this kind of model arising from the dual nature of attention and attending acts. Thus "attention" is at the same time the channel through which information flows, and a guiding system for directing this channel to one or another aspect of the environment. One is faced with the necessity for appealing to a hierarchy of "homunculi". This is unsatisfactory and casts doubt on the adequacy of all such models.

Messages in Broadbent's model may reach receptors simultaneously but be selectively blocked by some kind of filtering mechanism which stores on a short term basis some of the aspects of the stimuli which have been blocked. Clearly, there needs to be some kind of decision mechanism which depends upon long-term memory and selects a response which is appropriate to the information that has been filtered. The filter itself, in turn, must be guided by some kind of input from the decision system. In other words, any such model requires that it know what it is rejecting in order that the act of rejection can occur. The work on simultaneous listening to two messages, each presented to one ear, shows that various aspects of these input messages are successively discriminated, and that the different hierarchial levels of the decision channel operate successive filters until a single input message in fact gets through. Thus, the selective filters must consist of a series of operators which make comparisons between the various input channels and pass on certain messages for further inspection. The fact that messages which are repeated tend ultimately to be ignored suggests that attention is controlled in some way by the relative uncertainty of the different messages which are presented. However, this cannot be the only mechanism which controls attentional shift since changes in motivation of the observer, perhaps brought about by instruction, can cause him to observe signals which are less novel and less uncertain instead of those which are more so. Broadbent does point out that the general behaviour of a system which passes only novel stimuli will cause the filter to shift to new channels as habituation with any particular stimulus increases. He arrives at a conclusion that some finite time is required to shift from one channel to another.

Other evidence from Moray (8) shows that with careful timing of the stimuli subjects can alternate between trains of stimuli at rates higher than those achieved by Broadbent's subjects. Rabbitt, (9), performed a similar study in which similar results were obtained for two aspects of a single visual stimulus, shape and color, for example. Thus the selection mechanisms could apply either to different sense modalities, to different sense organs in the same sense modality, or to different stimulus aspects in the same organ.

A. A Discussion of Attention

One man understands what another man means when he says either that he gave his attention to something or that something caught his attention. If something *catches* your attention, you attend to it, i.e., focus on it, and examine it. Perhaps the *giving* of attention prior to its being caught is a different phenomenon, in which the incentive comes from within, rather than from without. In either event the process ends with an examination of the thing attended to. During the process of examination one assumes that attention is being "given" to the object; in a sense, though, it is the transition or the switching of attention from one object to another which is the most manifest character of attention. Attention is still a difficult thing to define and to work with. Woodworth (3) says: "In spite of its functional genuineness, the psychological status of the concept of attention has become more and more dubious." However, the experiments and thought given by previous investigators to the subject of attention are both of importance in our approach to the problem at hand.

Fundamental to our investigation is the question of whether a man can do two things at once. The formal name for this area of inquiry is "the division of the attention". Again to quote Woodworth, "Division of attention would mean a simultaneous focusing upon two *separate* activities. If one of them is automatic and goes forward smoothly without conscious control, no division of attention is required. If both are combined into a single integrated performance, no division of attention is required." And further, "if two activities, while carried on simultaneously in a loose sense, are kept going by rapid shifting of attention frcm one to the other and back again, there is in a strict sense no division of attention." Thus, in the classical sense, *division of attention* means the strictly simultaneous division of the attentive capacity of a man. Man always does more than one thing at a time, at least so it would appear to the casual outside observer. The entire autonomic nervous system functions (apparently) without interference from conscious activity. Walking apparently does not interfere with seeing and hearing.

An example of apparent division of attention is the kind of simultaneous performance investigated in 1887 by Paulhan (as cited by Woodworth). He was able to recite one familiar poem orally while writing another. The interference between the two was "minimal." He could also recite a poem while performing simple multiplication without interference. However, "An operation offering any difficulty was retarded even by so automatic a simultaneous performance as the recitation of a familiar poem." The early experiments of Binet involving motor acts of the two hands differently coordinated with auditory signals showed that there was interference between the two sets of activities. In general, the evidence is fairly clear that double performances result in a diminution of performance on either one or the other or both of the two components if the joint task cannot be combined into a single coordinated movement. When the two parts can be combined, there is, perhaps, no reason to expect diminution of performance on a component of the combined unitary task. The question as to whether two attentive acts can be done at the same instant still remained to be answered. The experiments of Mager 1920, and Pauli 1924, (both cited by Woodworth (3)), seem to indicate that simultaneous performance of two attentive acts of cognition did not often, if ever, occur. Thus their evidence suggests a unitary quality to attention and denies the general possibility of simultaneous non-alternating attention to two or more things.

More recently Hebb (10) states that the conclusion of unity of attention needs qualification. He denies that there is evidence to justify the general statement that learning never occurs without the help of attention and raises the anecdotal evidence that people seem to carry on two familiar activities at the same time, such as arguing and driving a car. He states that "neither seems possible without attention," and further: "it certainly seems that the unity of attention has been exaggerated". However, rapid alternation is frequently mistaken for simultaneity. Deutsch and Deutsch (11) raise some issues which relate to central neural and neurophysiological models for selective attention.

They briefly describe a mechanism "which assumes the existence of a shifting reference standard, which takes up the level of the most important arriving signal.", but this mechanism can easily be incorporated into an uncertainty model.

There have been numerous experiments on simultaneous listening to two different messages as well as on listening and reading at the same time. The results of these studies have made it necessary to erect a variety of models which first examine the nature of the material being presented and then select parts of the material for further consideration by some more centrally located mechanism. The reason there is a problem, of course, is that there is no way of comparing two or more streams of events and selecting one for its importance unless the individual streams of events are first evaluated. The evaluation requires some form of central nervous processing and yet this can't be simultaneous attention in the ordinary sense since the material which is not selected is not remembered, not learned, not responded to. It is a problem analogous to the question discussed by Boring (12) of whether a hypnotized person who has been instructed not to see anything which is red can be said to be blind to red. In a sense, in order to state that a red object is not "there" he must have first seen that it was red and then subsequent to that perception, analysis, and identification, performed an act of rejection.

One fairly general finding of simultaneous listening--or listening and looking--experiments is that interference is produced only when some arbitrary limit of task complexity has been passed. The work of Crossman (13) suggests that there is an upper information transmission rate beyond which "more or less simultaneous" processing of two streams of data cannot be performed. Many investigators in England, Welford (4), for example, have shown that one can predict the results of simultaneous input signals on the assumption of a single channel somewhere in the central nervous system, and Crossman states (13) "the most plausible view here seems to be that there is indeed only one central organizing channel for new external information, but the feedback from the subjects' own actions may sometimes be processed in parallel with it. However, the available time is very efficiently shared between various demands in a complex task."

The analysis which follows is predicated on the notion that attention is directed by a need on the part of the observer to reduce uncertainty about the information source which is attended to. Thus, it makes no difference whether uncertainty is generated by an external time-varying process or by an internal process of forgetting it or a combination of both. In all cases, as uncertainty increases, the necessity for its reduction grows until an attentive act is demanded. Thus, although I do not necessarily deny the possibility of a completely voluntary act of attending, I would argue that the analysis of such behaviour brings one to the dilemmas of free will and determinism.

Thus various events or sequences of events in the perceptual environment from time to time "demand" attention from the observer. For certain classes of sequences of events the timing of this demand can be estimated. The magnitude of the demand can be estimated both on the basis of objective physical characteristics of the time series of events and of subjective states and characteristics of the observer. The product of the frequency and magnitude of the demand will be a measure of the "attentional demand" made by that information source, or signal, on the observer.

Attention will be considered to be unitary and capable of dealing with one demand at a time. The frequency with which it can alternate between various time series of events may be sufficiently high so that apparent simultaneity of processing will be observed. Whether apparent simultaneity will be observed is calculable on the basis of the physical characteristics of the time series involved. Looked at in this light one may consider the attention of an observer to be a channel which processes in sequence, never simultaneously, information arriving from many outside sources.

What might be the basis on which an information source demands attention from an observer, or, alternatively, the basis on which an observer decides to direct his attention to an information source? It is reasonable to treat these as examples of the same general process. This process is one of uncertainty reduction. In other words, the observer who directs his attention to some information source voluntarily, does so in order to reduce his uncertainty about the nature of the information presented. This uncertainty could arise in either one or two ways. First, if the information source, or place in the visual field, is a dynamic time-varying one, uncertainty as to the value of the variable presented must have accumulated since the last observation of that source, and, second, even if the process is a static process, unchanging in time, there is an internal time-based change in the observer, i.e., forgetting, which results in an increase in uncertainty about the nature of the information displayed. As the observer forgets, he does not instantaneously become totally uncertain about the nature of the thing he has seen. Instead, there appears in the observer merely an increase in the possible range of values which might be identified as the one previously seen. Such an increase in the range of possible values could be computed as an increase in entropy or uncertainty. There is no reason for treating this internal growth in entropy as being different from that associated with the dynamic time-varying process which is being observed. Thus the observer attends to the information source whenever the uncertainty as to the value presented rises above some critical level. From his point of view, the increasing uncertainty, whether internally or externally generated, has meaning only as a function of what the observer is trying to do. If the observer is interested in being aware of the magnitude of the time-varying process being observed, at every point in time, then his behaviour will be quite different from that which will be exhibited by someone engaged in "check-reading." The former person is doing a quantitative read-out of the magnitude of the process at every moment in time. The latter person is engaged in making a three-part decision about the values, i.e., it is above acceptable limits; it is below acceptable limits; or, it is within acceptable limits; and the numerical value of the name of the item presented is of no consequence. Further this latter observer will have some cost associated with the act of observing and will have some cost associated with a value outside acceptable limits. These give rise to a calculable threshold probability for the observer. Then he will observe when the probability of going outside of acceptable limits has exceeded that threshold value.

Further complications may exist if there is some probability distribution of acceptalbe limits instead of sharp, well-defined upper and lower limits; and instead of some arbitrary probability of exceeding acceptable limits, there is some variable probability. However these complications give rise for the most part merely to increases in the complexity of the calculation rather than to changes in the forms of the equations.

If an observer had only one information source to tend to, and literally no distracting or attention-demanding internal events occur, then his probability of detection of events of interest on that source would approach unity. That is to say, whenever that source required attention it could be attended to without delay. On the other hand, if there exist two or more information sources, each demanding attention and uncorrelated with one another, then there exists a probability that simultaneous demand will occur. That is, the observer will be attending to one source and satisfying a requirement for uncertainty reduction when the other source demands attention. Under these conditions the second source must wait. If it must wait, and if, as defined earlier, the probability of an event of importance has risen above some arbitrary value, then there is a finite probability that the event of interest will occur and will be missed. Thus, if there were N information sources, we could compute the probability of simultaneous demand upon the observer and therefore an overall probability that signals will be missed.

We are led to the idea that the single channelness of the observer causes information sources to queue up and wait their turn. The analysis of attention can then be approached as a problem in queueing theory. From queueing theory we can arrive at estimates of the probability distribution of simultaneous demands, the probability distribution of waiting times of information sources, and estimates of the probability that events of interest will be missed.

B. A Queueing Model of Attention

We have suggested that an information source will from time to time demand attention from the observer, and, that if he is able, the observer will "pay attention" to that information source. Either on the basis of purely theoretical considerations or on the basis of actual observations of observing behaviour, we could construct a probability distribution of attentional demands made by information sources. Although for actual calculations there is a question of how one deals with such very short intervals that the observations overlap, the general argument is not affected. If one accepts the notion of overlapping but distinct demands, then the probability function has some non-zero value at t=0. If it is assumed that demands are always separated by periods of non-attention, then the probability function has the value 0 at t=0. In general, as t increases, the probability of a new demand increases to a maximum and then diminishes monotonically to 0. It is conceivable that an information source would demand attention on a completely periodic basis; the distribution for that source would merely be a point, p=1, at that interval. In general, however, information sources will demand attention at intervals which depend on characteristics of the sources and of the observer's task using that source.

We can similarly calculate or measure the distribution of durations of attentive acts or observations of the various information sources. Observation takes time, so the probability of an observation of 0 duration is 0. In general, the probability of a duration will increase with increasing duration to a maximum and then decline monotonically to zero at some very large duration. Again, if an information source were so constructed as to require a constant observation time, then the distribution would shrink to a point with a p=1 at that duration. For most information sources there will be a distribution of durations of observation which will depend on the characteristics of the information source and the observer's task in using that source.

It is immediately clear that if the two distributions $d_i(t)$ and $o_i(t)$, for all the information sources in aggregate, do not overlap, no degradation of performance should be encountered or observed on any of the information sources as compared with the performance on it when it is dealt with alone. This follows since whenever one source demands attention, it will be dealt with no delay, since no other demands are being made on the operator. Since, in general, the distributions will overlap, one can compute the probability that there will be interference, i.e., the human observer will be busy observing one source when another source demands attention. This probability is

$$\int_0^\infty d(t) \cdot o(t) \, dt \qquad (1)$$

where d(t) and o(t) are the combined functions for all sources. As either of two events occur--either an increase in the frequency with which demands are made by one or more sources, or an increase in the duration of observation times, resulting perhaps from an increase in complexity of signal to be observed,--the amount of interference will increase in accord with the amount of overlap of the two probability density distributions. The value of this integral is P_{sd}, _the probability of simultaneous demand_.

If we wish to examine the process in detail we can imagine that for each information source, _i_, it is possible to measure or calculate the probability distributions $d_i(t)$

as well as the observing distributions $o_i(t)$. Then the probability of simultaneous demand will be the weighted sum of the integrals of equation 1 computed for all i.

$$P_{sd} = \sum_i p_i \int_o^\infty d_j(t) \cdot o_i(t)\ dt \quad i \neq j \qquad (2)$$

where p_i is the probability that source i is being observed, and $d_j(t)$ is the probability distribution of demand by each of the other non-i sources.

The value of the two integrals, of course, would be the same. The virtue of the more explicit statement of the second is that we can see at this point a possibility of computing, on the basis of known characteristics of the information source, the distributions o_i and d_i for each source, as well as the probabilities p_i that each of the information sources will be observed at all. Therefore we can see an analytical solution to the calculation of P_{sd} for certain classes of information sources. (The results of previous investigations on simultaneous listening, or simultaneous listening and looking, which suggest a "competition between sources as a function of the redundancy or predictability of the sources" are fairly well in accord with the results of this simple analysis.) We will attempt in later sections to make a more rigorous calculation of the relationship between the information flow rate from each of the sources and the distributions of intervals of attentional demand and of durations of attending.

To recapitulate briefly, I assume that the operator or observer is a single channel device and the demands are made upon this device by sources of information in the environment; that the sources, in a sense, arrive at the single channel device and form a queue; and the length of the queue formed by the information sources at any time is a direct measure of the degree of interference which will exist in any experiment involving "simultaneous attending to two or more sources of information." The length of the queue is a distribution function. It can be calculated on the basis of the probability of simultaneous demand. The notion of the probability of simultaneous demand serves as the basis for a rational attack on questions of perceptual overload and of workload calculations. The various components of the theory which will be advanced in the following sections are intended to apply to behaviour in the limiting case where the operator is at peak loading and subject to potential overload. The questions which are raised by the underloaded case are more difficult to analyze and for many practical applications of less importance.

C. Theoretical Sampling Behaviour

Senders (14, 15, 16) attempted to apply a very much simplified sampling theory to human scanning behaviour. A great many assumptions were made in arriving at the simple solutions and for many real situations these assumptions were not well founded. The ideal observer which was discussed in that paper was assumed to be interested in reading out, or reconstructing, the signal on the basis of the samples which he had taken. If that were the case then the calculations which were used would hold. However, as indicated earlier, most real observers engaged in real tasks are not concerned with signal reconstruction. Instead the observer attempts only to be aware of a departure of the signal from some arbitrarily chosen value by some arbitrarily chosen amount. That is to say, as mentioned briefly earlier, most real observers are engaged in "check-reading." The data presented in Reference (16) show remarkably good approximation to the theoretical values. This was particularly so for the transition probabilities, and was sufficiently close for the sampling frequencies themselves to permit useful estimation of the "Attentional Demand" imposed by each of the four independent signals. The task which was set to those observers, however, was not in fact the task of signal reconstruction. Instead it was a check-reading task. The data conform because the powers of the signals and the magnitudes of the significant deviations were the same for all signals. There was a logical necessity, therefore, for the sampling frequencies to be in proportion to the bandwidths, and in fact the data were in accord with this prediction. It was not pointed out in that earlier paper that if the powers had not been equal, or, given that they were equal, the magnitudes of the significant deviations were not equal, then the sampling frequencies would not have been proportional to the bandwidths. It is, however, constructive to follow the original reasoning because under certain operational conditions the operators are in fact engaged in the reading of signals. The theory holds quite well for these conditions and in a sense, the behaviour of the subjects is forced.

The following material is taken verbatim from Reference 16.

Frequencies and Durations of Sampling

Some general (and simple) theoretical notions about the sampling behavior of human monitors are presented here. It is impossible to estimate the information presented by a continuously varying instrument if consideration is given only to the instrument itself, apart from its use. It is still more difficult to estimate the total information flow from a display consisting of a multiplicity of instruments differing from one another in a variety of ways. Let us consider first the case of the single instrument (among many) as it is used by an ideal observer.

1. *The Single Instrument*: An instrument, i, will generate (under given system conditions) a sequence of pointer positions in time, $f_i(t)$. From $f_i(t)$ we can compute a power density spectrum $\Phi_i(\omega)$. Assume that $\Phi_i(\omega)$ has a maximum frequency (cutoff frequency) of W_i. The minimum sampling rate for periodically taken samples of the function $f_i(t)$ will be $2W_i$, if $f_i(t)$ is to be specifiable from the samples. We can also calculate the rate at which the instrument is generating information, if we specify a permissible rms error of readout by the observer, and the rms amplitude of the signal (17). For $f_i(t)$, with a cutoff frequency of W_i, an rms amplitude of A_i, and a permissible rms error of E_i, the information generation rate is

$$\text{Eq. (1)} \qquad \dot{\bar{H}}_i = W_i \log_2 \frac{A_i^2}{E_i^2} \text{ bits/sec.} \qquad (3)$$

Our ideal observer samples at a rate which permits the reconstruction of the signal from the samples. Therefore, he must sample with a fixation frequency FF_i, which is at least equal to $2W_i$. If FF_i is exactly equal to $2W_i$, then the average amount of information which he must assimilate at each sampling, $\bar{H}_i$, is

$$\text{Eq. (2)} \quad \bar{H}_i = \log_2 \frac{A_i}{E_i} \text{ bits} \qquad (4)$$

Reaction time has been shown by Hick (18) and Hyman (19) to increase with increasing stimulus information. For some stimulus conditions, the relationship has been shown to be linear. If we assume that our ideal observer has a fixed input channel capacity, then the duration of each fixation, D_i, should also be linearly related to the amount of information to be taken in at each observation. Therefore, we can calculate $\bar{D}_i$ to be

$$\text{Eq. (3)} \quad \bar{D}_i = K \log_2 \frac{A_i}{E_i} + C \text{ sec,} \qquad (5)$$

where K has the dimensions of time per bit, and C (with the dimensions: time per fixation) is a constant to account for movement time and minimum fixation time. This is an intuitively satisfying result: A_i is related to the possible range of values which the instrument could present, and E_i is a measure of the accuracy to which the instrument must be read. For the conditions specified, the attentional demand or work load placed on our observer by instrument i is clearly the product T_i of the fixation $\overline{FF}_i$ and fixation duration $\bar{D}_i$.

$$\text{Eq. (4)} \quad T_i = \overline{FF}_i \times \bar{D}_i = 2KW_i \log_2 \frac{A_i}{E_i} + 2W_i C \text{ sec/sec.} \quad (6)$$

T_i, the proportion of total time spent on instrument i, is, as it should be, related to the information generation rate of the instrument $\bar{H}_i$.

If the fixation frequency is greater than $2W_i$, the samples will be correlated and the amount of information to be taken in at each sample will be less than $\log_2 A_i/E_i$.

Since $\overline{H}$ is a property of the signal and not of the sampling process, it will be constant as the fixation frequency increases. Thus

$$\text{Eq. (5)} \quad \overline{H}_i = \frac{2W_i}{FF_i} \times \log_2 \frac{A_i}{E_i} \text{ bits} \qquad (7)$$

and

$$\text{Eq. (6)} \quad \overline{D}_i = \frac{2KW_i}{FF_i} \times \log_2 \frac{A_i}{E_i} + C \text{ sec.} \qquad (8)$$

Because of the additive constant C, the percentage total time spent on an instrument is minimized by making $FF_i = 2W_i$, as in Eq. (4).

2. Multiple Instrument Displays: For a complex of m instruments, we can calculate the total work load placed on the ideal observer by summing the individual work loads of the m instruments. For each instrument, we calculate or measure W_i and A_i/E_i. From these we calculate the product $\overline{FF}_i \times \overline{D}_i$, and sum across instruments. The sum would be the minimum utilization time per unit time for the m instruments,

$$\text{Eq. (7)} \quad \text{Min } T_m = 2 \sum_{i=1}^{i=m} W_i \left[K \log_2 \frac{A_i}{E_i} + C\right]. \qquad (9)$$

This result can be used in the design of instrument panels. For example, if a decision must be made about the addition of an instrument, we might proceed as follows: let T be the unit time; then, if $T > \text{Min } T_m$, one can try to add instrument j to the set of instruments. W_i and A_i can be determined or estimated from known parameters of the system to be monitored or controlled; E_i can be determined or estimated from the system requirements. Therefore, the decision to add or not to add can be made rationally: if $T_i + \text{Min } T_m \leq T$, add.

Fixation Sequences

"As a consequence of the sampling performed by the observer on the various instruments of a set, transitions will be made from one instrument to another and frequency distributions of such transitions will be generated.

"*Transition Probabilities*: We can examine the consequences of the assumption that the sequence of transitions is a random series constrained only by the relative frequencies of fixation of the instruments involved in any transition. We assume that a transition starting from instrument i may end on any instrument, including instrument i in accord with the probabilities of fixation on each instrument. Over a sufficiently long time interval, the relative number of fixations on each instrument will be an estimate of the probability of fixation on that instrument, and this in turn reduces to the equating of the relative frequency of fixation to the probability of fixation. Thus,

$$\text{Eq. (8)} \qquad P_i = \frac{T \times FF_i}{T \sum_{i=1}^{N} FF_i} = \frac{FF_i}{\sum_{i=1}^{N} FF_i} \qquad (10)$$

"The probability of a transition between instrument a and instrument b is P_aP_b; the probability of transitions in both directions, $P_{\overline{ab}}$, is

$$\text{Eq. (9)} \qquad P_{\overline{ab}} = 2P_aP_b \qquad (11)$$

"It is clear that if P_a and P_b are large, many transitions will perforce be made between them. However, it is also obvious that, as the probabilities of the various instruments approach one another, the freedom of path through the set of instruments increases and is maximal when all are equal. Thus, as the restraints of relative frequency diminish, there is greater opportunity for logical patterns of scanning to occur. We expect, however, that much of what has been observed about transition probabilities can be calculated on the basis of the sampling frequencies.

Measurable Data

"If the observer is looking at instrument a, there is a probability P_a that the next observation will also be on instrument a. This fact very much affects the empirical data which will be obtained from measurements of a multi-instrument task. In the first place, the measured frequency of observation will fall short of that predicted by P_a x FF_a samples per second. The observable frequency of observation of instrument a, FF_{oa}, must be corrected:

$$\text{Eq. (10)} \quad FF_{oa} = FF_a(1 - P_a) - 2W_a(1 - P_a) \text{ if } FF_a = 2W_a. \quad (12)$$

"The numerical deficiency will be proprotionally larger for the instruments with the greater bandwidths (and higher frequencies of fixation) and as a result, the relative frequencies for the instruments with the lesser bandwidths will be increased.

"In the second place, the pair of observations of a constitutes an unobservable transition from a to a which occurs with probability P_a^2. Therefore the observable probability of transition from a to b, $P_{\overline{oab}}$, must be corrected:

$$\text{Eq. (11)} \quad P_{\overline{oab}} = \frac{2P_aP_b}{1 - \sum_{i=1}^{N} [P_i^2]} \qquad (13)$$

Therefore, the observable transition probabilities will be larger than those calculated on the basis of Eq. (11).

"By the same process, the distribution of observable durations of fixation will be skewed toward larger values, and the observable mean duration of fixation $\overline{D}_{oa}$ must be corrected:

$$\text{Eq. (12)} \quad \overline{D}_{oa} = \left(\frac{1}{1 - P_a}\right)\left(K \log_2 \frac{A_a}{E_a} + C\right)." \qquad (14)$$

For the situation with equal signal powers and equal significant deviations the foregoing is an adequate description and can be used, although with caution, in the analysis of real systems. However, the estimates which can be made are only estimates of the means of distribution of intervals between observations. There is nothing which permits an estimate of the standard deviation or the probable range of these distributions.

That there are such distributions is apparent both from the data and from purely logical considerations. Aperiodicity in the sampling of any one instrument would result from almost any configuration of instruments and bandwidths except for cases where all the signals had identical bandwidths and identical significant deviations, or in certain other equally unlikely cases where a totally periodic scanning process was possible. There are at least two different ways of approaching the analysis of aperiodic sampling and we will consider these in turn.

PART II--THE INTERACTION BETWEEN REQUIRED ACCURACY AND EFFECTIVE BANDWIDTH AND SOME PLAUSIBLE APERIODIC SAMPLING MODELS

Shannon has pointed out (17) that a function of time limited to a band from 0 to W cycles per second can be completely determined by giving the ordinates of the function at a series of discrete points spaced 1/2W seconds apart, or the minimum frequency of sampling necessary for complete determination of such a function of band is 2W. However, as he further points out, for a continuous function, the information transmission rate would be infinitely large unless there is some error permitted between the output of the source and the signal which is recovered at the receiving end. In particular, he shows that the rate of information generation for a white noise source of power Q_i and band W_i with some permissible mean square error N_i is equal to $W_i \log_2 Q_i/N_i$, and, secondly, that the rate for any source (not necessarily white noise) of band W_i is bounded by $W_i \log Q_i/N_i$ and $W_i \log Q/N_i$ where Q is the average power of the source, Q_i its entropy power and N the allowed mean squared error. The entropy power is the power of an equivalent white noise limited to the same band of frequencies and having the same entropy as the signal in question. Whether, for human observers, the task of monitoring or tracking two signals of the same entropy powers would be of equal difficulty has not yet been tested. However, it can be assumed that there will be some agreement between the difficulty of an observing or tracking task and the entropy power of the signal which is observed or tracked. (The subjects in the experiments which are described later in this report were monitoring signals whose entropy power was less than their average power. When they were exposed to signals of the same bandwidth but with a higher entropy power, they expressed the opinion that these latter signals were more difficult. This point must be kept in mind in evaluating the workload which a system places on an observer.)

If it is desired to use a single transmitting channel to transmit information from a number of sources the channel must commutate between or among these sources at a rate at least equal to 2 x W_i for source i where W_i is the maximum frequency for the source i. In addition, if the channel has some capacity C, then W_i x $\log Q/N_i$ must be equal to or smaller than C for each of the cases in question. It must be remembered, however, that the criterion chosen, i.e., that of reconstructing to some error the value of the function which is being sampled, is not necessarily the only criterion, nor is it the only useful criterion to be considered. Let us examine the case of a human (or inhuman) monitor of a multi-degree-of-freedom process. Such a monitor may serve not as a channel for the transmission of a complete time function but rather as a channel for the transmission of a dichotomized (or poly-chotomized) time function or signal. For any function one might assume that there is a limit to the value of the function which calls for the transmission of a message, and all values of the function below this limit call for no transmission of the message. This is analogous to stating that the monitor observes the time functions and does nothing so long as they remain within a "safe" interval. When a function exceeds the limits of safe operation the monitor emits a signal which may be the present value of the function. We may now ask what the appropriate sampling strategy will be for the monitor. How accurately must the function be read if signals are to be sent properly? It is easy to see that if the permissible error, between the function as presented and the function as read, is equal to the amplitude of the function, no observation is needed. Similarly, if the permissible error approaches 0 then the information to be absorbed per sample increases and a longer time will be required for the monitor to accept and transmit the information. What is the appropriate strategy for selection of an interval between observations? If the function at the moment of observation has a value 0 (i.e., its mean), then the next sample may be deferred until such time τ as the probability of the function's exceeding the limits of safe operation exceeds some arbitrarily set probability. In particular, if the limit of safe operation is some ℓ standard deviations, then as τ increases, the correlation decreases, the variance increases and there will come about a point where the probability of the function's exceeding the limit is equal to or greater than the arbitrarily set probability. At that point a sample would be taken. If the function when observed is greater than 0, i.e., is some fraction of the way toward the limit, then the point at which the probability reaches or exceeds the arbitrarily chosen probability will in general come sooner and the sample must be taken after a shorter interval. In the limit, as the observed value of the function approaches the limit, the acceptable sampling interval approaches 0.

The following analysis provides a means of calculating the interval for any observed value, granted that the autocorrelation function of the signal is known.

Conditional Sampling I*

Assume that each sample of the signal gives us perfect information about the magnitude of the signal at the sampling instant, but no information about its derivatives. Assume, also, that we want to minimize the number of samples of the signal that have to be taken, or, equivalently, maximize the interval between successive samples. We are willing to accept some small probability, q, that we will not detect the fact that the signal exceeds limit L. Assume that the signal has zero mean and a standard deviation, σ_y.

First, let us establish some notation. We represent the signal by y(t). The autocorrelation function of the signal is $R(\tau)$. The normalized, or autocorrelation, function is represented by $p(\tau)$ and is equal to $R(\tau)/\sigma_y{}^2$. We use E[ ] to indicate the expected value of a random variable.

As the process unfolds, we sample it. Presumably, the closer the signal is to the limit L, the more likely it is to exceed L during some subsequent interval τ. Thus, if we sample the signal and discover that its sampled value is close to L, it would be wise to make the next sampling interval short. On the other hand, if the sampled value of the signal shows that it is remote from L, we could probably tolerate a fairly long interval before we sampled the signal again. Thus, the interval between successive samples is dependent upon the value of the signal observed at the previous sampling instant.

Since we have assumed that the sampling process gives us only the magnitude of the signal and none of its rates of change, we can use the autocorrelation function of the signal to account for the relation between samples of signal magnitude. This we can show as follows. Since the process is gaussian, the best prediction of the future value of the signal is obtained by a linear operation. Since we have measured only the magnitude of the signal at a sampling instant, t, its magnitude at some future time, $t+\tau$, is best predicted by the relation

$$y(t+\tau) = k(\tau)y(t) \qquad (15)$$

* This section is due to J. Elkind of Bolt Beranek and Newman Inc.

where k is the coefficient of regression. It can be shown simply that the regression coefficient is given by the relation

$$k(\tau) = \frac{E[y(t+\tau)y(t)]}{E[y(t)^2]} = \rho(\tau) \qquad (16)$$

Thus, the regression coefficient is equal to the normalized autocorrelation function of the signal.

Now, $\rho(\tau)^2$ is the fraction of the variance of $y(t+\tau)$ that is linearly correlated with $y(t)$. It is the fraction of the variance of $y(t+\tau)$ that is predicted by the term $ky(t)$ in Eq. (15). The fraction of the variance $y(t+\tau)$ that is uncorrelated with $y(t)$ is $1-\rho(\tau)^2$.

We make use of these facts to determine how to sample. If at time t we sample the signal and obtain a sample value, Y, the expected value of $y(t+\tau)$ is given by the relation

$$E[y(t+\tau)|y(t)=Y] = \rho(\tau)Y. \qquad (17)$$

This is the best prediction we can make of the future value of $y(t+\tau)$ given the magnitude at t. The variance of $y(t+\tau)$ with respect to the expected value given by Eq. (17) is just the variance of the part of $y(t+\tau)$ that is not linearly correlated with $y(t)$. Thus

$$\sigma_\epsilon^2(\tau) = [1-\rho^2(\tau)]\ \sigma_y^2 \qquad (18)$$

where $\sigma_\epsilon^2(\tau)$ is the variance of $y(t+\tau)$ about its expected value, $\rho(\tau)Y$.

We want the probability that $y(t+\tau)$ is equal to, or exceeds the limit L given that $y(t)$ is equal to Y, to be small, say q. This requirement may be written

$$\text{Prob}\left\{ y(t+\tau) < L|y(t)=Y\right\} = p = 1-q, \qquad (19)$$

where p is the probability that $y(t+\tau)$ will be less than the specified limit L.

Since the process is gaussian, we can rewrite Eq. (19) in terms of σ_ϵ.

$$E[y(t+\tau)] + n\sigma_\epsilon = L, \tag{20}$$

where n is chosen to give the desired value of p in Eq. (19). By making use of Eqs. (17) and (18), we may write for Eq. (20)

$$\rho(\tau)Y + n\sqrt{1-\rho^2(\tau)}\;\sigma_y = L,$$

or (21)

$$z_y\rho(\tau) + \sqrt{1-\rho^2(\tau)} = z_L,$$

where

$$z_y = Y/n\sigma_y \quad \text{and} \quad z_L = L/n\sigma_y,$$

We can solve Eq. (21) for $\rho(\tau)$

$$\rho(\tau) = \frac{z_y z_L}{1+z_y^2}\left[1 + \frac{1}{z_y z_L}\sqrt{z_y^2 - z_L^2 + 1}\right]$$

(22)

$$\rho(\tau) = \frac{z_y z_L}{1+z_y^2} + \frac{1}{1+z_y^2}\sqrt{z_y^2 - z_L^2 + 1}.$$

τ is the sampling interval we wish to determine. To minimize the number of samples that have to be taken, τ should be made as large as possible. The smallest value of τ, for which the left side of Eq. (21) is equal to the limit L, is the desired maximum sampling interval. If we know the autocorrelation function of the signal y(t), the sampling interval between each sample of the signal can be determined by using Eq. (22).

It will be noted that we have selected only the principal root of Eq. (22) since we are interested in the smallest value of τ for which Eq. (22) is satisfied. The smallest value of τ will, in general, correspond to the largest value of ρ. Thus, Eq. (22) can be used directly to find the value of the autocorrelation function and, therefore, the value of τ that maximizes the sampling interval.

To check the correctness of Eq. (22), let us work through a few examples. First, assume that z_y, the normalized present sampled value of the signal, is zero. In this case, we find from Eq. (22) that $\rho(\tau)$ is

$$\rho(\tau) = \sqrt{1-z_L{}^2} \qquad (23)$$

If z_L is unity, we need not sample the signal again until ρ is zero, which corresponds to a sampling interval of infinity. This result makes sense since the limit L has been placed at the $n\sigma_y$ point of the distribution of y, and even without sampling we can be assured that the probability will be p that the signal will not exceed the limit. As a second example, consider the case in which y is equal to L. ($z_y = z_L$), the case in which the observed value lies exactly on the limit. In this case, Eq. (22) reduces to

$$\rho(\tau) = 1 \qquad (24)$$

Since the smallest value of τ for which $\rho(\tau)$ equals one is zero, Eq. (24) implies that the sampling interval must be infinitesimal, a result that is entirely consistent with the condition that Y = L.

Now for a more realistic example, assume that the signal Y is obtained by passing white noise through a simple RC low-pass filter with time constant of α seconds. It is well-known that the autocorrelation function of the signal obtained from such a filter is

$$\rho(\tau) = e^{-|\tau|/\alpha}. \qquad (25)$$

Substitute Eq. (25) into Eq. (22). By taking the ln of both sides, we obtain a direct solution for the sampling interval.

$$\tau/\alpha = - \ln \left[\frac{z_y z_L}{1+z_y^2} + \frac{1}{1+z_y^2} \sqrt{z_y^2 - z_L^2 + 1} \right] \qquad (26)$$

These equations can be used to compute the sampling interval τ. They apply to the case a single valued limit, L, not a symmetrical pair of limits, +L and -L. It is also assumed that some fixed probability of a miss can be tolerated. To compute the intervals one must know or be able to calculate $\rho(\tau)$, σ_y^2, L and Y.

However, if "n is chosen to give the desired value of p in equation 19", one is involved in the solution of the transcendental equation:

$$\int_{-\infty}^{\infty} e - \frac{x^2}{2} \, dx = p \sqrt{2\pi} \qquad (27)$$

Ideally one would desire to be able to compute some results from this theoretical model and compare these with the obtained distributions. Unfortunately, the transcendentality of Eq. (27) is a stumbling block to the analytical derivation of the distribution statistics of sampling intervals. Only if we make some simplifying assumptions about the choice of p, or if we do not choose a fixed value for p do we obtain tractable equations. In particular, as the succeeding analysis shows, if we choose our sample moment when p is maximum, for example, we obtain results of some interest.

Conditional Sampling IIa*

A sample function, $y(t)$, of a random, gaussian, $(0,\sigma)$ process is sampled at time, $t=0$. The sample value, $y(0)=Y$, is compared with a threshold or limit, $L>0$. It is desired to try several strategies that could be used in sampling the waveform, bearing in mind that the cost of allowing $y(t)\geq L$ without noticing it is very high while the cost of taking a sample is smaller but not zero. In what follows we shall assume that the normalized autocorrelation function of the process, $\rho(t)$, vanishes for $t=T_n$ and remains negligible beyond that point. This is the same as to assume the process is bandlimited.

The double event that $y(0)=Y$ and $y(t<T_n)=y$ has a joint probability of occurrence given by:

$$p(Y,y) = \frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}} \, e^{-\frac{Y^2-2\rho Yy+y^2}{2\sigma^2(1-\rho^2)}} \qquad (28)$$

while

$$p(Y) = \frac{1}{\sqrt{2\pi\sigma^2}} \, e^{-\frac{Y^2}{2\sigma^2}}. \qquad (29)$$

Then, the probability density function of $y(t)$ given that $y(o)=Y$ will be:

$$p(y|Y) = \frac{p(y,Y)}{p(Y)} = \frac{1}{\sqrt{2\pi\sigma^2(1-\rho^2)}} \, e^{-\frac{(y-\rho Y)^2}{2\ ^2(1-\rho^2)}} \qquad (30)$$

which is gaussian $(\rho Y, \sigma\sqrt{1-\rho^2})$ and where the notation $\rho(t)=\rho$ has been used for brevity.

The probability of exceeding the limit, L, at any time $t\geq 0$, given that $y(o)=Y\leq L$, is

$$\int_L^\infty p(y|Y)\, dy = P(\rho,Y) \qquad (31)$$

[*] This section is due to M. Grignetti.

Intuition suggests that this function has a maximum for some value of $\rho=\rho_m$ $(0\leq\rho_m<1)$. If this is so, a valid strategy could be to sample at that particular instant (defined by $\rho(t_m) = \rho_m$ for which the probability of exceeding the limit is maximum. The value ρ_m is found by solving the equation:

$$\frac{d}{d\rho} \int_L^\infty p(y|Y)\, dy = 0. \qquad (32)$$

With the help of the results

$$\int x\, e^{-\frac{x^2}{2\sigma^2}} dx = -\sigma^2 e^{-\frac{x^2}{2\sigma^2}} \qquad (33)$$

$$\int x^2 e^{-\frac{x^2}{2\sigma^\sigma}} dx = -\sigma^2 x e^{-\frac{x^2}{2\sigma^2}} + \int \sigma^2\, e^{-\frac{x^2}{2\sigma^\sigma}} dx, \qquad (34)$$

it can "readily" be found that

$$\frac{\partial P(\rho,Y)}{\partial\rho} = \left[Y - \frac{\rho}{1-\rho^2}(L - \rho Y)\right] \frac{e^{-\frac{(L-\rho Y)^2}{2\sigma^2(1-\rho^2)}}}{\sqrt{2\pi\sigma^2(1-\rho^2)}} \qquad (35)$$

which vanishes for $\rho_m = \frac{Y}{L}$ and for $\rho=1$, while at $\rho=0$ it amounts to

$$\left(\frac{\partial P}{\partial\rho}\right)_{\rho=0} = Y\, \frac{e^{-\frac{L^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}. \qquad (36)$$

We can see that no intermediate maximum exists if $Y\leq 0$.

As for P(ρ,Y) itself,

$$P(\rho,Y) = \frac{1}{2} [1 - \Phi (\frac{L - \rho Y}{\sigma\sqrt{1 - \rho^2}})] \qquad (37)$$

where Φ is the normal probability integral.

Figure I summarizes our results so far.

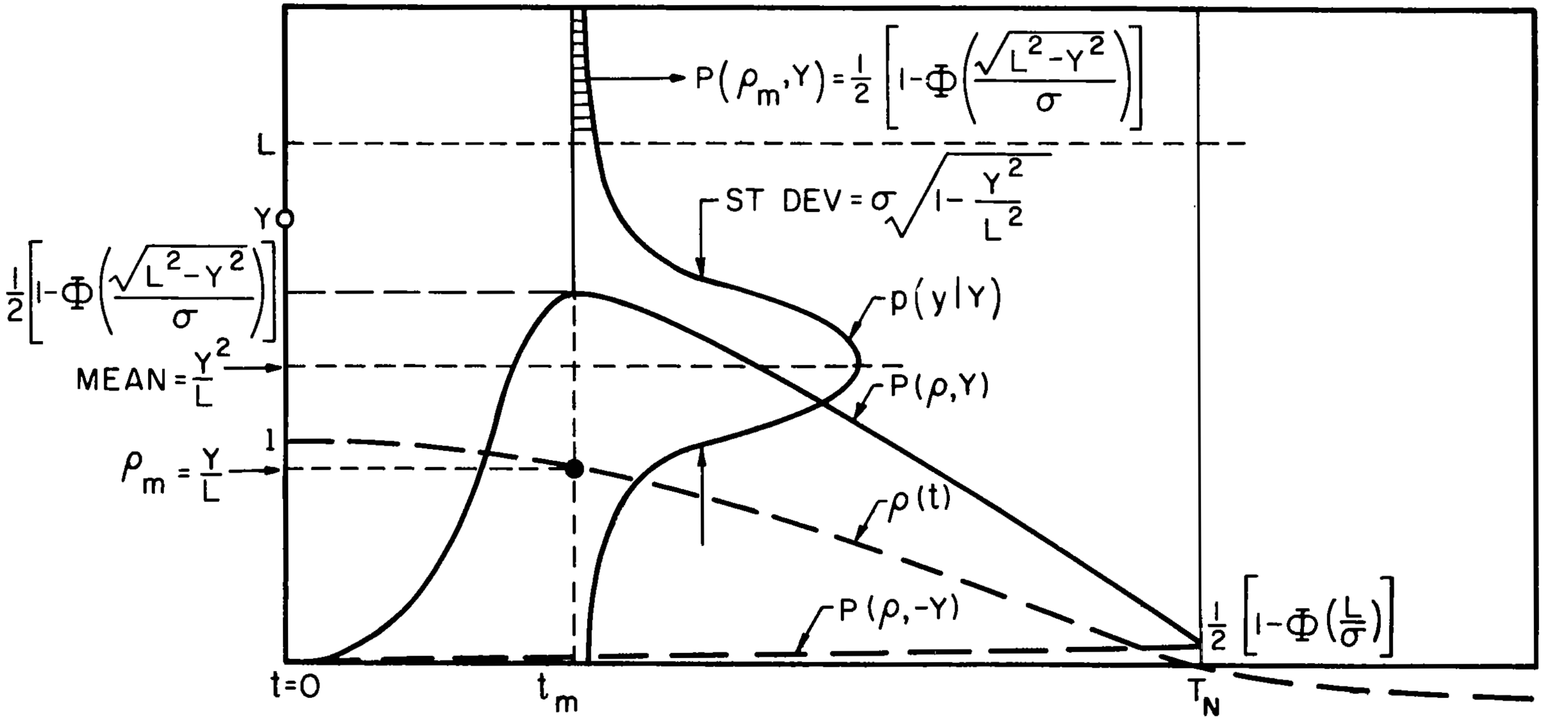

$P(\rho_m, Y) = \frac{1}{2}\left[1-\Phi\left(\frac{\sqrt{L^2-Y^2}}{\sigma}\right)\right]$
L
Y
$\text{ST DEV} = \sigma\sqrt{1-\frac{Y^2}{L^2}}$
$\frac{1}{2}\left[1-\Phi\left(\frac{\sqrt{L^2-Y^2}}{\sigma}\right)\right]$
$p(y|Y)$
$\text{MEAN} = \frac{Y^2}{L}$
$P(\rho, Y)$
1
$\rho_m = \frac{Y}{L}$
$\rho(t)$
$P(\rho, -Y)$
$\frac{1}{2}\left[1-\Phi\left(\frac{L}{\sigma}\right)\right]$
t=0
t_m
T_N

FIG. 1 PROBABILITY OF EXCEEDING L AS A FUNCTION OF TIME

We desire to compute the mean and variance of the autocorrelation values under the assumption of sampling at t_m.

$$\text{Since } \rho_m = \begin{cases} Y/L & \text{for } 0 \leq Y < L \\ 0 & \text{for } Y < 0 \\ 1 & \text{for } Y \geq L \end{cases} \tag{38}$$

and Y is gaussianly $(0,\sigma)$ distributed, the probability density function $p(\rho_m)$ looks like Fig. 2.

Analytically:

$$p(\rho_m) = \frac{1}{2} u_o(\rho_m) + \left[u_{-1}(\rho_m) - u_{-1}(\rho_m - 1)\right] \frac{L}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\rho_m L)^2}{2\sigma^2}}$$

$$+ \frac{1}{2} [1-\Phi(\tfrac{L}{\sigma})]\ u_o(\rho_m - 1). \tag{39}$$

The mean value, $\overline{\rho_m}$, is:

$$\overline{\rho_m} = 0 + \int_0^1 \frac{L}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\rho L)^2}{2\sigma^2}} \rho d\rho + \frac{1}{2} [1-\Phi(\tfrac{L}{\sigma})] \tag{40}$$

$$\overline{\rho_m} = \frac{\sigma}{L\sqrt{2\pi}} \left(1-e^{-\frac{L^2}{2\sigma^2}}\right) + \frac{1}{2} \left[\ 1-\Phi(\tfrac{L}{\sigma})\ \right] \tag{41}$$

In the same way, the 2^{nd} moment, $\overline{\rho_m{}^2}$, is:

$$\overline{\rho_m{}^2} = 0 + \int_0^1 \frac{L}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\rho_L)^2}{2\sigma^2}} \rho^2 d\rho + \frac{1}{2} \left[1-\Phi(\tfrac{L}{\sigma})\right] \tag{42}$$

$$= \int_o^L \frac{1}{L^2\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} x^2\, dx + \frac{1}{2}\left[1-\Phi(\tfrac{L}{\sigma})\right] \tag{43}$$

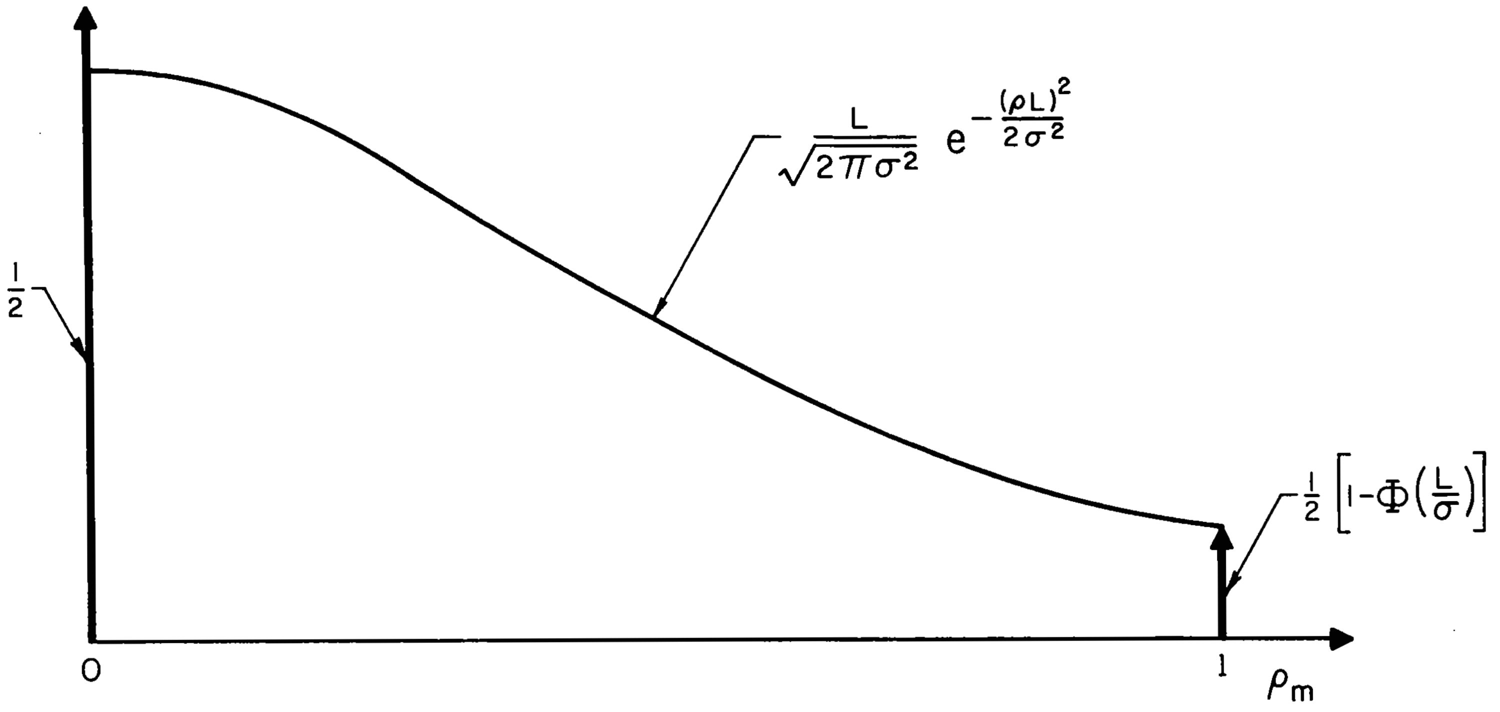


FIG. 2 PROBABILITY DENSITY OF AUTOCORRELATION FACTOR ρ_m

$$\overline{\rho_m^{\,2}} = \frac{\sigma}{L\sqrt{2\pi}} \left[\frac{\sqrt{2\pi\sigma^2}}{2L} \Phi\left(\frac{L}{\sigma}\right) - e^{-\frac{L^2}{2\sigma^2}} \right] + \frac{1}{2}\left[1-\Phi\left(\frac{L}{\sigma}\right)\right] \quad (44)$$

Where L is much less than σ, we can use the following approximations:

$$\text{a.} \quad e^{-\frac{L^2}{2\sigma^2}} \cong 1 - \frac{L^2}{2\sigma^2} \quad (45)$$

$$\text{b.} \quad \Phi\left(\frac{L}{\sigma}\right) = \frac{1}{\sqrt{2\pi}} \int_{-L/\sigma}^{L/\sigma} e^{-x^2/2}\, dx \cong \frac{2L}{\sqrt{2\pi\sigma^2}} . \quad (46)$$

Substituting in our expressions for $\overline{\rho_m}$ and $\overline{\rho_m^{\,2}}$ we get:

$$\overline{\rho_m} \cong \overline{\rho_m^{\,2}} \cong \frac{1}{2}\left(1 - \frac{L}{\sqrt{2\pi\sigma^2}} \right) \simeq \frac{1}{2}. \quad (47)$$

For other values of L/σ, the following table might be useful:

L/σ	$\overline{\rho_m}$	$\overline{\rho_m^{\,2}}$	$\sqrt{\overline{\rho_m^{\,2}} - \overline{\rho_m}^{\,2}}$
.1	0.48	0.48	0.5
.2	0.46	0.44	0.48
.5	0.41	0.37	0.45
1.0	0.28	0.22	0.37
2.0	0.20	0.08	0.20
5.0	0.08	0.02	0.12

Conditional Sampling IIb*

In the foregoing we have proposed a mathematical model for the behaviour of human monitors while performing a certain task, namely: to monitor a waveform against its amplitude exceeding a given limit L, by means of aperiodic sampling.

The model was based on the assumption that the waveform, after being sampled at time t_o where its amplitude is $y(t_o)=Y$, will only be sampled again either at the particular instant of time for which the probability of exceeding the limit is maximum or at the following generalized Nyquist instant, whichever is less.

The model was developed to the point where the probability distribution of the autocorrelation values corresponding to the sampling time intervals, as well as the mean and s.d. were calculated.

In this section we attempt to derive the same results after changing our basic assumption. Instead of letting the model wait until the probability of exceeding the limit L is maximum we make the model sample the waveform when this probability exceeds a certain threshold. This can be better explained with the help of Fig. (3), where a family of curves for $P(\rho,Y)$ have been represented.

$P(\rho,Y)$ is the probability of exceeding the limit L, t seconds after the last sampled amplitude, Y. For normalization purposes, t does not appear explicitly; the value of the autocorrelation at time t is used instead.

As shown in Fig. (3) the shape of $P(\rho,Y)$ changes considerably with Y, but some features remain fixed, among them the initial and the end point. The end point represents the next Nyquist instant, and therefore $P(\rho,Y)$ is independent of the previous value, Y.

It seems natural then to adopt this value of $P(\rho,Y)$ as a threshold. For the curve labeled $P(\rho,Y)$ this occurs at time t_T and our next task will be to find the corresponding value of ρ, ρ_T.

For that we have from Eq. (37):

$$P(\rho_T,Y) = \frac{1}{2}\left[1-\Phi\left(\frac{L-\rho_T Y}{\sigma\sqrt{1-\rho_T^2}}\right)\right] = \frac{1}{2}\left[1-\Phi(\frac{L}{\sigma})\right] \qquad (48)$$

* This section is due to M. Grignetti of Bolt Beranek and Newman Inc.

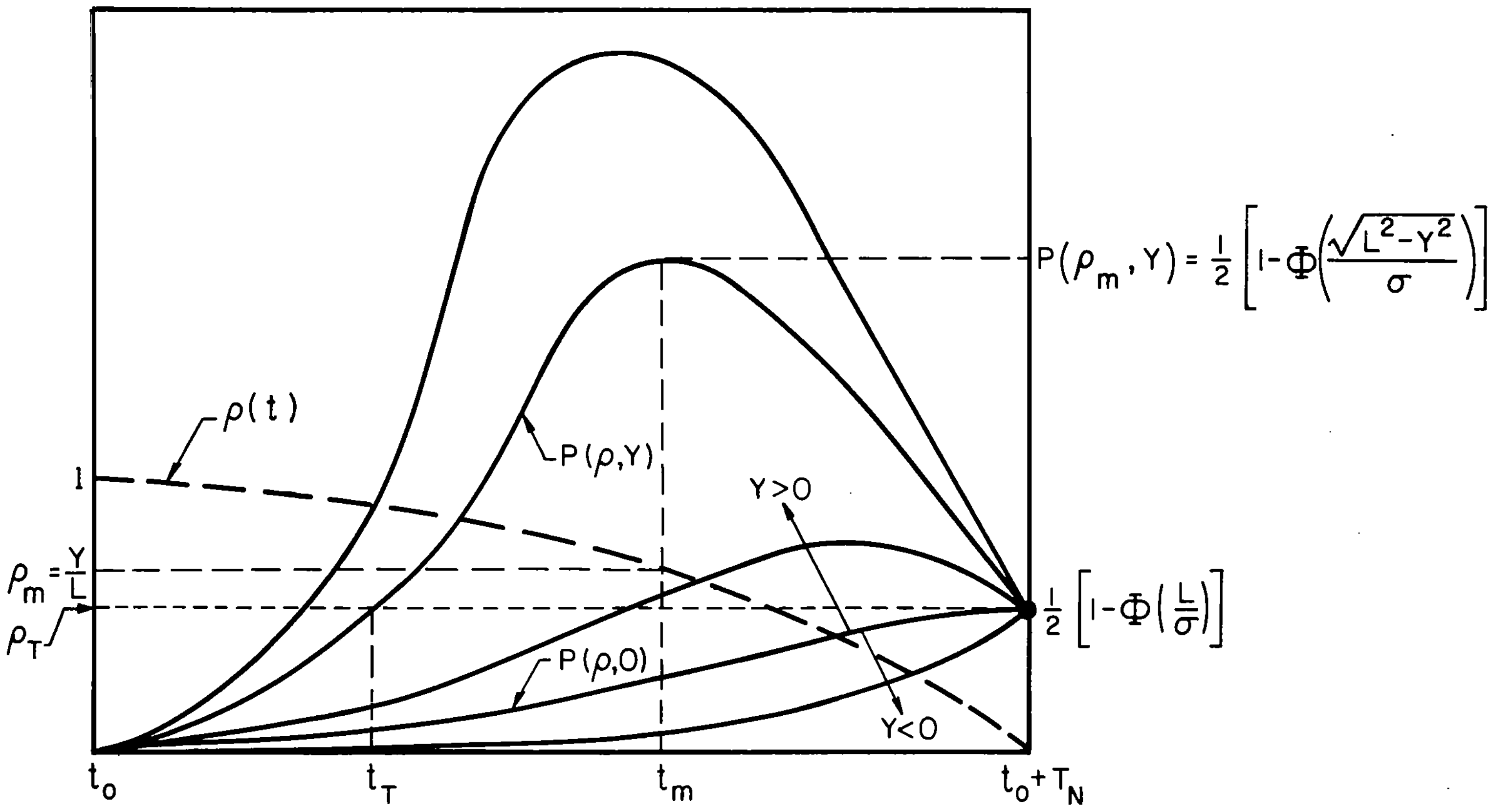


FIG. 3 PROBABILITY OF EXCEEDING L FOR VARIOUS VALUES OF Y

Solving for ρ_T we get:

$$\rho_T = \begin{cases} \frac{2LY}{L^2+Y^2} & \text{for } 0 \leq Y \leq L \\ 0 & \text{for } Y < 0 \\ 1 & \text{for } Y > L \end{cases} \qquad (49)$$

The mean and variance of this random variable are given by expressions analogous to those shown in Eqs. (41) and (44). They are:

$$\overline{\rho_T} = \int_0^L \rho_T \, p(Y) \, dY + \frac{1}{2} [1-\Phi \, (\frac{L}{\sigma})] \qquad (50)$$

$$\overline{\rho_T{}^2} = \int_0^L \rho_T{}^2 \, p(Y) \, dY + \frac{1}{2} [1 - \Phi \, (\frac{L}{\sigma})] \qquad (51)$$

Equation (50) can be reduced to tabulated functions. The result is:

$$\overline{\rho_T} = \frac{1}{\sqrt{2\pi}} \, \frac{L}{\sigma} \, e^{L^2/2 \, \sigma^2} \int_{\frac{L^2}{2\sigma^2}}^{\frac{L^2}{\sigma^2}} z^{-1} \, e^{-z} \, dz \qquad (52)$$

where the integral is known (and tabulated as the "exponential integral").

Equation (51) was approximated by Simpson's rule.

Numerical results are given in the following table.

$\frac{L}{\sigma}$	$\overline{\rho}$	$\overline{\rho^2}$	$\sqrt{\overline{\rho^2} - \overline{\rho}^2}$
.1	.487	.680	.665
.2	.474	.639	.643
.5	.433	.516	.572
1	.379	.333	.435
2	.310	.141	.212

Conditional Sampling III

A. Variable "Nyquist Interval" Model

We can consider the general case of a Gaussian signal with a power spectrum S(f) which diminishes with increasing frequency (beyond some frequency); and a permissible error power N(f). Shannon (20) defines a Rate Distortion Function R(D), as the minimum channel capacity required for the transmission of a signal with a distortion no greater than D. If the criterion D is a mean square error criterion, then, as shown by Kolmogorov (21), the rate distortion function becomes:

$$R(D) = \int_0^\infty \log \frac{S(f)}{N(f)} \, df \qquad (53)$$

which can be broken into two parts:

$$R(D) = \int_0^f \log \frac{S(f)}{N(f)} df + \int_{f_o}^\infty \log \frac{S(f)}{N(f)} \, df \qquad (54)$$

If, at and beyond f_0, the value of S(f) is equal to N(f), then the second part is equal to 0, and R(D) is what would be required for the transmission of a function limited in frequency to the range 0 to f_0. Thus, if samples are taken at a frequency of $2f_0$, the information contained in the signal will be transmitted with error no greater than D. The portion of signal with frequency greater than f_0, and with S(f) equal to N(f) makes no contribution to the signal and no demand upon the transmission channel. If now, N(f) varies for any reason, thus varying D, the frequency f_0 will vary. In general, as N(f) decreases, (which is another way of saying that the accuracy requirements increase), D decreases, and f_0 will increase requiring an increase in sampling frequency.

If the magnitude of N(f) varies as some function of the observed value of the signal being monitored, then a distribution of sample intervals will be generated which will depend on the form of S(f), and upon the rule which governs the relation between the observed value of the monitored signal and the value of N(f). If no other process were operating to produce aperiodicity in the sampling behaviour, then the process described above would generate a succession of "Nyquist intervals" of variable duration.

Since, following an observation of a signal value close to the limit, there would be another sample taken after a relatively short interval, the distribution of observed values would no longer be Gaussian, but will be rectangularized. The samples taken shortly after a deviant sample will have a higher probability of being deviant than samples taken at random. Since the interval between samples is inversely proportional to the value of the sample, there will be a non-gaussian distribution of intervals. Likewise the durations of observations will depend on the value of the signal which is sampled. The closer the sample value to the limit, the longer will be the observation time for that sample. Thus we might expect observation times and the durations on the succeeding intervals to be inversely related.

Let us now consider our original time function to be sampled and assume that it has a monotonically decreasing power as a function of frequency, and ask for the sampling strategy. The permissible error may be considered to be a function of the observed value of the process. A suitable criterion will be that the square error will be equal to some proportion of the squared difference between the value of the function and the limit, whatever it may be. Thus, $E^2=K(L-X)^2$ (where L and X are in terms of σ). If X on any observation is other than 0, the interval between that observation and the next will be determined by examination of the power function of frequency of the underlying process and as a function of K. K merely sets the arbitrary probability level that the signal will in fact exceed the set limit.

Of course, if X=L, then E^2 is 0, and the sampling interval is 0. When X=0, E^2 = K(L) and the sampling interval is maximum. A simple example will clarify this. If one were to have a band limited Gaussian signal and were to add to this a low amplitude signal in a band of frequencies well outside the random signal, then if, and only if, the value of the random signal itself approaches the limit does the high frequency "modulation" become of significance. Thus, if the permissible error is less than the amplitude of this high frequency signal, it becomes necessary to sample within an interval appropriate to the high frequency signal. However, if the observed value of the process is at or near 0 then the high frequency signal cannot make a significant contribution, i.e., send the process over the limit. Consequently the sampling interval can be adjusted to the low frequency part of the spectrum.

The models which have been presented for conditional sampling are not necessarily mutually exclusive. In particular, the last of these probably operates simultaneously with one of the others to make up the whole sampling behaviour. The approach to be used is clearly to be a function of what the monitor is trying to do, and a selection on any other grounds will surely be inappropriate.

Whatever model or set of models is chosen for the monitor, the distributions which result are to be inserted into the queueing theory equations presented earlier. Then the queue statistics can be calculated and a complete picture of the hypothetical behaviour generated. The coalescing of the various parts of the complete queueing model will be reported at another time.

Attached as an appendix to this report is a discussion, due to R. Smallwood, of Markov models for the human monitor. This appendix was issued as a separate BBN Report No. 1121. It suggests an alternative approach to the problem of analysis of the human monitor of multi-degree-of-freedom systems. In any final solution of the monitoring problem all of the notions presented both in this paper and in the appendix will almost certainly be included.

PART III--AN EXPERIMENTAL INVESTIGATION OF VISUAL SAMPLING BEHAVIOUR

Contract No. NAS1-3860 was entered into between Bolt Beranek and Newman and the NASA Langley Research Center on 10 April 1964. The contract called for the performance of five experiments in the course of twelve months. The objective of the contract was stated in Part I, Section B, Statement of Work, "The objective of this contract shall be a means of estimating to some precision the relationship of the human observer/controller to any definable system to the extent that if a system can be described in detail of mission requirements and information-processing requirements, then the degree to which such system operation loads the attentional demand of the human observer/controller can be calculated before simulation or prototype construction, and estimates can be made of the effect of variations in the system." The foregoing statement is, of ccurse, a very broad long-range goal of this contractual piece of work and of subsequent pieces of work to follow. The particular contractual requirements are as follows: "In performance of this contract, the contractor shall perform the series of experiments, as set forth below, designed to explore the relationship between visual attention, observer/controller workload and the information-theoretic characteristics of a display system. (The term 'theoretical' shall be defined as those analytic functions of mathematical structure called 'stochastic,' or commonly known as Markov processes.) It is, therefore, possible to define the work to be performed as a test of the hypothesis that either of these functions will, when used as a model, permit the prediction of observer/controller workload, visual attention and/or time and frequency of visual fixation when the signal characteristics are known.)"

The plan for these experiments had its genesis in a series of prior works of the principal investigator. Those had grown out of considerations of sampling theory and information theory as enunciated by Weiner and Shannon. The basic notion is generally this: it is evident that human controllers and monitors of systems must fixate their attention (their eyes) on a succession of instruments (information sources) within the cockpit of the vehicle or work station. This is true whether the system is fully manual, fully automatic, or semi-automatic. The major part of the continuous activity, particularly for the case of the monitor of the automatic system, and to a lesser degree for any monitor or controller, consists of observing the behaviour of the state variables of the system and, in anything other than a fully automatic system, correcting the state variables by means of appropriate input devices whenever necessary: i.e., when they exceed the limits or significantly depart from desired values. Thus, the task of the controller is some high percentage of monitoring and some low percentage of controlling. The more nearly automatic the system is, the higher the percentage of the total time spent in monitoring and evaluating the behaviour of the system.

To do this task the human monitor and controller must look at a variety of displays in a variety of locations. He must move his point of fixation from one instrument to another in order to be able to take in the information which is presented on the various instruments. It goes almost without saying that if an instrument is totally unrelated to the particular task at hand, it will not be fixated. It goes almost without saying that if an instrument varies very, very slowly, it will be fixated very, very infrequently. Conversely, instruments whose readings are of vital importance to the task at hand will be examined in great detail; and instruments whose readings vary rapidly will be examined often; and the more unpredictable is the signal, the more often will it be looked at. These self-evident statements are verbal analogs of the sampling theorem (as well as other parts of information theory). If an instrument must be read in detail and varies its reading often, then it will be looked at often and of necessity at the expense of other instruments. The original theoretical notions (14) are briefly summarized on pages 13-18 of this report.

In 1958 the results of some preliminary experiments were presented at a symposium at Wright-Patterson Air Force Base (15). These results showed that, in general, there was conformity between the actual behaviour of subjects and the predicted behaviour based on a very simple periodic sampling model. By 1963 there had been additional theoretical work which extended the earlier periodic model to one involving the use of Markov processes (16) to describe and analyze the behaviour; the data had been further analyzed and found to exhibit strong conformity with this more sophisticated approach to the problem (16). It was at that time that the experiments which were to be conducted under this project were planned.

In brief summary then, the experiments which were selected were based on a development of a theory or model first propounded as descriptive of human visual monitoring behaviour in 1953, and supported by data gathered in 1954 and 1955. In the light of this, the goal of the present experiments was twofold: (1) to re-confirm the conformity (of behaviour with theory) exhibited by the subjects in the earlier experiment, and (2) to obtain empirical data relating to correlated or coupled information displays and to discrete translations of continuous variables. There was no good analytical solution or prediction for either of the two latter cases at the time of their formulation; nor is there now.

Three experiments were designed as confirmatory of the earlier work. Their goals were: to explore the relationships between (a) signal bandwidth and frequency of duration, (b) required accuracy of reading and duration of observation, and (c) simultaneous variation of bandwidth and required accuracy of reading on the one hand, and frequency and duration of observation on the other--explored as concomitant rather than separated variables. The contract sets forth these five experiments precisely as follows:

1. Measure the relationship between observation time and required accuracy of reading and compare these results with theoretical predictions.

2. Measure the relationship between signal bandwidth and frequency of observation and compare these results with theoretical predictions.

3. Measure the effect of combined variations of bandwidth and required accuracy on frequency and duration of fixation and compare these results with the theoretical relationship obtained from the model equations.

4. Explore the relationship between the signal bandwidth and attention when the signal is quantized and displayed as a set of binary variables and fit these data to the theoretical model.

5. Explore the effects of signal dependency through correlation of signals and/or systems coupling.

In the material which follows these experiments will be identified as numbers 1, 2, 3, 4 and 5.

The first experiment done was No. 2. This was followed by Nos. 5, 4, 1 and 3, in that order.

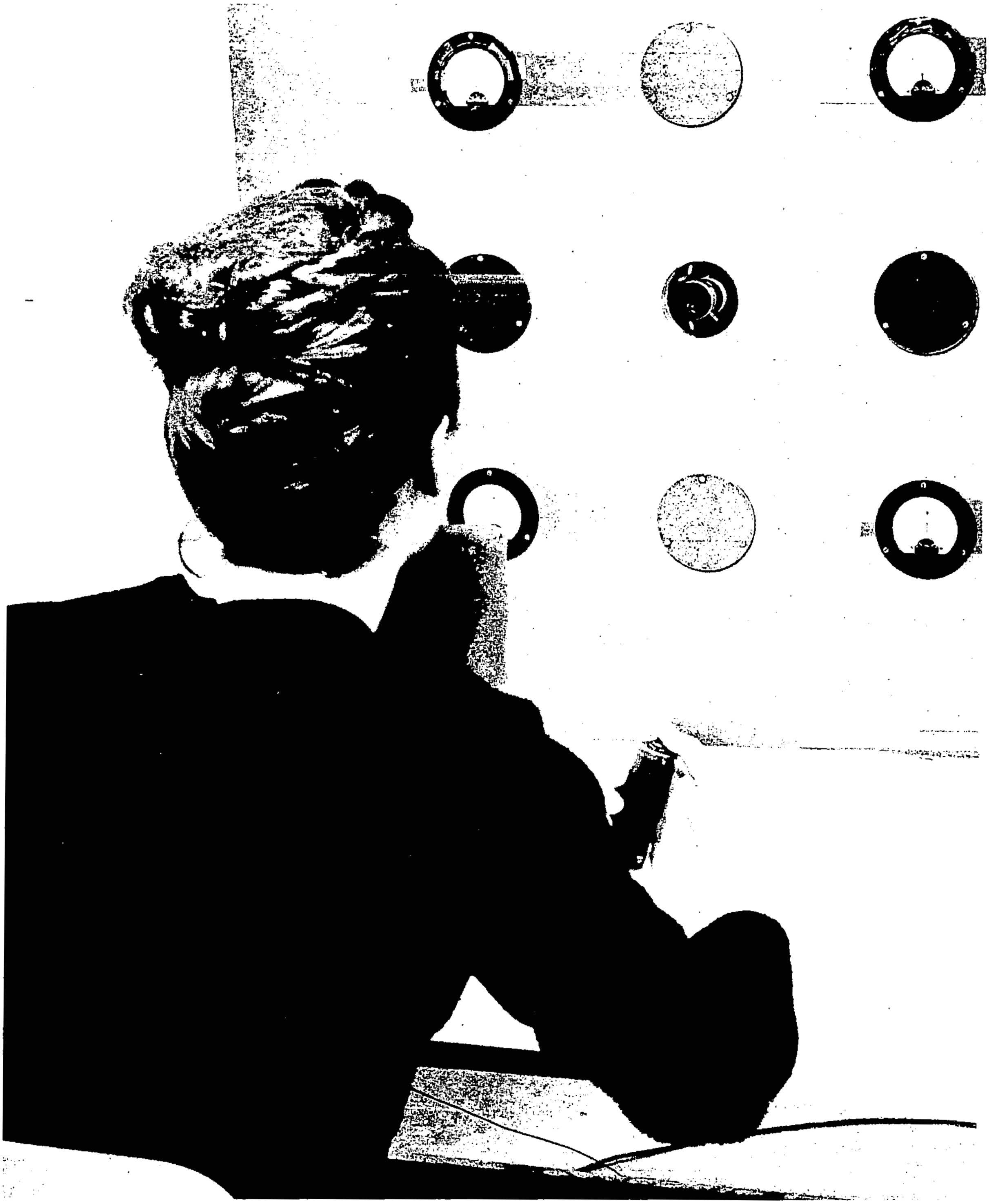
The Experimental Situation

The subjects were high school students in their fourth year at the Belmont High School. They were all upper-level students, selected by the school. Three were male and two female. All had adequate vision, although it was considered unnecessary that the vision be adequate when uncorrected, unless the wearing of eyeglasses interfered with the recording of eye movements. Photograph No. 1 shows the general layout of the experimental room and the five subject booths. Photograph No. 2 shows a close-up of the scene as watched by each of the observers. The subjects sat on chairs and adjustable chin rests were provided. The chair heights and chin rests were so set as to place the eyes of the observer at camera level at the center of the screen, and equidistant from the two sides. As seen in Photograph No. 3, there were six microammeters whose readings could range from -50 to +50, arranged in a row of three above the center of the field and a row of three below. The instruments were mounted six inches on center, or approximately 12 degrees apart at the 30 inches viewing distance. Each of the meters in a set was driven at a different bandwidth. The five meters corresponding to one bandwidth were connected in series through the five positions. As a result uniformity of deviation (within the accuracy of the microammeters) was possible without necessity of adjustment for differences in meter resistance. The meters' positions were varied in a quasi-random way in order to achieve as much counterbalancing as possible, since the theoretical model which was to be tested did not consider the factor of arrangement of signals of various frequencies. The instruments themselves were connected through a variable series resistor to the source of current. The series resistor permitted minor adjustment of signal amplitude to meet the requirements of the experiment.

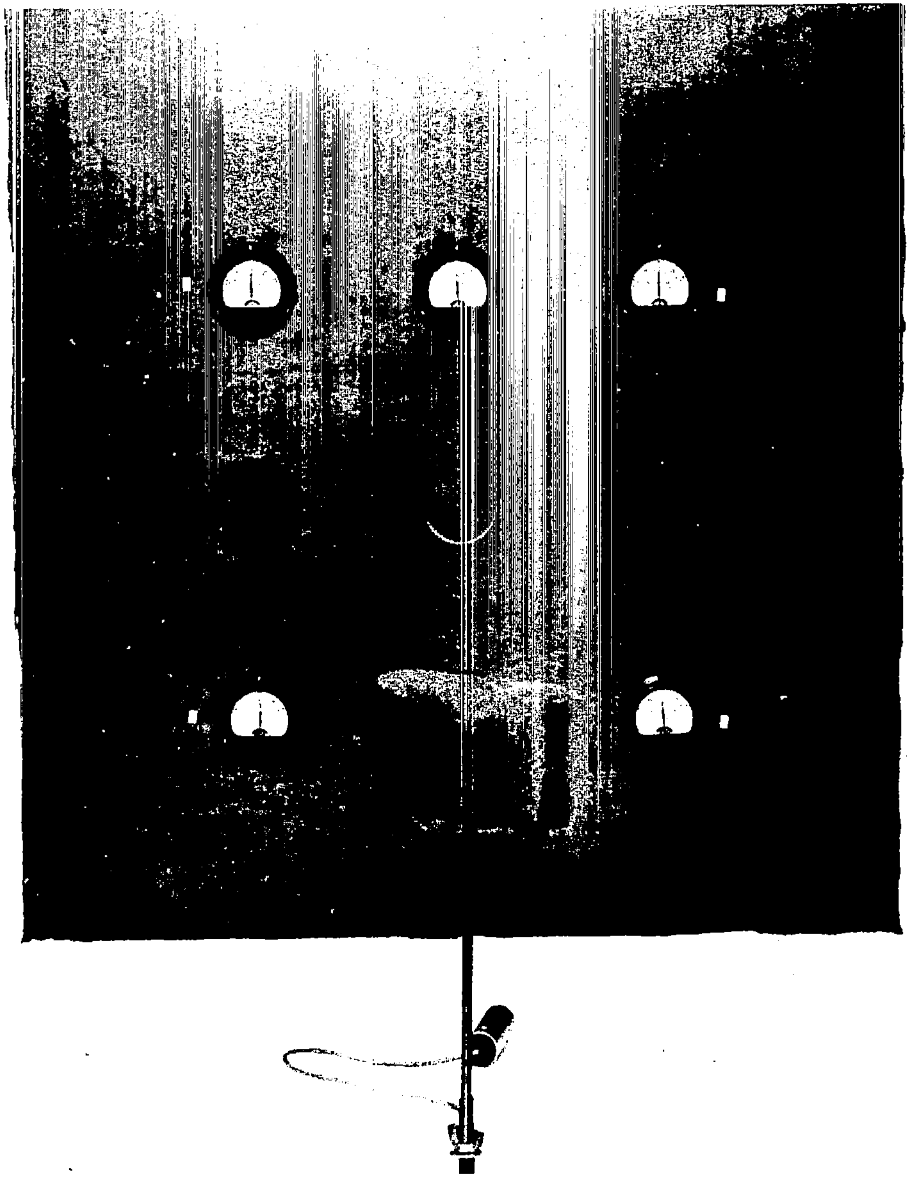
The Signals

The signals for the first three experiments, Experiments 2, 4 and 5, consisted of quasi-random sums of sines which had been used in the past for tracking work; the zero order distribution of these signals was approximately Gaussian, and the signals were flat from some relatively low frequency to the indicated cutoff point. Since the recorded signal amplitudes were not all equal, the series resistor mentioned above was used to adjust the power of the signals. One hour of recorded signals on six channels with a Mnematron tape recorder was available. It was felt that the complexity of the task would probably preclude learning during the course

Photograph 1: The Experimental Environment--The Subjects Seated at the Observation Posts.



Photograph 2: View Over Subject's Shoulder, Experiment 3, Motion Picture Camera in Place.



Photograph 3: Observation Post--Signal switch, Chin Rest--
Six Dials Task.

of the experiment. In fact, different "passes" of the signals "through" the subjects would be based on observations made at different times. The observers would not necessarily be aware of the fact that the signals were recorded and, therefore, completely repeated on successive set of trials. In order to lessen the possibility of the recorded signals being learned, a different starting point was chosen each day, except in those instances where changes in the level of performance were being checked.

The signal bandwidths chosen were .48, .32, .20, .12, .05 and .03 cycles per second. The sum of these is 1.20 cycles per second. These values were, in fact, chosen somewhat arbitrarily, the goal being to provide something approximating a 100 per cent workload for the subjects.

The peak monitoring capacity of a human monitor can be estimated crudely on the basis of observed facts about the durations of fixation on instruments. The mean duration of fixation on instruments of the sort used in aircraft and in these experiments, taken from all experiments of which I have knowledge, is about .40 seconds, including the transition times from one instrument to another. This means that such a monitor can make no more than 2.5 fixations per second. If such a monitor were presented with a task involving the monitoring of a set of signals the sum of whose bandwidths was 1.25 cycles per second, then such a task would constitute a full load. The point is that samples would have to be taken on each of the signals, if they are to be taken at all, at a frequency no less than double the frequency of each signal. The sum of these must equal or exceed 2.5 samples per second. Therefore, the task of monitoring a set of signals the sum of whose frequencies was 1.20 cycles per second is very nearly a full load for the monitor.

The task of the subjects did not constitute an identical test of the theory either in the earlier experiments of 1954, or in the current ones of 1964. The original relatively simple theory which was being tested was concerned with the bandwidth of the signal and with the relative accuracy, i.e., the ratio of mean-square amplitude to mean-square error, permitted in the readout of the signal. Since these signals were meaningless, i.e., they had no relationship to the real world, instead of requiring differential readout accuracy which would be needed for strict conformity to the sampling theorem model, the subject was required to respond by pushing a switch whenever the signal exceeded a value of 40 microamperes. The value of 40 microamperes was chosen to provide a suitably high, but not catastrophically high, output rate.

The subjects were seated in front of the instrument panel, and at a signal from the experimenter began to observe and to operate a (silent) switch on the end of a flexible cable whenever any meter-pointer went over |40| microamperes. However, they were told that they would receive bonuses which would be based on how closely their score came to the actual number of times the signals exceeded |40|. They were, in fact, rewarded in a random way: the amounts, which ranged from $.50 to $1.50 per pay check, were assigned randomly, and were not based on performance. The subjects were given no further instruction or knowledge of results.

The subjects performed their monitoring task for ten minutes and then received a rest of two minutes. This was repeated for one hour. This daily schedule was then repeated for ten days in order to bring them up to some consistent level of performance prior to the taking of data. It must be remarked here that in the earlier study of 1954 two things were evident. First, the efficiency of the subjects as detectors did not reach approximate asymptote until after 20 hours of training. Second, the frequencies of fixation of the signals approached a stable and very nearly theoretically correct level after as little as two or three hours of training. In this experiment we compromised. We assumed that since our goal was to investigate the fixation frequency, rather than detection efficiency ten hours would be sufficient time to permit the measurement of stable performance. At the end of this time the data were taken.

Recording

A Bolex Reflex H-8 Camera was used with an electric motor drive operating at 12 frames per second. The frame speed was based on a calibrated internal governer of the camera. This, in turn, was checked by photographing a stop watch and was found to be more than accurate enough for the task at hand. One-hundred-foot reels of film were loaded into the camera, and when the signals had been started and monitoring had begun as evidenced by the recording of responses of the subjects (which, presumably, were the result of detection), the camera was turned on by remote control and allowed to run until it ran out of film. This took, at 12 frames to a second, approximately 11 minutes. The subjects were then given a rest period of approximately 5 minutes, during which time the magazine was reloaded into the camera and the whole made ready for a recording of another subject. Thus, each subject provided approximately 10 or 11 minutes of data at the conclusion of approximately 10 hours of monitoring behaviour. The subjects were photographed, therefore, monitoring different sections of the signals, and their individual behaviours will, indeed, reflect the individual characteristics of the segments of tape signal which were being observed during the recording process.

The films were analyzed on a Gerber Digital Data Reduction System model number GDDRS-3B in conjunction with a Gerber Scanner S-10-C, and a Projector S-10-P. This system allows direct conversion of distance to digital readout onto punched cards. The process is as follows: the film is projected onto the surface of the analyzer screen and the hairline placed adjacent to the sprocket hole of the first frame in which the subject is looking a particular direction. The code for the direction of fixation is set in and the punch operated, providing a digital record of the location of fixation and a number associated with the first moment of that fixation. The distance to the sprocket hole at the end of the fixation is measured, converted to a digital readout and again the punch operates, punching in the number of the last frame of the fixation. Since the frame speed is known to be 12 frames per second, an immediate conversion to time is possible. The cards themselves are then analyzed statistically by simple computer processes.

A. Experiment 2: Comparison of Signal Bandwidth and Frequency of Observation

Experiment 2, which was a continuation of the earlier experiments of 1954, was performed under the conditions as described above. The hoped for relationship would have been the one predicted by Eq. (12) of this report. A number of things happened during this experiment. We started with five subjects. On one subject, the first one to be recorded, it was ascertained at the end of the recording process that the .48 cycles-per-second signal had not been presented throughout the entire recording. These data then were treated separately. This provided us with an inadvertent test of behaviour in an underloaded situation. The subject who was being recorded had a great deal of spare time. In particular, since the instrument was the highest frequency one, i.e., .48 cycles-per-second, the amount of time available was approximately one fixation-per-second. What the subject did with this spare time is shown in Fig. 4. This figure shows the frequency of fixation in fixations-per-second as a function of bandwidth. The ordinate is plotted on double

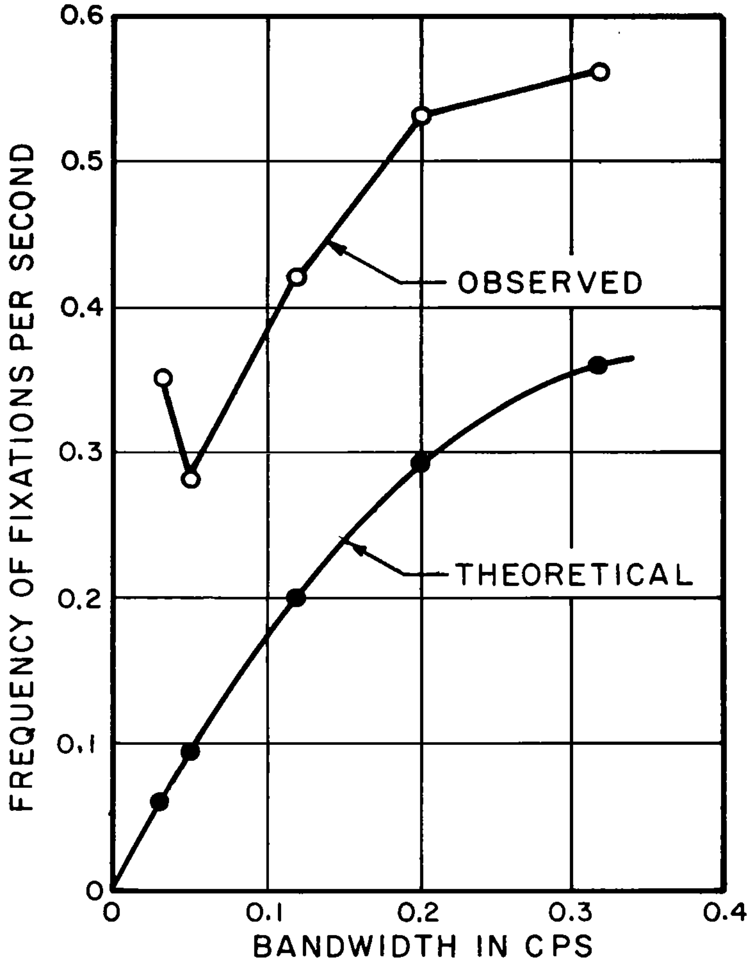


FIG.4 FREQUENCY OF FIXATIONS AS A FUNCTION OF BANDWIDTH

the scale of the abscissa. The lower curved line is the predicted frequency of observable fixations based on the notion of the zero order Markov process for transition probabilities in accord with Eq. (12). The upper points are the data obtained from the subject. The raw data are shown in Table 1.

There were 7639 frames of film read, or 636 seconds of film at 12 frames per second. Thus, if the subject had been working at peak load for 636 seconds at 2.5 fixations per second, there would have been 1590 fixations instead of 1422, but this difference is remarkably small considering the simplicity of the theoretical peak-load calculation.

Taking the various numbers of fixations on the five instruments which were in fact operating, we can calculate the fixation frequency (observable) for each of these as shown in Table 2.

The near constancy of the differences indicates that the surplus time which became available as a result of the failure of the one instrument was distributed more or less uniformly over the other five. The sum of the surplus is 1.14 looks per second, which compares very well with the required .96 for the missing instrument. Alternatively, 1.14 times 636 seconds equals 725 looks on these five, more than would have been expected on the basis of the theoretical full-load conditions. The deviant point for the .03 per second signal is not too surprising considering the relatively short period over which the data were collected. That is, 636 seconds is only about (636 x .03) 19 cycles long.

The results on observation duration are shown in Table 3. Equation (14) predicts that the duration of observation will be:

$$\left(\frac{1}{1-P_j}\right) \quad \left(K \log_2 \frac{A_j}{E_j} + C\right) \text{ seconds.} \qquad (55)$$

Where P_j is the probability that an instrument will be observed, and, in turn, is equal to $BW_j/\Sigma BW$, i.e., the relative frequency of the signal; A_j, and E_j are the mean-signal amplitude and fidelity criterion (or permissible error), respectively.

Table 1

BW cps	Number of Fixations
.03	225
.05	179
.12	264
.20	338
.32	357
.48	(59) but note discussion
Total Number of Fixations	1422

Table 2

Frequency of Fixation--Per Second

BW cps	Obtained	Theoretical	Difference fps
.03	.35	.06	+.29
.05	.28	.09	+.19
.12	.42	.20	+.22
.20	.53	.29	+.24
.32	.56	.36	+.20

(The theoretical values are calculated from Eq. (12).)

Table 3

Duration of Observation in Seconds

BW cps	Observed	Theoretical
.03	.30	.32
.05	.32	.33
.12	.37	.37*
.20	.43	.43
.32	.45	.55

* Anchor point for calculation.

(The theoretical values are calculated from Eq. (14).)

K is a constant of human information processing speed in seconds per bit; and C is a constant in seconds to cover movement time, etc. Since, by the nature of our experiment, we cannot explicitly state the ratio of A to E, we cannot hope to make an exact calculation to compare with the obtained values. However, we can assume that the central value is exactly in accord with theory and ask how deviant the others are. Figure 5 shows the result of this operation. However, since the subject is not in a fully-loaded state, and since we have no analysis at this time of the effects on duration of observation of surplus sampling, we can make no definitive statement as to the significance of the result. However, it is encouraging in that there is a monotonic increase in duration of sample with increasing bandwidth, and this, at least, is in accord with the theoretical predictions. Anchoring the equation on the observed data for a bandwidth of .12 cycles per second, we get the results shown in Table 3.

The large error for the .32 cycles per second (Table 3) may be the result of the large number of extra observations, or it may be that the theory does not hold out that far. Taking the values of observed duration and calculated probability of observation, as in Eq. (10), one can calculate the values for each bandwidth of the term

$$(K \log_2 \frac{A}{E}) + C \qquad (56)$$

See Table 4.

There is fair uniformity among the first four values; the fifth is deviant. We can achieve still further simplification by making the assumption that 5 bits per second is a reasonable value for human information processing in tasks of this sort. This makes K = .2 seconds per bit, and reduces Eq. (56) to:

$$.2 \log \frac{A}{E} + C \approx .3 \text{ seconds} \qquad (57)$$

or

$$.2H + C \approx .3 \text{ seconds} \qquad (58)$$

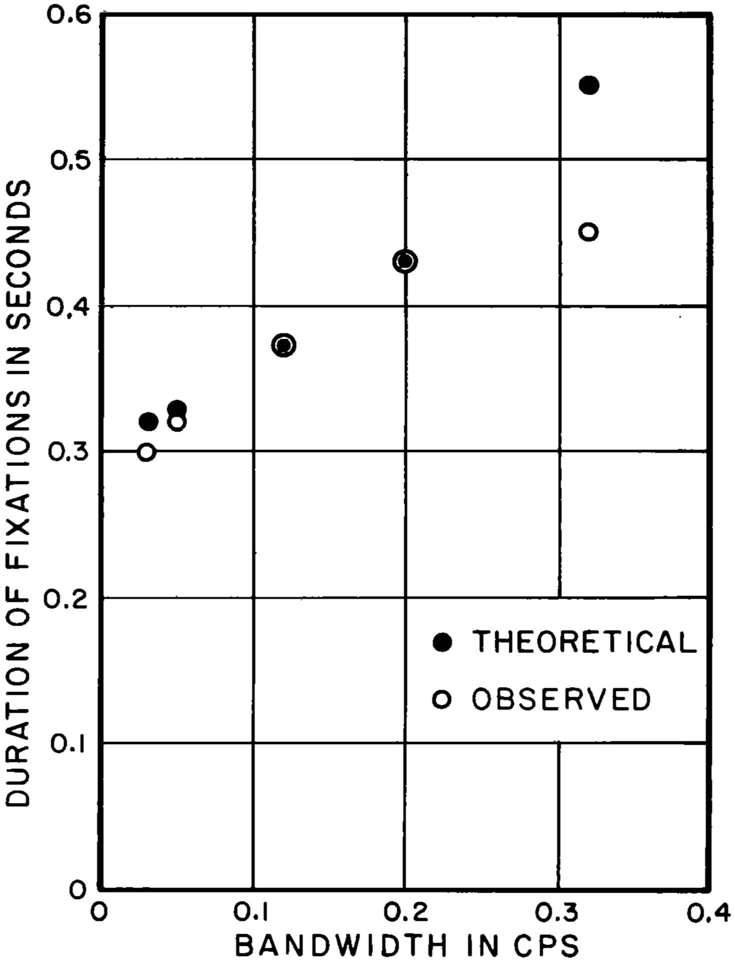


FIG.5 DURATION OF FIXATIONS AS A FUNCTION OF BANDWIDTH

Table 4

Value of $(K \log_2 \frac{A}{E}) + C$ in Seconds as a Function of Bandwidth

BW cps	$(K \log_2 \frac{A}{E})+C$
.03	.288
.05	.298
.12	.307
.20	.309
.32	.252

where H is the information taken in per observation, and the .2 is seconds per bit.

If C is equal to .1 second, then H per observation was 1 bit. If C is equal to 0 seconds, then H per observation was 1.5 bits.

Both of these values are reasonable and it suggests that there was a tendency for the observer to return to an instrument when the uncertainty about its reading reaches a constant level (except, of course, for the instrument which was presenting .32 cycles per second information).

Extinction of Observing Response

It is unfortunate that the sixth instrument failed during the data taking, but this inadvertent "experiment" did provide much that is interesting, confirmatory, and provocative.

The data on the sixth instrument (see Table 5) show what might be construed as an "extinction" curve of fixations as a function of time.

Apparently the subject did not instantly change her concept of the instrument as an information producer. Instead, the gradual reduction suggests that there was an expectation that something might happen which should be watched for. After ten minutes, this unsatisfied expectation apparently was extinguished. This result has implications for future work in this area. It will be recalled that in the 1953 study, I had trained my subjects for 30 hours; and I had observed that their detection rate did not asymptote until 20 hours of training had passed. I also observed that their sampling rates during the first hour were different only in minor detail from those at the end. The difference lay in the differentiation between the high and the low frequency signals. The lows were too often sampled and the highs too infrequently sampled at first, but the tendency disappeared in the course of five hours of training. In the future, it may be the case that we can use the same subjects and merely give them an hour of exposure to the new situation before taking data. It should be noted in this regard that all the early pilot eye-movement studies report different frequencies of fixation as a function of the state of the aircraft, the maneuver, and the external condition (i.e., day or night), and that these differences did not require extensive retraining in order to appear, but were, instead, a rapid

Table 5

Time (Minute)	Number of Fixations
1	20
2	12
3	9
4	5
5	6
6	2
7	3
8	1
9	1
10	0

adaptation to the circumstances. It is fairly obvious: the pilots looked at what they needed: to do what they had to do. In a sense, there was a drastic shift of the fidelity criterion for each instrument as a function of what the pilot was trying to do. Thus, for this one subject, although the circumstances of the actual recorded data did not correspond to those which were desired, there is some degree of general conformity to the theoretical predictions based on the report of 1963.

In the case of one of the remaining four subjects, there was a failure of the camera speed regulator such that the camera ran at a much higher speed. As a result, the absolute levels of the numbers obtained with this subject are not in accordance with those exhibited by the other subjects, and they will be presented separately. The slowness of operation was evidenced by the unnaturally long duration of blinks, far slower than those ordinarily exhibited by the subject. The remaining three subjects provided good data and these will be described below.

For these three subjects of Experiment 2, approximately 33 minutes of behaviour were studied. This involved the exposing of 300 feet of 8 millimeter film. Since there are 80 frames per foot of film, the data represent the results of analysis of 24,000 frames of film. The subjects made about 2 fixations per second at least so that there were a total of about 4,000 fixations identified and recorded. Since, the model being tested does not make predictions about variance, no calculations were made of anything other than the means of the distributions. Since the data are in digital form, additional calculations can be made cheaply at a later date. Thus, if it is desired at some other time to test other models of observing behaviour such as those proposed in the earlier portions of this report, it can be done.

Fixation Frequency

Table 6 presents the data on frequency on fixation. These were obtained from the films as previously explained. The means are formed only from data of subjects 2, 3, and 5, for reasons mentioned earlier. The right hand column has the sums of the fixation frequencies for subjects 1, 2, 3, and 5. (Subject 4 is eliminated because of the uncertainty about the actual values.) The means have been corrected in accord with Eq. (12) and appear at the bottoms of the columns. The same data are plotted in Fig. 6. The frequency of fixation is shown as a function of bandwidth of signal being monitored. The data are for subjects 2, 3, and 5 only. The solid line is that predicted by simple sampling theory.

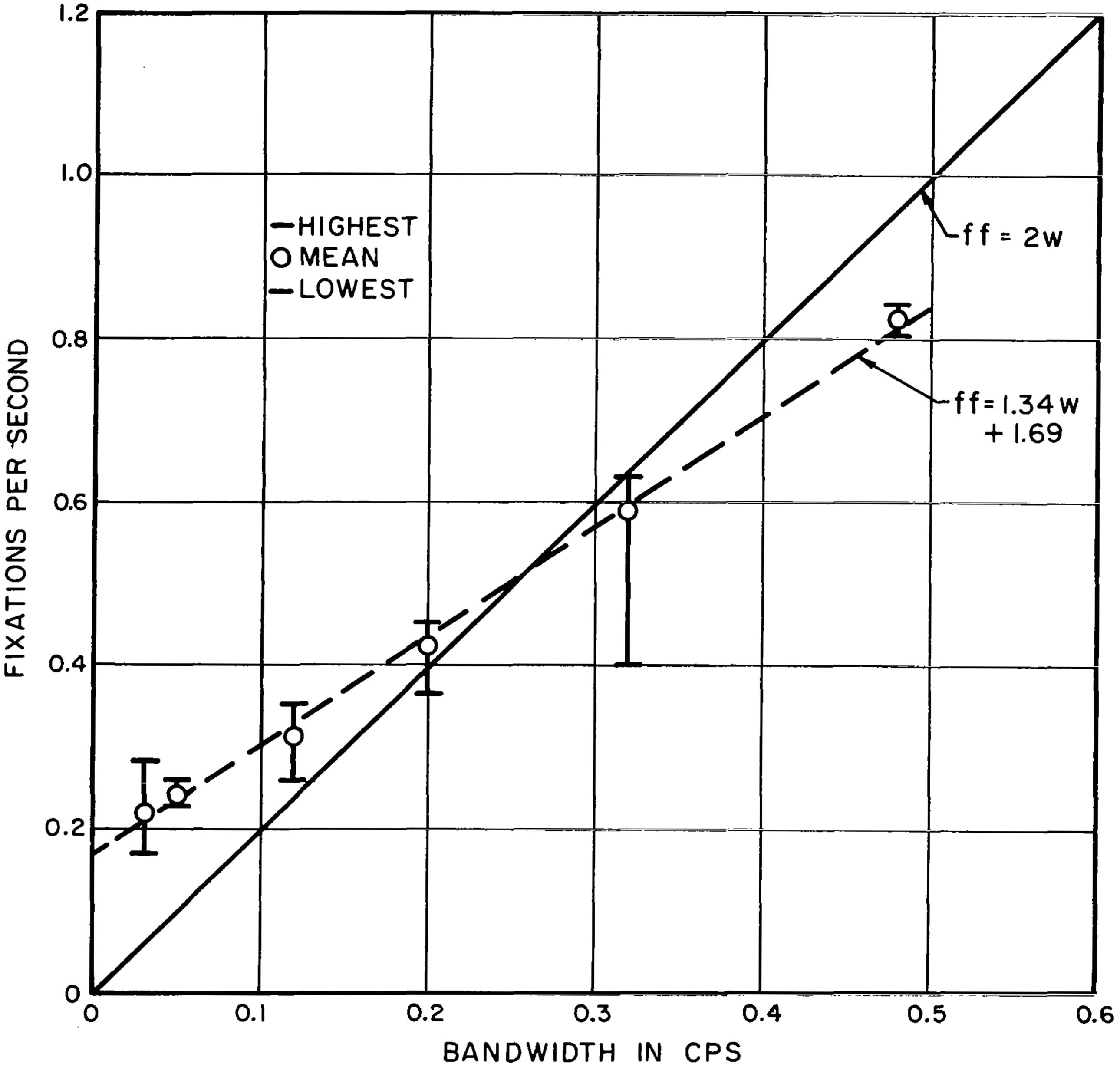


FIG. 6 EXPERIMENT II: FREQUENCY OF FIXATION VERSUS BANDWIDTH IN CPS

Table 6

Experiment 2

Frequency of Fixation versus Bandwidth in cps

							Σ
Bandwidth cps	.48	.32	.20	.12	.05	.03	1.20
Subject 1		.501	.507	.391	.327	.375	2.101
Subject 2	.495	.448	.377	.298	.252	.172	2.042
Subject 3	.505	.393	.302	.320	.239	.278	2.037
Subject 4	.274	.238	.238	.215	.167	.116	*
Subject 5	.483	.463	.379	.235	.218	.196	1.974
$s_2+s_3+s_5$	1.483	1.304	1.058	.853	.709	.646	
Mean (2,3,5)	.494	.435	.353	.284	.236	.215	sum (2,3,5) 2.018
Corrected Means	.823	.596	.425	.316	.246	.221	

* See Text

The data show a monotonic increase of fixation frequency with increasing bandwidth of signal. However, the slope of the fitted line, 1.34, is less than the theoretical value of 2.00. The discrepancy arises as a result of the oversampling at the low and the undersampling at the high end of the function.

There are many possible reasons for this kind of behaviour and clarification will depend on further experimentation. Three hypotheses might be advanced to explain the discrepancy. Firstly, the signal made up of sums of sine waves appeared to be more predictable than would be the case for a true random function. Secondly, for the very low frequency signals the theoretical intervals between samples would have been 15 seconds and 10 seconds approximately for the lowest and the next-to-lowest bandwidth. The drawing of attention to a signal may be related to the increasing uncertainty of the observer about the present state of the signal based on the last reading. In addition, there would be an increase in the uncertainty of the observer as to the nature of the last reading itself; i.e., forgetting would occur even during the short intervals between readings since the information is retained only in short-term "storage" soon to be supplanted by a new value. This increase in entropy within the observer, when added to the increase in entropy external to the observer, increases the total rate at which uncertainty grows, and will have to be included in any comprehensive theoretical formulation. Thirdly, the variances of the signals were probably not absolutely equated. It was not recognized in the most early simple formulation of the theory that this would be a highly significant factor. However, during the course of this program itself, more extended theorizing about the nature of the behaviour of the operator has led to the conclusion that the ratio of the significant value to the signal variance is the most important determiner of fixation frequency. Thus, if the variance of the high frequency signal is somewhat lower than it should be, the probability that this signal will exceed the limit as a function of time from the previous reading, is lower than would be the case had its variance been greater. Similarly, slightly larger variances of the low frequency signals would increase the probability that they would exceed the limit in a shorter length of time, and, therefore, require more frequent observation. Thus, the experiment (like the other experiments in this group) is attempting to test a theory which itself has been supplanted by a more advanced and comprehensive approach. The conformity of the earlier data of 1954 to the theory of that time, which was even more constrained, can only be the result of a nearly absolute equality of signal power fed to the instruments being monitored by the subjects.

Thus the results indicate that the equations must take into account internal increases in uncertainty as well as external. The earlier research of reference (15), used frequencies of .64, .32, .16, and .08 cycles per second bandwidth. There was a slight tendency to oversample the lowest bandwidth but the extra sampling which was done was not sufficient to diminish the amount of sampling done on the highest bandwidth signal. As a result the best fitting line for those data had a slope of 2.4 for the five subjects taken in aggregate. A test of the explanation based on the predictability of the signals was performed and is reported here as Experiment 6. The results strongly support the idea that the use of signals from a random function generator elicit different behaviour from the observers.

Fixation Duration

Table 7 presents the data on duration of fixation. The means are formed only from the data of subjects 2, 3, and 5 for reasons earlier given. At the bottom of the columns are the means corrected in accord with Eq. (14). The corrected data are plotted as a function of signal bandwidth in Fig. A line fitted by least squares is drawn in. The theoretical line should have a slope of 0, and an intercept based on the information acceptance rates of the subjects. In general, an intercept of .40 seconds may be considered in accord with past experience, although the exact value is not too important except to provide a reasonable visual anchor for the comparison to be made between theory and data.

The durations of fixation on the higher frequency signals are too low. One would have expected to find longer fixations based on the frequency data, assuming that the uncertainty of the subject about the value of the signal displayed would be greater since the samples were taken further apart in time. However, the predictability of the signals apparently permitted shorter samples to be taken. Again, the data of Experiment 6 will clarify this matter.

TABLE 7

Experiment 2

Duration of Fixation versus Bandwidth in Seconds/Look

(with Bandwidth in cps)

Bandwidth cps	.48	.32	.20	.12	.05	.05
Subject 1	.33 see text	.55	.46	.44	.35	.40
Subject 2	.58	.43	.44	.49	.49	.43
Subject 3	.48	.44	.53	.45	.47	.38
Subject 4	.80 see text	.75	.65	.64	.69	.66
Subject 5	.42	.50	.49	.45	.48	.44
$s_2+s_3+s_5$	1.48	1.37	1.46	1.39	1.44	1.25
Mean (2,3,5)	.49	.46	.49	.46	.48	.42
Corrected Mean	.29	.34	.41	.41	.46	.41

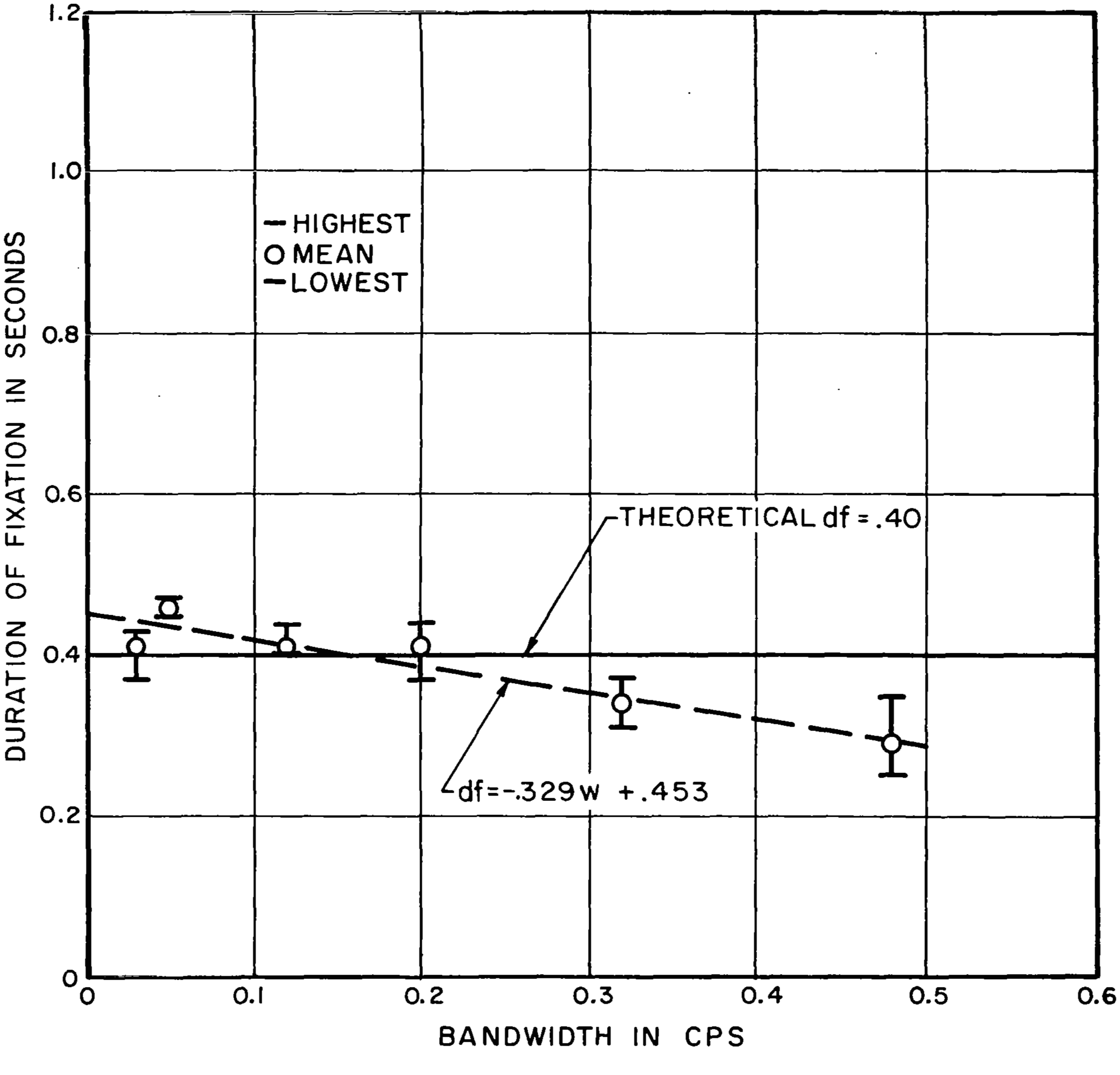


FIG. 7 EXPERIMENT II: DURATION OF FIXATION VERSUS BANDWIDTH IN SECONDS/LOOK

Experiment 5: Correlated Signals

Experiment 5 was next performed. Five subjects were trained with the same arrangement of dials in the panel as was described in Experiment 2, but with the following difference in the signals presented: Channel 4, which ordinarily carried a signal of .12 cycles per second, was furnished with a signal composed of the .20 signal multiplied by a gain of .707 added to the .12 signal multiplied by a gain of .707. Thus, there existed a correlation of .707 between this new signal and the .20 cycles per second signal. The new signal consisted of the .20 cycles per second bandwidth signal with a .12 cps signal superimposed upon it, and with overall power equated to that of the other signals.

It was clear from preliminary trials that no easily detectable changes occurred as a result of the correlation between signals. This result is not too surprising since prior research has shown that for two indicators side by side and moving in one degree of freedom, the absolute threshold for the perception of correlation between them is of the order of .4 to .5. Thus, these signals, which were correlated .7, and imbedded in a matrix of other signals, would be fairly unlikely to be detected by the subjects as being related. At the same time that the new group of subjects was.being trained with a .7 correlation between signals, the original group of highly trained subjects was being trained with a correlation of .9 between the same pair of signals.

The study was then carried out using these highly-trained subjects who had performed in Experiment 2. In this study the .20 cycles per second signal was multiplied by .9, and added to the .12 signal which had been multiplied by .44. Thus the resultant signal consisted of approximately 81 per cent of a .20 cycles per second bandwidth signal, and 19 per cent of a .12 cycles per second bandwidth signal. The coefficient of correlation between this new signal and the .20 cycles per second signal was .9.

Results: This experiment involved the reading and identification of approximately 42,000 frames of film recorded from 3,500 seconds of behaviour of the five subjects in aggregate. The results are summarized in Tables 8 and 9.

Table 8 presents the data on frequency of fixation as a function of signal bandwidth and Table 9 presents the data on duration. The data are plotted in Figs. 8 and 9 respectively.

TABLE 8

Experiment 5

Frequency of Fixation vs. Bandwidth

Bandwidth cps	.32	.20	.48	.12 .20	.05	.03	Σ
Subject 1	.237	.254	.363	.160	.175	.206	1.395
Subject 2	.340	.334	.337	.175	.202	.166	1.554
Subject 3	.326	.202	.372	.220	.083	.085	1.288
Subject 4	.451	.462	.553	.363	.172	.095	2.096
Subject 5	.385	.252	.360	.187	.176	.249	1.609
Σ	1.739	1.504	1.985	1.105	.808	.801	
Mean	.348	.301	.397	.221	.162	.160	1.589
Mean Corrected by Eq. (12) Theoretical 1	.477	.363	.662	.246	.169	.164	
Theoretical 2	.464	.357	.635	.262	.169	.164	

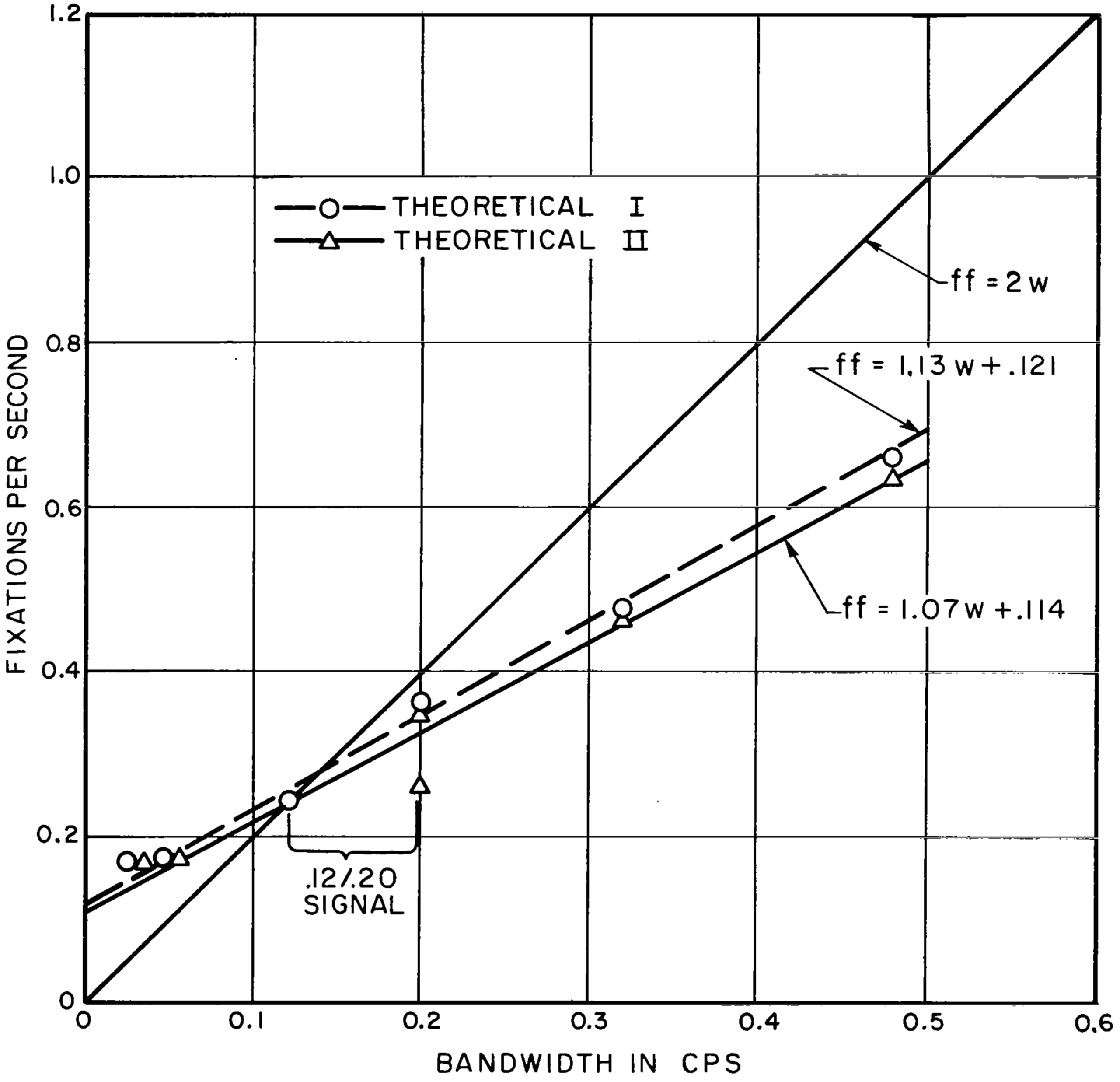


FIG. 8 EXPERIMENT V: FREQUENCY OF FIXATION VERSUS BANDWIDTH

TABLE 9

Experiment 5

Duration of Fixation Seconds vs. Bandwidth

Bandwidth cps	.32	.20	.48	$\frac{.12}{.20}$	.05	.03
Subject 1	.86	.77	.71	.49	.58	.49
Subject 2	.72	.73	.62	.44	.50	.48
Subject 3	.74	.99	.77	.50	.70	.54
Subject 4	.47	.59	.47	.32	.41	.47
Subject 5	.63	.70	.70	.43	.51	.52
Σ	3.42	3.78	3.27	2.18	2.70	2.50
Mean	.68	.76	.65	.44	.54	.50
Mean Corrected by Eq. (14) I	.50	.63	.39	.37	.52	.49
II	.51	.64	.41	.37	.52	.49

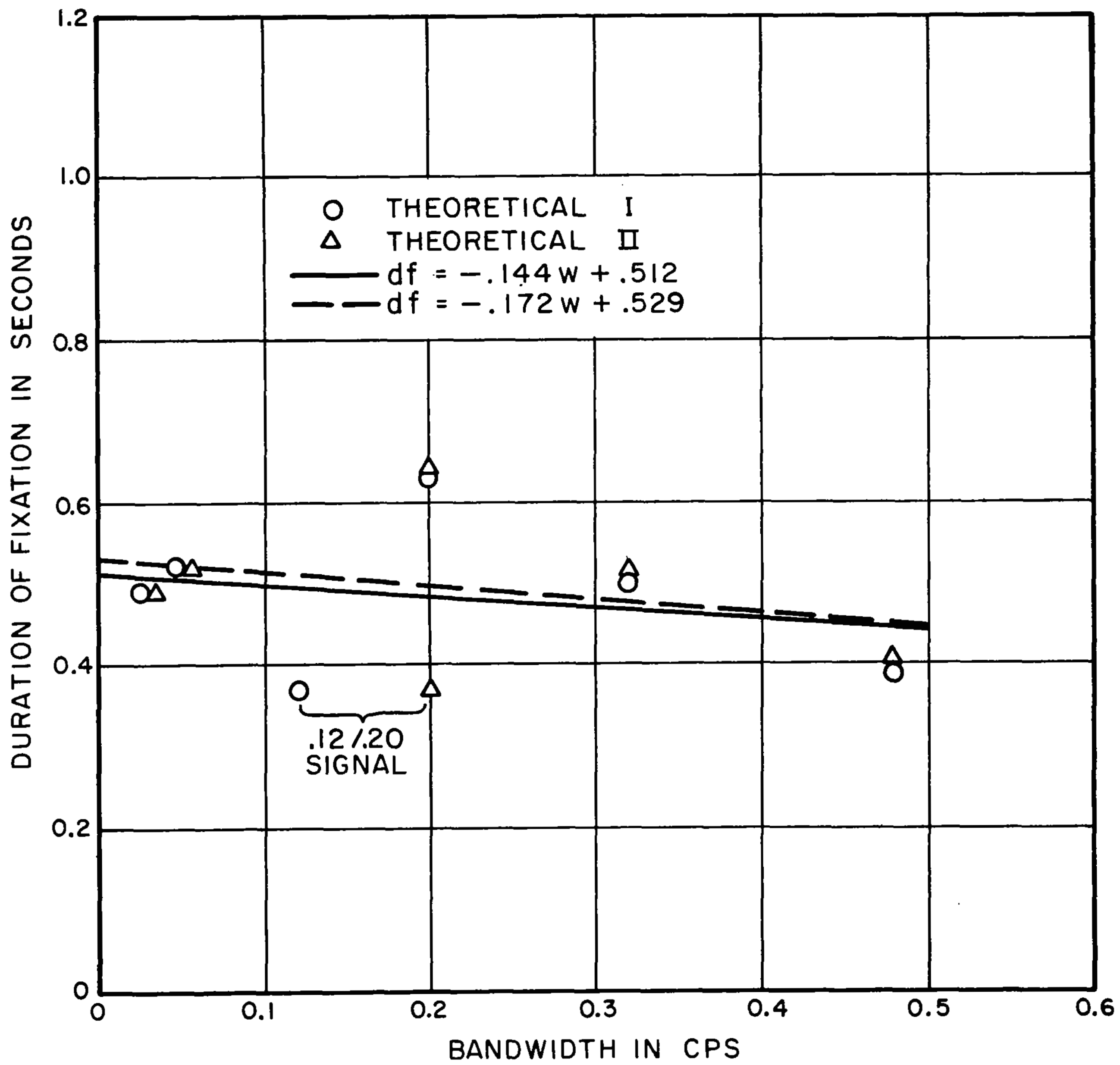


FIG. 9 EXPERIMENT V: DURATION OF FIXATION IN SECONDS VERSUS BANDWIDTH

Frequency of Fixation

Figure 8 shows the obtained frequencies of fixation as a function of the signal bandwidth. The results are not markedly different from those on Experiment 2 in that there is a strong tendency to oversample the lesser bandwidth signals and to oversample the greater bandwidth signals. What does appear to be clear is that the data for the signal composed of a mixture of the .20 cycles per second and the .12 cycles per second are appropriate for .12 cycles per second and not for .20 cycles per second.

The plot shows two points for each of the signal bandwidths. These are the consequence of the fact that the application of Eq. (12) to correct the observed fixation frequencies gives a result that depends on "p", and "p" in turn is a function of the frequencies which the subjects are monitoring. There are two different ways of dealing with the mixed signal: it can be considered to be a .12 cycles per second bandwidth signal or it can be considered to be a second .20 cycles per second bandwidth signal. The values of the corrected sampling frequency for each assumption are plotted for each bandwidth in this figure. The best fitting straight lines are labeled as theoretical I and theoretical II, corresponding to the two ways of considering the signal.

Figure 9 shows the corrected mean duration of observation plotted as a function of signal bandwidth. The theoretical line would have a slope of 1.00 and an intercept which is a function of the information processing rate of the observers in accord with Eq. (14). The data for the two correlated signals are deviant. The durations are excessively long for the .20 cycles per second bandwidth signal and short for the .12/.20 cycles per second bandwidth signal. The mean of the times spent on these two signals is approximately equal to the best fitting line at either bandwidth. Thus, taken together, the total time spent on these two signals is what it would have been if the .20 cps bandwidth and the .12 cps bandwidth signal had been independently presented.

Let us consider what this instrument is and how it might be treated. If an observer were to look at the .20 cycles per second bandwidth instrument first and then look at the "correlated" instrument, and if he were aware of the correlation, he would have a smaller range of possible values within which to make his observations; ie., given a reading of 30 microamperes on "Instrument 5", the reading on "Instrument 4" would consist of 27 microamperes plus or minus 44 per cent of some other random number whose mean value is zero. Thus, the possible range within which the pointer of that instrument can fall is reduced. This is analogous to a reduction in the mean-square power of the signal which has to be read. It would appear that this factor probably is operating here in that the subjects did observe the signal far less frequently than would be required for a .20 cycles per second bandwidth signal. In fact, they observed it almost exactly as often as would be required for a .12 cycles per second instrument, but observed it for a shorter time--presumably because of an ability to utilize the information obtained through the correlation with the other signal. They treated it as a signal whose bandwidth were .12 and whose entropy or uncertainty was less than would otherwise be appropriate for such an instrument. Otherwise, the results of this study are largely in accord with the results of Experiment 2.

We do not necessarily expect that these same results will be obtained with experienced and knowledgeable pilots controlling vehicles. In the first place, the nature of the coupling of these two signals, a synthetic one, is a far less meaningful one than the intrinsic coupling which exists between the various degrees of freedom of an aircraft or a space vehicle. In this experimental situation there was no preferred order of reading. The observer might well look at the mixed instrument before looking at the .20 cycles per second bandwidth instrument, and the frequency with which he would look at one or the other presumably was determined by some sort of random or quasi-random Markov process. In the space vehicle, on the other hand, the detection of a limit indication on one indicator will give rise to a series of attentive acts based on the coupling or the physical connection between that indicator and the other instruments which contribute to that particular indication. Thus, we would expect, on a moment-to-moment basis, that a pilot will make use of the intrinsic correlations existing within the vehicle and will, therefore, alter his sequences of looking and his patterns of scanning as a function of the instantaneous values which he reads.

In summary, presenting a signal which is highly correlated with another signal does, in fact, cause the total amount of attention given to the two signals to be less than would be the case if they were independent. The two correlated signals in question were looked at for a total of 2437 frames. Had they been independent signals of .20 cycles per second bandwidth, they would have been looked at approximately 2618 frames. The difference is difficult to evaluate. If the .12/.20 signal is treated as a signal with a bandwidth of .20 cycles per second, then there was a clear reduction of 500 frames or about 40 per cent below the theoretical level.

On the other hand, there was also an apparent oversampling of 319 frames on the .20 cps signal with which it was correlated, or an apparent "over-attending" of about 25 per cent.

On the other hand, one can more easily accept the idea that the instrument was treated as if it had a bandwidth of .12 cycles per second. Thus, the residual uncertainty of this signal had a bandwidth of .12 cycles per second, and apparently the signal was sampled on that basis with a slight reduction in duration of observation as shown in Fig. 9 .

Perhaps the most general conclusion is that even a correlation of .9 between two signals does not lead to a marked reduction in the attentional demand imposed by the two signals. Even though the observer samples about as often as he should as an ideal sampler, and in some cases somewhat more often, it may be that he is unaware of the correlation between the signals because of the time interval between his observation of signal i and signal j. This was true for four of the five subjects.

Subject 1, who contributes the lowest end of the range of the readings on the ".12/.20 signal", detected the existence of the correlation during his first exposure to it. On the other hand, the remainder of his data are not strikingly at variance with those of the other subjects; nor is there any consistent pattern which comes to light. Subject 4 had data which very nearly corresponded to theory at the high end and at the very bottom, but markedly oversampled in the middle range. Subject 3 corresponded very well at the very low end with marked under-sampling at the high end. It is not clear whether these deviations and individual differences are the result of random variables, of unidentified individual characteristics, or of the fact that each subject's data came from different segments of the signals.

The interpretations of Experiment 2 and of Experiment 5 must be considered in the light of the more extended theoretical ideas which have been presented in Part II. We must expect that relatively small differences in the signals will make large differences in the frequency with which they will be monitored. In these experiments, although an effort was made to equate the powers of the signals to some small error, it was not considered necessary to make them absolutely the same by all possible tests. As a result, there is a very strong possibility that the powers of some of the signals may have differed significantly from those of the other signals;

and this is offered as a possible explanation for the depression at the high end of the scale and the elevation at the low end of the scale of the observers' actual behaviour compared to that predicted by theory.

There are alternative explanations. For the two smallest bandwidth signals, the intervals between observations which would be expected from an ideal sampling system would probably be reduced because of failures of short-term memory. If the monitor forgets, then there are two sources of uncertainty about whether the signal would exceed the limit. One of these, of course, is that the signal itself generates uncertainty during the time that it is not observed. The other is increasing uncertainty in the subject about what he last saw. We suggest that: (a) when the probability that the signal exceeds or approaches the arbitrary limit is greater than some threshold, then the observer will look; and (b) as the memory trace fades, the uncertainty of the position of the needle at the time of the last observation contributes to this growth of uncertainty and to the growth of the probability that the signal may, in fact, have approached the limit. Thus, the interval between observations of a low-frequency signal would be expected to be shorter, and we would expect to find a reduced frequency of observation on the high-frequency signals since the calculations to be made are relative to one another within the overall capability of the observer. If the observer is fully loaded, then he must under-sample the high-frequency signals and over-sample the low. In addition, since the ability of the observer to detect velocity is also a direct function of the magnitude of velocity, we would expect that for rapidly moving signals more information will be gotten per observation with a resulting diminution of required frequency of observation. It has been shown by Vogel (23) that the taking of two simultaneous derivatives of a signal reduces by half the number of samples to be taken over-all. For the human observer who at times will take note of and utilize velocity information, the frequency of observation may be decreased. For very slow-moving signals, there will be many occasions when the signal will have a very small derivative. For the short period of fixation available, this derivative may be below the observer's threshold.

Experiment 4: An Empirical Investigation into the Effects on Observing Behaviour of Dichotomization of a Continuous Signal

Experiment 4 involved the substitution of a discrete indication for one of the continuous indications of the preceding experiments. The notion underlying this empirical investigation was simply to find out how the behaviour of observers confronted with a dichotomized signal would be altered as compared with their behaviour when confronted with the same signal in its continuous form.

There are at least two quite opposite initial hypotheses. The first might be: if the observers, when the entire signal is available to them, make use of derivative information in estimating the probability that the signal will exceed the critical level, then the dichotomization would eliminate this type of information and make it necessary for the observers to sample more often than would otherwise be the case. Thus, we would expect an elevation of the sampling frequency for this instrument over that previously exhibited. A second hypothesis, and one which is equally tenable, is that if an observer ordinarily takes "count" only of the position of the indicator, and notes that it is above or below the critical level without regard for its velocity, then the dichotomization of the signal into 0's and 1's (i.e., above the critical level) will facilitate his observational process and require at the very most the same number of fixations, and perhaps fewer. One would also expect a reduction in the duration of fixation required in both cases. It is possible that some observers deal with such dichotomized instruments in one way, and others deal with them in the other way.

The experimental procedure was the same as described earlier. The subjects were those used in the earlier experiments and had become well practiced in the observation of the six-dial matrix. The bandwidths used were the same as before: i.e., .03, .05, .12, .20, .32, and .48 cycles per second. The .12 bandwidth was the one chosen for dichotomization. The signals with the larger bandwidths appear to have little room in which to demand a higher frequency of observations, except at great expense to the other indicators of the system. The very small bandwidth dials apparently are already being largely over-sampled so that it would be difficult to obtain any estimate of the effect of the dichotomization upon observing behaviour. The previously observed behaviour toward the .12 cycles per second bandwidth signals lies on, or close to, the theoretical functions, and we might, therefore, expect it to be easier to detect departures from this.

The subjects were presented with the new signal condition for a series of training trials, each training session being one hour long. The behaviour of the dichotomized signal was explained to them as being a new signal which required the same kind of detection as the other five signals (all continuous) of the same experiment.

After five days of practice with one hour of observing per day under these new conditions, the data were taken for ten-minute runs on each of the five subjects. The films were read as earlier described. The data are taken from some 40,204 frames of film. These, in turn, can be reduced to approximately 6,100 fixations of the subjects' eyes. The tables and figures present the data in summary form.

Frequency of Fixation

Table 10 presents the data on frequency of fixation for each of the five subjects for each of the six bandwidths. The lower figure in each cell is the value as corrected by Eq. (12). These data are plotted in Fig. 10 as a function of bandwidth of signal in cycles per second. The individual subjects' data are identified by number. Subject 3 looked more often at the dichotomized .12 cps bandwidth signal than at the .20 cps or the .05 cps bandwidth signals. The reverse is true for Subjects 1, 2, and 5. Subject 4 exhibited less regular behaviour than any of the other subjects in this experiment and the reasons for this are not clear. The percentage variability of the data on the .12 cycles per second bandwidth signal is greater than that for any other signal and this may be the result of the adoption by subjects of individual attitudes toward the dichotomized signal. The subjects taken as a group showed a reduction of fixation frequency on this signal.

Duration of Fixation

Table 11 presents the data on duration of fixations for each of the five subjects for each of the five bandwidths. The lower figure in each cell is the value of duration as corrected by Eq. (14). These data are plotted in Fig. 11 as a function of bandwidth of signal in cycles per second. The individual subjects are identified by number. The subjects as a group, as well as individually, show no deviant behaviour toward the dichotomized signal. Thus there was no compensation for the reduction in fixation frequency shown in the data. The best fitting line is shown in the figure based on the means of the corrected durations. The slope is .002 as compared with a theoretical slope of zero.

TABLE 10

Experiment 4

Frequency of Fixation vs. Bandwidth in cps

Bandwidth cps	.48	.32	.20	.12	.05	.03	Σ
Subject 1	.412	.350	.343	.128	.262	.245	1.740
Corrected	.687	.480	.413	.142	.274	.251	
Subject 2	.579	.469	.279	.116	.208	.191	1.842
Corrected	.965	.643	.336	.129	.217	.196	
Subject 3	.532	.339	.344	.411	.273	.263	2.162
Corrected	.887	.464	.414	.457	.285	.270	
Subject 4	.438	.488	.156	.175	.204	.063	1.524
Corrected	.730	.669	.188	.195	.213	.065	
Subject 5	.413	.410	.269	.285	.242	.233	1.852
Corrected	.688	.562	.324	.206	.253	.239	
Mean	.475	.411	.278	.223	.238	.199	
Mean Corrected by Eq. (12)	.792	.563	.335	.248	.249	.204	

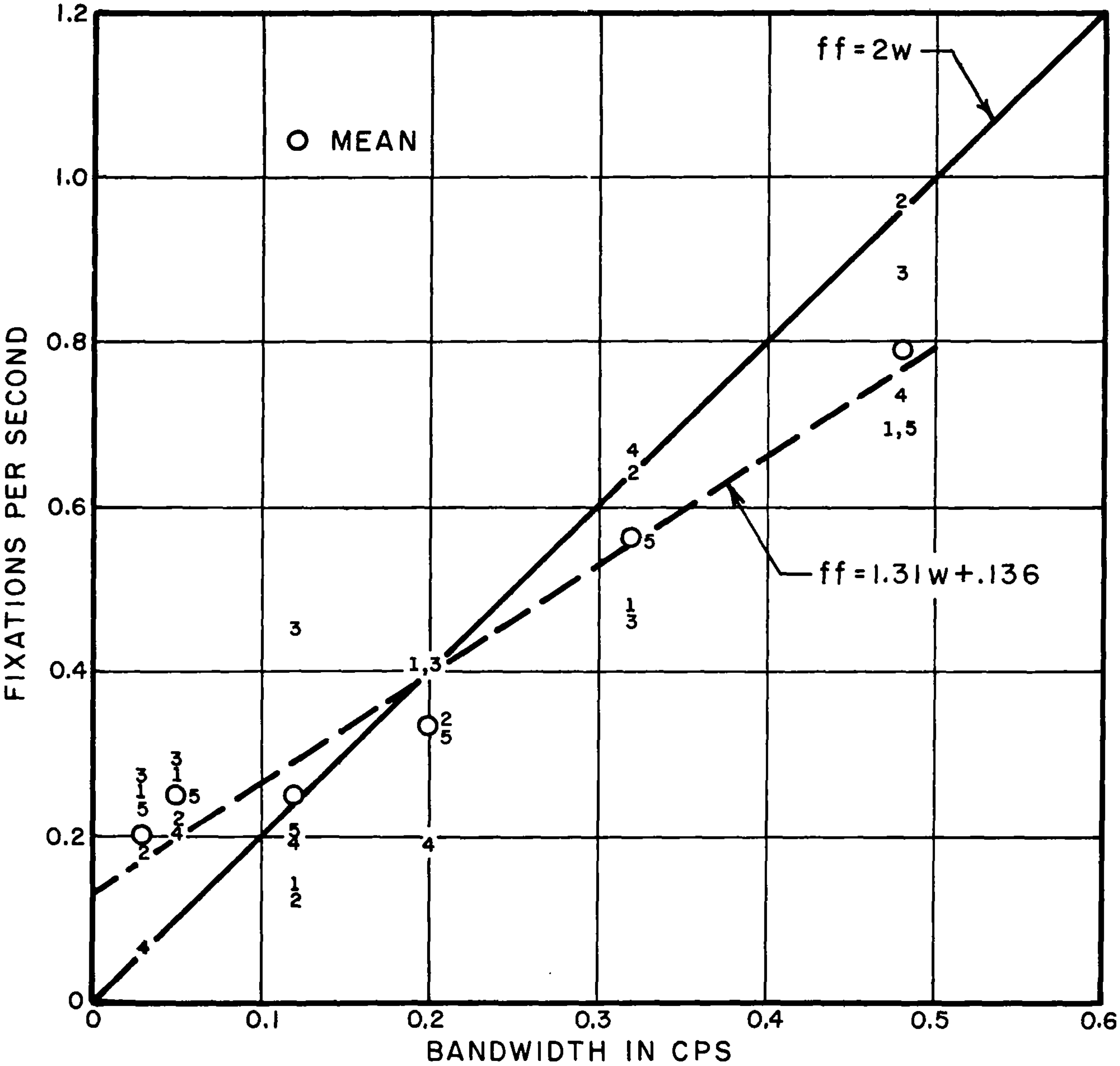


FIG. 10 EXPERIMENT IV: FREQUENCY OF FIXATION VERSUS BANDWIDTH

TABLE 11

Experiment 4

Duration of Fixation in Seconds vs. Bandwidth in cps

Bandwidth cps	.48	.32	.20	.12	.05	.03
Subject 1	.72	.59	.56	.44	.54	.40
Corrected	.43	.43	.46	.40	.52	.39
Subject 2	.68	.58	.35	.55	.40	.41
Corrected	.41	.42	.29	.49	.38	.40
Subject 3	.42	.54	.54	.34	.42	.51
Corrected	.25	.39	.45	.31	.40	.50
Subject 4	.81	.65	.44	.45	.55	.45
Corrected	.49	.47	.37	.41	.53	.44
Subject 5	.55	.48	.64	.55	.49	.42
Corrected	.33	.35	.53	.49	.47	.41
Σ	3.18	2.84	2.53	2.33	2.40	2.19
Mean	.64	.57	.51	.47	.48	.44
Mean Corrected by Eq. (14)	.38	.42	.42	.42	.46	.43

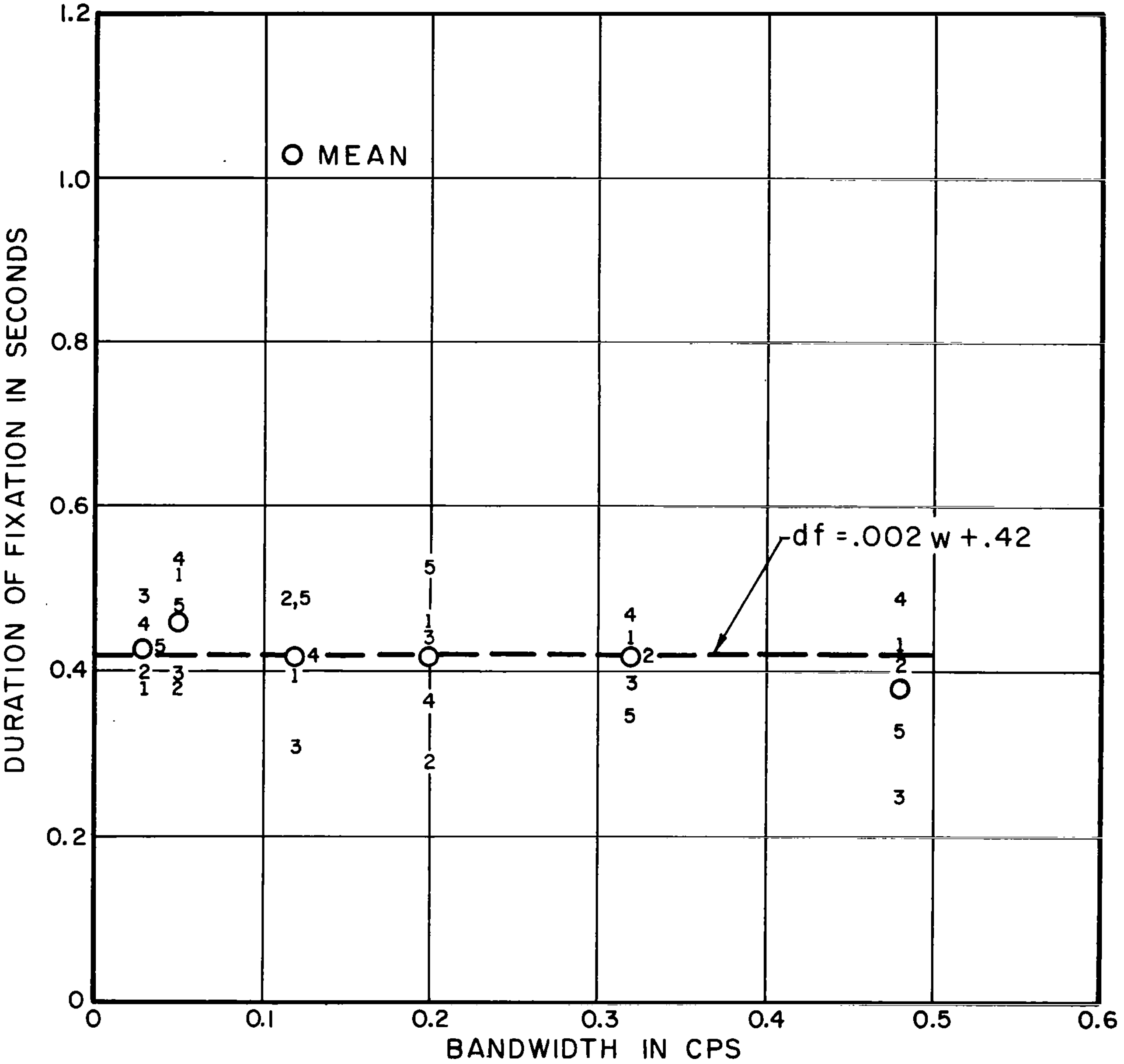


FIG. 11 EXPERIMENT IV: DURATION OF FIXATION IN SECONDS VERSUS BANDWIDTH IN CPS

The results seem to indicate that the use of a dichotomized presentation of a continuous signal somewhat reduces the workload associated with that portion of the monitoring task but that this is not always the case for all subjects. Whether specific instructions about the nature of the task associated with this instrument would have produced a more consistent pattern of behaviour must wait for other experiments, designed with that goal, to answer. For the analysis of a real system, one could well take the position that there would be no great error if one were to treat the dichotomized signal as if it were continuosuly presented. Of course, if the signal were to be transformed into another form which was capable of attracting the attention even in peripheral vision, the required fixation frequency would probably be reduced, at the cost of deleting the actual quantity by which the signal exceeded the limit, as was done here.

Experiment 1: The Relation of Required Accuracy of Reading and Duration of Fixations

Experiment 1 was aimed at the problem of determining whether, for a given signal bandwidth, a change in required accuracy would result in a change of duration of fixation independent of frequency of fixation.

We had assumed that an alteration of the "critical limit" for a particular signal would be tantamount to an alteration of the required accuracy. However, the theoretical work which was undertaken early in the program made it clear that although there would, indeed, be marked modifications of behaviour with alterations of the critical limit, but that these alterations of behaviour would no wise correspond to those predicted on the basis of simple sampling theory. Thus, the experiments, besides being difficult (if not impossible) to do, also had little to offer since they had been outstripped by theory before they had been performed.

The original idea was that if the required relative accuracy were reduced, there would be a relatively smaller reading time required. This hypothesis originates in earlier psychological work relating to stimulus information and response latency. The theoretical work of Part II suggests that for the human observer the significant factors are the size of the standard deviation of the signal and the number of standard deviations away from the mean that the critical limit is; and that the relationship between sampling interval and the duration of sample on the one hand, and distance of limit from mean, on the other hand, is not simple and is not the one originally suggested by the simple theory.

There remained also the question of how the experiment was to be performed. Since a simple alteration of the "Z Score" of the critical limit would not do, the only way in which one could test the original notion would be to require the observers actually to read the value of the instrument upon each observation, and to vary for one or more instruments the required accuracy with which the needle must be read upon each observation. Thus, one might require one signal to be read with an error no greater than ± 5 and another with an error no greater than ± 1, and require the observer to call out the reading in accord with these limits each time that the instrument was fixated. However, the speed with which the eye movements occurred would have required an exceedingly rapid and long stream of speech and it was clear that the task would have been too difficult for our subjects to handle. The times required to make each reading and to emit it vocally would have so reduced our sampling rates as to overload the monitors.

We attempted to perform the experiment by changing the number of scale marks, and requiring the monitor to interpolate in order to make his decision as to whether the signal had exceeded the limit. Thus, if the marks on one set of dial faces were separated by 25 spaces and those on another set separated by 1 space, it might be expected that more time would be taken in making a reading, to the same accuracy, on the former than on the latter. The experiment was combined with Experiment 3 since it was virtually impossible to separate the two in practice.

Experiment 3: The Effect of Simultaneous Variation of Required Accuracy of Reading and Signal Bandwidth on Duration and Accuracy of Reading

Experiment 3 was intended as a demonstration that the separate parts of the monitoring process: frequency of fixation, and duration of fixation, which would be separately studied in Experiments 1 and 2, could be estimated from a knowledge of both the bandwidth of the signal being monitored and the required accuracy of reading to be made by the monitor. However, Experiment 1 was difficult (if not impossible) to do in its original form. As a consequence, Experiment 3 had far less point than when it was originally formulated.

It was decided that the two experiments: 1 and 3 might be combined and run as a single study.

The experiment was run in a manner similar to those earlier described. Only four signals were presented to the four subjects available from the trained group used in the earlier studies. Two of these signals had bandwidths of .32 cycles per second, uncorrelated. The other two had bandwidths of .16 cycles per second, uncorrelated. The subjects were trained for 5 hours to monitor the new signals and were recorded on the fifth day. Two of the dial faces presented to the subjects were masked by new scales. The new scales were marked only at 0, 25, and 50 microamperes. The task of the subjects was identical to that described earlier: to report by a switch closure whenever a signal exceeded the limit of 40 microamperes.

Approximately 30,000 frames of film were read to obtain the data.

Results

The mean frequencies of fixation of the masked dials were less than those for the unmasked dials for both bandwidths. The durations of fixation show no difference between masked and unmasked dials. The experiment did not succeed in its purpose nor did the results provide any insights into the complexities of monitoring behaviour.

It is difficult to evaluate the reasons for this discrepancy. One would have expected the frequency of observation for the higher frequency dial to be much higher than for the low frequency dial, whereas the difference is, in fact, quite small. Similarly, one would have expected that the durations of observations for the masked dials would have been longer than those for the unmasked dials. The subjective reports of the subjects indicated that they did, indeed, think that this was the case. However, the data (as read and analyzed by computer) indicate that this was, in fact, not the case. A more extensive analysis or additional experiments will be required to solve the problem if, indeed, there is a solution to be had. It may be that the very earliest efforts of the subjects to read the masked dials were less efficient and that the fixations were, in fact, longer then with experience, the subjects' strategies might have altered so as to reduce or eliminate the difference (and perhaps even reversed it).

There has long been controversy as to the effect of the number of scale marks on the ease of reading of instruments.

It is possible that the instruments which we were using, which possess marks at two microampere intervals, were not maximally easy to read. The masked and the unmasked dials might have been equally difficult to read.

There was perhaps an additional factor which operated to prevent our obtaining interesting differences between conditions on this experiment. The problem of motivation in an experiment is a difficult one, nor does cash payment always substitute for basic interest. The subjects had been engaged in dial reading nearly every day for approximately two months and very possibly had exceeded their limits of willingness to monitor a set of relatively meaningless indicators.

It is perhaps ultimately the case that this experiment can be validated only in an operational or quasi-operational, situation wherein a pilot flying a vehicle can be given instructions requiring him to report certain data at various times within the flight. Then the required accuracy of reporting could be varied in a meaningful, mission-related way and the behaviour toward that instrument examined. This would not be too unlike the activity of a test pilot who makes oral reports of data read from his instrument panel. A task could be so set up as to be meaningful and well within the range of tasks to which a pilot is accustomed. The totally synthetic laboratory situation obviously lacks motivating power and lacks that degree of meaningfulness which is almost essential to a highly-efficient performance.

Experiment 6: A Check of the Hypothesis that the Characteristics of the Signals Used Affected the Results Obtained in Experiments 2, 4, and 5

It will be recalled that the signals which were used in Experiments 2, 4, and 5 consisted of pseudo-random time functions composed of a very large number of sine waves recorded onto the same tape track. The amplitude distribution of such a signal is very nearly gaussian and signals of this sort have been used in servo-analysis work with success. However, these signals did not appear subjectively to have the same quality of randomness which other signals characterized as "random" appeared to have, and it was felt that an apparent pendulosity, which was so characteristic of these signals, would very much affect the required sampling intervals since the predictability of the signal would be greater than for a random signal. If the signals did have serial redundancy then the rate of increase of uncertainty between samples would be less and the necessary intervals between samples increased. The subjects would, in a sense, be under-worked, rather than working at the limit. The earlier data showed that the subjects quite characteristically spent some appreciable percentage of their time looking at places other than at the dials themselves. This percentage varied from 2 to 8 percent depending on the subject and on the experiment. Since the calculated workload was very close to 100 percent, this observation supports the suspicion that the characteristics of the signal influenced the behaviour of the subjects in an unexpected way.

Experiment 2 was repeated using signals generated by a "random noise generator". The Zener noise which was produced was filtered by a three section Butterworth filter to provide the signals which were then recorded on a six channel Mnemotron recorder. The filter parameters were changed only once during the recording process since changes in tape speed provided the variations in bandwidth for all but one signal. The signals, even if the bandwidths are not exactly as calculated, are in an appropriate ratio to one another.

Four subjects from the earlier studies were used. It is interesting to note that although the sum of the bandwidths of this study is only .06 cycles per second greater than the sum of the bandwidths used in Experiment 2, the subjects reported that the signals were "much harder" to monitor, and expressed themselves as being more fatigued after one hour of these new signals than had been the case with the old.

Results

The subjects monitored the new signals for five hours and then were recorded on film. For the 2 subjects whose data are reported here (the other 2 were not analyzed at the time of this writing), 15,832 frames were read. These in turn showed 2,800 fixations. These latter are the data discussed below.

Frequency of Fixation

Table (12) presents the data for 2 subjects monitoring the six signals. The raw data are also shown corrected by Eq. (12). The corrected data are shown in Fig. 12. A line fitted to the means by least squares is also shown. The slope of the line is greater than that found in Experiment 2 and more nearly approaches the theoretical value. Of particular interest is the fact that only about .5 percent of the time spent by these two subjects was spent looking at anything other than the dials which were to be monitored. This should be compared with the figure of about 5 percent for all the other experiments combined. The subjects were less able to waste observation time during the monitoring task. There is still an oversampling of the lower bandwidth signals, but it is reasonable to assign this to forgetting during the relatively long intervals between looks at these signals. The data represent the behaviour of subjects toward two segments of the signal tape. Presumably the addition of the data from the other two subjects would provide a more consistent picture of the monitoring behaviour toward these signals. The new signals clearly elicited behaviour which was different from that elicited by the old signals. This finding, that subtle variations in the temporal characteristics of the signals appears to affect the monitors' behaviour quite strongly, is an important one.

Duration of Fixation

Table 13 presents the data for the duration of fixations on the signals of various bandwidths. The mean data are shown corrected by Eq. (14). The same corrected data are shown in Fig. 13. A line has been fitted by least squares to the corrected data. The results strongly support the simple Markov transition model. The slope of the best fitting line is -.11 whereas if there were no relation between bandwidth and duration of fixation, the slope would have been -.30. This last number is obtained by assuming that there is no change in duration of fixation with bandwidth, applying the correction which arises from Eq. (14), and calculating the slope of the line which would fit those data.

The results of this experiment support the notion that the signal characteristics did in fact alter the results obtained in the other experiments and cause those results to depart more from the predictions of simple sampling theory than would have been the case for truly random signals.

TABLE 12

Frequency of Fixation

Signal Bandwidth	.64	.32	.16	.08	.04	.02
Subject 1	.504	.502	.265	.303	.215	.233
Corrected	1.024	.673	.297	.323	.222	.237
Subject 2	.639	.445	.423	.253	.263	.127
Corrected	1.298	.597	.485	.270	.272	.129
Mean	.572	.478	.344	.278	.239	.180
Corrected Mean	1.161	.635	.391	.297	.247	.183

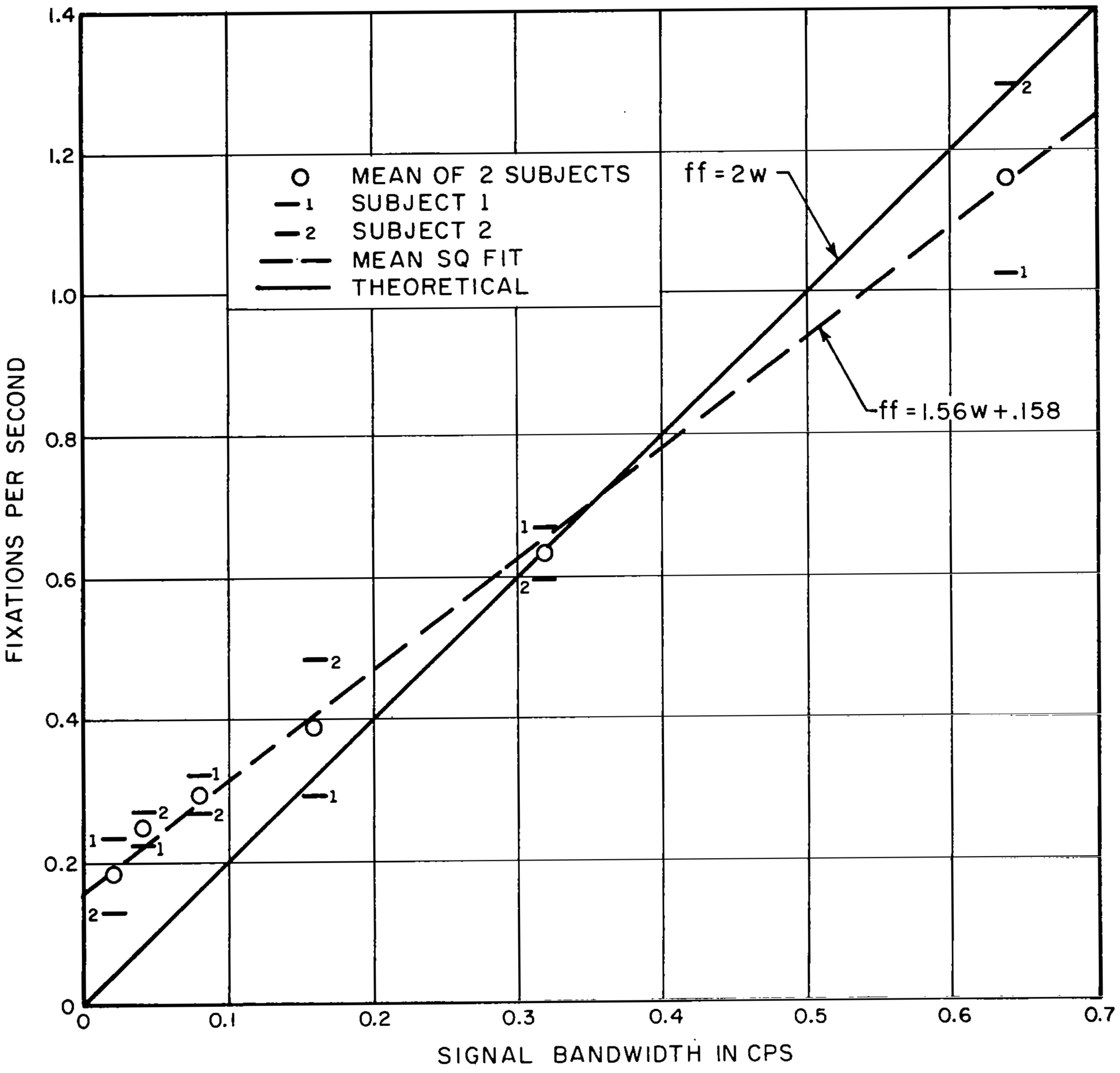


FIG. 12 EXPERIMENT VI: CORRECTED FREQUENCY OF FIXATION AS A FUNCTION OF BANDWIDTH IN CPS

TABLE 13

Duration of Fixation

Signal Bandwidth	.64	.32	.16	.08	.04	.02
Subject 1	.73	.39	.43	.39	.41	.47
Corrected	.36	.29	.38	.37	.40	.46
Subject 2	.63	.45	.39	.37	.33	.41
Corrected	.31	.33	.34	.34	.32	.40
Mean	.68	.42	.41	.38	.37	.44
Corrected Mean	.33	.31	.36	.35	.36	.43

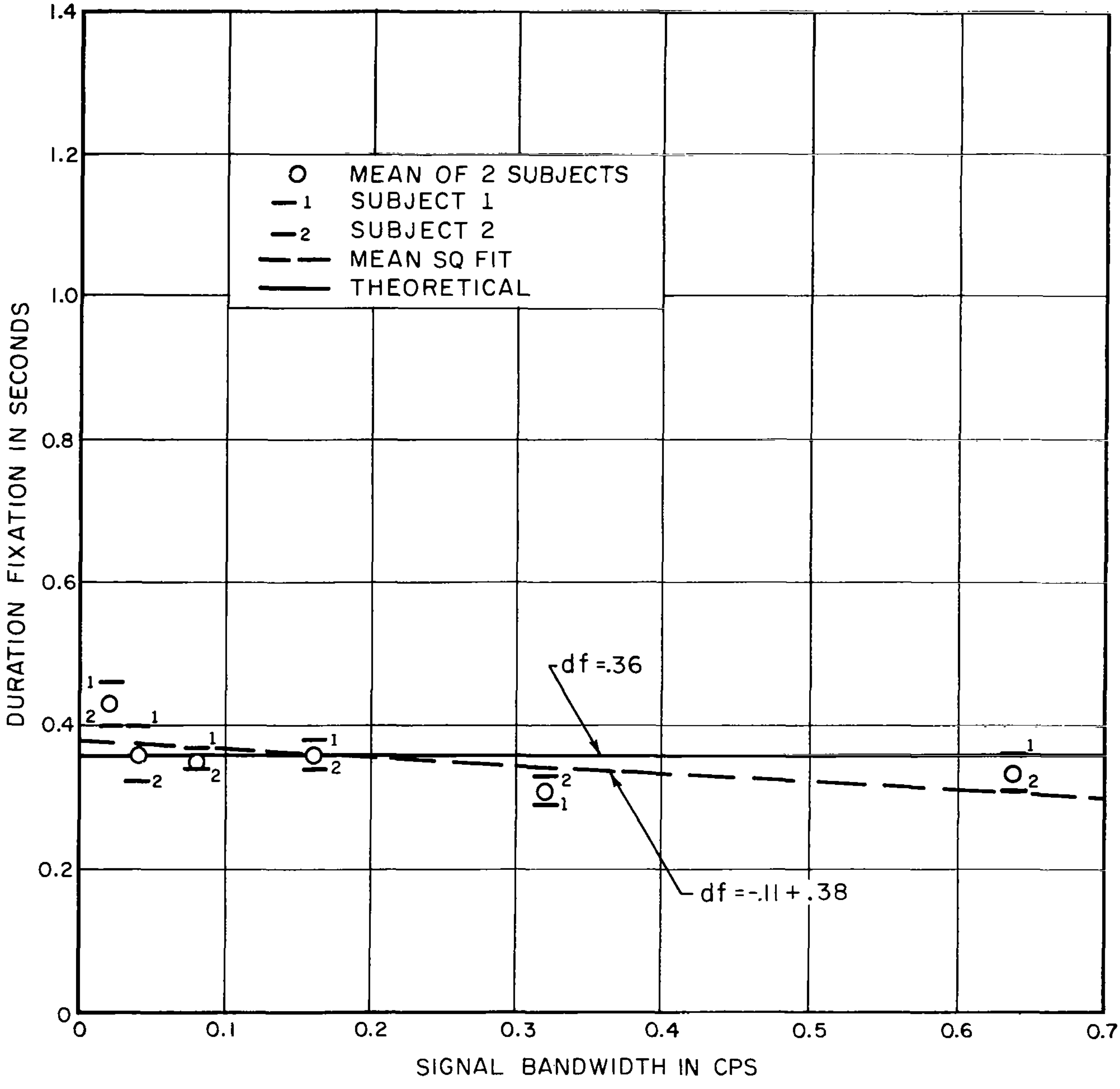


FIG. 13 EXPERIMENT VI: CORRECTED DURATION OF FIXATION AS A FUNCTION OF BANDWIDTH IN CPS

PART IV--DISCUSSION AND CONCLUSIONS

Taken in the aggregate, the results of the experiments give very strong support to the theory that frequency of observation is largely determined by the signal characteristics. In particular, when signal powers are equalized, the correlation between signal bandwidth and frequency of fixation will be very nearly 1.0. The regression of fixation frequency on signal bandwidth does not have a slope constant of 2.0 as predicted by the simple sampling theory originally suggested. The data in this respect are somewhat at variance with those data obtained in 1953 and 1954. However, Experiment 6 suggests that the difference may lie in the signal sources used rather than in anything more basic. In addition, these experiments carried the functions down to frequencies below .08 cycles per second bandwidth which was the lowest bandwidth of the earlier studies.

The simple Markov model for transitions, which gives rise to the equations for observable data (Eqs. (12) and (14)) appears to hold very well for these 6 independent signals much as it did for the 4 independent signals of the earlier study.

The experiments intended to study the effect of permissible error were unsuccessful largely because they were founded on a misconception as to the function of a limit indication on an instrument. A totally different kind of experiment would have to be run if the original hypotheses were to be tested. However, the large majority of operational situations are of the kind studied here, and the results of these studies are more widely applicable than would be the case for the others. In a sense, the theories proposed in Part II of this report suggest that the more important characteristic of a signal is the use to be made of it. Certainly, in hindsight, this seems like a most reasonable finding.

The experiment on the effect of correlation between signals produced results which will not greatly affect the application of the theories presented. However, the essential meaninglessness of the correlation may have seriously influenced the behaviour of the observers. In an operational vehicle, the relations which exist between instruments on the instrument panel are meaningful. The signals detected on one will dictate, in part, the next to be observed. Studies will have to be performed in simulated flight if one is to find out how general the results of this experiment (largely negative) are. There was a slight hint that the subjects were making use of the correlation. Presumably, this utilization could have been enhanced by specific instructions as to fixation sequence and the nature of the relationship between the instruments. However, that was not the purpose of this experiment. Here we were interested in the effects of correlation without special instructions or knowledge on the part of the subjects. Future experiments can be devised to explore the effect of instructions or visual sampling behaviour.

The dichotomization of a continuous signal produced changes in the observing behaviour of the subjects. For one subject there was a sharp increase in the amount of attention paid to the dichotomized signal. For the others this was not the case; for three of these four there was a sharp reduction in the attention paid. In the aggregate, there was a slight reduction in attentional demand of the dichotomized instrument. However, the magnitude of this change in observing behaviour is not apparently large enough to make an important change in the calculations one might wish to make of an operational system.

The overall conclusion of the studies taken as a whole is that the bandwidth of the signal which is being monitored is the single most important factor influencing the frequency of fixation on that signal, granting that the power of the signal is set at some fixed level. The interaction between frequency of fixation and duration of fixation suggested by the simple Markov model is also strongly supported by the data taken as a whole as well as by the results of the individual experiments. The influence of reading accuracy on duration of fixation is still moot and remains for some future test. The simple Markov model also fits very well the data on observable frequency of fixation. However, this finding must be looked at with caution. The more thorough analysis of Markov models presented in Appendix I suggest other models which probably apply in other situations. The data obtained in all these studies are relevant to a situation in which a number of indicators present signals of nearly equal power, of nearly identical "limits", having no correlation (except for one instance), and having no logical relationship between the readings on one indicator and the readings on another. In an operational space vehicle or aircraft, all of these conditions will be different for the various indicators. As a result the more sophisticated notions presented in Part II of this report might be expected to afford better predictions of what will happen to the pilot in a monitoring task.

The results of this study support and reconfirm the results obtained in 1953 and 1954 on a set of 4 instruments in a similar monitoring task. Such reconfirmation and the extension of the very favourable results of those prior studies to a larger number of instruments was one of the main objectives of this study. The variance between the earlier and the later results, while suggesting that more sophisticated theory and experimentation are required, is not great enough to vitiate the applicability of the simple model. However, in such application, care must be taken to examine the meaningfulness of the relations between indicators, the criticality of the limits, and the relation of the signal variance and the distance of the limit from the mean. These factors, for the time being, must be dealt with on an intuitive basis until specific research has been undertaken to determine the magnitude of their effects. In future studies of this type, subjects experienced in the monitoring of meaningful instruments should probably be used. It is our belief that only when cost and gain are introduced into the monitoring task will the behaviour of the subjects be like that exhibited by pilots in operational situations.

The use of human readers to analyze the films has been very successful. However, it may be the case that this task will be excessively difficult when the number of instruments becomes very much larger than the 6 that were used here. Particularly when one wishes in the future to examine the behaviour of persons engaged in the monitoring of large numbers of instruments requiring head-as well as eye-movements, some other means of obtaining information about the visual axis will be required.

The inadvertent results obtained on one subject for the underloaded monitoring situation suggest that experiments should be done to find out if the behaviour of the one subject is an individual characteristic or will be found generally in the population.

For the future, it would be of interest to investigate the effects of instruction, training and knowledge of results on the behaviour of monitors. A detailed study of the effects of signal characteristics on monitoring behaviour appears to be most probably fruitful, as does a study of the effects of the value of an observation and the cost of a failure to detect. The kind of monitoring behaviour which the subjects were engaged in is like any other behaviour in that it will be influenced by a wide variety of situational characteristics. What we have attempted to do here is to show that certain limits are placed on behaviour by non-psychological aspects of the situation. This has been amply demonstrated by the data.

REFERENCES

1. Boring, E.G., A History of Experimental Psychology, Second Edition, Appleton-Century, Crofts, Inc., New York, 1950.

2. Teichner, W.H., Recent Studies of Simple Reaction Time, Psychological Bulletin, 51, 2, 128-149, March 1954.

3. Woodworth, R.S., Experimental Psychology, Henry Holt and Company, New York, 1938.

4. Welford, A.T., The 'Psychological Refractory Period' and the Timing of High-Speed Performance--A Review and a Theory, British Journal of Psychology, 43, 2-19, 1952.

5. Telford, C.W., The Refractory Phase of Voluntary and Associative Responses, Journal of Experimental Psychology, 14, 1-36, 1931.

6. Kristofferson, A.B., Discrimination of Successiveness: A Test of a Model of Attention, Science, 1393550, 112-113, January 1963.

7. Broadbent, D.E., Perception and Communication, Pergamon Press, Oxford, 1958.

8. Moray, N., Broadbent's Filter Theory: Postulate H and the Problem of Switching Time, Quarterly Journal of Experimental Psychology, 12, 4, 214-220, 1960.

9. Rabbitt, P.M., Nature, London, 195, p. 102

10. Hebb, D.O., Organization of Behavior, John Wiley and Sons, Inc., New York, 1949.

11. Deutsch, J.A. and D. Deutsch, Attention: Some Theoretical Considerations, Psychological Review, 70, 1, 80-90, January 1963.

12. Boring, E.G., A note made during a classroom lecture.

13. Crossman, E.R.F.W., Information Processes in Human Skill, British Medical Bulletin, Vol. 20, 1, 32-37, 1964.

14. Senders, J.W., Man's Capacity to Use Information from Complex Displays, in Information Theory in Psychology, H. Quastler, Ed., The Free Press, Glencoe, Illinois, 1955.

15. Senders, J.W., Information Input Rates to Human Users: Recent Research Results, W.A.D.C. Symposium on Air Force Flight Instrumentation Program, Wright-Patterson Air Force Base, Ohio, 1958.

16. Senders, J.W., The Human Operator as a Monitor and Controller of Multi-Degree of Freedom Systems, Transactions of the Prof. Group on Human Factors in Electronics, IEEE, HFE-1, 1, September 1964.

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

18. Hick, W.E., On the Rate of Gain of Information, Quarterly Journal of Experimental Psychology, 4, 11-26, 1952.

19. Hyman, R., Stimulus Information as a Determinant of Reaction Time, Journal of Experimental Psychology, 45, 3, 188-196, 1953.

20. Shannon, C., Coding Theorems for a Discrete Source with a Fidelity Criterion, in Information and Decision Processes, R.E. Machol, Ed., McGraw-Hill Book Company, Inc. New York, 1960.

21. Kolmogorov, A.N., On the Shannon Theory of Information Transmission in the Case of Continuous Signals, Transactions on Information Theory, IRE, December 1956.

22. Fogel, L.J., A Note on the Sampling Theorem, Transactions of the Prof. Group on Information Theory, IRE, March 1955.

23. Senders, J.W. and K. Stevens, A Re-Analysis of the Pilot Eye-Movement Data, BBN Report No. 1136, May 1964.

APPENDIX I*

SOME MODELS FOR THE HUMAN INSTRUMENT MONITOR

I. Introduction

Much work has been done in the general area of the development of instrument panels for the human controller. Figure 1a portrays a block diagram of a rather idealized situation for the design of an instrument panel. In this situation the human instrument monitor observes the instrument panel and decides on an appropriate course of action which in turn affects the system. The loop is completed when these inputs to the system produce a change that is reflected in the instrument outputs. The design portion of the diagram in Fig. 1a is shown in dotted lines. The instrument readings and monitor's actions are evaluated on the basis of some performance criterion and the instrument panel is then changed in order to improve this performance.

Unfortunately this process becomes impractical for any physical system of any complexity, so the designer is forced to use simulations and/or some of the more general results from the area of human engineering. The purpose of this paper is to describe a model for the human instrument monitor that will hopefully add to this latter category and provide a more detailed structure for the overall description of any system. In particular, the goals of the model are to provide a framework for:

(1) The interpretation of theoretical and experimental results in the area of human instrument monitoring.

(2) The design of future experiments.

* This section is due to R. Smallwood of Bolt Beranek and Newman Inc. now at Stanford University.

(3) The analysis of present and proposed physical systems.

(4) A more efficient and effective design of future systems.

The proposed functional block diagram of the human instrument monitor is shown in Fig. 1b. There is a large body of experimental evidence to support the thesis that the human monitor of several instruments divides his attention among the different instruments, observing only one instrument at a time. This is the justification for portraying the eye in Fig. 1b as a commutated single-channel input switch.

The reading of the instrument that the eye is looking at is not observed correctly, of course, and this imperfect perception is represented by the data channel in Fig. 1b. The data channel output plus the monitor's set of instructions or goals are used to arrive at the decision that determines the monitor's action.

As shown in Fig. 1b, we assume that a sample selector is deciding what instrument to observe according to some sampling criterion which in turn is determined by the initial set of instructions and the output of the data channel. We shall be quite concerned with the sampling characteristics of the human instrument monitor because we can observe the movements of the eye by various techniques and thus measure his sampling behavior (as represented in Fig. 1b).

In Section II we shall present some Markovian models for the sample selector in Fig. 1b, and in Section III a simple model for the data channel will be discussed and combined with one of the models of Section II to give a measure of the average information transmitted from the instruments to the decision process. In Section IV some brief remarks on sampling criteria will be made; the paper concludes with a description of some proposed experiments for the validation of the models.

II. Sampling Models

The situation to be described by the sampling models is the following: The human observer has before him N instruments that he is to monitor. We shall say that the monitor is in state i at the time t, if his attention is focused on the i^{th} instrument at that time; or more concisely, $s(t) = i$. The sampling models, then, will be concerned with describing the function, $s(t)$.

One of the simplest sampling models is the periodic one proposed by Senders (8). In this model, the monitor repeats the same scan of the instruments periodically so that $s(t) = s(t-nT)$ for some value of the period T and integer n. It is very likely that the human monitor does not scan the instruments in a periodic fashion; however, as we shall see in Section III, the periodic model has the virtue of insuring very simple calculations for the information transmitted. An example of s(t) for the periodic model is shown in Fig. 2.

In order to allow for more random variations in the sampling behavior of the instrument monitor we shall relax the requirement that his state be deterministic. Instead, we define the probability, p_{ij}:

$$p_{ij} = \Pr\,[\text{next instrument monitored is } j \,|\, \text{present instrument monitored is } i] \tag{1}$$

where a vertical bar ($|$) is to be read as "given that" or "conditioned upon." We shall assume that the behavior of the human monitor is completely determined by the p_{ij}'s; that is, we shall assume that the probabilities for the next instrument to be observed are dependent only upon the monitor's present state and not upon any of the instruments observed in the past. Stochastic processes with this property are called simple or first order Markov processes and the assumption mentioned above, the Markovian assumption. A convenient graphical representation of a simple Markov model is shown in Fig. 3a. Each of the nodes represents a state or instrument and the branch from state i to state j represents the probable transition of the monitor's attention from the $i^{\underline{th}}$ to the $j^{\underline{th}}$ instrument. The probabilities, p_{ij}, of each transition are written above the transition branches. Because the monitor must shift his attention to some instrument, we have:

$$\sum_{j=1}^{N} p_{ij} = 1 \text{ for all } 1 \leq i \leq N. \qquad (2)$$

An alternative matrix representation of a Markov process is shown in Fig. 3b. Equation 2 demands that the elements in each row sum to one. The elements in the $i^{\underline{th}}$ row of P correspond to the transition branches leaving the i^{th} node of Fig. 3b, while the elements of the $j^{\underline{th}}$ column correspond to the branches entering the $j^{\underline{th}}$ node.

The matrix of probabilities (we shall call them transition probabilities) defined by Eq.(1)describe the transitions of the monitor from instrument to instrument, but say nothing about the times when these transitions occur. The various methods for the description of these times will be the main differences between our Markov models.

The simplest assumption that can be made concerning the transition times is that transitions can occur only at discrete intervals of time, i.e., at $t = n\Delta$, where $n = 0, 1, 2....$ An example of the state of such a system is shown in Fig. 4. It is necessary for this discrete case to allow transitions back to the same instrument; that is, $p_{ii} \neq 0$. If this were not allowed, then the time spent on any one observation of an instrument would be independent of the instrument and just equal to Δ. The probability of spending $n\Delta$ seconds observing the i^{th} instrument during a single observation is:

$$h_i(n\Delta) = p_{ii}^{\,n-1}(1-p_{ii}) \qquad n \geq 1 \qquad (3)$$

and the mean observation or dwell time for a single look is:

$$\bar{t}_i = \sum_{n=1}^{\infty} (n\Delta)h_i(n\Delta) = \frac{\Delta}{(1-p_{ii})} \qquad (4)$$

An important statistic for the models we are considering is the distribution of first arrival times for each instrument. The first arrival time, τ_{ij}, is defined as the time for the monitor's first arrival at the j^{th} state if his present state is i. Some examples of the τ_{ij}'s are shown in Fig. 3. The derivation of a method for calculating the distribution of this statistic is carried out in Appendix B. These first arrival times are related to the sampling rate of the monitor and so are very important when we start calculating in the next section, the information absorbed by the monitor.

The distribution of τ_{ij} will be labelled $g_{ij}(n)$:

$$g_{ij}(n) = \Pr[\tau_{ij} = n\Delta] \qquad (5)$$

There is a large class of Markov processes (called **ergodic** processes) that eventually approach a limiting "state probability distribution" that is independent of the starting state of the system. In equation form:

$$\lim_{n \to \infty} \Pr[s(n\Delta)=j \,|\, s(0)=i] = \pi_j \tag{6}$$

We shall call π_j the steady-state probability that the discrete system is in state j. In the physical terms of the real world situation we are modeling, Eq. (6) states that if we observe the monitor looking at a particular instrument and then wait for a long time (i.e., long relative to Δ), the probability that the monitor will be looking at the $j^{\underline{th}}$ instrument is independent of where he was looking originally. For this discrete case π_j also represents the fraction of time over a long time interval that was spent monitoring the $j^{\underline{th}}$ instrument. All of the models that we will consider will be ergodic, so we shall talk freely of the steady-state probabilities.

It is shown in Appendix A that the steady-state probabilities satisfy the following set of equations:

$$\sum_{i=1}^{N} \pi_i p_{ij} = \pi_j \text{ for } 1 \leq j \leq N \tag{7a}$$

$$\sum_{i=1}^{N} \pi_i = 1 \tag{7b}$$

The N equations in Eq. (7a) are linearly dependent; hence, the need for Eq. (7b).

The steady-state probabilities can be used to calculate another important statistic - important in the sense that it is readily observable. This statistic is the fraction of all real transitions in the steady-state that go from state i to state j; we shall label this quantity q_{ij}. By a "real transition" we mean a transition between two different states; transitions between the same states will be called "virtual transitions." Thus, by definition $q_{ii} = 0$ we can derive an expression for q_{ij} using Bayes rule as follows:

$$q_{ij} = \Pr[\text{Next transition is from } i \text{ to } j \mid \text{real trans.}]$$

$$= \frac{\Pr[s(t)=i]\ \Pr[\text{next trans. is real and to } j \mid s(t)=i]}{\Pr[\text{next trans. is real}]}$$

$$= \frac{\pi_i\ p_{ij}(1-\delta_{ij})}{1- \sum_{k=1}^{N} \pi_k\ p_{kk}} \qquad (8)$$

where δ_{ij} is the Kronecker delta function. The q_{ij}'s are equivalent to the "link probabilities" discussed by Senders (8).

There are two important disadvantages to the discrete time Markov model of the human instrument monitor. The first of these concerns the introduction of virtual transitions (i.e., $p_{ii} \neq 0$) in order to circumvent the stringent requirements on the transition time imposed by the discrete time assumption. These virtual transitions are almost certainly not observable from an experimental point of view and there is a serious question as to whether or not they exist in the real world situation being modelled.

The second disadvantage is allied to the first and concerns the experimental determination of the transition interval Δ. The only requirement on Δ is that all times between real (i.e., observable) transitions be some integer multiple of Δ There appears to be little evidence that a human monitor does shift his attention only at discrete time intervals; and if he does not, then Δ will have to be quite small in order to fit any sizeable amount of data. This will force the p_{ii}'s to be close to unity (see Eq. (4)). And, of course, no matter what value of Δ is decided upon, any integral fraction of that Δ will satisfy the data equally as well. There are, however, experimental situations in which there is a natural value of Δ that is determined by the experiment. The recording of sampling behavior with a movie camera is a good example of such a situation.

In the face of this indeterminary of the transition interval Δ, we shall generalize our model by allowing the time between transitions to be a continuous rather than a discrete variable. The most common example of such Markov process is the so-called "stationary continuous time Markov process." In this type of process, the time spent observing the $i^{\underline{th}}$ instrument (i.e., the dwell time) during any one look is assumed to be exponentially distributed with a mean that is dependent only upon i. However, we shall use a more general model that includes the stationary continuous time Markov process as a special case; this class of Markov processes is called semi-Markov processes. The semi-Markov process requires only one further extension of the discrete time Markov process we have been discussing up to now. In this model we shall assume that the monitor is operating as follows: Upon entering the $i^{\underline{th}}$ state, he then decides what state j, he will go to next according to the probabilities,

p_{ij}, defined by Eq. (1). Once this is decided, the time spent in the $i^{\underline{th}}$ state is selected from a probability density function, $h_{ij}(t)$, that depends on both i and j. Therefore, the semi-Markov process is completely defined by the transition probabilities, p_{ij}, in Eq. (1) and the density functions, $h_{ij}(t)$. An example of s(t) for a semi-Markov model is illustrated in Fig. 5.

Since the times between transitions are not fixed for the semi-Markov model, it will no longer be necessary to allow for virtual transitions. Hence, p_{ii} will be set equal to zero for our semi-Markov model.

For this model the average dwell time on the $i^{\underline{th}}$ instrument for subsequent transitions to j is just the mean of $h_{ij}(t)$:

$$\bar{t}_{ij} = \int_0^\infty t h_{ij}(t)\, dt \qquad (9)$$

And the average dwell time on instrument i over all subsequent transitions is:

$$\bar{t}_i = \sum_{j=1}^{N} p_{ij}\bar{t}_{ij} \qquad (10)$$

It is easy to see how the discrete and continuous time Markov models discussed earlier are contained within the class of semi-Markov models. For the discrete Markov model we just have

$$^{s}p_{ij} = \frac{^{d}p_{ij}}{(1-{}^{d}p_{ii})} \quad (1-\delta_{ij}) \qquad (11)$$

$$h_{ij}(t) = {}^{d}p_{ii}^{n-1} \; (1-{}^{d}p_{ii}) \; \delta(t-n\Delta) \qquad (12)$$

where the superscript d refers to the discrete time transition probabilities and s, the semi-Markov transition probabilities. The quantity $\delta(t)$ is the Dirac-delta function.

The continuous time Markov model is represented in the semi-Markov model by a matrix of transition probabilities with $p_{ii} = 0$ and a matrix of probability density functions, $h_{ij}(t)$, of the form:

$$h_{ij}(t) = \lambda_i e^{-\lambda_i t} \tag{13}$$

A special case of the semi-Markov model is the one in which the dwell time is independent of where the monitor looks next, i.e., when $h_{ij}(t) = h_i(t)$. The discrete and continuous time processes are obviously examples of this simpler type of model.

For the semi-Markov model, the π_i defined by Eqs. (7) still retains its property of being the total fraction of transitions entering the $i^{\underline{th}}$ state after the process has been operating for a long time (long compared to the $\bar{t}_{ij}$'s). However, since the times between transitions now depend upon the states of the system, π_i must be weighted with $\bar{t}_i$ in order to find the average fraction of time spent in the $i^{\underline{th}}$ state over a long period of time. If we label this quantity, ϕ_i, we have:

$$\phi_i = \frac{\pi_i \, \bar{t}_i}{\sum\limits_{j=1}^{N} \pi_j \bar{t}_j} \tag{14}$$

The quantity ϕ_i also represents the probability of finding the monitor's attention on the $i^{\underline{th}}$ instrument if the system is started and allowed to run for a long time before observing it again (i.e., if it is in the steady-state). Equation (14) is derived more rigorously in Appendix C.

For the semi-Markov model the values of the q_{ij}'s defined by Eqs. (7) and (8) are still valid. However, since we are assuming that $p_{ii} = 0$ for this model, Eq. (8) can be simplified to:

$$q_{ij} = \pi_i p_{ij} \tag{15}$$

The quantity, q_{ij}, still represents, of course, the fraction of all real transitions that proceed from i to j.

The first arrival times, τ_{ij}, for each instrument (see Fig. 5) are now continuous random variables rather than the discrete random variables mentioned in Eq. (5) for the discrete Markov model. The definition of τ_{ij} must also be made more specific; the definition of τ_{ij} for a semi-Markov process will be the time for the process's first arrival at the j^{th} state if the process has _just arrived at the i^{th} state_. The probability density functions, $g_{ij}(\tau)$, for these first arrival times are derived in Appendix D.

A word should be said concerning the experimental validation of these external models and the estimation of the model parameters from experimental data. First of all, the semi-Markov model is a very versatile one and should be able to handle a very general class of situations. For example, it may be true that the probability of going to instrument j in the next transition is dependent not only upon the present state, but also upon the last k states. Such a process is called a $(k+1)^{th}$ order Markov process and can obviously be reduced to a first order (or simple) Markov process by defining all of the N^{k+1} possible sequences of past and present states as the states of the first order Markov process. As a very simple illustration of such a process Fig. 6 presents a semi-Markov model for the periodic process of Fig. 2. The first state has been broken into two states to account for the (3→2→1) and (4→2→1) transitions while state 2 has been broken into three states in order to describe the (3→2), (1→2), and (4→2) transitions. Since this process is deterministic all of the transition probabilities will be either 1 or 0, and the dwell time density functions, $h_{ij}(t)$, will be unit impulses at the appropriate values of t_{ij}. Bartlett (1) has described a technique for testing the order of a Markov process.

Much work has been done on the estimation of the transition probabilities of a Markov model. Billingsly (2) presents a survey of the work done in this area up to 1961 and has a very extensive list of references. Silver (10) in his excellent report has developed a Bayesian approach to the problem that is more intuitively satisfying.

All of the models in this section have been assumed stationary with time, that is, the transition probabilities and dwell time density functions have been assumed to be unchanging with time. In the terms of Fig. 1b, this is equivalent to assuming that the inputs to the "sampling criterion" block are not of such a nature to cause any changes in the sampling behavior. For the more realistic cases in which the sampling behavior changes with time -- or more particularly, with the data channel output -- the appropriate extension of the Markov sampling models is to allow the p_{ij}'s and $h_{ij}(\cdot)$'s to vary with time also. Some analytical work has been done on time-varying discrete and continuous Markov processes. The sampling criterion operation will be discussed further in Section IV.

III. Data Channel Models

In this section we shall discuss a very simple model for the description of the information perception by the human monitor. This model will attempt to describe in a very simple way the transformation that occurs between the true reading of the instrument and the reading perceived by the monitor. In terms of the diagram of Fig. 1b, we shall attempt to model the data channel between the eye and the internal decision mechanism.

The model chosen is the very simple one shown in Fig. 7. If the true instrument reading is x then the perceived instrument reading is characterized as a single sample of a continuous random variable, y, whose probability density function for a given input x is $f_c(y|x)$. In terms of communication theory we are modelling the data channel as a continuous, memoryless, noisy communication channel. The conditional density function, $f_c(y|x)$ is assumed to specify the model completely. More complex models could be assumed, of course, such as allowing the density function to depend on all past inputs from that channel or even on all the past inputs from all channels. In the interests of algebraic simplicity, we shall restrict our discussion here to memoryless channels.

A logical candidate for the conditional density function, $f_c(\circ|x)$ is the normal density function with mean x and variance N:

$$f_c(y|x) = f_N(y|x,N) \triangleq (2\pi N)^{-\frac{1}{2}} \exp[-\frac{(y-x)^2}{2N}] \qquad (16)$$

This is the particular form that we shall use for $f_c(\circ|x)$ in all the following examples.

In general we should expect N to vary with the amount of time spent observing the instrument reading. For example, if the monitor looks at the instrument for a very short amount of time (t_{ij} in Fig. 5), then we would expect the variance of y to be quite large, and to decrease as he spends more and more time observing x. (We are assuming here that x does not vary very much while the monitor is observing it; i.e., the bandwidth of x is small compared to t_{ij}^{-1}.) This idea of N being a function of time is a very useful concept.

Let us now combine our data channel model with the periodic sampling model to find the average amount of information transmitted by each instrument to the monitor. First of all, however, some more assumptions are necessary. We shall assume for our periodic sampling model that the monitor samples each instrument only once during eaeh period, T, and the dwell time for the ith instrument is t_i. Thus we have:

$$\sum_{i=1}^{N} t_i = T \qquad (17)$$

Secondly, we shall assume that the readings, x(t), of each instrument are bandlimited, zero mean, white Gaussian noise with a bandwidth W and a mean square value of S_i for the ith instrument. Finally, we shall assume that successive samples of $x_i(t)$ are uncorrelated (i.e., $T>>W^{-1}$). With these restrictions we find that the average amount of information, $I_i(X;Y)$absorbed by the monitor from the i$^{\underline{th}}$ instrument per period is:

$$I_i(X;Y)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(x,y)\ \log[f_c(y|x)/f(y)]dx\ dy = \frac{1}{2}\log[1+\frac{S_i}{N}] \qquad (18)$$

and the total information absorbed per unit time from all instruments is:

$$J(X;Y) = \frac{1}{T} \sum_{i=1}^{N} I_i(X;Y) = \frac{1}{2T} \sum_{i=1}^{N} \log[1+ \frac{S_i}{N_i} \qquad (19)$$

Equations (18) and (19) are well known relations in information theory (4,9).

An interesting speculation can be derived from Eq. (18). Let us assume that the amount of information absorbed by the monitor during any one observation is proportional to the time, t_i, spent looking at the instrument. Then if k is the proportionality constant, we have:

$$\frac{1}{2} \log[1 + \frac{S}{N}] = kt \qquad (20)$$

which gives:

$$N(t) = \frac{Se^{-2kt}}{1-e^{-2kt}} \qquad (21)$$

for the variation of the data channel variance with signal strength and dwell time. This function is plotted in Fig. 8. The speculation represented by Eq. (21) is dependent, of course, on all of the assumptions of the previous paragraph as well as this one.

The results of Eqs. (18) and (19) are quite simple, but are based on some very stringent assumptions concerning both the instrument readings, $x_i(t)$, and the sampling behavior of the monitor. For the sake of completeness, a very general relation for J(X;Y) will be derived. A semi-Markov process will be assumed for the sampling model. The data channel model will just be the simple one of Fig. 7. We will consider the situation for n observations of the $i^{\underline{th}}$ instrument as shown in Fig. 9 where τ is the time between successive observations of the instrument. The subscripts for x, y, t, and τ will refer to the number of the observation rather than the instrument; the i subscript for the instrument will be understood.

The joint distribution of t_k and τ_k in Fig. 7 can be written as the sum of the joint distributions over all possible transitions that can occur at the end of t_k. These joint distributions must, of course, be weighted by the probability, p_{ij}, of their applicability.

$$g_i(t_k,\tau_k) = \sum_{j=1}^{N} p_{ij}h_{ij}(t_k)\ g_{ji}(\tau_k) \tag{22}$$

Since the successive random variables, t_k, t_{k+1} and τ_k, τ_{k+1} are independent, we can write the joint distribution over all n of the t's and (n-1) of the τ's as:

$$g_i(\underline{t_n},\underline{\tau_n}) = \prod_{k=1}^{n-1} g_i(t_k,\tau_k) \quad \sum_{j=1}^{N} p_{ij}h_{ij}(t_n) \tag{23}$$

where $\underline{t}_n$ is the label for the n dimensional vector, $\{t_1, t_2 \ldots t_n\}$ and τ_n is the (n-1) dimensional vector, $\{\tau_1, \tau_2 \ldots \tau_{n-1}\}$.

The most general representation of n instrument readings is just the joint distribution of n x_i's for the particular times. We shall represent this by:

$$f_i(\underline{x_n}|\underline{t_n},\underline{\tau_n}) = f_i[x_1(t), x_2(t+t_1+\tau_1), x_3(t+t_1+t_2+\tau_1+\tau_2, \ldots, x_n(t+\mathop{}_{k}^{n-1} t_k + \mathop{}_{k}^{n-1} \tau_k)] \tag{24}$$

where $\underline{x_n}$ is the n dimensional vector of the n instrument readings, $\{x_1,x_2,\ldots,x_n\}$. The joint density function for the n data channel outputs is:

$$f(\underline{y_n}|\underline{x_n},\underline{t_n}) = \mathop{\pi}\limits_{k=1}^{n} f_c(y_k|x_k,t_k) \qquad (25)$$

where $\underline{y_n}$ is the vector $\{y_1,y_2,\ldots,y_n\}$. The dwell time, t_k, for the k^{th} observation is included in $f_c(y|x,t)$ since the distribution of outputs is assumed to depend on the dwell time as well as the instrument reading, x. An example of this is the function N(t) used for the variance in Eq. (16).

For any one observation in Fig. 9, the information supplied by y_k about x_k is just:

$$I_1(x_k;y_k) = \log\,[f(x_k|\underline{y_k})/f(x_k|\underline{y_{k-1}})] \qquad (26)$$

That is, it is the logarithm of the density function for x_k conditioned on all the y's up to and including y_k, divided by the density function for x_k conditioned on all the y's up to and including y_{k-1}. Applying Bayes rule to $f(x_k|\underline{y_k})$ we have:

$$I_1(x_k;y_k) = \log\,[f(y_k|x_k,\underline{y_{k-1}})/f(y_k|\underline{y_{k-1}})] \qquad (27)$$

Now, if the channel is a memoryless one, then the output y_k is dependent only on x_k and so:

$$f(y_k|x_k,\underline{y_{k-1}})=f_c(y_k|x_k,t_k) \qquad (28)$$

which allows us to write Eq. (27) as:

$$I_1(x_k;y_k) = \log\,[f_c(y_k|x_k,t_k)/f(y_k|\underline{y_{k-1}})] \qquad (29)$$

Equations 22-24 and 28 can be used to find the expected amount of information to be absorbed from the $i^{\underline{th}}$ instrument per observation in the n observations:

$$I_i\ (X_n;Y_n) = \frac{1}{n}\int_0^\infty g_i(\underline{t_n},\underline{\tau_n})\ d\underline{t_n}\,d\underline{\tau_n}\ \int_{-\infty}^{\infty} f_i(\underline{x_n}|\underline{t_n},\underline{\tau_n})\ d\underline{x_n}\int_{-\infty}^{\infty} f(\underline{y_n}|\underline{x_n},\underline{t_n})\ d\underline{y_n}$$

$$\sum_{k=1}^{n} I_i(x_k;y_k) \qquad (30)$$

The expected rate of information absorbed from the i^{th} instrument in the n observations is:

$$J_i\ (X_n;Y_n) = \int_0^\infty g_i(\underline{t_n},\underline{\tau_n})d\underline{t_n}d\underline{\tau_n}\ \int_{-\infty}^{\infty} f_i(\underline{x_n}|\underline{t_n},\underline{\tau_n})d\underline{x_n}\int_{-\infty}^{\infty} f(\underline{y_n}|\underline{x_n},\underline{t_n})d\underline{y_n} \qquad (31)$$

$$\Big[\sum_{j=1}^{n-1}(t_j+\tau_j)+t_n\Big]^{-1}\ \sum_{k=1}^{n} I_i(x_k;y_k)$$

Equations (30) and (31) will obviously be quite difficult to evaluate if the forms of g_{ij}, h_{ij}, and $f(x_n|t_n,\tau_n)$ are very complicated. In such a situation, these equations can most easily be solved numerically by simulating the process on a computer and using Monte Carlo techniques to evaluate the means. In Appendix E, an illustrative example is presented in which an analytical expression for $I_i(X_n;Y_n)$ and $J_i(X_n;Y_n)$ is derived.

It is not true, of course, that the total rate of information from all instruments is equal to the sum of the individual rates in Eqs. (30) and (31). This is due to the dependence of the dwell times and interobservation times upon the next state as well as the present state. However, this would not be a serious limitation in any Monte Carlo evaluation of the total rate.

IV. Sampling Criteria

In the previous two sections mathematical models for the sample selector and data channel in Fig. 1b have been proposed. One of the underlying assumptions of the semi-Markov sampling model was that the transition probabilities and dwell time density functions be stationary with time. There are obvious situations, however, when this will not be true; and in these situations a more general sampling model such as a time-varying Markov process may have to be used. In these non-stationary situations the sampling criterion block shown in Fig. 1b is using the observed outputs of the instruments to alter the sampling behavior according to some criterion.

An example of such a non-stationary situation is the typical threshold experimental situation in which the monitor is instructed to take some specific action, such as pushing a button, if the instrument reading exceeds a certain threshold. In this situation one would expect the sampling behavior to shift toward a preference for those instruments whose readings were near the threshold. Thus, the sampling behavior would be a function of the previous data channel outputs.

If it is possible to measure this sampling criterion and develop a model for the way in which it affects the sampling behavior, then a very complete description of the human instrument monitor should be possible. A factor that may be very useful in describing a particular sampling criterion is the information rate of a particular instrument. This is defined by Shannon (9) to be the minimum amount of information needed to specify the instrument reading to within a given measure of fidelity. For example, suppose that the reading of the i$\underline{\text{th}}$ instrument is the output of a bandlimited white noise source of bandwidth W_i and average power S_i, and that it is necessary to specify the instrument readings to within a mean square difference of E. That is, if y is the estimate of the instrument reading, x, then:

$$\iint (x-y)^2 f_i(x,y)\, dx\, dy \leq E \qquad (32)$$

In this case Shannon (9) has shown that the information rate of the source is:

$$R_i = W_i \log(S_i/E) \qquad (33)$$

The important quantity in this measure of information rate is, of course, the measure of fidelity (represented by E in the example of Eqs. 32-33).

A reasonable assumption for the sampling criterion is that the sampling rate for an instrument will depend on the information rate of the instrument. However, one would expect the information rate of an instrument to vary with time due to variations in the measure of fidelity -- that is, due to variations in the degree of precision that the monitor believes is necessary to specify the instrument readings for his purposes. This variation will depend not only upon the instructions and goals given the monitor but may also depend on the last observed output of the data channel. For example, consider the mean square measure of fidelity in Eq. (32). If the monitor is instructed to be sensitive to a threshold level of the instrument (e.g., the oil pressure gauge in a car), then his measure of fidelity will most likely be a function of the true instrument reading x and look like E(x) in Fig. 10a. If he is attempting to keep an instrument reading within a narrow region (e.g., a compass) E(x) will have the form of Fig. 10b; and if he is only interested in reading x to within a certain error independent of where it is (e.g., air speed), then E(x) will be uniform as in Fig. 10c.

Senders (8) has proposed a sampling criterion for a periodic sampling model in which the fraction of time spent on any instrument is proportional to R_i. There is some experimental evidence to support this proposed criterion within the limitations of the periodic sampling model.

For the experimental situations in which one is testing a sampling model or data channel model, it is most desirable that the sampling behavior be stochastically stationary. If the sampling criterion is indeed dependent on the R_i's, then for this situation one should strive by instructions to give the subject a fidelity measure, E(x), that is uniform (see Fig. 10c).

One possible sampling criterion that might explain the sampling behavior of a human instrument monitor is:

$$C_l = \sum_{i=1}^{N} \left(\frac{J_i}{R_i}\right) \qquad (34)$$

Since J_i represents the rate at which information is absorbed from the i^{th} channel and R_i the necessary information transmitted per unit time, Eq. (34) represents the sum of the relative efficiencies of each of the instruments. It is possible, of course, for these relative efficiencies to exceed unity since R_i is only a minimum information rate and since the information absorbed may specify the channel input to a better precision than represented by R_i.

Two other possible sampling criteria are:

$$C_2 = \sum_{i=1}^{N} (J_i - R_i)^2 \qquad (35)$$

$$C_3 = \sum_{i=1}^{N} J_i \qquad (36)$$

The criterion of Eq. (35) has the disadvantage of counting $(J_i - R_i) = a$ equally as bad as $(J_i - R_i) = -a$ which seems rather unrealistic. The criterion of Eq. (36) is unrealistic because it is independent of the information rates of the instruments.

V. Recommendations and Conclusions

In the previous sections an over-all model (Fig. 1b) for the human instrument monitor has been presented along with some models for the individual components of the over-all model. The validity of these models can be verified only through controlled experiments. Some experiments for this purpose are described below.

Sampling models -- Experiments for the testing of stationary, sampling models should be designed so that the sampling criterion does not cause any changes in behavior that are dependent on the instrument readings. In terms of the previous section this corresponds to insuring a constant fidelity measure, E. An example of such an experiment would be one in which the subject is instructed to monitor several instruments and occasionally asked to specify the latest reading of each of them. It should be possible, then, to study the sampling criterion by observing changes in behavior as the instrument information rates are changed (e.g., changing W_i and S_i in Eq. (32)).

Data channel models -- The data channel model of Section III is more difficult to investigate experimentally because of the decision function that is interposed (see Fig. 1b). This difficulty can be somewhat reduced by making the decision function as simple as possible. As an example, consider the experiment in which the subject is allowed to view a single instrument for a short amount of time through a shutter arrangement. If he is then asked to specify the instrument reading either verbally or manually, the distribution of the error between true and estimated instrument reading would give some measure of the channel characteristics such as $f_c(y|x)$ for the memoryless case. If the Gaussian channel of Eq. (16) proves to be a good model, then by varying the viewing time, the function N(t) can be measured.

It is a dangerous business, of course, to propose models for the real world with only one's intuition and other people's experiments as justification. Still, it is felt that the Markov models of Section II are general enough to model very well any stationary sampling behavior of an instrument monitor. As mentioned earlier, it may be necessary to go to time-varying Markov process for non-stationary behavior.

The simple memoryless channel model for the data channel in Section III is more open to question. It is most likely a good model for those situations in which $x(t)$ does not vary much during the observation times (t_j in Fig. 9) and in which successive samples of x (x_j in Fig. 9) are statistically independent. When these conditions are not satisfied a more complex data channel model will probably be necessary. A good example of the latter limitation is one in which $x(t)$ is a constant for many observations and is then suddenly increased. In this situation the perceived reading, y, will very likely be influenced by the unvarying past values of x; and thus, a model with some form of memory will be appropriate.

There are some important uses for the sampling models of Section II even without the rest of the models. For example, it may be possible to characterize the skill of individual monitors by the parameters of their sampling model (e.g., the P and H matrices for the semi-Markov models). These results could then be used to direct the training procedures for the monitors. Furthermore, if it is found that skilled monitors generally have similar transition characteristics, then an initial design for an instrument panel could be based on a minimization of the average distance that the eyes must move per second in monitoring the instruments.

REFERENCES

1. Bartlett, M., "The frequency goodness of fit test for probability chains," Cambridge Philosophical Society Proceedings, 47:86-95, 1951.

2. Billingsley, P., "Statistical methods in Markov chains," Annals of Mathematical Statistics (U.S.), 32:12-40, 1961.

3. Feller, W., An Introduction to Probability Theory and Its Applications, Vol. 1, Wiley and Sons, New York, 1950.

4. Fano, R., Transmission of Information: A Statistical Theory of Communication, M.I.T. Press, Cambridge, Mass. 1961.

5. Howard, R., Dynamic Programming and Markov Processes, Technology Press and Wiley and Sons, New York, 1960.

6. Howard, R., Dynamic Probabilistic Systems, in preparation.

7. Mason, S. and Zimmerman, H., Electronic Circuits, Signals, and Systems, Wiley and Sons, New York, 1960.

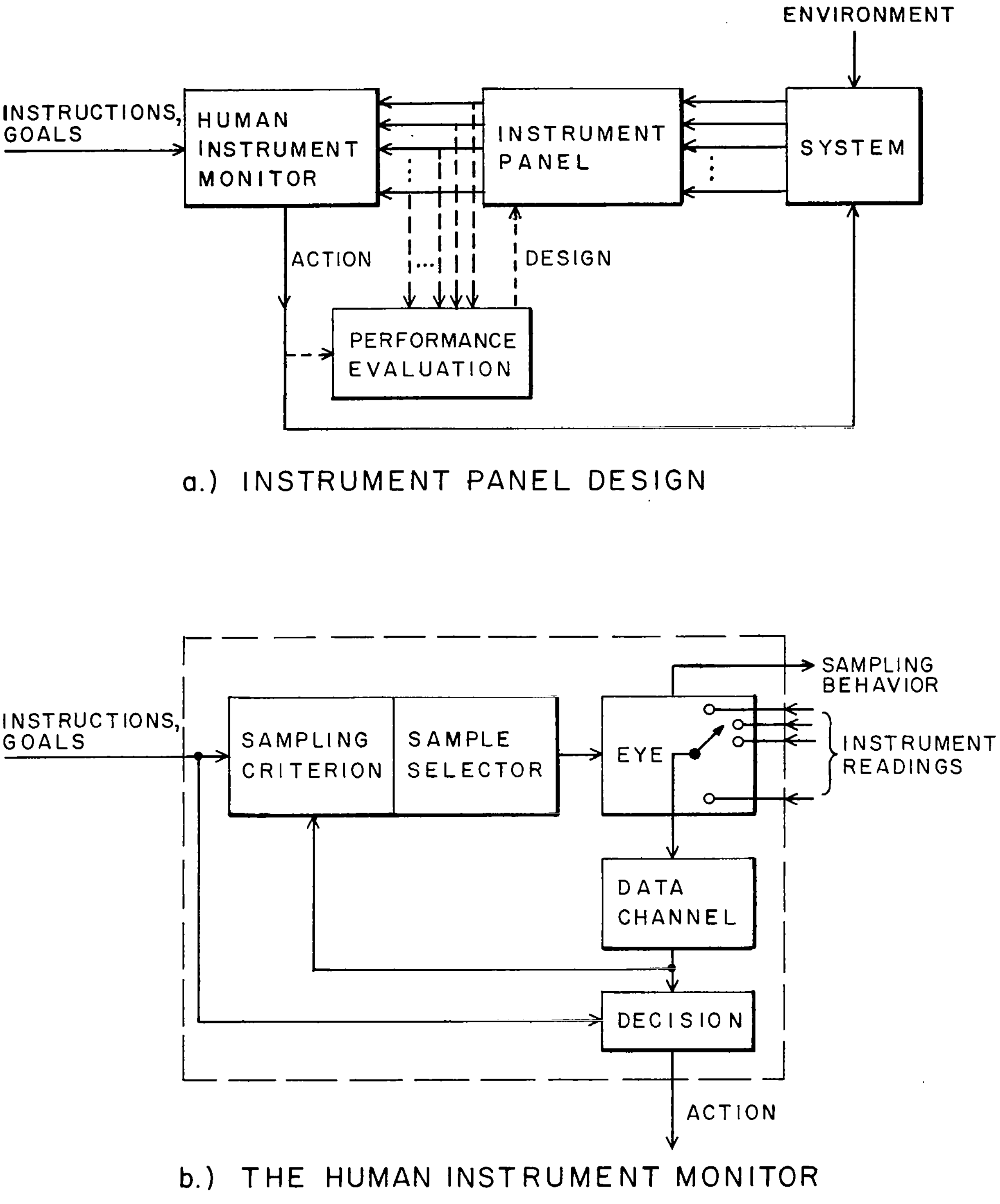
8. Senders, J.,"The human operator as a monitor and controller of multi-degree-of-freedom systems,"Paper presented at the Fourth National Symposium on Human Factors in Electronics, May 2-3, 1963.

9. Shannon, C.,"A mathematical theory of communication, Bell Systems Technical Journal, 27: 379-423, 623-656, 1948.

10. Silver, E., "Markovian decision processes with uncertain transition probabilities and rewards," Sc.D. Thesis,

REFERENCES (continued)

M. I. T., August 1963; also Interim Technical Report No. 1, Research in the Control of Complex Systems, Contract Nonr-1841(87), ONR, Operations Research Center, M. I. T., August 1963.



a.) INSTRUMENT PANEL DESIGN

b.) THE HUMAN INSTRUMENT MONITOR

FIG. 1 A BLOCK DIAGRAM OF THE PROBLEM

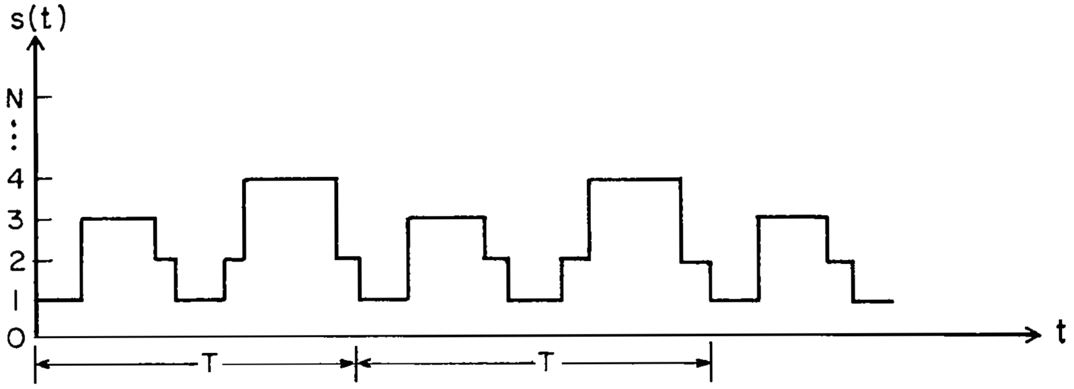
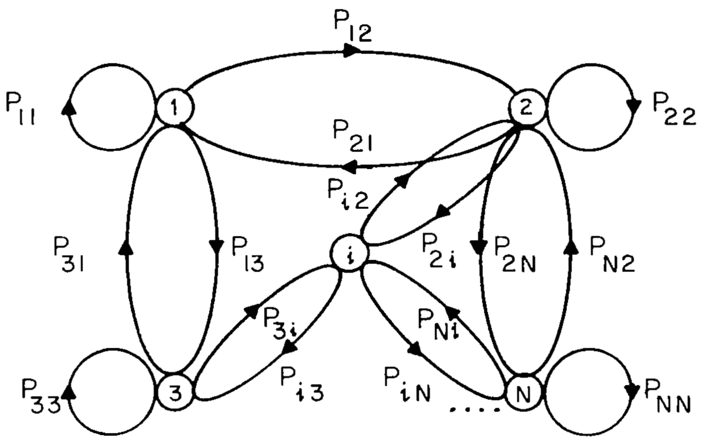


FIG. 2 s(t) FOR THE PERIODIC MODEL



a.) GRAPHICAL REPRESENTATION OF A MARKOV PROCESS

$$P = \begin{bmatrix} P_{11} & P_{12} & - - - - - - & P_{1N} \\ P_{21} & P_{22} & & \\ | & & & \\ | & & & \\ | & & & \\ P_{N1} & & & P_{NN} \end{bmatrix}$$

b.) MATRIX REPRESENTATION OF A MARKOV PROCESS

FIG. 3 MARKOV PROCESS REPRESENTATION

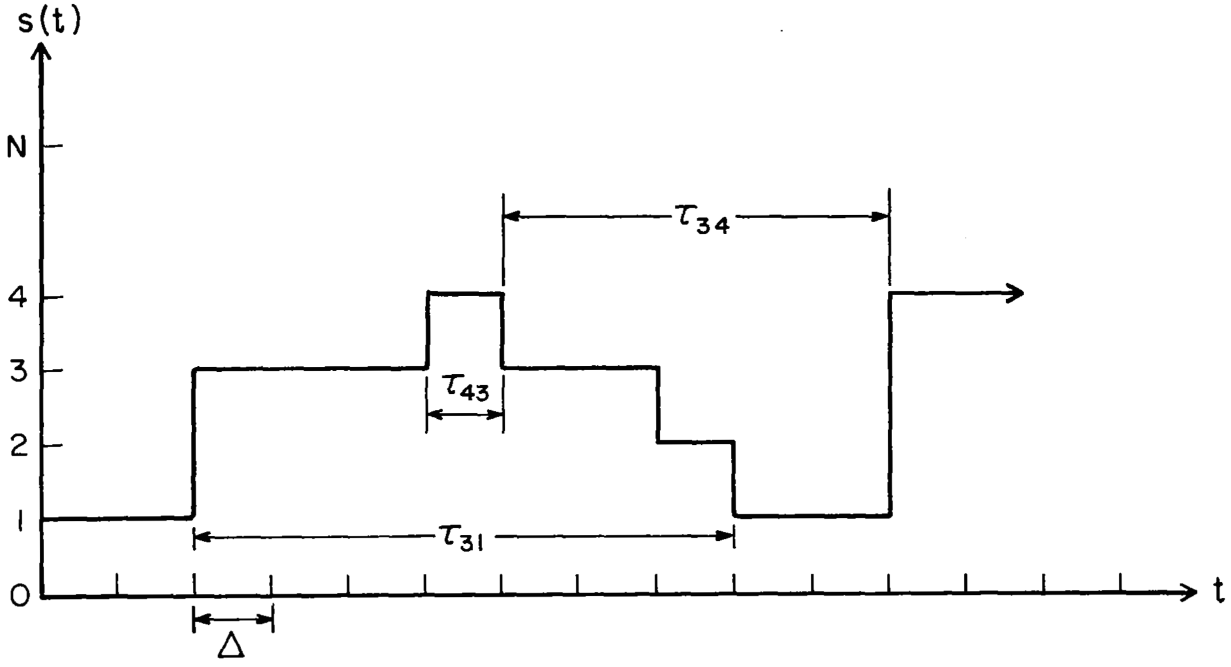


FIG. 4 s(t) FOR THE DISCRETE TIME MARKOV MODEL

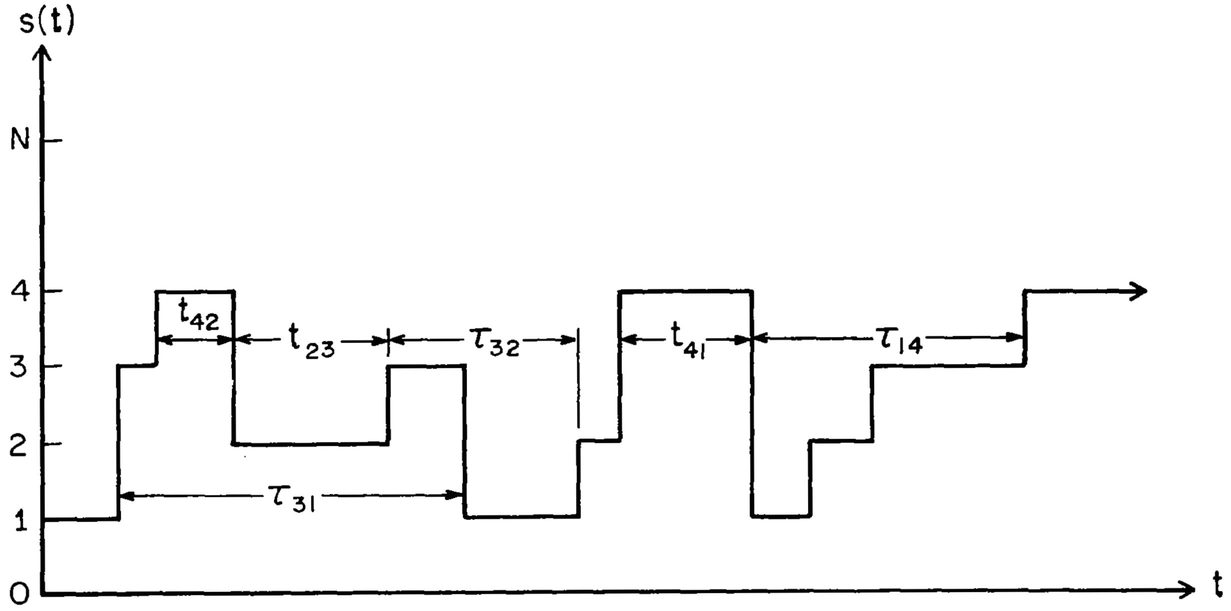


FIG. 5 s(t) FOR THE SEMI-MARKOV MODEL

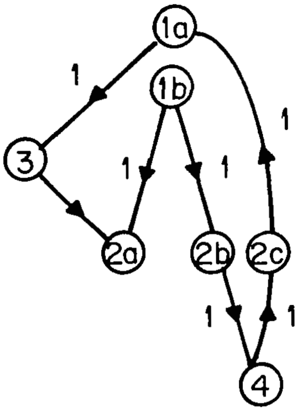


FIG. 6 MARKOV MODEL FOR THE PERIODIC MODEL OF FIG. 2

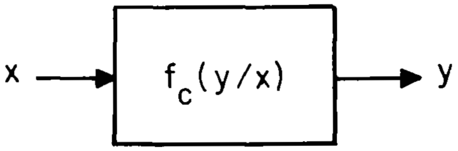


FIG. 7 THE DATA CHANNEL MODEL

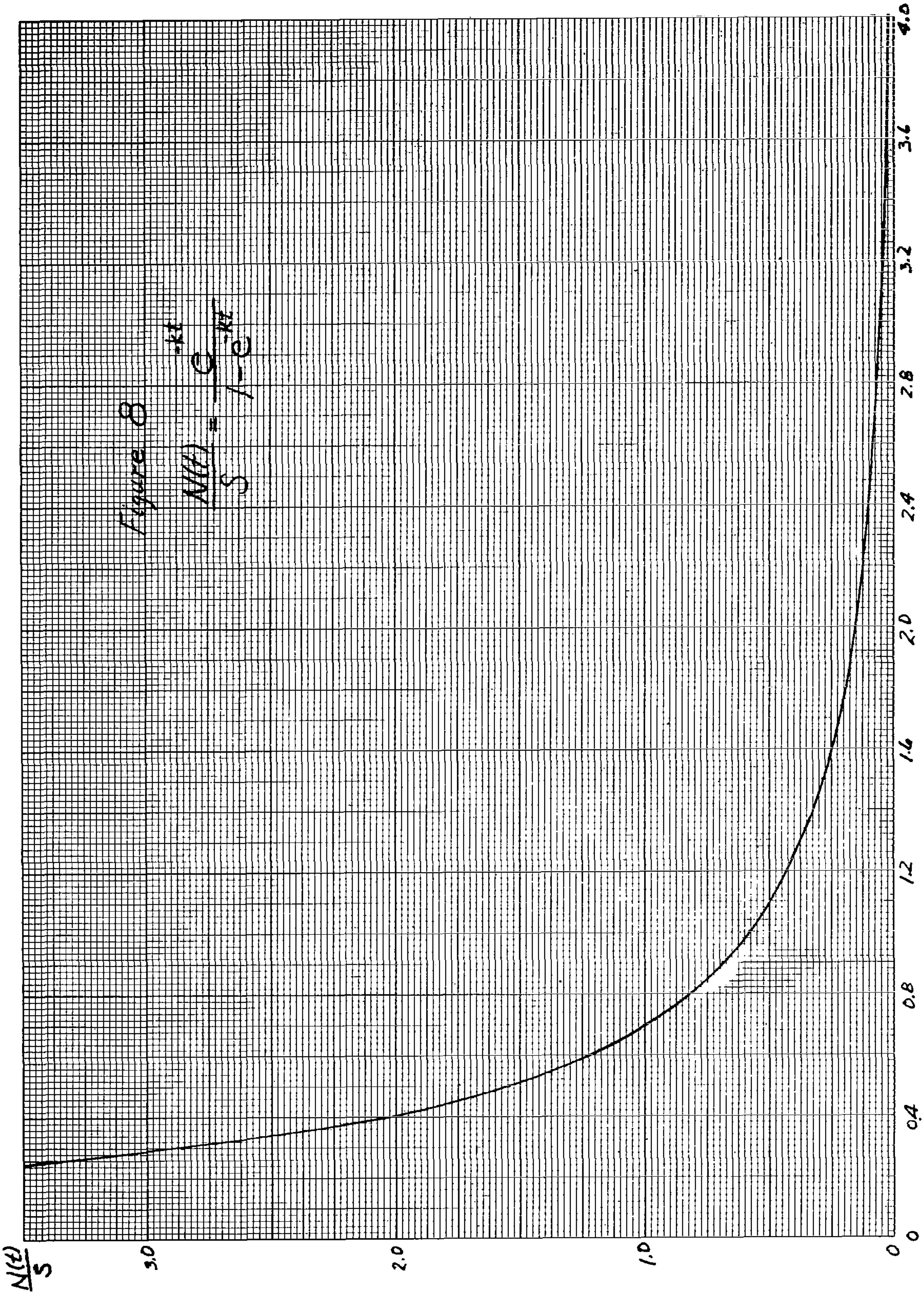


Figure 8

$$\frac{N(t)}{S} = \frac{e^{-kt}}{1-e^{-kt}}$$

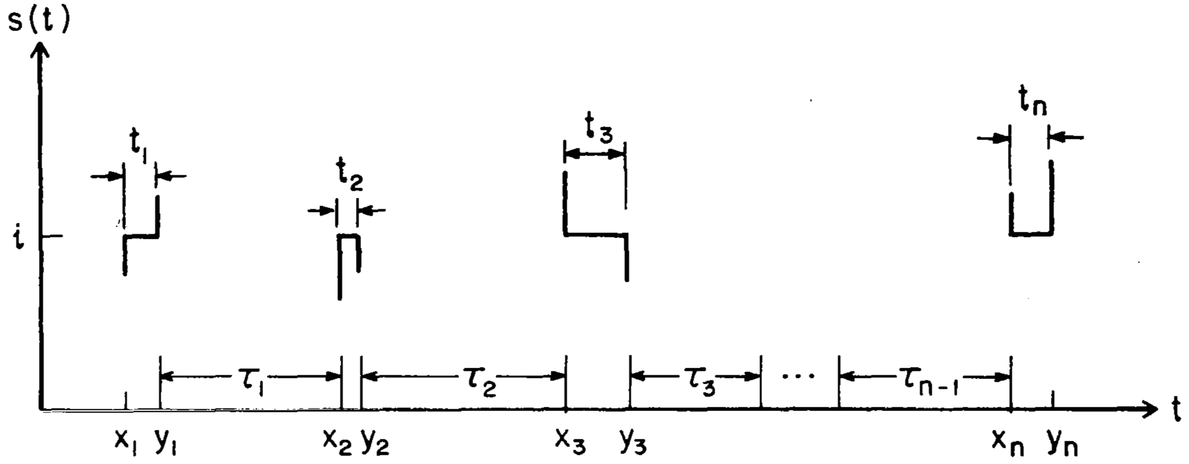
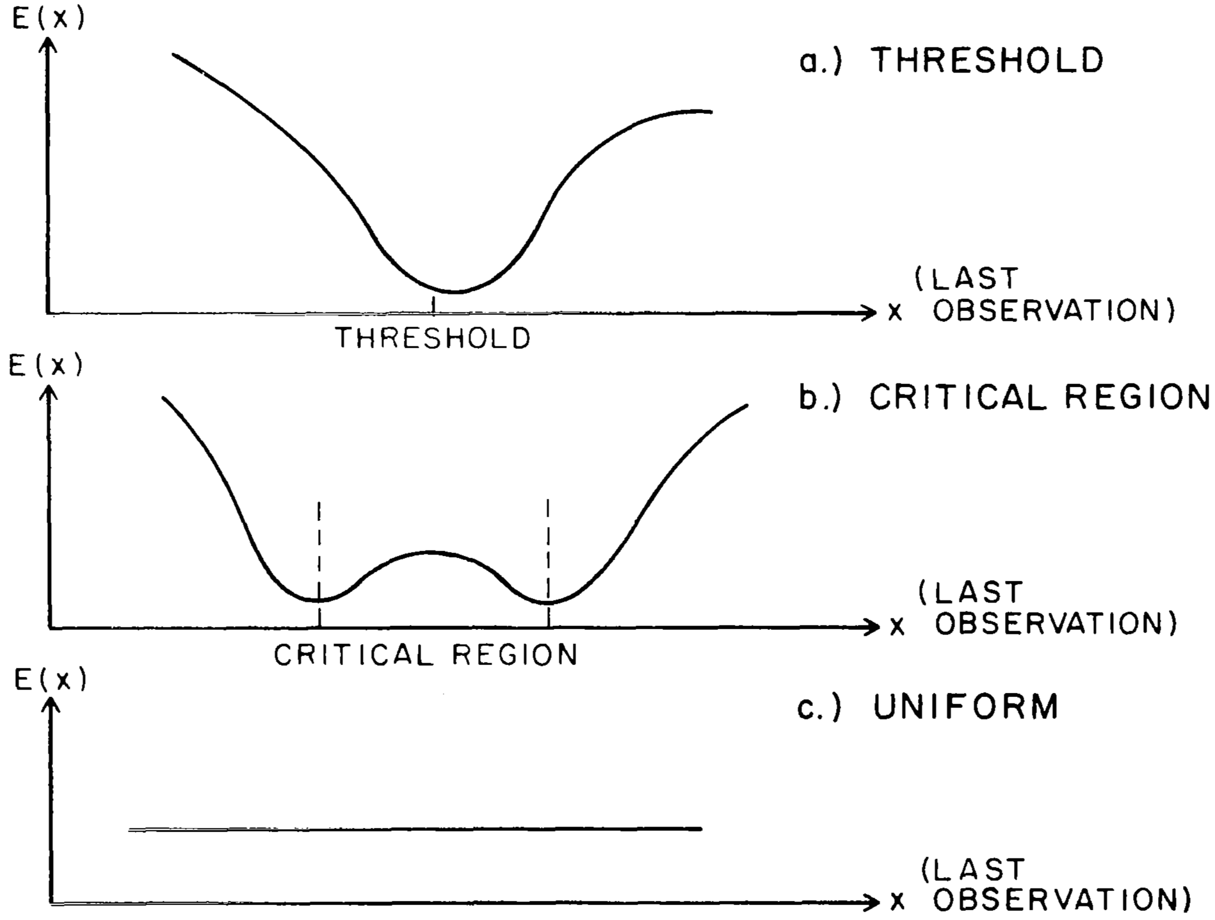


FIG. 9 n OBSERVATIONS OF THE iTH INSTRUMENT


a.) THRESHOLD

b.) CRITICAL REGION

c.) UNIFORM

FIG. 10 SOME FIDELITY MEASURES

APPENDIX A

THE DISCRETE TIME MARKOV PROCESS

In Appendices A-D some of the results of the theory of discrete time and semi-Markov processes will be summarized. It is impossible, however, in this short space to present anything approaching a comprehensive coverage of the subject. We shall merely develop the tools and ideas necessary for this particular application of Markov processes. For a very complete and lucid discussion of both discrete time and semi-Markov processes, the reader is referred to Howard's new book (6). Markov processes are also discussed in (3) and (5).

The quantity that is basic to the analysis of a discrete time Markov process is the multi-step transition probability defined by:

$$\phi_{ij}(n) = \Pr[s(n\Delta) = j|s(0) = i] \qquad (A.1)$$

That is, $\phi_{ij}(n)$ is the probability that the system is in the j^{th} state n transitions after it is in the i^{th} state. For n=0 and n=1, $\phi_{ij}(n)$ is easy:

$$\phi_{ij}(0) = \delta_{ij} \text{ and } \phi_{ij}(1) = p_{ij} \qquad (A.2)$$

We can also write a recursion relation for $\phi_{ij}(n)$:

$$\phi_{ij}(n) = \sum_{k=1}^{N} \Pr[s(n\Delta-\Delta)=k,s(n\Delta)=j|s(0)=i]$$

$$= \sum_{k=1}^{N} \Pr[s(n\Delta-\Delta)=k|s(0)=i] \Pr[s(n\Delta)=j|s(n\Delta-\Delta)=k,s(0)=i]$$

$$\phi_{ij}(n) = \sum_{k=1}^{N} \phi_{ik}(n-1)p_{kj} \qquad n\geq 1 \qquad (A.3)$$

The final step is valid because of the Markovian assumption. In matrix notation we have

$$\Phi(n) = \Phi(n-1)\, P \qquad n \geq 1 \qquad (A.4)$$

and by combining Eqs.(A.2) and (A.4) we have:

$$\begin{aligned} \Phi(0) &= I \\ \Phi(1) &= P \\ \Phi(2) &= P^2 \\ \Phi(n) &= P^n \end{aligned} \qquad (A.5)$$

where I is the N by N identity matrix.

Thus, Eq. (A.5) allows us to write $\Phi(n)$ for any value of n if we are patient enough to perform (n-1) matrix multiplications. There is an easier method, however, that allows us to find a closed form expression for $\Phi(n)$. In order to present this method, we define the z transform of a discrete time function, f(n), as:

$$f^T(z) = \sum_{n=0}^{\infty} f(n)\, z^n \qquad (A.6)$$

A table of useful transforms is presented in Table A-1. While a more rigorous definition of the z transform would have to consider contour integration for the inverse operation, we shall be satisfied with Table A-1 and the knowledge that its use will never give us erroneous results.

If we now multiply Eq. (A.3) by z^n and sum from n=1 to n=∞ we get:

$$\phi^T_{ij}(z) - \delta_{ij} = z \sum_{k=1}^{N} \phi^T_{ik}(z)\ p_{kj}$$

or

$$\phi^T_{ij}(z) = \delta_{ij} + \sum_{k=1}^{N} \phi^T_{ik}(z)\ (zp_{kj}) \qquad (A.7)$$

where we have used the fact that $\phi_{ij}(0) = \delta_{ij}$.

In matrix form Eq. (A.7) becomes:

$$\Phi^T(z) = I+z\Phi^T(z)\ P$$

or

$$\Phi^T(z) = [I-zP]^{-1} \qquad (A.8)$$

where $\Phi^T(z)$ is the matrix of the transforms of the $\phi_{ij}(n)$'s. Thus, a closed form expression for $\Phi(n)$ can be found by taking the inverse transform of the individual components of the right hand side of Eq. (A.8).

As an aside, it is worth noting that one can also use the flow graph technique (5, 6, 7) to solve the set of equations in Eq. (A.7). The pertinent flow graph for the 2 state example in the next paragraph is shown in Fig. (A.1). This technique is especially useful if only one of the ϕ_{ij}'s is needed.

A simple example of a two state process is:

$$P = \begin{bmatrix} .8 & .2 \\ .3 & .7 \end{bmatrix} \qquad (A.9)$$

Applying Eq. (A.8) we have:

$$\Phi^{T}(z) = \begin{bmatrix} 1-.8z & -.2z \\ -.3z & 1-.7z \end{bmatrix}^{-1} = \frac{1}{(1-z)(1-.5z)} \begin{bmatrix} 1-.7z & .2z \\ .3 & 1-.8z \end{bmatrix}$$

$$= \frac{1}{1-z} \begin{bmatrix} .6 & .4 \\ .6 & .4 \end{bmatrix} + \frac{1}{(1-.5z)} \begin{bmatrix} .4 & -.4 \\ -.6 & .6 \end{bmatrix}$$

and using Table A.1 we have

$$\Phi(n) = P^{n} = \begin{bmatrix} .6 & .4 \\ .6 & .4 \end{bmatrix} + (1/2)^{n} \begin{bmatrix} .4 & -.4 \\ -.6 & .6 \end{bmatrix} \qquad (A.10)$$

Notice in Eq. (A.10) that $\phi_{ij}(n)$ approaches a steady state value for large n that is independent of the starting state, i. This will generally be true of all the processes used for sampling models (more rigorously it is true of all ergodic processes (3, 5)), and we shall label these steady state probabilities as:

$$\pi_j = \lim_{n \to \infty} \phi_{ij}(n) \qquad (A.11)$$

A simple method for calculating the steady state probabilities can be derived from Eq. (A.3) by letting $n \to \infty$ in that equation to yield:

$$\pi_j = \sum_{j=1}^{N} \pi_k p_{kj} \qquad (A.12)$$

The N equations of Eq. (A.12) are not independent since the sum of any (N-1) of them yields the $N^{\underline{th}}$ one. Therefore, to Eq. (A.12) must be added the obvious requirement:

$$\sum_{i=1}^{N} \pi_i = 1 \qquad (A.13)$$

and these equations can then be used to find the steady state probabilities.

Applying Eqs. (A.12) and (A.13) to our example we have:

$$\pi_1 = .8\pi_1 + .3\pi_2$$

$$\pi_1 + \pi_2 = 1$$

The solution of these equations is:

$$\pi_1 = .6 \quad , \quad \pi_2 = .4$$

which is obviously consistent with our previous results in Eq. (A.10). It is worth noting that the Eqs.(A.12) and (A.13) will have a unique solution if and only if a steady state behavior makes any sense for the process (i. e., if it is an ergodic process).

	TIME FUNCTION	**z TRANSFORM**	
1.	$f(n)$	$f^{\tau}(z) = \sum_{n=0}^{\infty} f(n)\, z^{n}$	
2.	$nf(n)$	$z \frac{d}{dz}\, f^{\tau}(z)$	
3.	$a^{n}f(n)$	$f^{\tau}(az)$	
4.	$f(n+k)$	$z^{-k}[f^{\tau}(z) - \sum_{\ell=0}^{k-1} z^{\ell} f(\ell)]$	$k>0$
5.	$f(n-k)$ *	$z^{k}\, f^{\tau}(z)$	$k>0$
6.	$\sum_{m=0}^{n} f(m)\, g(n-m)$	$f^{\tau}(z)g^{\tau}(z)$	
7.	$\sum_{m=0}^{n} f(m)$	$(1-z)^{-1}f^{\tau}(z)$	
8.	a^{n}	$(1-az)^{-1}$	
9.	1 (unit step)	$(1-z)^{-1}$	
10.	n	$z(1-z)^{-2}$	

* $f(n)$ is assumed equal to zero for $n<0$.

Table A-1. A Short Table of z Transforms

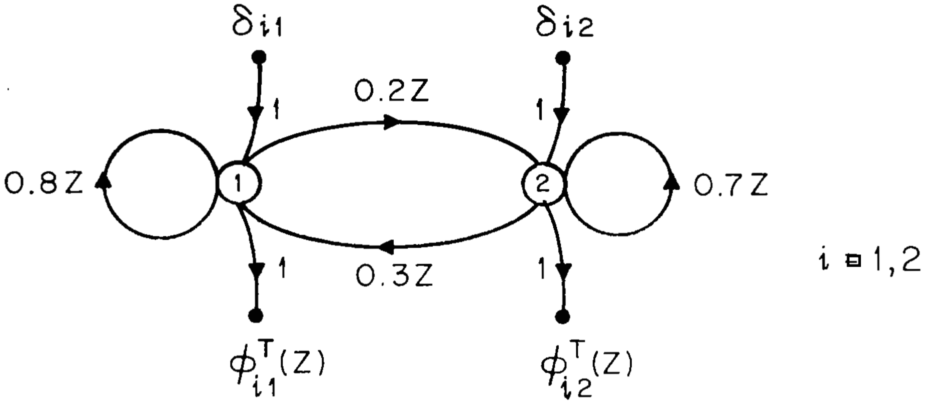


FIG. A-1 FLOW GRAPH FOR A DISCRETE TIME MARKOV PROCESS

APPENDIX B

FIRST PASSAGE TIMES FOR THE DISCRETE TIME MARKOV PROCESS

In this appendix we shall describe a method for finding the probability distributions for the first passage times of a discrete time Markov process. The first passage time, τ_{ij}, is defined as the time for the process to occupy the $j^{\underline{th}}$ state for the first time if it is in the $i^{\underline{th}}$ state now. We shall be interested in finding the probability function, $g_{ij}(\circ)$ for this random variable:

$$\begin{aligned} g_{ij}(n) &= \Pr[\tau_{ij}=n\Delta] \\ &= \Pr[s(n\Delta)=j,\ s(k\Delta)\neq j \| s(0)=i] \quad \text{for } 0<k<n \qquad \text{(B.1)} \end{aligned}$$

There is some ambiguity in this definition for $i=j$, and we shall assume that $\tau_{ii}=1$ if the system remains in the $i^{\underline{th}}$ state for the next time interval; i. e. $g_{ii}(1)=p_{ii}$.

We can write the multi-step transition probability, $\phi_{ij}(n)$, in terms of these probability functions

$$\begin{aligned} \phi_{ij}(n) &= \sum_{\ell=1}^{n} \Pr[s(\ell\Delta)=j,s(k\Delta)\neq j | s(0)=i] \\ &\qquad \Pr[s(n\Delta)=j | s(0)=i,s(\ell\Delta)=j,s(k\Delta)\neq j] \quad \text{for } 0<k<\ell \\ &= \sum_{\ell=1}^{n} g_{ij}(\ell)\phi_{jj}(n-\ell) \quad \text{for } n\geq 1 \qquad \text{(B.2)} \end{aligned}$$

and since $g_{i\ell}(0)=0$, Eq. (B.2) can also be written as:

$$\phi_{ij}(n) = \sum_{\ell=0}^{n} g_{ij}(\ell)\, \phi_{jj}(n-\ell) \qquad \text{for } n\geq 1 \qquad \text{(B.3)}$$

Taking the z transform of both sides of Eq. (B.3) we have:

$$\phi_{ij}{}^{T}(z) - \phi_{ij}(0) = g_{ij}{}^{T}(z)\, \phi_{jj}{}^{T}(z) \qquad \text{(B.4)}$$

where we have used Item. 6 in Table A-1 and the fact that $g_{ij}(0) = 0$. Solving Eq. (B.4) for $g_{ij}{}^{T}(z)$ gives:

$$g_{ij}{}^{T}(z) = \frac{\phi_{ij}{}^{T}(z) - \delta_{ij}}{\phi_{jj}{}^{T}(z)} \qquad \text{(B.5)}$$

Or in matrix form:

$$G^{T}(z) = [\Phi^{T}(z) - I]\ [\Phi^{T}(z)\Box I]^{-1} \qquad \text{(B.6)}$$

where the box notation in Eq. (B.6) denotes term by term matrix multipliciation. That is,

$$A\Box B = C \ \text{ implies } a_{ij}b_{ij} = c_{ij} \qquad \text{(B.7)}$$

Thus, Eq. (B.6) gives a method for calculating the probability distributions of the first passage times.

For the example in Appendix A we have:

$$G^T(z) = \frac{1}{(1-z)(1-.5z)} \begin{bmatrix} .8z-.5z^2 & .2z \\ .3z & .7z-.5z^2 \end{bmatrix} \begin{bmatrix} (\frac{1}{1-.7z}) & 0 \\ 0 & (\frac{1}{1-.8z}) \end{bmatrix} (1-z)(1-.5z)$$

$$= \begin{bmatrix} \frac{8z-.5z^2}{(1-.7z)} & \frac{.2z}{(1-.8z)} \\ \frac{.3z}{(1-.7z)} & \frac{.7z-.5z^2}{(1-.8z)} \end{bmatrix} \qquad (B.8)$$

and taking the inverse transform of Eq. (B.8) yields

$$\begin{aligned} G(0) &= 0 \\ G(1) &= P \\ G(n) &= \begin{bmatrix} .06(.7)^{n-2} & .2(.8)^{n-1} \\ .3(.7)^{n-1} & .06(.8)^{n-2} \end{bmatrix} \quad \text{for } n\geq 2 \qquad (B.9) \end{aligned}$$

APPENDIX C

THE SEMI-MARKOV PROCESS

In this appendix we shall derive some of the properties of semi-Markov processes that are analogous to those discussed in Appendix A. For the semi-Markov process, the quantity corresponding to the multi-step transition probability of Eq. (A.1) is:

$$\phi_{ij}(t) = \Pr[s(t)=j \mid \text{just } \underline{\text{entered}}\ i^{\underline{th}} \text{ state at } t=0] \qquad (C.1)$$

For the sake of generality we shall allow virtual transitions in our discussion (i.e., $p_{ii}\neq 0$); thus, it is possible in Eq. (C.1) for the system to have entered via a virtual transition. A relation for $\phi_{ij}(t)$ can be written as:

$$\begin{aligned}\phi_{ij}(t) = {} & \delta_{ij}\Pr[\text{stay in } i \text{ for greater than } t] \\ & + \sum_{k=1}^{N} \Pr[\text{next transition (real or virtual) is to } k] \\ & \int_{o}^{t} \Pr[\text{transition occurs at } \tau]\ \phi_{kj}(t-\tau) \qquad (C.2) \\ = {} & \delta_{ij} \sum_{k=1}^{N} p_{ik} \int_{t}^{\infty} h_{ik}(\zeta)d\zeta + \sum_{k=1}^{N} p_{ik} \int_{o}^{t} h_{ik}(\tau)\ \phi_{kj}(t-\tau)\ d\tau\end{aligned}$$

where δ_{ij} is the Kronecker delta function.

For continuous time processes the Laplace transform is the important transform:

$$f^{T}(s) = \int_{0}^{\infty} f(t)e^{-st} \; dt \qquad (C.3)$$

A table of useful Laplace transforms is presented in Table C-1.

Taking the Laplace transform of both sides of Eq. (C.2) gives:

$$\phi_{ij}{}^{T}(s) = \delta_{ij} \frac{1}{s} [1- \sum_{k=1}^{N} p_{ik}h_{ik}{}^{T}(s)] + \sum_{k=1}^{N} p_{ik}h_{ik}{}^{T}(s) \; \phi_{kj}{}^{T}(s) \qquad (C.4)$$

In order to write Eq. (C.4) in a convenient matrix form it is necessary to introduce the random variable t_i, the unconditional dwell time. This is the time that the system spends in the $i^{\underline{th}}$ state during any one occupancy unconditioned by any subsequent state. The probability density function for t_i is

$$w_i(t_i) = \sum_{j=1}^{N} p_{ij}h_{ij}(t_i) \qquad (C.5)$$

that is, it is just the sum of the conditioned dwell time density functions weighted with the probability of their relevance. The expected value of t_i was written in Eq.(10) and is:

$$\overline{t}_i = \int_{0}^{\infty} \zeta \, w_i(\zeta) \, d\zeta = \int_{0}^{\infty} \zeta \sum_{j=1}^{N} p_{ij}h_{ij}(\zeta)d\zeta = \sum_{j=1}^{N} p_{ij}\overline{t}_{ij} \qquad (C.6)$$

We shall define an N by N matrix, W(t), for which the diagonal terms are just the $w_i(t)$'s and the off diagonal terms are zero:

$$W(\cdot) = \left[w_{ij}(\cdot) = \delta_{ij}w_i(\cdot)\right] \tag{C.7}$$

The matrix form of Eq. (C.4) can now be written as:

$$\Phi^{T}(s) = \frac{1}{s}\,[I - W^{T}(s)] + [P\Box H^{T}(s)]\;\;\Phi^{T}(s) \tag{C.8}$$

where H(t) is the matrix of conditioned dwell times, $h_{ij}(t)$, and the box operation was defined in Eq. (B.7) of Appendix B. Solving Eq. (C.8) for $\Phi^{T}(s)$ we obtain:

$$\Phi^{T}(s) = \frac{1}{s}\,[I - P\Box H^{T}(s)]^{-1}\,[I - W^{T}(s)] \tag{C.9}$$

Thus, Eq. (C.5) and Eq. (C.9) can be used to find the Laplace transform of $\Phi(t)$ from the P and H(t) matrices. The inverse transforms of the individual terms of $\Phi^{T}(s)$ will then yield a closed form expression for $\Phi(t)$.

As in the case for discrete Markov processes, it is also possible to find any or all of the $\phi_{ij}{}^{T}(s)$'s by using flow graphs to solve the set of equations in Eq. (C.4).

The semi-Markov process also exhibits a steady state behavior (again under the assumption of ergodicity). We can find the steady state value of $\phi_{ij}(t)$ by applying the final value theorem to Eq. (C.9).

$$\Phi = \lim_{s \to 0} [s\Phi^T(s)] = \lim_{s \to 0} \left\{ s[I-P\Box H^T(s)]^{-1} \frac{1}{s}[I-W^T(s)] \right\} \tag{C.10}$$

The last term in Eq. (C,10) is a diagonal matrix, M, with terms:

$$\delta_{ij} \lim_{s \to 0} \frac{1}{s}[1-W_i^T(s)] = \delta_{ij} \lim_{s \to 0} \frac{1}{s}\,[1- \sum_{k=1}^{N} p_{ik}h_{ik}^T(s)]$$

$$= \delta_{ij} \lim_{s \to 0} \frac{-\sum_k p_{ik}(dh_{ik}^T(s)/ds)}{} = \delta_{ij} \sum_k p_{ik}\bar{t}_{ik}$$

$$= \delta_{ij}\bar{t}_i \tag{C.11}$$

where we have used L'Hopital's rule and Item 2 of Table C-1. The first term of Eq. (C.10) can be evaluated by defining the matrix:

$$A(s) = s[I- P\Box H^T(s)]^{-1}$$

Than we have:

$$A(s)\ [I - P\Box H^T(s)] = sI$$

$$A(s) = sI + A(s)\ (P\ \ H^T(s))$$

which has the limiting value:

$$A(0) = A(0)\ P \tag{C.12}$$

Eq. (6.12) can be expressed in the individual terms of A(0) by

$$a_{ij} = \sum_{k=1}^{N} a_{ik}\, p_{kj} \tag{C.13}$$

Each value of i in Eq. (C.13) produces a set of equations identical in form to Eq. (A.12). Thus, the solution is:

$$a_{ij} = K\pi_j \tag{C.14}$$

where the π_i's are the solution of Eqs. (A.12) and (A.13). For the semi-Markov process π_i represents the fraction of all <u>transitions</u> in the steady state that enter (or leave) the i^th state. Combining Eqs. (C.10), (C.11), and (C.14) we can write the steady state probabilities for the semi-Markov process as:

$$\phi_j = K\pi_j\bar{t}_j \tag{C.15}$$

and evaluating K we have:

$$\phi_j = \frac{\pi_j\bar{t}_j}{\sum_{k=1}^{N} \pi_k\bar{t}_k} \tag{C.16}$$

The ϕ_j in Eq. (C.16) represents the probability of finding the process in the j^th state after the process has been running for a long time. As represented by Eq. (C.15) ϕ_j is proportional to the probability that the last transition was to the j^th state and the expected time-spent in the state before the next transition.

	TIME FUNCTION	LAPLACE TRANSFORM
1.	$f(t)$	$f^{\tau}(s) = \int_{0}^{\infty} f(t)e^{-st}\,dt$
2.	$tf(t)$	$-\frac{df^{\tau}(s)}{ds}$
3.	$f(t+\tau)$	$e^{s\tau}[f^{\tau}(s)-\int_{0}^{\tau} f(t)e^{-st}dt]$ $\tau>0$
4.	$f(t-\tau)$ *	$e^{-s\tau}f^{\tau}(s)$
5.	$\int_{0}^{t} f(\zeta)g(t-\zeta)d\zeta$	$f^{\tau}(s)g^{\tau}(s)$
6.	$\int_{0}^{t} f(\zeta)d\zeta$	$\frac{1}{s} f^{\tau}(s)$
7.	$e^{-\lambda t}$	$\frac{1}{s+\lambda}$
8.	1 (unit step)	$\frac{1}{s}$
9.	t	$\frac{1}{s^{2}}$

* $f(t)$ is assumed equal to zero for $t<0$.

Table C.1 A Short Table of Laplace Transforms

APPENDIX D

FIRST PASSAGE TIMES FOR A SEMI-MARKOV PROCESS

In this appendix we shall describe a method for obtaining the probability density functions of the first passage times. For the semi-Markov process, the random variable, τ_{ij}, is defined as the time for the process to enter the j^{th} state for the first time if it has just entered the i^{th} state. For i=j, we shall count a virtual transition as an entry into the state. The density function for τ_{ij} will be denoted by $g_{ij}(\cdot)$.

Following the procedure of Appendix B we can write an expression for $\phi_{ij}(t)$ in terms of $g_{ij}(t)$:

$$\phi_{ij}(t) = \int_0^t g_{ij}(\tau)\, \phi_{ij}(t-\tau)\, d\tau + \delta_{ij} \sum_{k=1}^{N} p_{ik} \int_t^{\infty} h_{ik}(\zeta) d\zeta \quad (D.1)$$

The last term in Eq. (D.1) is just the probability of remaining in the i^{th} state longer than t, and must be added in if i=j. Using Items 5 and 6 in Table C.1, and the fact that $h_{ik}(t)$ is a probability density function, we can write the Laplace transform of $\phi_{ij}(t)$ as:

$$\phi_{ij}^{\tau}(s) = g_{ij}^{\tau}(s)\, \phi_{jj}^{\tau}(s) + \delta_{ij} \frac{1}{s} [1-w_i^{\tau}(s)] \qquad (D.2)$$

Solving Eq. (D.2) for $g_{ij}^{\tau}(s)$ we have:

$$g_{ij}^{\tau}(s) = \frac{1}{\phi_{jj}^{\tau}(s)} \left[\phi_{ij}^{\tau}(s) - \frac{\delta_{ij}}{s} [1-w_i^{\tau}(s)]\right] \qquad (D.3)$$

or in matrix form:

$$G^T(s) = \left[\Phi^T(s) - \frac{1}{s}[I - W^T(s)]\right] [\Phi^T(s) \square I]^{-1} \qquad (D.4)$$

where $G^T(s)$ is the matrix of the transforms of the $g_{ij}(\circ)$'s and $W(\circ)$ is defined in Eqs. (C.5) and (C.7). The box notation is defined in Eq. (B.7). Thus, Eq. (D.4) gives us a method for calculating $G^T(s)$ and the inverse Laplace transform of the elements of this matrix will yield the required density functions.

APPENDIX E

AN ILLUSTRATIVE EXAMPLE

In this appendix we shall calculate $I_i(X_n;Y_n)$ and $J_i(X_n;Y_n)$ for a simple but non-trivial situation. The assumptions that will be made concerning the sampling model, data channel model, and instrument readings are as follows:

Sampling model - A periodic sampling model with period T will be used. Furthermore, it will be assumed that each instrument is sampled only once per period. The dwell time for the $i^{\underline{th}}$ instrument will be denoted t_i; thus,

$$\sum_{i=1}^{N} t_i = T \tag{E.1}$$

where N is the number of instruments. The value of τ_i is also fixed:

$$\tau_i = T-t_i \tag{E.2}$$

Data channel model - The continuous memoryless Gaussian channel of Eq. (16) will be used for the data channel model:

$$f_c(y|x,t) = f_N(y|x,N(t)) = [2\pi N(t)]^{-1/2}\exp[-\frac{(y-x)^2}{2N(t)}] \tag{E.3}$$

We will also define:

$$N(t_i) = N_i \tag{E.4}$$

Instrument readings - Each of the instrument readings will be assumed to originate from a continuous Gaussian, Markovian source. The average power and correlation coefficient for the $i^{\underline{th}}$ source will be denoted by S_i and $\rho_i(t)$, respectively. In other words if $x(\zeta)$ is the reading of the $i^{\underline{th}}$ instrument at time, ζ:

$$f_i(x) = f_N(x \,|\, 0, S_i) \tag{E.5}$$

$$f_i[x(\zeta_n)|x(\zeta_{n-1}),x(\zeta_{n-2}),\ldots x(\zeta_2),x(\zeta_1)] = f_N[x(\zeta_n)|\rho x(\zeta_{n-1}),(1-\rho^2)S_i] \tag{E.6}$$

where $\zeta_{j+1} > \zeta_j$ for $1 \leq j < n$ and $\rho = \rho_i(\zeta_n - \zeta_{n-1})$

In simpler terms the instrument readings will be assumed Gaussian with means and variances that are completely determined by the last observed instrument reading. The following notation will be used for the correlation coefficients:

$$\rho_i(\tau_i) = \rho_i \tag{E.7}$$

where τ_i is $(T-t_i)$ (Eq.(E.2)).

Under these assumptions the joint density function for the n instrument readings in Eq. (24) can be written as:

$$f_i(\underline{x_n}|\tau_i) = f_N(x_1|0,S_i) \prod_{k=2}^{n} f_N(x_k|\rho_i x_{k-1},(1-\rho_i{}^2)S_i) \tag{E.8}$$

The expression for $I_i(X_n;Y_n)$ in Eq. (29) reduces to:

$$I_i(X_n;Y_n) = \frac{1}{n} \int_{-\infty}^{\infty} f_i(\underline{x_n}|\tau_i)\, d\underline{x_n} \int_{-\infty}^{\infty} f(\underline{y_n}|\underline{x_n},t_i) dy_n$$

$$\sum_{k=1}^{N} \log\,[f_c(y_k|x_k,t_i)/f(y_k|\underline{y_{k-1}})] \qquad \text{(E.9)}$$

We shall concentrate our efforts on evaluating the $k^{\underline{th}}$ term in the summation of Eq. (E.9). We shall label this term $I_i^{(k)}$. Physically it corresponds to the expected amount of information to be absorbed in the $k^{\underline{th}}$ observation of the $i^{\underline{th}}$ instrument. We can integrate out the expressinns involving $(x_{k+1},y_{k+1},\ldots x_n,y_n)$ since the argument of the logarithm for $I_i^{(k)}$ will not involve these random variables. The result is:

$$I_i^{(k)} = \int_{-\infty}^{\infty} f_i(\underline{x_k}|\tau_i)\, d\underline{x_k} \int_{-\infty}^{\infty} f(\underline{y_k}|\underline{x_k},t_i)\, d\underline{y_k} \log[f_c(y_k|x_k,t_i)/f(y_k/\underline{y_{k-1}})]$$

$$= \int_{-\infty}^{\infty} f_i(\underline{x_k}|\tau_i)\, d\underline{x_k} \int_{-\infty}^{\infty} f(\underline{y_k}|\underline{x_k},t_i)\, d\underline{y_k}[\log f_c(y_k|x_k,t_i)-\log f(y_k|\underline{y_{k-1}})]$$

(E.10)

The first term in Eq. (E.10) can be evaluated by using Eq. (E.3) to integrate with respect to y_k first. The result of this fir integration is independent of x_k and so the rest of the integrations yield no further changes in the term. The result is:

$$I_i^{(k)} = -\frac{1}{2}\log 2\pi e N_i - \int_{-\infty}^{\infty} f_i(\underline{x_k}|t_i)\, dx_k \int_{-\infty}^{\infty} f(\underline{y_k}|\underline{x_k},t_i)d\underline{y_k}\log f(y_k|\underline{y_{k-1}})$$

(E.11)

The problem now is to determine the probability distribution, $f(y_k|\underline{y_{k-1}})$, of the observed reading y_k conditioned on all the previous readings. We can write this denisty function as

$$f(y_k|\underline{y_{k-1}}) = \int_{-\infty}^{\infty} f(x_k,y_k|\underline{y_{k-1}})dx_k = \int_{-\infty}^{\infty} f(x_k|\underline{y_{k-1}})f(y_k|x_k,\underline{y_{k-1}})dx_k \tag{E.12}$$

Since we have assumed that the data channel is memoryless, the distribution of y_k is completely defined by x_k. Therefore, the second term under the integral in Eq. (E.12) is just $f_c(y_k|x_k,t)$ and our problem has been reduced to finding $f(x_k|\underline{y_{k-1}})$. The following relations for this distribution can be proved:

$$f(x_k|\underline{y_{k-1}}) = f_N(x_k|m_k,v_k) \tag{E.13}$$

where

$$m_k = \rho_1 \frac{y_{k-1}v_{k-1}+m_{k-1}N_1}{v_{k-1}+N_1} \tag{E.14}$$

$$v_k = \frac{S_1(1-\rho_1^2)N_1+\rho_1^2 v_{k-1}N_1+v_{k-1}S_1(1-\rho_1^2}{v_{k-1}+N_1}$$

$$= S_1(1-\rho_1^2) + \frac{\rho_1^2 N_1 v_{k-1}}{v_{k-1}+N_1} \tag{E.15}$$

The distribution of x_k for k=1 is given by Eq. (E.5), so that

$$m_1 = 0 \text{ and } v_1 = S_1 \tag{E.16}$$

Thus, Eqs. (E.13) to (E.16) completely define the distribution of x_k conditioned upon the k-1 previous observations. An important property of the variance in Eq. (E.15) is that it is independent of the k-1 observations (y's). As a check, we also notice that all the $v_{k's}$ are equal to S_1 if $\rho_1=0$. This is as one would expect and corresponds to the white noise example discussed in Section III.

The proof of Eqs. (E.14) and (E.15) proceeds by induction; we shall show that if $f(x_{k-1}|y_{k-2})$ is a normal distribution with mean m_{k-1} and variance v_{k-1}, then Eqs. (E.13) to (E.15) are true. First of all we use Bayes rule to express $f(x_k|y_{k-1})$ in the form:

$$f(x_k|\underline{y_{k-1}}) = \frac{f(x_k, y_{k-1}!\ \underline{y_{k-2}})}{f(y_{k-1}|\underline{y_{k-2}})} \tag{E.17}$$

$$= \frac{1}{f(y_{k-1}|\underline{y_{k-2}})} \int_{-\infty}^{\infty} f(x_k, y_{k-1}, x_{k-1}|\underline{y_{k-2}})\, dx_{k-1}$$

$$= \frac{1}{f(y_{k-1}|\underline{y_{k-2}})} \int_{-\infty}^{\infty} f(x_{k-1}|\underline{y_{k-2}}) f(y_{k-1}|x_{k-1}, \underline{y_{k-2}})$$

$$f(x_k|x_{k-1}, \underline{y_{k-1}})\, dx_{k-1}$$

But the second term under the integral in Eq. (E.17) is just $f_c(y_{k-1}|x_{k-1}, t)$ by the same argument used for Eq. (E.12). The third term reduces to $f_N(x_k|\rho_1 x_{k-1}, (1-\rho_1^2)S_1)$ because of the Markovian assumption concerning the source (See Eq. (E 6)). The denominator in Eq. (E.17) is just a normalizing factor and can be found by integrating the numerator with respect to x_k. Thus, Eq. (E.17) reduces to:

$$f(x_k|y_{k-1}) = \frac{\int_{-\infty}^{\infty} f_N(x_{k-1}|m_{k-1}, v_{k-1}) f_N(y_{k-1}|x_{k-1}, N_1) f_N(x_k|\rho_1 x_{k-1}, (1-\rho_1^2)S_1)\, dx_{k-1}}{\int_{-\infty}^{\infty} [\text{numerator}]\, dx_k} \tag{E.18}$$

after a moderate amount of algebra, Eq. (E.18) yields the relations expressed in Eqs. (E.13) to (E.15).

Equations (E.12) and (E.13) can be combined to give:

$$f(y_k|\underline{y_{k-1}}) = f_N(y_k|m_k, v_{k+} N_1) \tag{E.19}$$

If we now use the results of Eq. (E.19) in Eq. (E.11) and integrate with respect to y_k first we find that because v_k is independent of $\underline{y_{k-1}}$ and $\underline{x_k}$ the result of this first integration is also independent of $\underline{y_{k-1}}$ and $\underline{x_k}$. Thus we find:

$$I_1^{(k)} = \frac{1}{2} \log \left(1+ \frac{v_k}{N_1}\right) \tag{E.20}$$

And if we now insert this result back into Eq. (E.19) for $I_1(X_n;Y_n)$ we get:

$$I_1(X_n;Y_n) = \frac{1}{n} \sum_{k=1}^{N} I_1^{(k)} = \frac{1}{2n} \sum_{k=1}^{N} \log\left(1+ \frac{v_k}{N_1}\right) \tag{E.21}$$

As a further check, if we set $\rho_1 = 0$, Eqs. (E.20) and (E.21) reduce to:

$$I_1^{(k)} = I_1(X_n;Y_n) = \frac{1}{2} \log \left(1+ \frac{S_1}{N_1}\right) \tag{E.22}$$

which is identical with Eq. (18) as one would expect, since this is just the case of independent samples of white Gaussian noise. In this case v_k and $I_1^{(k)}$ are independent of k because of the independence of the instrument readings, $(x_1, x_2, \ldots x_k)$.

We are now in a position to look at the steady state value of $I_1(X_n;Y_n)$, that is, as n becomes very large. To do this, it will be necessary to find the limiting value of v_k in Eq. (E.15) as k becomes large. We can prove that v_k will always converge to this limiting value by observing that:

(a) $v_k > 0$ for $S_1 > 0$. This is obvious from Eq. (E.15)

(b) v_k is not oscillatory, i. e., the sign of $(v_{k+1}-v_k)$ is independent of k. This can be proved by observing that:

$$\frac{(v_{k+1}-v_k)}{(v_k-v_{k-1})} = \frac{\rho_1^2 \; N_1^2}{(N_1+v_k)(N_1+v_{k-1})} > 0 \qquad (E.23)$$

Thus, v_k is either forever increasing or forever decreasing with k. For $v_1 = S_1$ it will be forever decreasing.

(c) v_k is bounded above by S_1. This can be proved by writing Eq. (E.15) as:

$$v_k = S_1(1-\rho_1^2) + \frac{N_1}{N_1+v_{k-1}} (\rho_1^2 v_{k-1}) < S_1(1-\rho_1^2)+\rho_1^2 v_{k-1}$$

$$< S_1(1-\rho_1^2)[1+\rho_1^2+\rho_1^4 \ldots + \rho_1^{2k-4}]+\rho_1^{2k-2}v_1$$

$$= S_1(1-\rho_1^{2k-2}) + \rho_1^{2k-2}S_1 = S_1 \qquad \text{for } k>1$$

Thus: $$v_k < S_1 \qquad \text{for } k > 1 \qquad (E.24)$$

These three conditions are sufficient for the convergence of v_k.

This limiting value of v_k can be found by setting v_k and v_{k-1} equal to v in Eq. (E.15) and solving:

$$v^{(i)} = \begin{cases} \frac{1}{2}\left|S_i - N_i\right|(1-\rho_i^2)\left\{1+\left[1+\frac{4N_iS_i}{(S_i-N_i)^2(1-\rho_i^2)}\right]^{1/2}\right\} & S_i \neq N_i \\ S_i\sqrt{1-\rho_i^2} & S_i = N_i \end{cases} \qquad (E.25)$$

Thus in the steady state we have:

$$I_i(X_n;Y_n) = \frac{1}{2}\log\left[1+\frac{v^{(i)}}{N_i}\right] \qquad (E.26)$$

where $v^{(i)}$ is given by Eq. (25) and the superscript refers to the ith instrument. For this periodic model, we can add the information from each of the instruments we obtain:

$$I(X_n;Y_n) = \sum_{i=1}^{N} I_i(X_n;Y_n) \qquad (E.27)$$

where Eq. (E.27) is valid for all n. Similarly the rate of information for this periodic case is:

$$J_i(X_n;Y_n) = \frac{1}{T}\, I_i(X_n;Y_n) \qquad (E.28)$$

and

$$J(X_n;Y_n) = \frac{1}{T}\, I(X_n;Y_n) \qquad (E.29)$$

Again these equations are true for all n.

With these results it is now possible to find the optimum values of the dwell times relative to some sampling criterion. In general, this will be difficult, but a numerical solution should be possible for any functional forms of N(t) and $\rho_1(t)$.

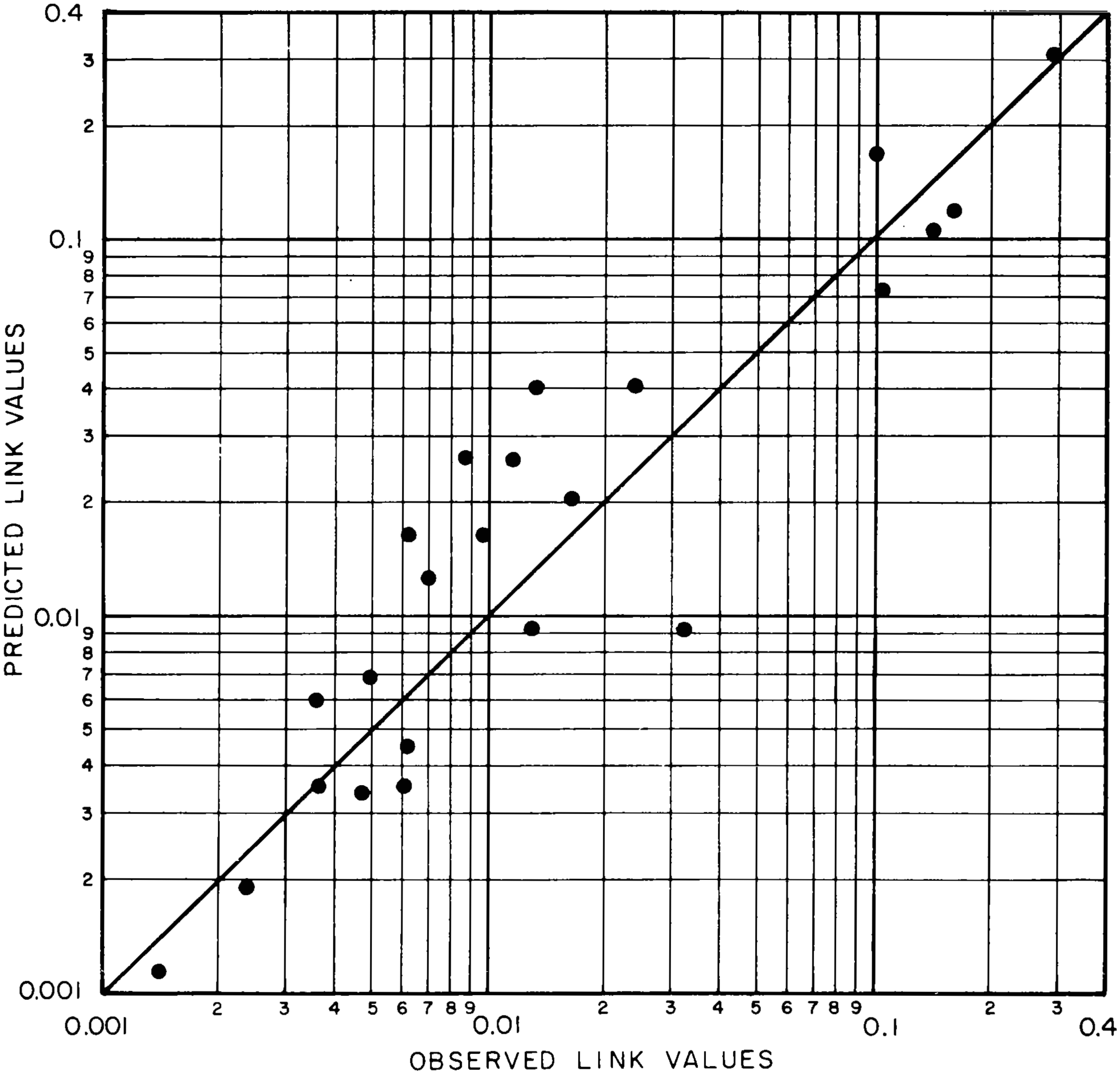
APPENDIX II

A Partial Application to Operational Data

The application of parts of the simple Markov transition model and the simple sampling model can be done even now with profit. We have attempted a reanalysis of the data obtained from the "Pilot Eye-Movement Studies" (23). Those experiments had made no records of the instrument readings which were observed by the pilots. However, on the basis of the measured frequency and duration of fixation of the various instruments, it was possible to calculate the "link values" which should have been observed between instruments if the pilots had been behaving as simple Markov processors. The results are shown in the figures which are taken from Reference (23).

The predicted (transition) probabilities are plotted against those obtained. Although there is no exact correspondence, the predicted results would not, in our opinion, have led to different practical conclusions about instrument panel design than the measured values. Each point on these plots represents a probability of transition in either direction between two instruments. All possible combinations are represented. If a point falls on the solid line, the predicted and the obtained are in exact agreement. The probabilities of the events at the lower left corner of these graphs are very small indeed. The major contribution to design decisions will be made by the instrument-to-instrument transitions represented by the points near the upper right hand corner. These latter account for more than 60 percent of the transitions in nearly every case.

The figures are shown for a variety of flight conditions. The condition under which the data were gathered is indicated at the bottom of each figure.



PREDICTED LINK VALUES
0.4
3
2
0.1
9
8
7
6
5
4
3
2
0.01
9
8
7
6
5
4
3
2
0.001
2
3
4
5
6
7
8
9
0.01
2
3
4
5
6
7
8
9
0.1
2
3
0.001
0.4
OBSERVED LINK VALUES

FIG. 1 ILAS

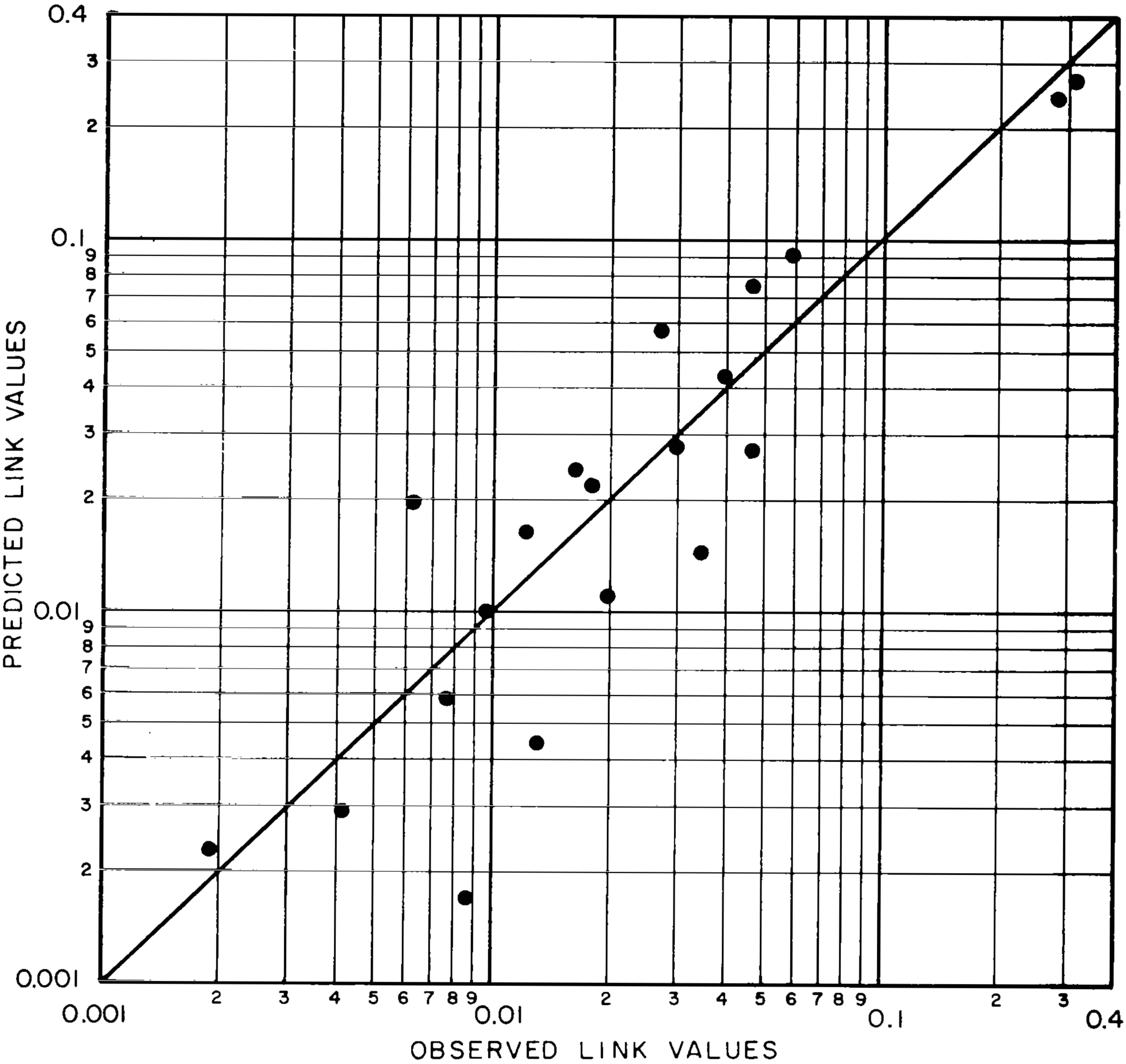

0.4
3
2
0.1
9
8
7
6
5
4
3
2
0.01
9
8
7
6
5
4
3
2
0.001
PREDICTED LINK VALUES
2
3
4
5
6
7
8
9
0.001
0.01
2
3
4
5
6
7
8
9
0.1
2
3
0.4
OBSERVED LINK VALUES

FIG. 2 GCA

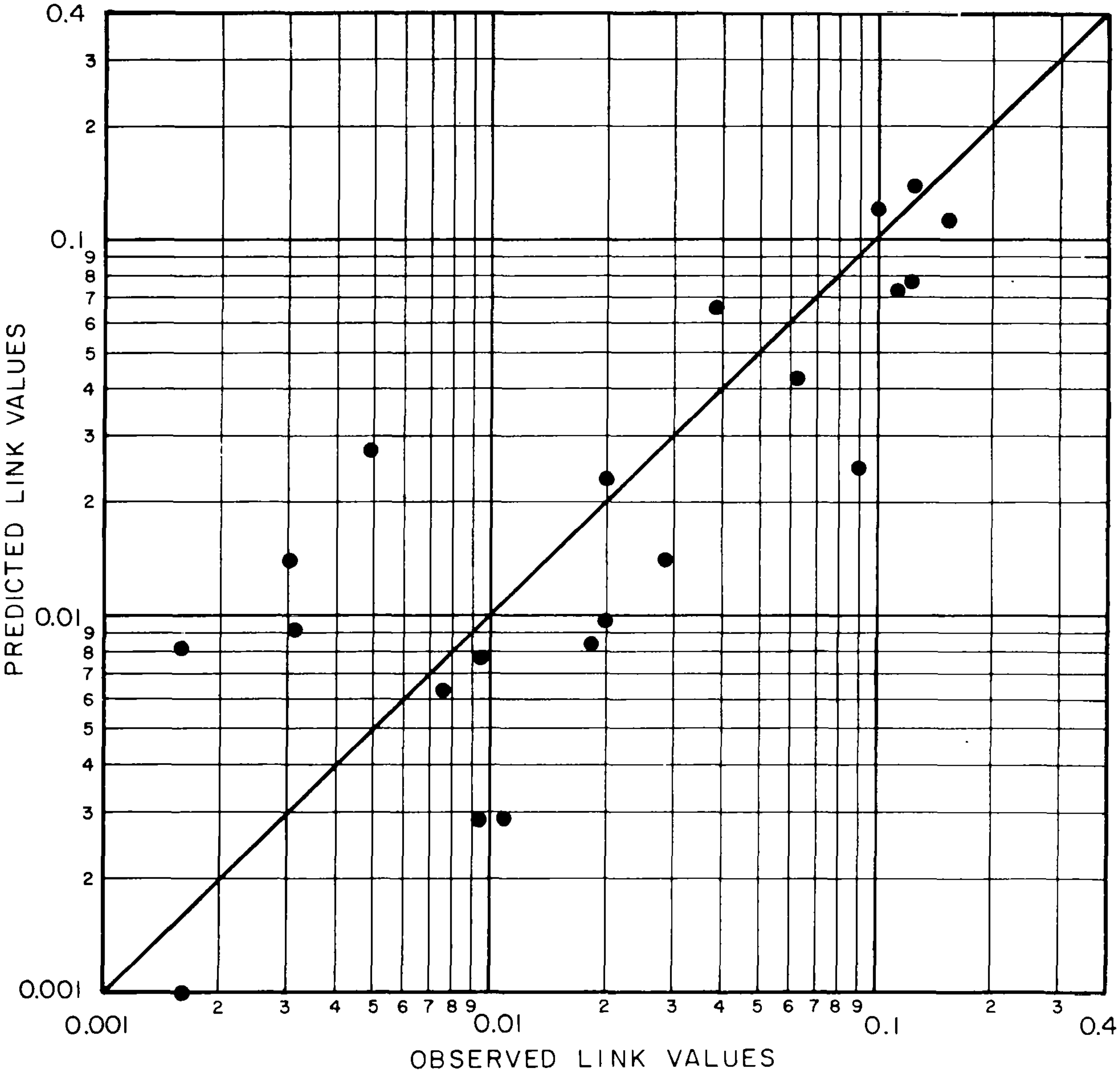


FIG. 3 STANDARD RATE TURNS

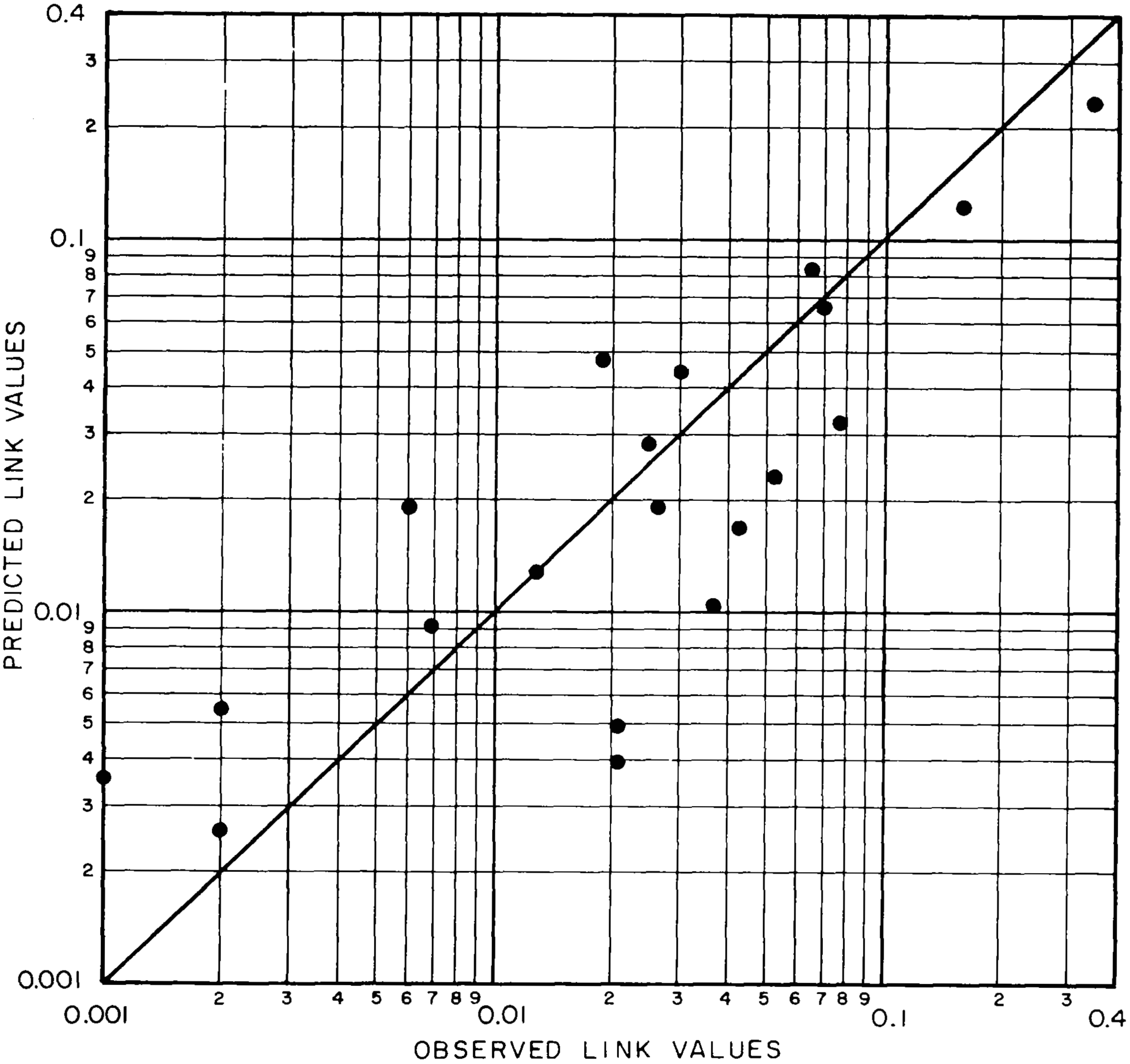


FIG. 4 STRAIGHT AND LEVEL FLIGHT

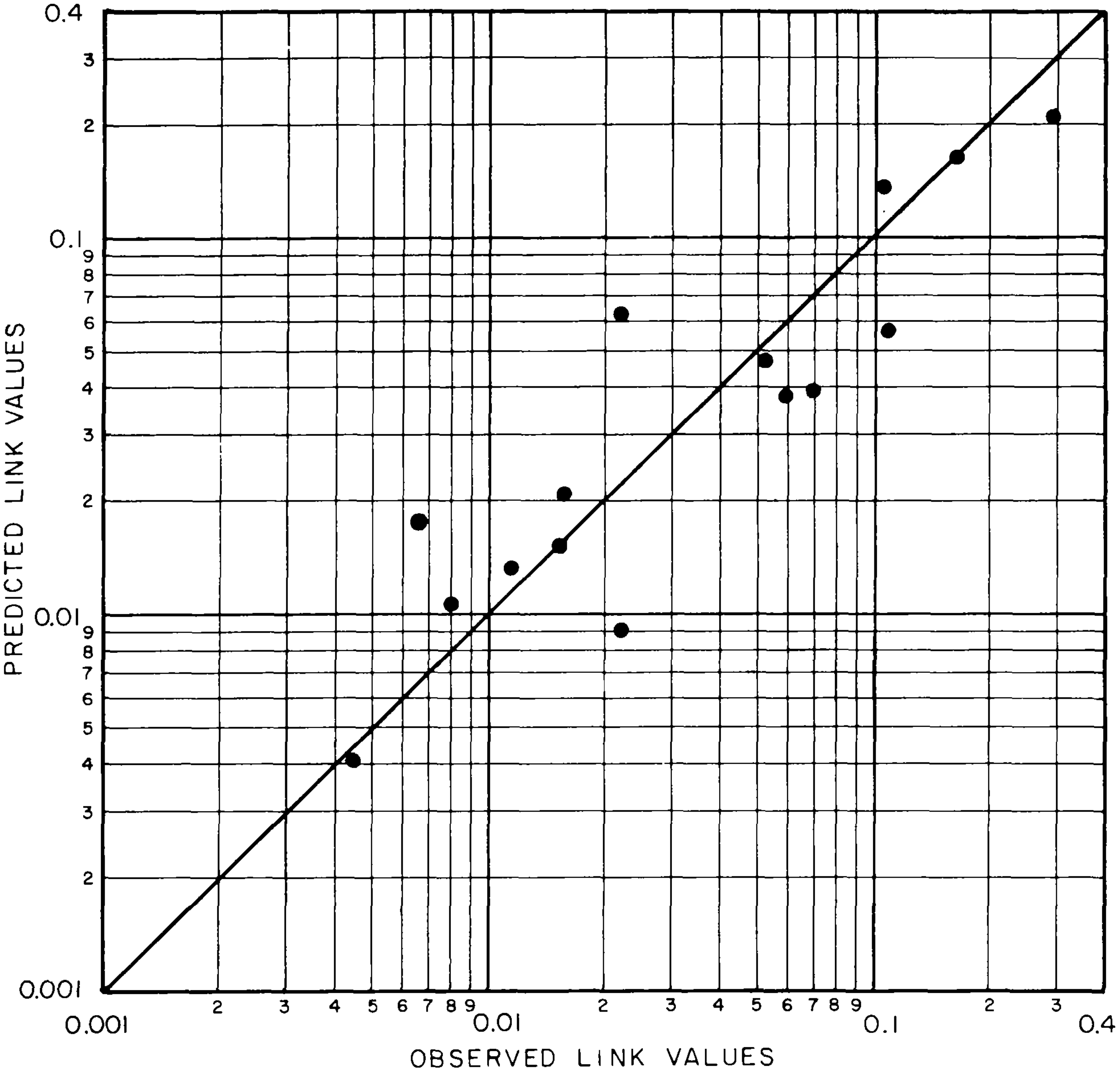

PREDICTED LINK VALUES
0.4
3
2
0.1
9
8
7
6
5
4
3
2
0.01
9
8
7
6
5
4
3
2
0.001
2
3
4
5
6
7
8
9
0.01
2
3
4
5
6
7
8
9
0.1
2
3
0.001
0.4
OBSERVED LINK VALUES

FIG.5 DAY GCA

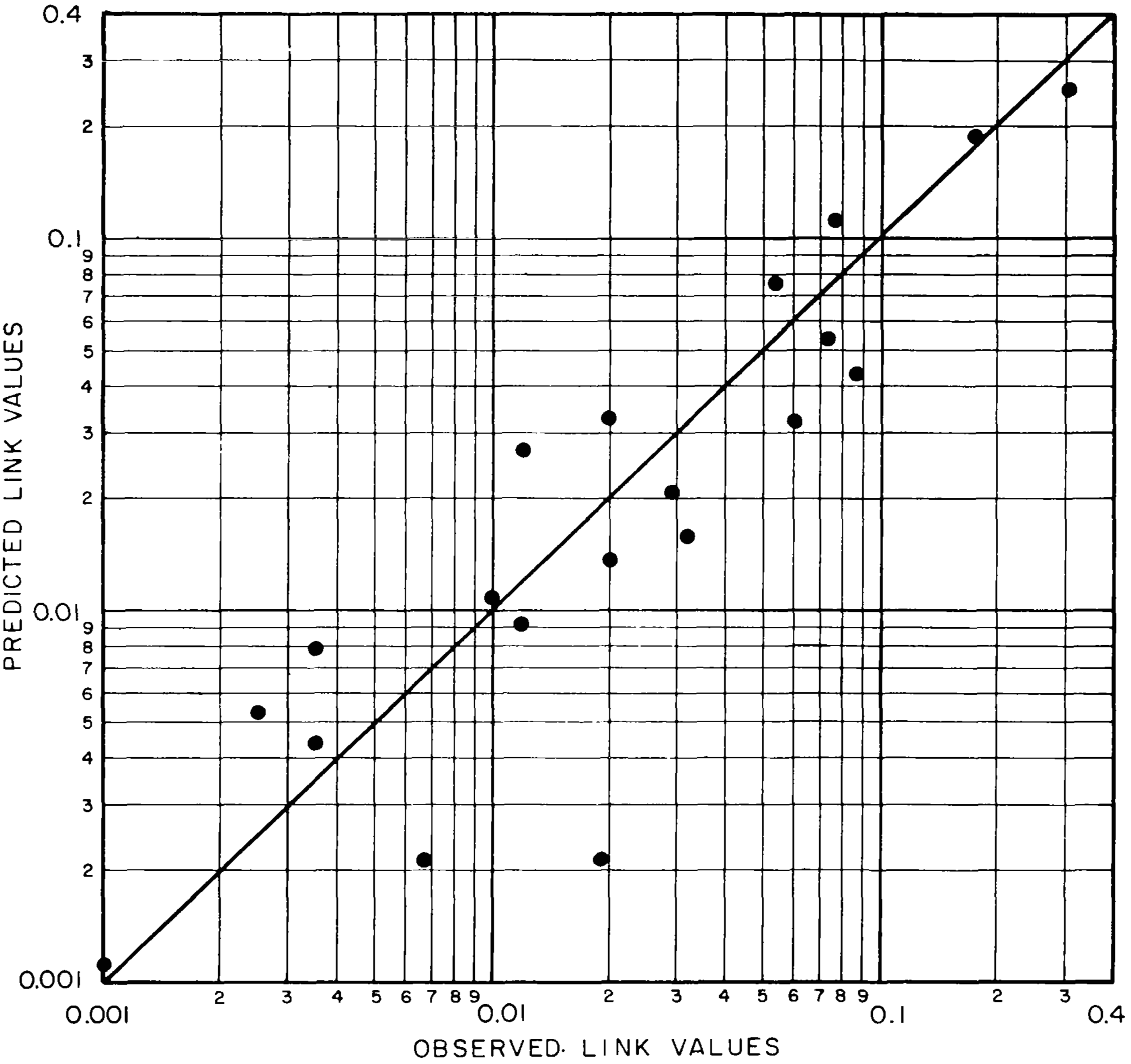


FIG. 6 NIGHT GCA

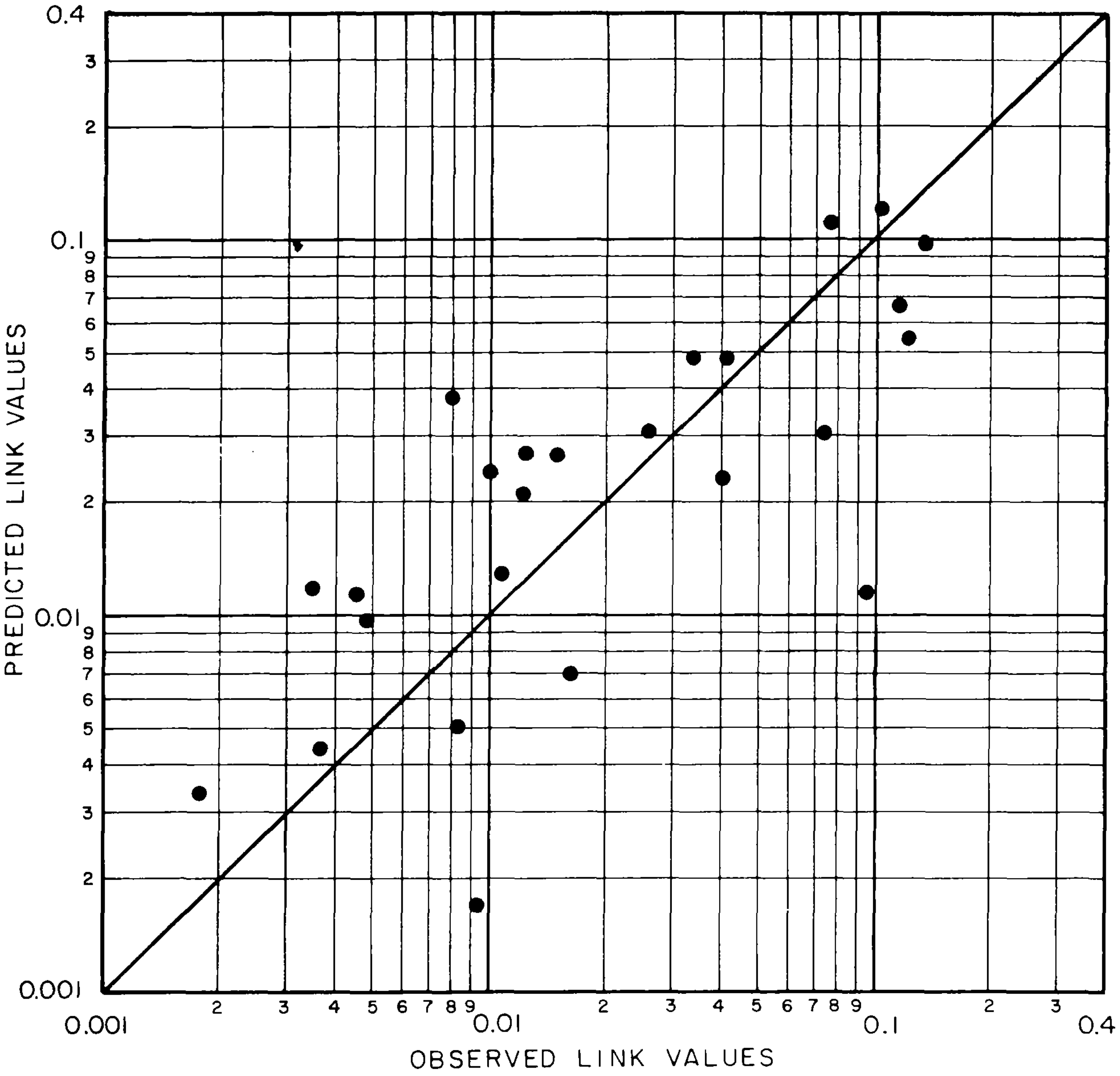


FIG. 7 DAY LEVEL TURNS

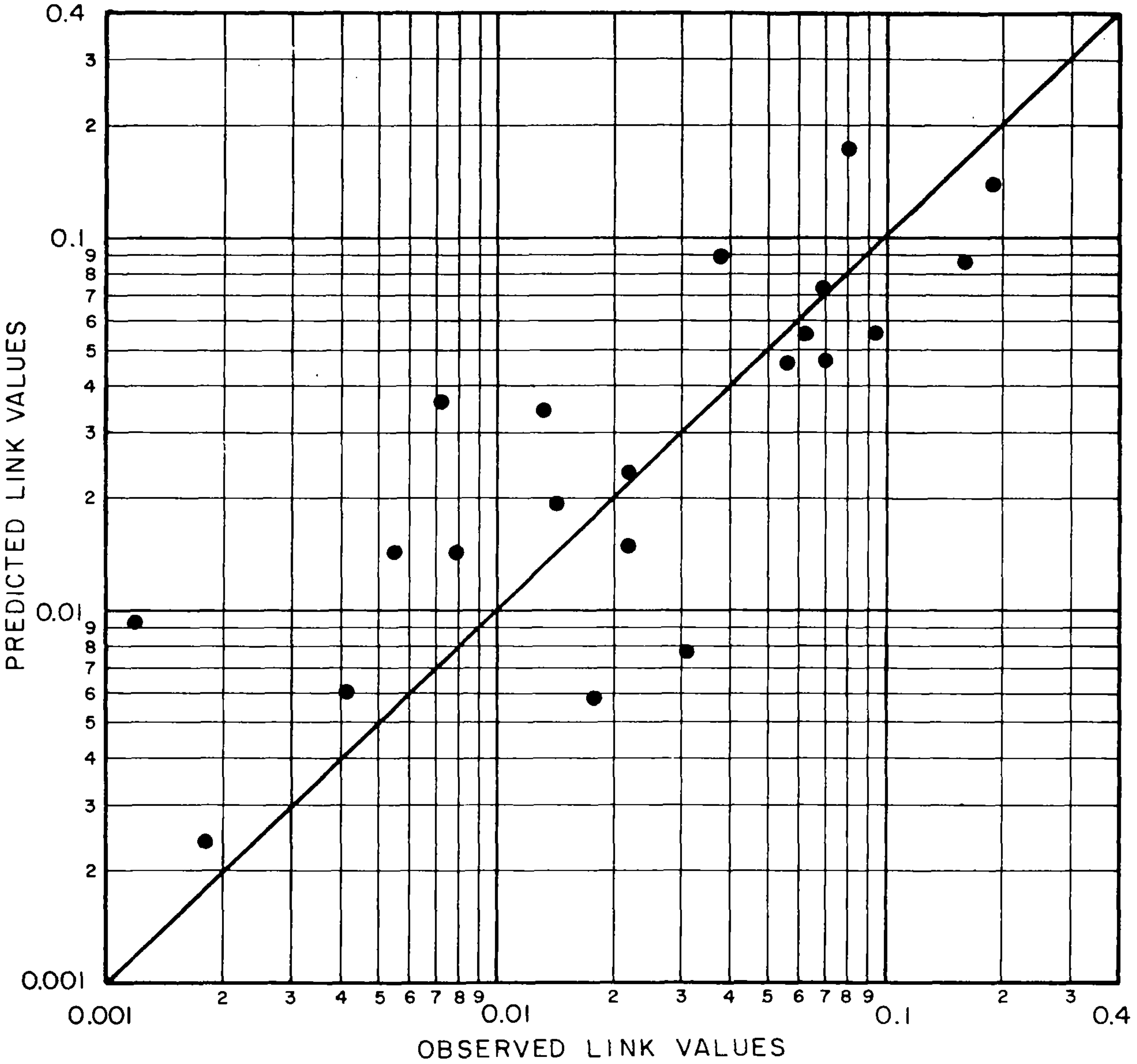


FIG.8 DAY STRAIGHT AND LEVEL

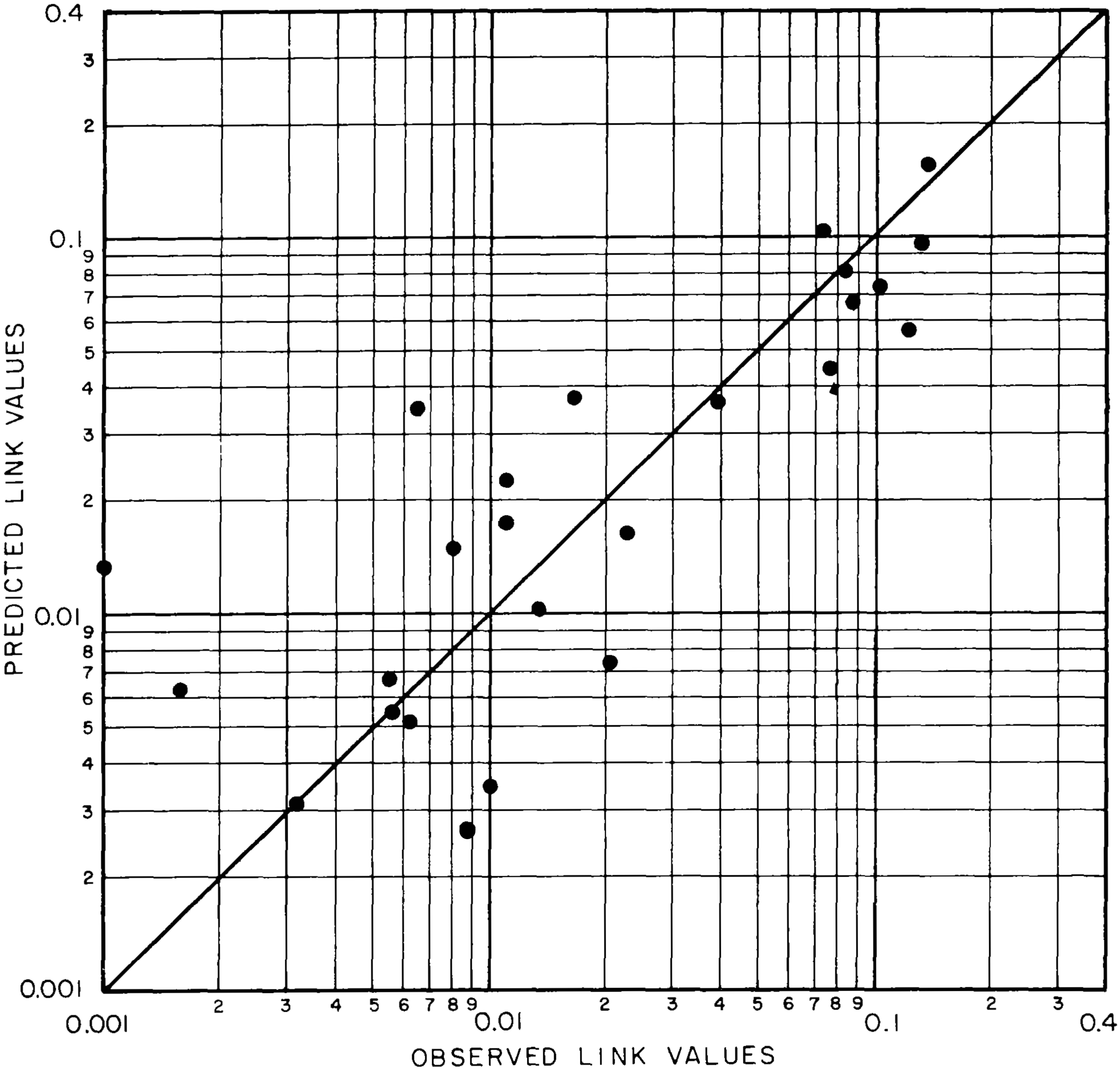


FIG. 9 NIGHT LEVEL TURNS

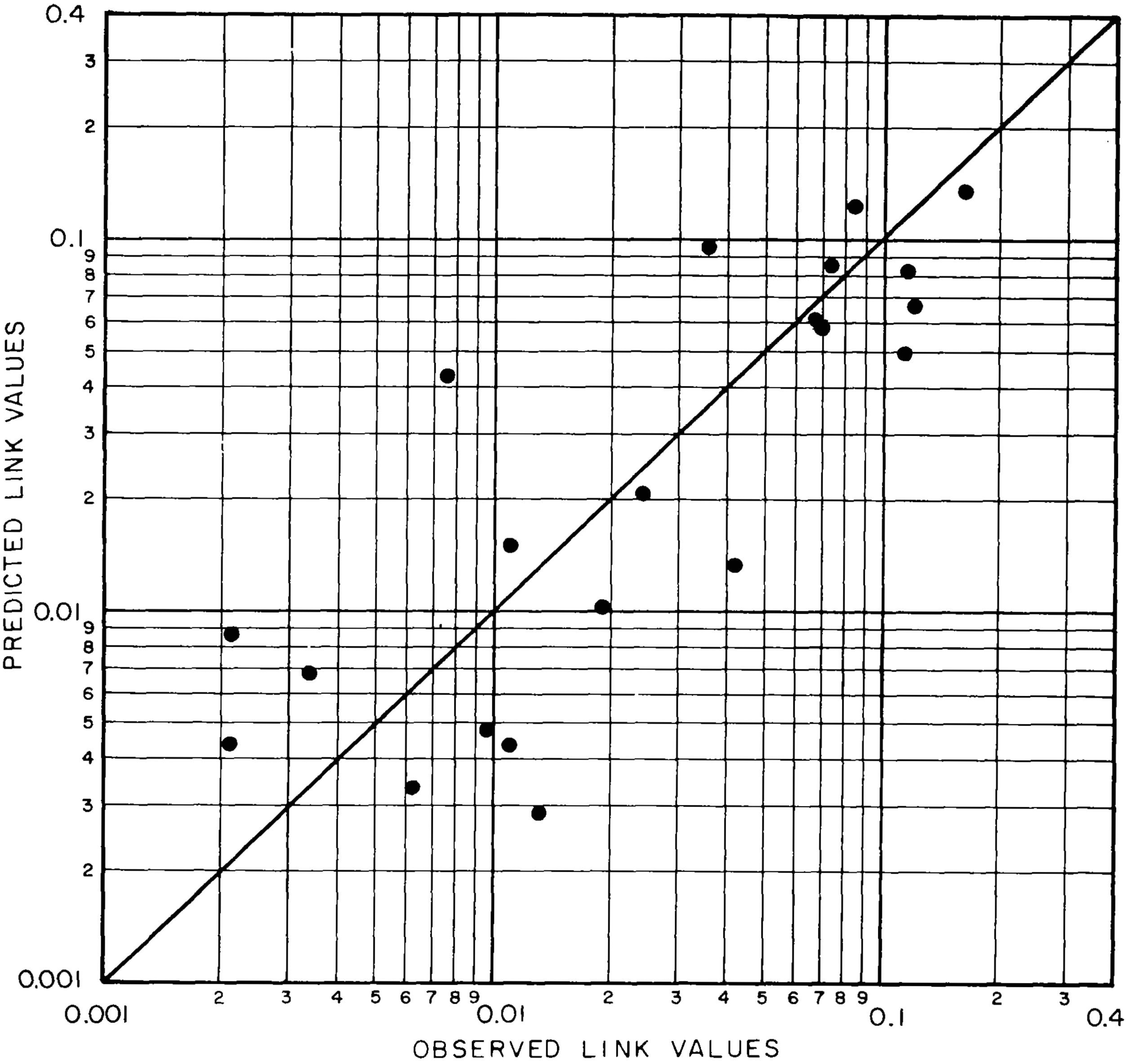


FIG. 10 NIGHT STRAIGHT AND LEVEL FLIGHT