FOREWORD

The research described in this report, *An Investigation of Sampled Data Models of the Human Operator in a Control System,* by George A. Bekey, was carried out under the technical direction of C. T. Leondes and G. Estrin and is part of the continuing program in Control Systems Theory.

This project is conducted under the sponsorship of Aeronautical Systems Division, Flight Control Laboratory, Project Engineer, Charles Harmon, Wright-Patterson Air Force Base. Submitted in final fulfillment of Contract Number AF 33(616)-7139, Project Number 8225.

This report was submitted as a dissertation in partial satisfaction of the requirements for the degree of Doctor of Philosophy in Engineering at the University of California, Los Angeles.

ABSTRACT

The mathematical models most commonly used to represent the human operator in a closed-loop tracking system are linear differential equations whose parameters depend on the characteristics of both the input signal and the controlled process. This report presents an analytical and experimental study of a new class of human operator models which are based on discrete rather than continuous operations. While intermittent processes in human tracking have been hypothesized in the past, this research is the first systematic study of the implications of intermittency by means of the theory of sampled-data control systems. The resulting models are shown to be consistent with the large body of experimental evidence concerning tracking. For the inputs considered in this study, the outputs from the sampled-data models have certain characteristics which approximate experimental data more closely over a wider range of frequencies than those obtained from the quasi-linear continuous models.

Both analytical and experimental techniques are employed in the study. Systematic procedures for construction of the proposed sampled-data model are presented, beginning with the measurement of power spectra and cross-spectra of the system. The frequency characteristics of the sampled-data models with stationary random inputs are analyzed by means of z-transform techniques. Closed-form expressions are derived for computation of the power spectral density of the model output and error signals and the resultant values are compared both with data in the literature and with the experimental phase of the work. In addition to using conventional zero-order and first-order "hold circuits" for reconstruction of sampled data, a "modified first-order hold circuit" is developed and analyzed.

A preliminary analysis of transient response and stability of sampled data systems with variable sampling rates is presented, as an introduction to the study of adaptive sampled-data models.

An experimental program was designed to measure the power spectral density of the tracking error under a variety of conditions. Measurements were made using both analog and digital computer techniques. The results of the experimental program were used to compare with predicted values obtained by analysis. A second phase of the experimental program showed that the use of intermittent displays yields results which are also consistent with the models proposed. The experimental work involved fairly extensive use of analog computer equipment and some novel techniques for simulation of sampled-data systems were explored.

The implications of the new sampled-data models for the design of man-machine systems are discussed and a number of suggestions for extensions of this work are presented.

PUBLICATION REVIEW

This report has been reviewed and is approved.

FOR THE COMMANDER:

C. R. Bryan

C. R. BRYAN
Technical Director
Flight Control Laboratory

Acknowledgments

The author gratefully acknowledges the help and encouragement given him by the members of his Doctoral Committee. Particular thanks are due to Professor John Lyman, without whose guidance, enthusiasm, sense of humor and gentle pushing all along the way, this work could not have been completed. Many thanks are also due to Professor C. T. Leondes for his advice and support and for making it possible for the author to work in the Adaptive Controls Research Group at UCLA. It was a privilege to be associated with this group.

Some of the academic background and orientation for this study was provided by the Human Tracking Project in the Biotechnology Laboratory at UCLA.

This research was supported in part by the United States Air Force under Contract No. 33(616)-7139, from the Aeronautical Systems Division, Air Force Systems Command. The work was also supported in part by Space Technology Laboratories, Inc. The author is particularly grateful to Dr. R. K. Whitford and Mr. N. W. Trembath for their continuing support of his work and for making available the analog computation facilities of the Control Systems Department at STL.

Many people were extremely helpful at various stages of the project and their help is gratefully acknowledged. Mrs. Bernice Hanson performed many of the early hand computations. Mr. Robert Lant of the STL Computation and Data Reduction Center skillfully guided the data through the maze of the STL spectral program on the IBM 7090. Mr. Jack Distaso of the Digital Technology Research

Group at UCLA did the programming required to evaluate the power spectral density equations. The latter computations were performed at the Western Data Processing Center at UCLA.

The author would also like to express his thanks to his colleagues and fellow students both at UCLA and at STL for the many enlightening discussions which accompanied the course of this research. Special gratitude is due to the many persons who helped with the preparation of the final report.

Last but definitely not least, the author is grateful to his wife for her patience and encouragement as he walked about the house mumbling about sampled-data models and bumping into the furniture.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

APPENDICES

APPENDICES (Continued)

LIST OF FIGURES

LIST OF FIGURES (CONT.)

LIST OF FIGURES (CONT.)

LIST OF FIGURES (CONT.)

LIST OF FIGURES (CONT.)

LIST OF TABLES

Chapter 1

INTRODUCTION AND BACKGROUND

1.1 General Statement of the Problem

This report is concerned with the mathematical representation of the input-output behavior of the human operator in a control system. A block diagram of such a system, in which the operator's function is called "compensatory tracking", is shown in Figure 1.1.

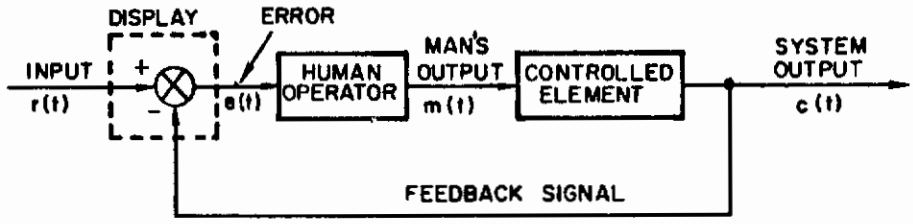


Figure 1.1 Block Diagram of Tracking Loop

The display acts as a subtraction device and the operator attempts to reduce the error signal to zero. The mathematical models most commonly used to represent the human operator in a system such as Figure 1.1 may be termed continuous and "quasi-linear"; i.e., they consist of linear differential equations whose coefficients depend on the bandwidth of the input signal and on the dynamics of the particular controlled process. For a particular class of inputs and a particular system the model for a trained operator would be a linear differential equation with constant coefficients. The present status of such representations is analyzed in detail in Chapter 2.

The major objective of this report is to present an analytical and experimental study of a new class of mathematical models of the human operator which are based on discrete rather than continuous operations. While intermittent processes in human tracking have been hypothesized in the past, this research is the first systematic study of what has become known in the literature as the "hypothesis

Manuscript released by the author February 1, 1962 for publication as an ASD Technical Documentary Report.

ASD-TDR-62-36

of intermittency". The major tools and methods utilized in the study are those of sampled-data control system theory.

The purpose of this introductory chapter is to give a brief statement of the background for the investigation, outline its major objectives, describe the limitations of the research program, and present in detail the organization of the body of the report.

1.2 Background

The development of mathematical models for human operators began late in World War II when human trackers were widely used in target tracking for anti-aircraft guns and similar devices. The first engineering approaches to the problem were reported by Tustin (1947)* in England and Ragazzini (1948)* in the U. S. Both investigators attempted to fit a linear input-output relationship to the tracker, since a human "transfer function" would have made possible the application of linear control system synthesis techniques to man-machine systems. However, the construction of an adequate model for the human operator, even for a particular task (such as tracking in one dimension) is extremely difficult. This difficulty arises not only from the inherent variability of human performance (and the consequent need to study many subjects under careful experimental conditions) but also from the fact that man's performance appears to be both non-linear and non-stationary. Furthermore, the variation from task to task indicates the high degree of adaptibility of the human operator. When tracking simple sine waves, the human operator is capable of learning their nature and predicting their future course sufficiently well to be able to track with his eyes closed. This behavior has been called "precognitive tracking" by McRuer and Krendel (1957). Consequently, it became apparent that experimental inputs must be either random or

* Names of authors followed by dates in parentheses refer to the Bibliography. If more than one work by one author in a single year is referenced, they are indicated by (1948a), etc.

at least appear random to the operator, and thus that superposition of responses to simple inputs is not applicable. However, a linear representation, for a particular input signal and a particular task, still was possible, with certain other restrictions to be discussed later.

Non-linear representations have also been explored. However, the lack of adequate analytical tools for dealing with complex non-linear systems has meant that such systems have been studied by analog computer simulation and their parameters selected by "brute force" techniques, such as adjustment of knobs until the response of a particular operator is matched as closely as possible. The major efforts in this direction were undertaken by Goodyear Aircraft Co. (1952). The use of phase-plane techniques was explored by Platzer (1955). Both of these approaches will be outlined in more detail in Chapter 2.

1.2.1 The quasi-linear continuous model. For random-appearing inputs (such as sums of sinewaves of non-harmonic frequencies) where the highest frequency does not exceed about 0.5 cps, the best linear fit to the trained operator's characteristics is given by a frequency domain relationship of the form

$$G_H(j\omega) = \frac{Ke^{-j\omega D}(1 + j\omega T_L)}{(1 + j\omega T_N)(1 + j\omega T_I)} \tag{1.1}$$

where ω is the natural frequency in rad/sec

T_N, T_L, T_I are time constants which depend both on input bandwidth and controlled element dynamics

K is the model gain, primarily a function of input bandwidth, and

D is the time delay associated with the data transmission and data processing ("reaction time") by the operator

The method of obtaining this relationship is described in Chapter 2. However, it can be noted here that the measurement of this "describing function" is based on a method suggested by Booton (1952, 1953) and that it yields the minimum-mean-squared-error approximation to the operator's output for a gaussian input process. Thus, the relation (1.1) does not attempt to explain all of the operator's output, but only that portion which can be explained by a linear, continuous operation on the input. Thus, the quasilinear model includes, in addition to the linear relation of eq. (1.1) a noise generator, as shown schematically in Figure 1.2. The work of a number of investigators, as summarized by McRuer and

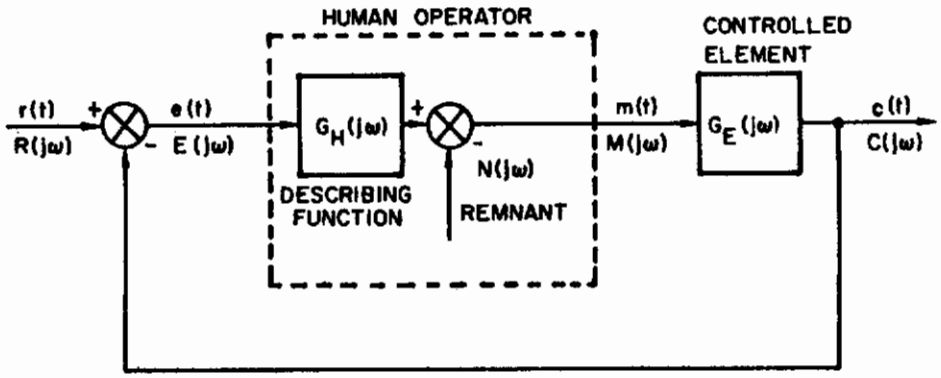


Figure 1.2 Quasi-linear Model of the Human Operator

Krendel (1957) has shown that the noise term accounts for less than 5% of output power when the input signal bandwidth is restricted to frequencies lower than about 0.5 cps. The variation of the model with input bandwidth and controlled element dynamics has been extensively studied.

1.2.2 Difficulties with the quasi-linear continuous model. The quasi-linear model gives impressive evidence of the nearly linear behavior of the human operator when tracking signals of low frequency. However, the model suffers from a number of drawbacks in addition to the frequency limitations. Among these are the following:

(a) Being linear and continuous, the model cannot generate frequencies beyond the bandwidth of the input signals (which are known to exist in human operator outputs).

(b) The model cannot account for a substantial body of experimental evidence which suggests intermittent behavior of the tracker.

(c) The model does not account for the predictive ability of the human operator.

The models proposed in this report attempt to remedy, at least in part, all the above difficulties.

1.2.3 Evidence for intermittency. The evidence for intermittent behavior of human operators comes both from tracking experience and from a number of psychological experiments. The evidence is summarized and documented carefully in Appendix 1. The evidence is based upon the following:

(a) Examination of tracking records reveals pronounced periodicities in the vicinity of 2 to 3 cps, regardless of input waveform.

(b) Tracking of series of discrete stimuli (such as steps) often reveals unexpectedly long delays in the response to a second stimulus if it follows closer than about 1/3 second behind the first.

(c) Prediction of target motion by human operators suggests that they are capable of extrapolating on the basis of recent samples of target velocity.

A number of other implications of the sampling theory is substantiated by the evidence and a detailed discussion is given in Appendix 1. They include such phenomena as perception of a number

of discrete events, the effect of perceptual delays on behavior, the study of eye movements in tracking, and other related areas.

The difficulties of the continuous model and the evidence for intermittency have led a number of investigators to suggest a mathematical model based on discrete operations. Only two attempts are known to this writer. North (1952) constructed a "difference model" by taking the linear continuous model and replacing all derivatives by finite differences. This method did not take into account the behavior of the system between the "sampling instants" and cannot be considered adequate. Ward (1958) used an analog computer to simulate a sampled-data model. However, he made no attempt to analyze the frequency characteristics of this model and he did not include a data-reconstruction device which would make extrapolation possible. His work was primarily a "brute-force" response-matching study in the time domain.

1.2.4 Characteristics of sampled-data models. The ability of a sampled-data model to meet some of the problems which face the continuous model can be seen intuitively by considering the following characteristics of sampled systems:

(a) The presence of the sampler limits the frequencies which can be reconstructed at its output to those not exceeding one-half the sampling frequency, by Shannon's sampling theorem (e.g., Goldman, 1952).

(b) The action of the sampler generates harmonics which extend over the entire frequency spectrum, even when the input is band-limited.

(c) The "hold" circuit which generally follows the sampler is a time-domain extrapolator, which reconstructs the signal based on information at the sampling instants. Consequently, it provides the model with some measure of "prediction".

(d) In the limit as the input frequencies decrease toward zero, the sampler-and-hold have less and less effect and the sampled system output should approach the continuous system output.

The above will be explored mathematically in Chapter 3, but they are introduced here only to provide a rational background to what follows. It should also be noted that the introduction of the sampler into a linear constant-coefficient system renders the system time-variable. Consequently, random processes in such a system are in general non-stationary even when the inputs are stationary. It is probably due to the difficulty of analysis of such systems that they have not been used to greater measure for human operator models in the past.

1.3 Objectives of the Study

The major objective of this study has been the construction and evaluation of a class of mathematical models of the human operator based on the "hypothesis of intermittency" in human tracking. This general objective can be restated in terms of a series of more specific objectives as follows:

(a) To examine the literature on human tracking and related fields carefully and on this basis to construct an *a priori* linear sampled-data model for study.

(b) To analyze the properties of the *a priori* model using the methods of sampled-data theory in order to ascertain frequency domain behavior and relation to continuous models used in the past.

(c) To perform an experimental study for the purpose of determining whether the characteristics revealed by analysis are substantiated in tracking experiments with several human operators.

(d) On the basis of both the previous analytical work and the experimental results, to propose extensions of the models to the case where the sampling frequency is variable.

(e) To investigate the analytical tools required to study the stability and transient behavior of variable rate sampled-data systems and determine their relevance both to human operator models and to control systems.

The specific study objectives outlined above have been organized essentially along the lines of the investigation itself. Thus, the study has proceeded from background through hypothesis to analytical investigation, hence to experiments and back to analysis.

1.4 Orientation of the Study

While the topic of this dissertation is interdisciplinary in the sense that it is relevant to control engineering, experimental psychology and human physiology, the orientation of the study has been that of control engineering. Thus, the concern of this work has been the development of input-output descriptions in mathematical terms which could be used for the analysis and synthesis of man-machine control systems. For this reason, frequency domain descriptions have been emphasized wherever possible. Portions of the study clearly have relevance for psychology and physiology, but the dominant point of view is that of systems engineering.

1.5 Limitations of the Study

A number of restrictions apply to the broad objectives stated above. The major ones will be stated here and some suggestions on removal of the restrictions will be given in the appropriate chapters below. The restrictions fall into three categories: those affecting the experimental situation, those affecting the mathematical model, and a general restriction on the interpretation of the entire study.

1.5.1 Restrictions on the situation. The study has been confined to compensatory tracking in a single dimension. The restriction to compensatory tracking means that the operator sees only the system error (say the displacement of a dot from the center of an oscilloscope screen). A second type of tracking in which both a "target" and a "pursuer" appear is known as "pursuit tracking" has been omitted entirely in order to restrict the channels of information to the operator to one. For the same general reason, the study has been confined to tracking in one dimension. Furthermore, the study has been restricted to the case where the dynamics of the controlled element are negligible and consequently expressible as simply a gain K_E, which provides for the appropriate transformation of units from the mechanical displacement of the operator's output to the electrical signal required as an input to the display device. The latter restriction has been added primarily to simplify the study, especially since the effect of controlled element dynamics on human tracking behavior has been studied by others (see for example McRuer and Krendel (1957) for a summary).

The above restrictions apply equally to the analytical work and to the experimental work. A more serious restriction applies to the experimental work only, and that is the limited number of subjects used. A total of ten subjects were used in various phases of the study. The data from eight of them is sufficiently extensive to be included in this report. Clearly, this is a small sample of the population. The justification and the reasons for choosing particular subjects will be discussed in more detail in Chapter 5. However, two comments may be relevant here: First, most previous studies have made use of many less subjects. Thus Elkind's extensive study was based almost entirely on one subject, as was much of Russell's work. Sheridan (1960) used 8 subjects for some experiments and only one or two for others. Ward (1958) used 3 subjects. The major justification given is usually that of Elkind, who indicated that the range of variation among well trained subjects for

sufficiently easy inputs is very small. Unfortunately, the second condition does not apply to the present study since the frequency range used in the inputs is such that all trackers have difficulty with it. The second comment is more practical: adequate statistical data cannot be obtained with volunteer subjects in a situation where the experimental study is only a portion of an over-all investigation of feasibility of a concept. This comment is not intended to disparage the experiments reported in Chapters 5 and 6 below, but merely to emphasize that the emphasis of the study is on the feasibility of a concept for the synthesis of human operator models, not on the establishment of population parameters.

1.5.2 Restrictions on the model. The majority of the work is based on the properties of linear sampled data systems, where the sampling pulse width is negligible. The major reason for this choice is the availability of analytical tools (the z-transform) for the convenient study of frequency-domain properties. In the concluding portions of the dissertation the restriction to linearity is relaxed by introducing systems with variable sampling frequencies, but still containing a linear "plant". The restriction to a linear model was based in part also on the desire to relate the present study to previous work in the field, in addition to the desire to use analytical methods wherever possible without relying solely on computer studies.

1.5.3 The synthesis problem. It should be noted here that while the purpose of the dissertation was to investigate a new mathematical model of the human operator and to compare it to other models, it was not its purpose to prove that the human operator is a sampled-data system, or that his brain operates as a digital computer with computation cycles of 2 to 3/second. It is characteristic of the synthesis problem that it cannot produce a unique result. That is, specifications (or measurements) at the input and outputs of a physical system cannot uniquely determine the physical compo-

sition of the system. All that can be said from an investigation such as this one is that the outputs of this model show a closer correspondence with experimental data for a certain class of inputs than do the outputs of other models. No attempt can be made to claim that this is the "correct" model. Furthermore, no attempt has been made, except in a speculative sense*, to correlate portions of the model with portions of the organism. Much more detailed experimental work would be required to establish such connections. This research does, however, present a new mathematical model and it shows the feasibility of using it to represent input-output characteristics of human operators in compensatory tracking tasks.

1.6 Organization of the Dissertation

The dissertation is organized into eight main chapters and includes eight Appendices. The present chapter has indicated the major objectives of the study, the restrictions which have been placed upon these objectives, and brief statement of general background information. The chapter will conclude with some comments on the importance and applicability of this research.

Chapter 2 considers the present status of human operator models. Linear, quasi-linear, and non-linear representations are reviewed along with the methods used for their determination and some comments on their generality and limitations. The background for intermittent models is presented and the two previous models (suggested by North and Ward) are reviewed.

Chapter 3 presents the theoretical considerations required for the analysis and experimental study of sampled-data systems. The major characteristics of sampled-data systems are reviewed briefly including the time and frequency descriptions of various "hold"

* See Section 8.3

circuits including a new "modified" hold circuit. The use of "hold" circuits for construction of the sampled-data models of the human operator is introduced.

In Chapter 4 the analysis of new models is begun. First, the background information and characteristics of sampled-data systems are used to postulate several models, based on zero-order, first-order, and "modified first-order" hold circuits. Analytic expressions for the power spectral density of the output and error signal in typical tracking loops with "white noise" inputs are derived for the three models. The behavior of the models in the frequency domain is compared with that of the quasi-linear models. Stability considerations are examined for the three models.

Chapter 5 presents the experimental program which was designed to relate measured properties of tracking power spectra to those computed for typical values in the previous chapter. The analog simulation of sampled-data systems is discussed and the techniques developed for this study are indicated. The analog measurement of spectral density is discussed briefly (with the major details relegated to an Appendix) and the experiments and experimental conditions are described.

The results of the experimental program are described in Chapter 6. The results are presented in three groups: the general characteristics of the measured power spectra, the effect of sampled displays on the spectra, and finally the use of experimental data to synthesize the sampled-data models. Some of the implications of the models for the design of man-machine systems are considered.

Chapter 7 returns to analysis again and considers the generalization of the models to the variable sampling-rate case. Some preliminary results on analytical tools for such studies are presented, including the use of variable-coefficient difference

equations for evaluation of transient response and an application of the Second Method of Lyapunov for the study of stability of an adaptive variable-rate sampled data system. The study of stability by the use of a variation of the root-locus (the "T-locus") is also discussed.

The conclusions of the work and some recommendations for future work are given in Chapter 8. In that chapter the removal of some of the restrictions of the present study is discussed.

The Appendices present detailed treatments of a number of topics. Appendix 1 is a summary of the psychological literature concerning the evidence for intermittency and includes an extensive bibliography. Appendix 2 presents a summary of the theory of sampled-data systems with random inputs, in order to place the definitions used in the study on a more firm foundation. The analog measurement of power spectra and the necessary analytical background are presented in Appendix 3. The analog simulation of sampled-data systems, including some novel techniques developed during this study, is discussed in Appendix 4. The program and parameters used for the digital computation of power spectral density are in Appendix 5, which also includes plots of digitally obtained spectra. Appendix 6 includes the "raw" experimental data presented in the form of graphs and numerical tables, while Appendix 7 includes the detailed schematics of the analog computer mechanization used in the experimental phase of the work. Appendix 8 includes the fitted "transfer functions" used in synthesis of the continuous models.

1.7 On the Applications of the Investigation

The results of this study may find some applicability in three areas. First of these is clearly the design of man-machine systems in which the man has control responsibilities. As the complexity and performance requirements of such systems increase, the limitations of the human component become increasingly important also. Adequate design of such systems in the future may require some

knowledge of the input-output characteristics of the human operator under conditions which approach his limits of performance. Furthermore, the operator is required to process many channels of information simultaneously in many present day systems. The present study has implications in the design of sampled display devices for time-sharing two or more information channels.

The study may also have some implications for the design of future control systems. The field of adaptive control, for example, was developed largely as an attempt to emulate the adaptability of organic systems. Thus, the type of adaptive sampled-data system proposed in Chapter 7 as a human operator model offers some interesting possibilities for space vehicle control systems where energy must be minimized.

But the major reason for research such as the present is not the practical one of providing tools or information for design purposes. Rather, it is hoped that this work will be a contribution to the purely scientific goal of furthering the understanding of human behavior on an objective and quantitative basis. Further, the study is an example of a methodology for the development of mathematical models which may find increasing usefulness in the future.

Chapter 2

THE STATUS OF HUMAN OPERATOR MODELS

2.1 Introduction: The Nature of Manual Tracking

The construction of mathematical models of the human operator is based on the assumption that the process relating the visual stimulus and the muscular output is, at least in part, determinate when the operator is an element in a closed-loop tracking situation. As mentioned in Chapter 1, the type of manual tracking situation considered here is known as "compensatory tracking" and the operator's function is to reduce to zero the error between a stationary "target" (such as the center position on a dial or screen) and a movable follower. (When both the "target" and "follower" move the operator's function is termed "pursuit tracking".) Consequently, the history of human operator models in manual tracking includes a series of attempts to subject the "human component" in the loop to experiments similar to those used for the determination of transfer characteristics of inorganic processes. In the present chapter some of the most successful attempts to obtain transfer characteristics and equivalent circuits will be reviewed, leading to the formulation of the model to be investigated in later chapters. It should be noted that the emphasis in the present treatment is on transfer characteristics (i.e., input-output behavior) and thus very little attention will be paid to the work of a number of investigators who have focussed their attention on the determination of "performance scores" or error criteria in human tracking. For the analysis of control systems the availability of input-output relationships is of primary interest. The material in this chapter is based heavily on the excellent summaries in the literature such as those of McRuer and Krendel (1957 and 1959), Licklider (1960) and Adams (1961) as well as a previous survey paper by this writer (Bekey, 1960).

The behavior of the operator in manual tracking situations can be described in a number of ways. It is characterized by a number

of physiological and psychological characteristics, such as perceptual ability, threshold phenomena, visual acuity, muscular movement accuracy and strength in different positions, kinesthetic feedback, and others. Discussions of some of these topics are given in such references as Broadbent (1958), Bates (1947) and Hick and Bates (1950). Of greater concern for the development of transfer relationships are the observed characteristics which refer to the man as a servo element. In particular, eight such characteristics can be cited here:

(a) Reaction time: The behavior of the operator is characterized by the presence of a pure time delay or transport lag, which can be clearly observed in the response to step function inputs.

(b) Low-pass behavior: Visual examination (and Fourier analysis) of tracking records reveals that the tracker tends to attenuate high frequencies, the amount of attenuation increasing as the frequency increases.

(c) Equalization ability: The operator has the ability to introduce various kinds of compensation into his transfer characteristics if required to do so by stability considerations or performance requirements. There are instances when trackers have generated up to 2nd order lead or lag terms.

(d) Adaptability: The tracker's ability to adapt appears in two forms: First, his performance changes with time as he learns, and secondly, he is capable of sensing changes in environmental parameters and controlled system parameters and adjusting his characteristics accordingly. There is evidence to indicate that the operator not only adapts but to some extent optimizes. McRuer and Krendel (1959) indicate that in regions where the operator's behavior is close to linear the phase margin observed in the tracking loop after proficiency is attained was in the region of 60 to 110 degrees.

(e) Prediction: The ability of the human operator to predict the course of a target based on past performance is well known. Some of the work of Gottsdanker (1952, etc.) is reviewed in Appendix 1. This ability to extrapolate is important in tracking since it means that tracking behavior is different with "predictable inputs" (such as sine waves or constant frequency square waves) than it is with random or random-appearing inputs. Tracking with a predictable input has been called "pre-cognitive" tracking.

(f) Nonlinearity: For a trained operator there is very little change in transfer characteristics when the input signal amplitude is changed, thus indicating that, within the range of conditions tested, a series nonlinearity is not an important effect (see Elkind, 1956, and Krendel and Barnes, 1954). A parallel nonlinearity does appear in the Goodyear studies (1952, 1957) referred to in section 2.2 below.

(g) Randomness: It has been mentioned above that the application of engineering techniques to human tracking studies was justified since tracking behavior appeared to be, at least in part, systematic and determinate, that is, obeying deterministic cause-and-effect relations. In addition to the deterministic portion of the tracker's output, there appears to be a component which has little correlation with the input. Whether this component is simply "noise" generated within the operator, or whether it represents a time-varying or non-linear component of response by the operator has not been determined.

(h) Intermittency: There is a considerable body of evidence which indicates that the human operator behaves as a discrete or sampling system in certain tracking operations. Since this evidence is central to the theme of this work, it is reviewed in detail in Appendix 1 and summarized below in connection with intermittent models in Section 2.6.

The various human operator models appearing in the literature are attempts to include the major characteristics listed above into a mathematical description, either in the frequency or time domains.

2.2 Survey of Human Operator Models

The earliest models constructed from the point of view of control engineering were those of Tustin (1947) and Ragazzini(1948). These models were postulated as linear and continuous with an additional disturbance of unknown origin. Tustin called the additional term which did not result from linear operations on the input the "remnant" (i.e., those components of the response at frequencies other than the input frequencies in this case). The name "remnant" is still used in the quasi-linear models to be discussed below. In block diagram form the human operator model may be indicated as shown in Figure 2.1 below.

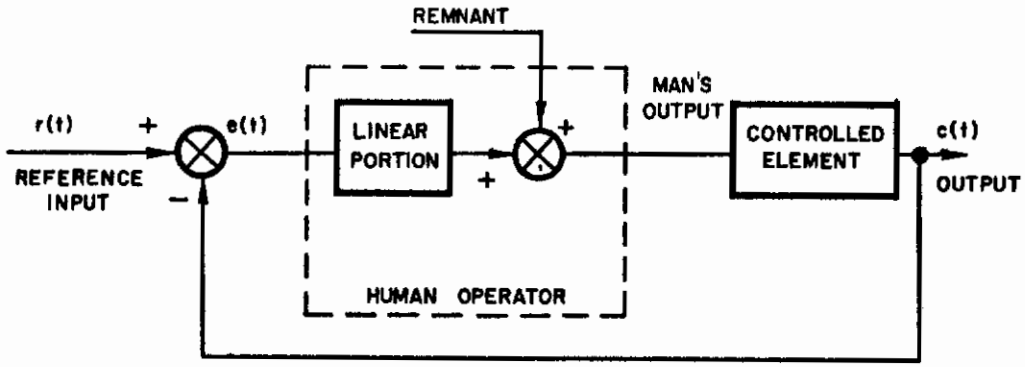


Figure 2.1 Operator Model with Output Noise Injection

The Tustin model was based on Fourier analysis of outputs when the input was a sum of 3 sinusoids. The resulting model was of the form

$$G_H(s) = K\left(\frac{K_1}{s} + K_2\right)e^{-Ds} \tag{2.1}$$

where s is the complex frequency variable, D is the time delay (taken as 0.3 sec by Tustin), and K, K_1 and K_2 represent parameters which depend on the choice of controlled element dynamics.

(Tustin's work was concerned with the motion of electrically controlled gun turrets in tanks.)

J. R. Ragazzini reported in 1948 in an unpublished paper delivered at the American Psychological Association that the transfer relationship may be of the form:

$$G_H(s) = K\left(K_1 s + K_2 + \frac{K_3}{s}\right) e^{-Ds} \tag{2.2}$$

In equations (2.1) and (2.2) the symbolism employed has been that of Laplace transforms as commonly applied to linear systems; i.e., the equations imply that the principle of superposition applies. However, we have seen above that the human operator's response to simple, predictable inputs is different from his response to complex inputs, and thus that superposition does not apply in the ordinary sense. Consequently, the input-output relationships of the type (2.1) or (2.2) will be termed quasi-linear transfer relationships. By quasi-linear in the present discussion we mean that the system has the following properties:

(1) The system is describable by a linear differential equation with coefficients which are dependent on the system configuration and the input signal bandwidth, but remain constant for a particular system, and

(2) The linear relationship determines only a portion of the system's output (the linear part). In addition, a random or uncorrelated component may exist. The validity of the quasi-linear model is dependent on the proportion of the system output which it specifies.

2.2.1 Quasi-linear Models. The definition of quasi-linearity given above was assumed by Tustin in his studies, even if not stated explicitly. It is interesting to review the methods used to arrive at various linear transfer relationships within this context.

Step-function response is traditionally used for the determination of transfer relationships in linear systems. Phillips (1947)

reports on a method used to determine aiding constants by minimizing mean-squared-error with a target motion containing step discontinuities. Mayne (1951) recorded responses to step inputs and found analytic fits to the resulting outputs. Hyndman and Beach (1958) used step responses to obtain open-loop transfer relationships, which do not necessarily apply well in a closed loop tracking situation. An obvious problem is that the parameters of the model will depend on the situation (step on or off, for example).

Response to simple sinusoids has been used as well, but, as previously noted, the response is expected to be different qualitatively from that obtained when tracking more complex signals. A number of interesting studies have been performed with sine-wave inputs, including those of Ellson, Gray et al (1948), Walston and Warren (1954) and Noble, Fitts and Warren (1955). Two major conclusions have been drawn from these studies: (1) if the frequencies are sufficiently low, a linear model adequately represents the major portion of the operator's output; and (2) as the frequency increases, the human tracker gradually adopts a different approach which does not appear to follow ordinary linear, continuous rules and is probably best explained by assuming the presence of a sampling operation.

Most of the significant work has been performed recently with complex signals consisting either of sums of sine waves or random noise of appropriate spectral characteristics. Tustin used a sum of 3 sine waves. Russell (1951) used a sum of 4 sine waves and also measured the output components at the same frequencies as those present in the input. Elkind (1957) performed an extremely thorough study of compensatory and pursuit tracking using input signals which consisted of a large number of sinusoids (as many as 144) of random phase. The method used by Elkind for the synthesis of the quasi-linear model will be described in the next section. Krendel (1951-54) in a series of studies used random noise to determine linear

relationships. The best quasi-linear results available to date are those of Elkind and Krendel.

2.3 Synthesis of Quasi-linear Models

The models described by Elkind are obtained by using a technique suggested by Booton (1953, 1954) for the analysis of nonlinear systems with Gaussian random inputs. The theory of the resulting "random input describing function" has been discussed in detail by McRuer and Krendel (1957) as well as by Booton and will not be repeated here. Basically the method consists of assuming that the system under test consists of a linear portion and a noise source which is uncorrelated with the input, as indicated in the block diagram of Fig. 2.1. To obtain the linear portion $G_H(j\omega)$ one can measure the cross-power spectral densities between input and error and input and operator output respectively. These spectral densities will be defined by the following relations

$$S_{re}(j\omega) = \frac{1}{1 + G_H(j\omega)\, G_E(j\omega)} \; S_{rr}(\omega) \tag{2.3}$$

and

$$S_{rc}(j\omega) = \frac{G_H(j\omega)}{1 + G_H(j\omega)\, G_E(j\omega)} \; S_{rr}(\omega) \tag{2.4}$$

where $S_{re}(j\omega)$ and $S_{rc}(j\omega)$ represent the input-error cross-spectral density and the input-output cross-spectral density. The noise spectrum $S_{nn}(\omega)$ does not appear in these relations since it is uncorrelated with the input, by hypothesis. Consequently, the "describing function" or quasi-linear transfer relationship can be obtained as the ratio of the two experimentally obtained quantities above, i.e.

$$G_H(j\omega) = \frac{S_{rc}(j\omega)}{S_{re}(j\omega)}$$

However, relation (2.5) represents a set of experimental points which define the amplitude and phase of the complex number $G_H(j\omega)$

When these quantities are plotted as a function of frequency, they can be fitted closely with a linear analytic relationship of the form:

$$G_H(j\omega) = \frac{Ke^{-j\omega D}(1 + j\omega T_L)}{(1 + j\omega T_N)(1 + j\omega T_I)} \tag{2.6}$$

where the time constants T_L, T_N, and T_I as well as the gain K and time delay D are functions of the particular situation. It is interesting to note that Jackson (1958) has shown that if one attempts to fit such experimental data with a general transfer relationship of the form

$$G(s) = K \frac{(T_{11}s + 1)(T_{12}s + 1)\cdots(T_{1n}s + 1)}{s^m(T_{21}s + 1)(T_{22}s + 1)\cdots(T_{2p}s + 1)} e^{-Ds} \qquad (m + p) > n \tag{2.7}$$

it is found that (2.7) reduces to (2.6) with negligible contributions from higher order terms. This indicates that the differences between the experimental behavior and linear model cannot be resolved merely by increasing the order of the linear model.

The "remnant" term is now represented by the noise injected at the operator's output in Fig. 2.1. If we define a "closed-loop quasi-linear describing function" $H(j\omega)$ as

$$H(j\omega) = \frac{G_H(j\omega)}{1 + G_H(j\omega)\, G_E(j\omega)} \tag{2.8}$$

and a closed-loop remnant spectrum as S'_{nn}:

$$S'_{nn}(\omega) = \left| \frac{1}{1 + G_H(j\omega)\, G_E(j\omega)} \right|^2 S_{nn}(\omega) \tag{2.9}$$

then the operator's output power spectrum can be described by the relation:

$$S_{mm}(\omega) = \left|H(j\omega)\right|^2 S_{rr}(\omega) + S'_{nn}(\omega) \tag{2.10}$$

Equation (2.10) shows clearly that the operator's output spectrum is composed of two portions, one resulting from a linear operation on the input spectrum S_{rr} and additional "noise" term. The degree

of validity of the quasi-linear model can now be related to the the proportion of total operator output power at each frequency which is predicted by the model. This can be done by defining the linear correlation between m and r (sometimes called the "coherence function") as

$$\rho^2 = \frac{\left| S_{rm}(j\omega) \right|}{\sqrt{S_{rr}(\omega)\ S_{mm}(\omega)}} \tag{2.11}$$

Elkind (1957) reported linear correlations of 0.9 or better for his studies when the input spectrum was flat with a sharp cutoff at or below about 0.75 cps. Based on these considerations, it can be concluded that the continuous quasi-linear model is an adequate description of the transfer characteristics of the human tracker, provided that the input frequencies are sufficiently low. Consequently, any new models should not be expected to deviate substantially from Elkind's models for low frequency inputs.

The technique described above for the construction of the quasi-linear model requires the measurement of power spectral density using a cross-spectral analyzer or equivalent and curve-fitting of the experimental results to obtain an analytic expression. Recently Ornstein (1961) has applied a modification of the method of steepest descent as developed by Margolis and Leondes (1959) to determine the model parameters automatically. This procedure requires that a form for the transfer characteristic be assumed known and a "parameter tracking servo" is used to adjust the parameter values to equal those of the system under measurement, using an appropriate error criterion. The technique appears to be promising but requires additional exploration.

2.4 Nonlinear Models of the Human Operator

The quasi-linear continuous models discussed above are able to represent a number of human operator characteristics, including reaction time, equalization ability, and low-pass behavior. The process of adaptation is not described, nor are predictive effects.

Nonlinear behavior is included only in the sense that it is a possible explanation of a portion of the "remnant" term.

A series of studies at Goodyear Aircraft Corp., reported in Goodyear reports (1952, 1957) and papers by Diamantides (1957, 1958) and others, have been aimed at the development of operator models which would account for a number of additional effects. Specifically, the Goodyear studies have investigated the effect of "dither" and the analog computer representation of some nonlinearities in tracking behavior.

Jet aircraft pilots are known to superimpose "dither" on their tracking motions, apparently in order to test the response of the vehicle to the rapidly changing environment. In the Goodyear studies the dither "mechanism," since it served no useful purpose in the laboratory simulator studies, was gradually extinguished by the subjects. The resultant models made no attempt to include it in the form of a parameter-adjustment loop, but merely as a noise input.

It has been noted above that the effect of a series nonlinearity in human tracking was not severe, since the transfer characteristics were not strongly amplitude-dependent. The representation of the operator developed by Goodyear had a perfect relay (signum function) in parallel with the remainder of the model, thus, in effect, giving the output a bias in accordance with the sign of the error signal. This feature gave the model a measure of anticipation and thus represented an attempt to include the predictive ability of the tracker in the model. The Goodyear model is shown in schematic form in Fig. 2.2.

It should be noted that the major disadvantage of such a model is that it must be simulated on a computer. While the Goodyear studies show good time-domain agreement with human tracking data, they are of limited usefulness since the lack of analytic expressions in closed form makes generalization of the results extremely difficult.

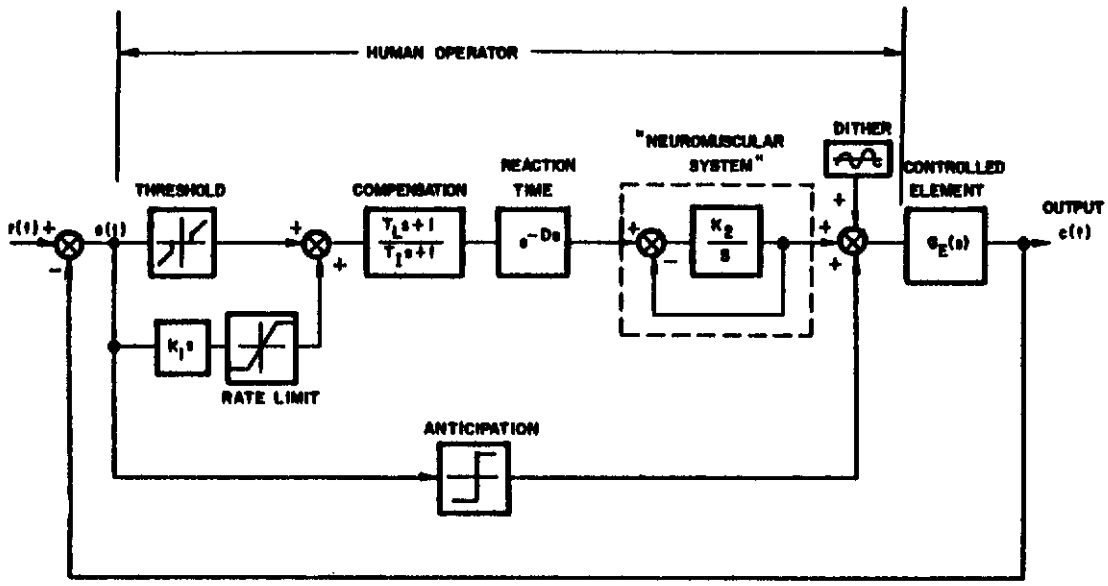


Figure 2.2 Nonlinear Model of the Human Operator

2.5 Time-varying Operator Model

The quasi-linear model takes into account the adaptive ability of the operator, but only after the adaptation process is complete. Thus, if a trained operator is subjected to a sudden change in the dynamics of the controlled element, he will gradually change the parameters of his transfer characteristics until he has achieved "optimum" equalization (to an unknown criterion) for the new situation. Following the adaptation, the quasi-linear model will again be applicable, with a new set of parameter values. During the adaptation process, however, the model does not apply.

The work of Sheridan (1960) is a careful experimental study of the variation of the frequency characteristics of the operator with time during "adaptation" to a change in the system. The method is also applied to observation of changes taking place in the operator's behavior with extreme fatigue. The results are obtained as time-variable frequency loci which represent the changes in gain and phase of the transfer relationship for each of 5 sinusoids which comprise the input waveform. The study is a valuable contribution to the knowledge of human tracking behavior, but since it does not correspond in form to any of the tracking models we have discussed previously, it will not be explored further in this connection.

2.6 Sampled-data Operator Models

All the models considered above have been continuous and thus have not been able to account for the possibility that the operator's behavior may be intermittent. Two previous studies have been addressed to this problem, both with serious limitations. In the present section the evidence for intermittency will be summarized and the studies of North and Ward reviewed.

2.6.1 Evidence for Intermittency. The details of the literature on this subject are given in Appendix 1, and thus in these paragraphs the major characteristics will be summarized without reference to sources. The reader is referred to Appendix 1 for an extensive bibliography.

The argument for intermittency is based on the following major experimental sources:

(1) Examination of tracking records reveals that the error curves have a pronounced periodicity in the vicinity of 2 cps even when this frequency is not very pronounced in the input. In one study it was found that over 80% of the wavelengths ranged between 0.2 and 0.6 seconds. Furthermore, tracking records inevitably reveal frequencies not present in the input.

(2) When a human operator tracks a series of steps which are spaced less than about $\frac{1}{2}$ sec apart, the effective "reaction time" in response to the second stimulus is often much greater than expected had the stimulus occurred in isolation.

(3) Pursuit tracking of pure sine waves of frequencies higher than about 2 cps often results in very close amplitude matching but frequency errors in the operator's output which are hard to detect and correct.

(4) The introduction of artificial time delays into the visual perception process of the order of 0.25 sec makes it nearly impossible to perform even very simple motor tasks.

Experiments such as the above (and many others) can be explained by assuming that the operator acts on samples of information taken from the perceptual input at discrete intervals. In addition, the predictive behavior of a human tracker (such as when a target disappears behind a cloud) can be explained by assuming that the operator extrapolates on the basis of past samples of the target's position and velocity. The "intermittency hypothesis" is also consistent with phenomena such as the perception of a number of discrete events, the visual perception of velocity, and others discussed in Appendix 1.

2.6.2 The Difference Equation Approach of North. The first mathematical approach to the intermittency hypothesis was made by J. D. North (1952, 1954). North examined the behavior of a typical quasi-linear differential equation model, of the type proposed by

Tustin. Then he replaced the derivatives in the equation by finite differences and examined the resulting linear difference equation. His final equations included not only the deterministic components of the tracker's response but also stochastic components which in the discrete case represent samples from "white noise." The approach of North is quite sophisticated and attempted to include many effects. However, his difference equation approach suffers from the fact that he is effectively concentrating on the system behavior at the "sampling instants" and the human operator's output is clearly continuous. That is, any discrete processes which may occur would be internal and a data-reconstruction element must appear in the model in order that the output be continuous. Consequently, while North's model has potentially similar features to those proposed in this report, in actuality they are not carried to same conclusion. For example, North evaluated the power spectral density of his model output, both in the continuous and the discrete case. However, since the time domain relation was valid for integral values of sampling intervals only, he was only able to say that the resulting spectral densities approach each other as the sampling interval approaches zero.

An "information transfer" model of the human operator has also been formulated by Fogel (1956, 1957). Once again, this model was not designed to provide time-domain behavior.

2.6.3 The Analog Computer Study of Ward. In a thesis submitted in Australia, Ward (1958) has discussed a study of a sampled-data model using an analog computer. Ward's model, as closely as this writer can interpret it, is given in Fig. 2.3. The results of Ward's study are given strictly in the form of computer output traces, compared with human output recordings for similarity of appearance. The matching was obtained by "brute force" or knob-adjustment techniques. The input functions consisted of the sum of 3 sine waves, the highest frequency being 0.2 cps and the lowest 0.01 cps. The apparent amplitude of the "noise" in both man and

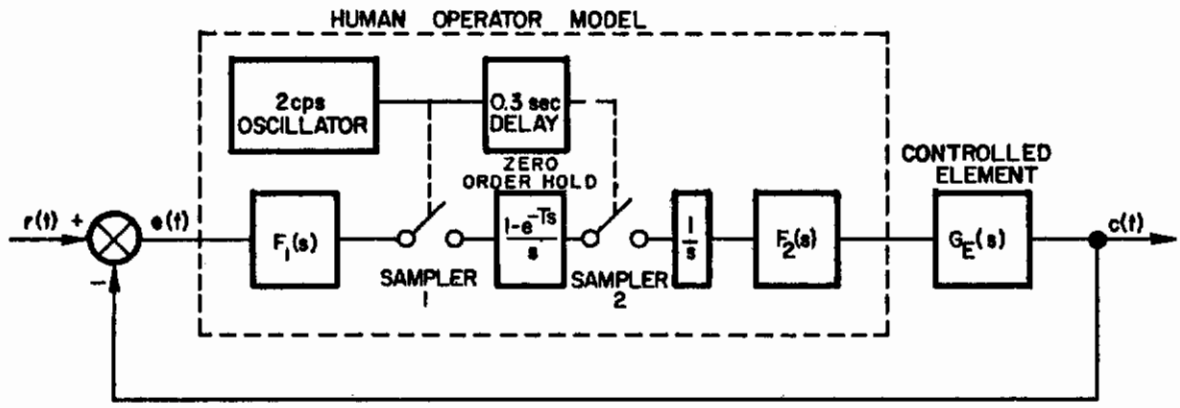


Figure 2.3 Sampled-Data Operator Model Due to Ward.

model is of the same order of magnitude. The study is severely limited limited for several reasons:

(1) The lack of analytical work. (Apparently Ward considered the system to be nonlinear merely due to the presence of the sampler); and (2) The poor choice of sampling and data reconstruction circuits made in the analog simulation. As can be noted in Fig. 2.4, the model includes a sampler and zero-order hold, followed by a delayed sampler (to account for reaction time), and then an integrator which reconstructs the very narrow pulses from S_2 into a continuous signal. However, the integrator is a non-resetting hold circuit, and consequently the reconstructed error signal e_R has no resemblance to the continuous input.

However, Ward's study, while limited, is useful in that it provides additional evidence to show that at least a portion of the "remnant" noise term required in continuous models may be accounted for by harmonics due to sampling. These effects of sampled-data systems will be reviewed in Chapter 3.

2.7 Summary and Evaluation

On the basis of the above discussion of various human operators, the following conclusions can be made:

(1) The transfer characteristics of the human operator can be adequately described by a quasi-linear differential equation plus a remnant term, provided that the input consists of a random-appearing signal of sufficiently low frequency.

(2) The quasi-linear continuous model fails to take into account several known characteristics of human tracking performance, including intermittency, adaptability, and ability to predict.

(3) Linear or quasi-linear models are preferable to nonlinear models since it is possible to use well known analytic techniques with linear systems. Furthermore, the open-loop parameters of the quasi-linear model can be deduced from closed-loop measurements for the compensatory tracking situation.

(4) Intermittent or sampling models have much supporting evidence but the work done on them to date is inadequate. There is a need for a study of sampled-data representations in an analytic form which makes generalization possible and which includes the operator's ability to extrapolate in the time domain as well as his frequency characteristics.

Chapter 3

THEORETICAL CONSIDERATIONS FOR THE SYNTHESIS OF A SAMPLED-DATA MODEL

3.1 Requirements and desirable features

In the previous Chapter the characteristics and some of the limitations of previous models of human operators have been reviewed. In this chapter we propose to present the major requirements for the new models, present the background from sampled-data theory, and develop the new model in a form suitable for analysis.

Based on the evidence and background from the previous Chapters, we can state the following requirements for the new models:

(1) The models must represent tracking behavior for random or random-appearing inputs

(2) For sufficiently low frequencies the models must agree with the quasi-linear continuous model since these models are excellent at low frequencies

(3) Any sampling frequencies present in these models must be of the same order of magnitude as those observed in various experiments on this subject.

In addition to this basic requirement, it is desirable that the model incorporate the two major features lacking in previous representations, i.e., extrapolation ability and adaptation. An additional desirable feature is that the models make use of the rate of change of error directly, in addition to the error signal itself. This ability to sense target velocity directly (and obtain position information by integration from it) would be a reasonable hypothesis based on recent studies of human visual perception (cf. Brown (1961)).

The approach taken to the synthesis problem here will be the following. In the first place, the validity of the quasi-linear continuous model will be assumed for low frequencies, since this model is precisely the linear system which minimizes the mean squared error of approximation to the actual tracker. The objective then will be to modify the continuous model in such a way that its low frequency behavior is not altered (at least in the limit as the frequency approaches zero), by the introduction of sampling and data reconstruction operations. In order to accomplish this objective, the next section will review the properties of sampled-data systems and these will be incorporated in a series of a priori models to be analyzed in future chapters.

The models proposed in this chapter will include the features of intermittency, extrapolation, and velocity sensing. However, they will not include continuous adaptation. The extension to the adaptive case will be discussed in Chapter 7.

3.2 Characteristics of Sampled-data Systems

The study of linear sampled-data systems is well established and several books on the subject are available(e.g. Tou, 1959, Ragazzini and Franklin, 1958). In this section we shall merely outline some of the major properties of such systems in order to use them for the construction of human operator models.

Sampling refers to the operation of converting a continuous signal $f(t)$ into a discrete signal $f^*(t)$ which carries information only at certain times. If the sampling operation is periodic and occurs every T seconds, the sampled signal carries information for a duration of h seconds during each sample period and equals zero at other times. Thus, sampling can be considered as a process of modulation of a continuous signal with a "sampling signal", a pulse train $p(t)$. We can indicate the sampler output as

$$f^*(t) = f(t)\, p(t) \tag{3.1}$$

where the sampling pulse train is given by

$$p(t) = \sum_{n=-\infty}^{+\infty} u(t-nT) - u\left[t-(nT+h)\right] \tag{3.2}$$

and $u(t)$ is the unit step function. The sampling operator or "sampler" is usually denoted as a switch, as indicated in Figure 3.1. If $h \ll T$ the pulse train can be represented by the train of Dirac delta functions of unity area:

$$p(t) = \sum_{n=-\infty}^{+\infty} \delta(t-nT) \tag{3.3}$$

and the resulting process is called "impulse modulation".# It should be noted that the sampler can be considered as a time-varying gain; thus, sampling is a linear operation. Since the sampler is a periodic amplifier, the resulting equations which describe the system operation will have periodic coefficients.

3.2.1 Frequency domain characteristics. When the impulse approximation is used, the sampler output becomes

$$f^*(t) = \sum_{n=0}^{\infty} f(nT)\delta(t-nT) \tag{3.4}$$

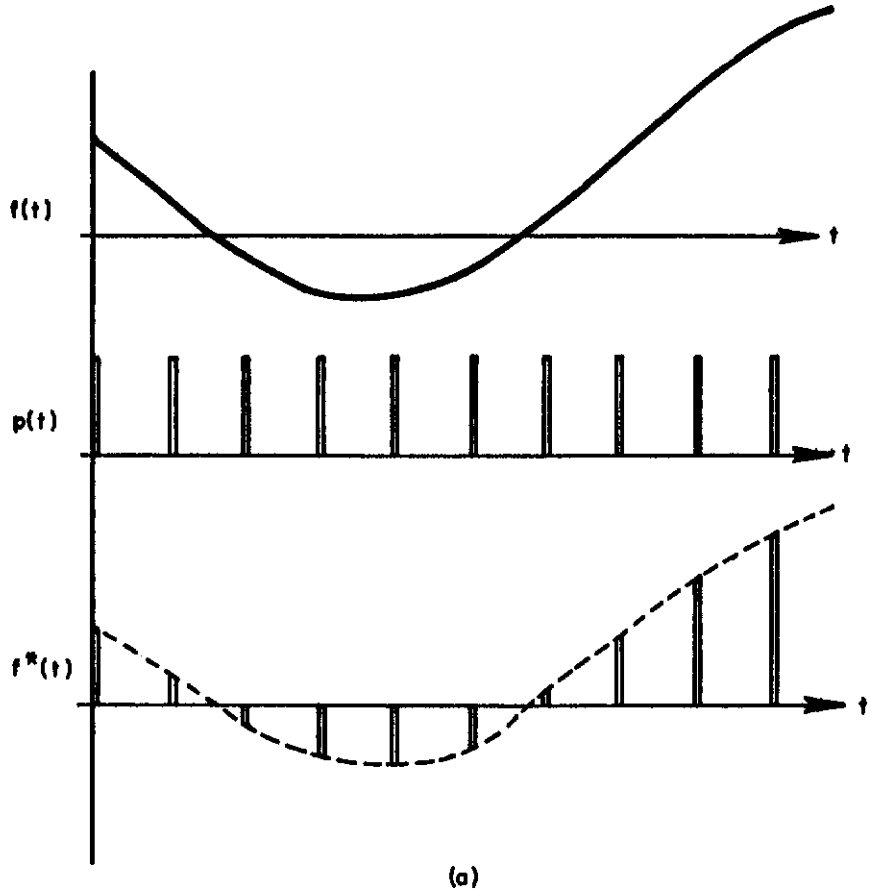
which represents a sequence of numbers $f(nT)$ modulating the train of impulses. The Laplace transform of (3.4) is

$$F^*(s) = \sum_{n=0}^{\infty} f(nT)e^{-nTs} \tag{3.5}$$

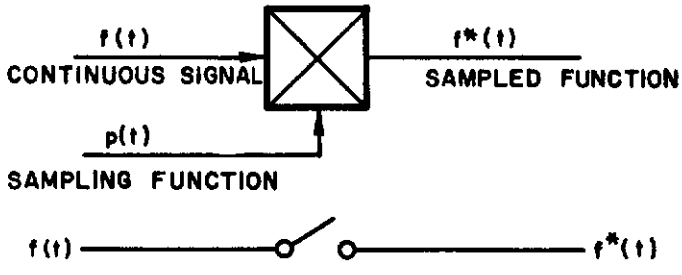
which clearly shows the number sequence, each number delayed appropriately to its position in the sequence by the operator

---

In engineering works on sampled-data systems the "delta function" is used formally and without rigor, thus creating some confusion. While the formal usage will be employed here also for simplicity, the development can be rigorized by reference to such works as Lighthill(1960).



(a)
SAMPLING OF CONTINUOUS SIGNAL



(b)
EQUIVALENT REPRESENTATIONS

Figure 3.1 Sampling of Continuous Signal

e^{-nTs}. The modulating pulse train can also be represented by a Fourier series (in the case of the impulse train this can be done by methods shown in Lighthill's book (1960)). As a result, the sampled signal is represented as

$$f^*(t) = \frac{1}{T} \sum_{n=-\infty}^{+\infty} f(t)\, e^{jn\omega_s t} \tag{3.6}$$

the Laplace transform of which is given by

$$F^*(s) = \frac{1}{T} \sum_{n=-\infty}^{+\infty} F(s - jn\omega_s) \tag{3.7}$$

where ω_s is the sampling frequency. Equation (3.7) indicates that sampling generates an infinite number of harmonics of the input frequency function $F(j\omega)$. The additional spectra are related to the amplitude of the original spectrum by the factor $1/T$ and separated by integral multiples of the sampling frequency. If the sampling pulse is of finite width it can be shown that the additional spectra are attenuated in amplitude increasingly as the frequency increases. (See Figure 3.2)

3.2.2 The sampling theorem. The fundamental theorem of sampling states that if a signal $f(t)$ has a frequency spectrum extending from zero to f_o cps, it is completely determined by the values of the signal (i.e. the samples) taken at a series of instants separated by $T = 1/2\ f_o$ sec, where T is the sampling period. (Goldman, 1953)

This theorem implies that if the sampling frequency is less than one-half the maximum signal frequency, then information is irretrievably lost in the sampling process. Any attempt to reconstruct a continuous signal from the samples will result in a distorted version of the original signal.

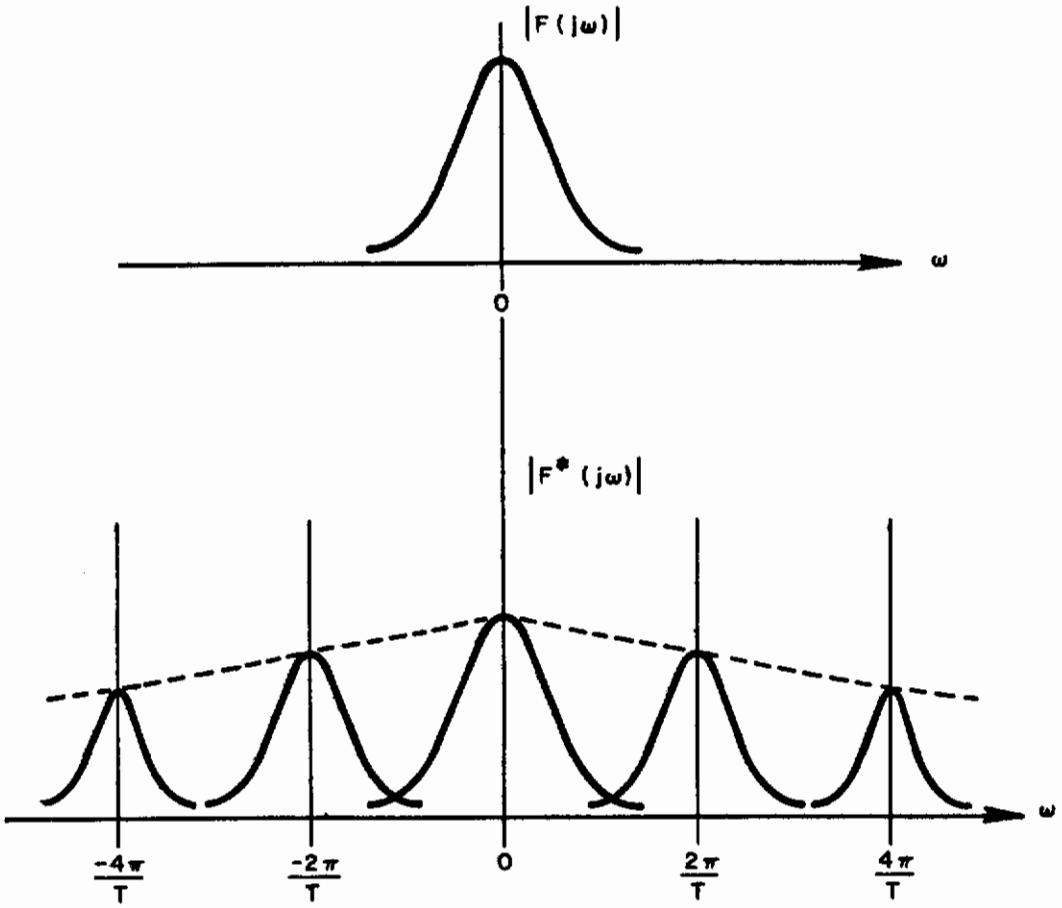


Figure 3.2 Frequency Characteristics of Sampled Functions

3.2.3 Data reconstruction. Various types of extrapolation devices are used to reconstruct a continuous signal from the sampled values f(nT). These devices are commonly called "hold" circuits.

The simplest or "zero-order" hold extrapolates between sampling instants by simply holding constant the last sample. In the time domain,

$$e(t) = e(nT) \quad , \qquad nT \leq t < (n+1)T \tag{3.8}$$

The next simplest hold circuit extrapolates with a constant slope between samples by using a slope estimate based on the last two samples, i.e.

$$e(t) = e(nT) + \frac{e(nT)-e\left[(n-1)T\right]}{T}(t-nT) \quad , \quad nT \leq t < (n+1)T \tag{3.9}$$

The action of these two circuits is illustrated in Figure 3.3.

The use of more past samples makes it possible to extrapolate with a higher-order polynomial. The frequency characteristics of the hold circuits will be discussed in the next section.

3.2.4 The z-transform. The study of linear sampled-data systems is greatly facilitated by the use of transform methods. The z-transform is particularly useful in systems where the impulse approximation is valid. The z-transform of an impulse sequence f(nT) is defined as

$$F(z) = \sum_{n=0}^{\infty} f(nT)z^{-n} \tag{3.10}$$

Comparison with (3.5) indicates the relation between the complex variable z and the complex variable s: i.e., by definition,

$$z \triangleq e^{Ts} \tag{3.11}$$

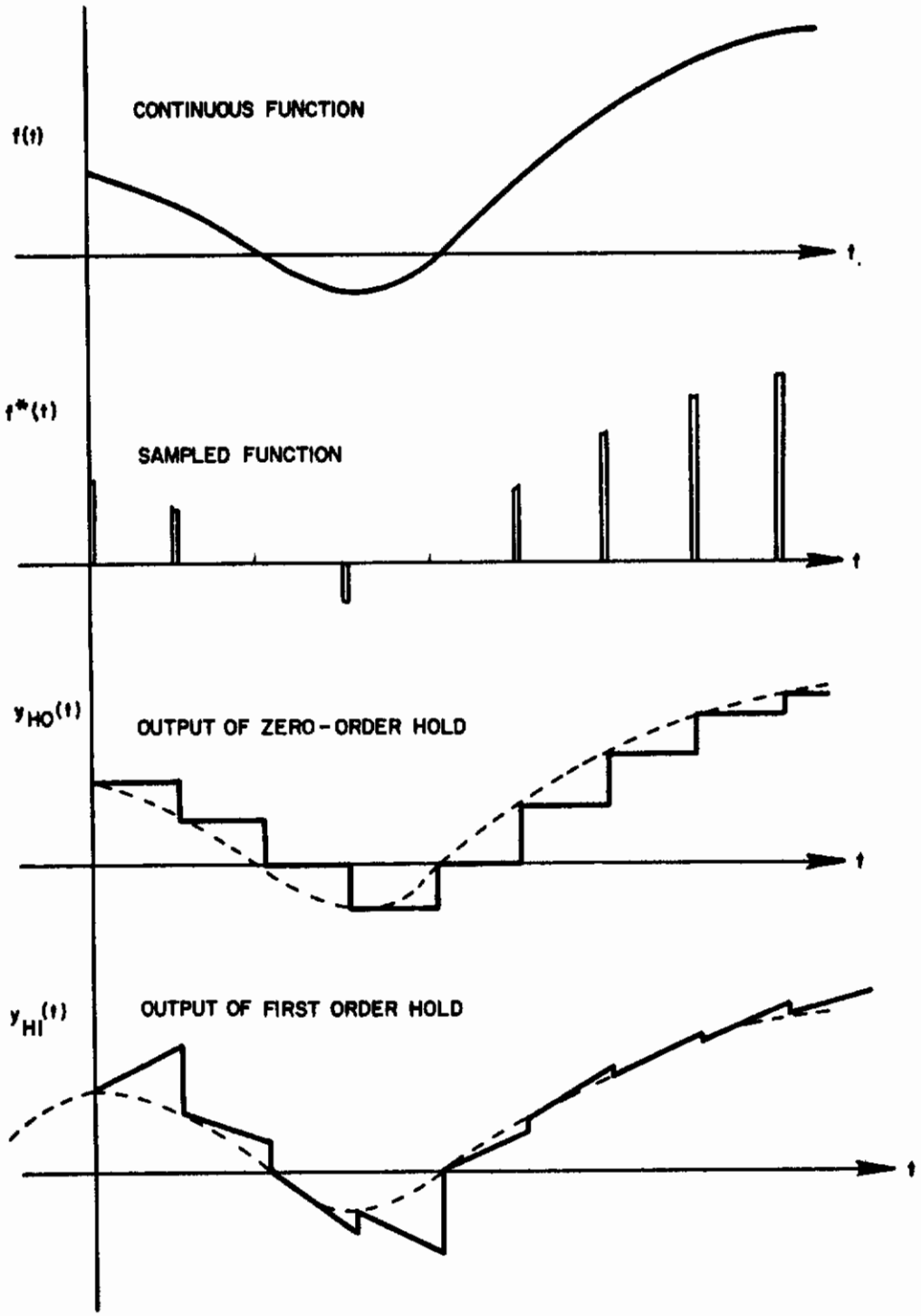


Figure 3.3 Zero and First Order Hold Circuits

In common usage the same letter is used for representing the Laplace transform of a time function and the corresponding z-transform. However, it should be noted that "taking the z-transform" implies a return to the time domain, finding the corresponding number sequence, and evaluating (3.10). In accordance with convention we shall indicate by G(z) the z-transform corresponding to a Laplace transform G(s). It can also be noted from (3.5) that

$$F(z) = F^*(s)\Big|_{s=T^{-1}\ell nz} \tag{3.12}$$

The periodicity of the z-transform is clearly seen by letting $s = j\omega$ in (3.11). The transformation $z = e^{Ts}$ is a mapping of the s-plane into a new z-plane, where the entire left-half s plane maps into the interior of the unit circle in the z-plane. Every strip of length $\omega_s = 2\pi/T$ of the $j\omega$ axis repeats a complete traversing of the unit circle in the z-plane.

3.2.5 Sampled-data systems. Consider the system of Figure 3.4, where G(s) represents a continuous, linear "plant". The continuous output y(t) can be obtained from the convolution summation as:

$$y(t) = \sum_{k=0}^{n} g(t-kT)\, x(kT), \qquad 0 < t < nT \tag{3.13}$$

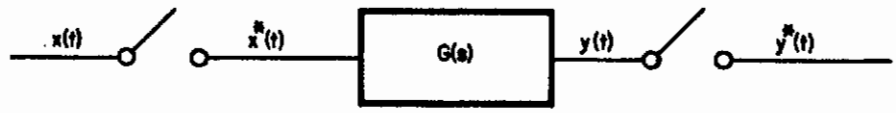


Figure 3.4 Open Loop Sampled Data System

where g(t) is the weighting function of the system. In terms of Laplace transforms,

$$Y(s) = X^*(s)\, G(s) \tag{3.14}$$

where X*(s) is obtained from (3.5). At the sampling instants we can compute the transform of the sampled output from

$$Y^*(s) = X^*(s)\, G^*(s) \tag{3.15}$$

where

$$G^*(s) = \sum_{n=0}^{\infty} g(nT)e^{-nTs} \tag{3.16}$$

or from the corresponding z-transform expression

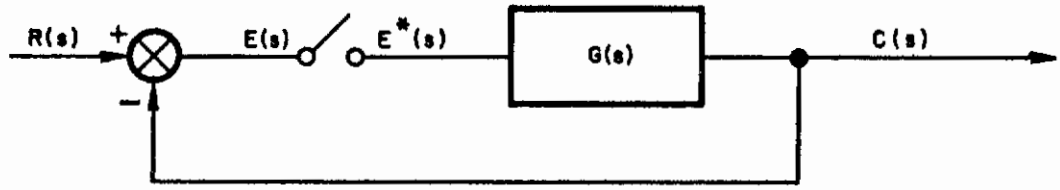
$$Y(z) = G(z)\, X(z) \tag{3.17}$$

The expression G(z) in (3.17) is termed the "pulse transfer function" corresponding to the continuous transfer function G(s). It should be noted that G(z) relates the frequency domain variables X(z) and Y(z) only at the sampling instants, and provides no information on behavior between these instants. Thus, inversion of Y(z) provides information only about y*(t), not y(t), (unless the sampling rate is sufficiently high compared to the highest frequency in y(t)).

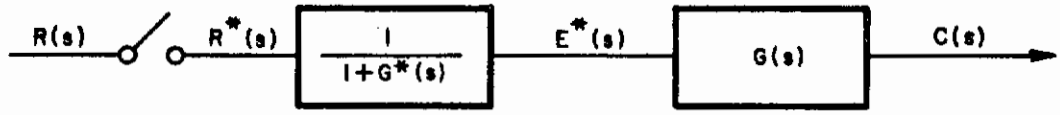
The corresponding closed-loop expressions may be obtained by referring to Figure 3.5(a) and its equivalent in 3.5(b). The expressions for the error and output transform for the error-sampled case are:

$$E^*(s) = \frac{R^*(s)}{1+G^*(s)} \tag{3.18}$$

$$C(s) = E^*(s)\, G(s) \tag{3.18}$$



(a)



(b)

Figure 3.5 Closed Loop System with Error Sampling

3.2.6 Summary of characteristics of sampled-data systems. The major features of sampled systems noted above may be summarized as follows:

(1) Unless the sampling rate is sufficiently high compared to the highest system frequencies, information is lost in the sampling process.

(2) The sampler can be considered a time-varying amplifier, thus, even if all the other elements in a sampled-data system are linear and invariant, the complete system is nonstationary.

(3) The frequency-domain properties of sampled-data systems are periodic and repeat in integral multiples of the sampling frequency.

(4) Sampling of a bandwidth-limited function generates harmonics which cover the entire frequency spectrum.

(5) In view of the fact that physically relizable filters do not have perfect cut-off characteristics, the reconstruction of sampled data always results in the addition of "ripple" between sampling instants.

3.3 Data Reconstruction Circuits

The time-domain behavior of zero-order and first-order hold circuits was illustrated above in Figure 3.3. In this section the frequency characteristics of these circuits will be reviewed and a new "modified" hold circuit will be introduced. The use of this "hold" circuit in human operator models will be considered in Chapter 4.

3.3.1 The zero-order hold circuit can be described by the transfer function

$$H_o(s) = \frac{1-e^{-Ts}}{s} \tag{3.20}$$

The frequency characteristics with sinusoidal inputs can be obtained by letting $s = j\omega$ to obtain

$$(3.21) \qquad H_o(j\omega) = T \left| \frac{\sin\omega T/2}{\omega T/2} \right| \angle -\omega T/2$$

From this relation it can be seen that the zero-order-hold exhibits a (sin x/x) amplitude characteristic (with zeros at frequencies $2n\pi/T$) and a phase shift proportional to frequency. Such a linear phase characteristic is exhibited by a pure time delay of T/2 seconds, i.e., one-half the sampling period.

3.3.2 The first-order-data hold. As indicated in Figure 3.3 and equation (3.9), the first-order hold extrapolates between sampling instants with a constant slope derived from the last two samples. The corresponding Laplace transform is

$$(3.22) \qquad H_1(s) = \left(\frac{1+Ts}{T} \right) \left(\frac{1-e^{-Ts}}{s} \right)^2$$

The frequency characteristics of this hold circuit are obtained by replacing s by $j\omega$ in (3.22) and simplifying the resulting expression to the following:

$$(3.23) \qquad H_1(j\omega) = T\sqrt{1+\omega^2T^2} \left(\frac{\sin\omega T/2}{\omega T/2} \right) \angle -(\omega T)+\tan^{-1}(\omega T)$$

The phase and amplitude response of the zero and first-order holds are plotted in Figure 3.6. From an examination of these curves, two observations can be made immediately:

(1) The first-order hold exhibits significant "peaking" in the amplitude characteristic, near one-half the sampling frequency. (Note that an ideal data hold would weight all frequencies equally).

(2) The zero-order hold exhibits more phase shift than the first-order hold at low frequencies (below $\omega T \cong 2.4$) and less phase shift at higher frequencies.

These observations are clearly of importance to the design of sampled-data control systems since they affect the amount of ripple introduced by the hold circuit and the stability of the system. For the synthesis of human operator models they give us clues to be sought in experiments. Thus, it will be of interest to examine human tracking records to seek for peaks near one-half the assumed sampling frequency; and phase-shift introduced by the hold circuit will have to be considered in relation to the total observed "reaction time".

3.3.3 Partial velocity (PV) hold. The excessive peaking observed in the first-order-hold can be reduced by using a data-hold with only partial velocity extrapolation. Consider the hold circuit transfer function in the form

$$(3.24) \qquad H_k(s) = \frac{k(1+Ts)}{T}\left(\frac{1-e^{-Ts}}{s}\right)^2 + (1-k)\left(\frac{1-e^{-Ts}}{s}\right)$$

When $k = 0$, (3.24) reduces to a zero-order hold; if $k = 1$ it becomes a first-order hold, and with $0 < k < 1$ it is denoted as a "partial-velocity-correction" hold. We shall abbreviate it as "PV hold" in this report. The frequency response of PV holds for $k = 0.5$ is also indicated in Figure 3.6.

3.3.4 Evaluation of hold circuits. If it is assumed that a reasonable sampling period for human operator models would be in the range of 0.25 - 0.50 sec (based on the experiments reviewed in Appendix 1), then some conclusions regarding data holds are possible. Typical "reaction times" for human operators in closed loop tracking situations with no controlled element dynamics, according to McRuer and Krendel (1957), are in the range of 0.12 to 0.16 seconds. Consequently, if zero-order hold circuits are used in the model, the phase shift may be excessive. For first order holds, it may be too small for low frequencies and too high for high frequencies, the "effective time delay" being such as to require the addition of a predictor (or time advance element) to the model.

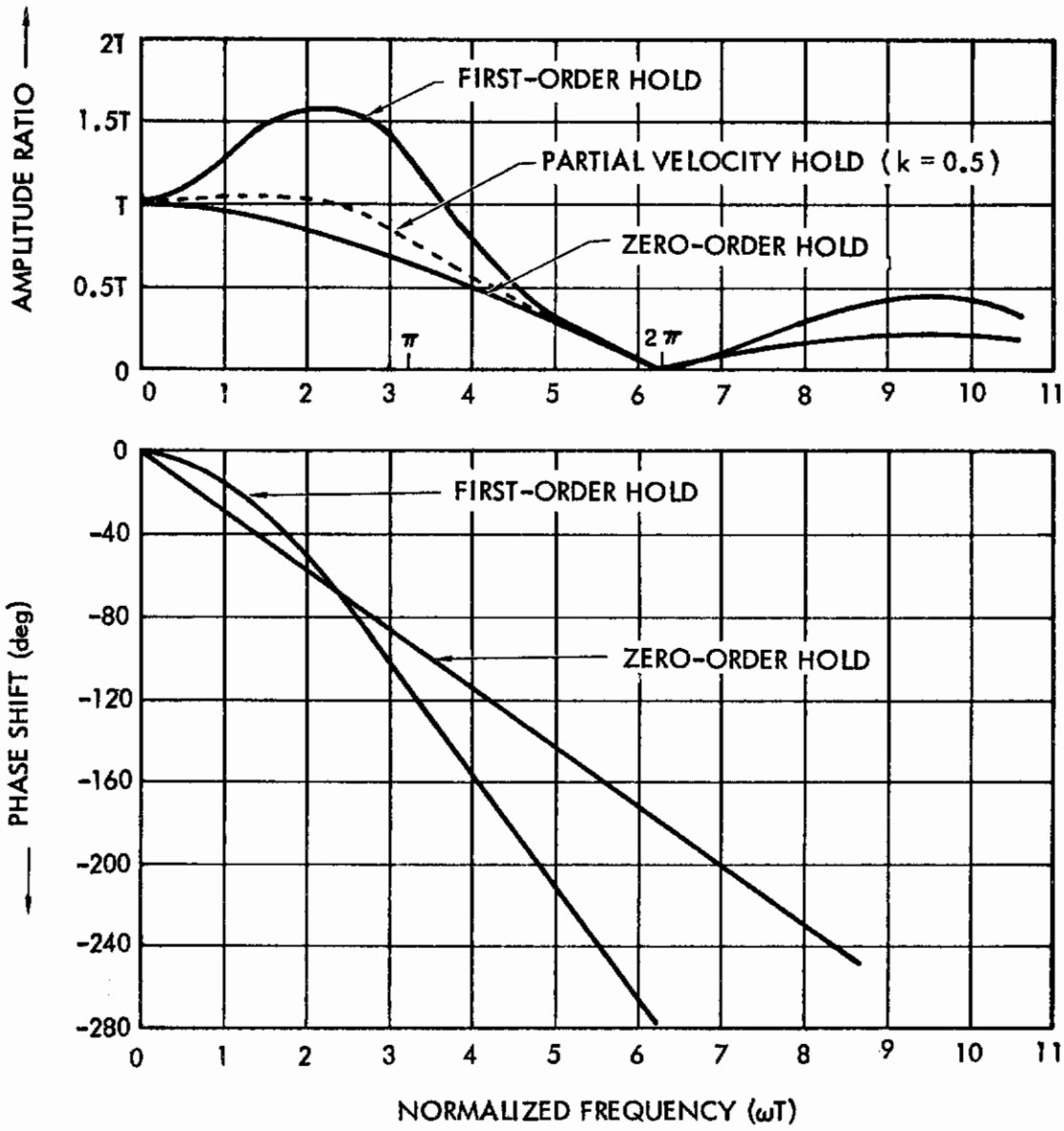


Figure 3.6 Frequency Characteristics of Zero and First-Order Hold Circuits

On the basis of known data on tracking, the position extrapolator does not appear to be a reasonable assumption. However, the velocity extrapolator (or first-order hold) which does appear reasonable uses a measurement of velocity based on past samples which causes it to exhibit considerable phase lag. In the following section a hold circuit based on measurement of current (rather than past) velocity is analyzed. This circuit apparently does not exist in the literature.

3.4 A Modified Partial Velocity Hold

As outlined in Paragraph 3.3.2 above, the transfer function of a first order hold is given by equation (3.22) which is repeated here:

$$H_1(s) = \left(\frac{1+Ts}{s}\right)\left(\frac{1-e^{-Ts}}{s}\right)^2 \tag{3.22}$$

This expression being obtained from the impulse response of a circuit which has the desired time domain performance (cf. Ragazzini and Franklin, 1958). Expression (3.22) can be visualized more clearly if it is diagrammed as shown in Figure 3.7,

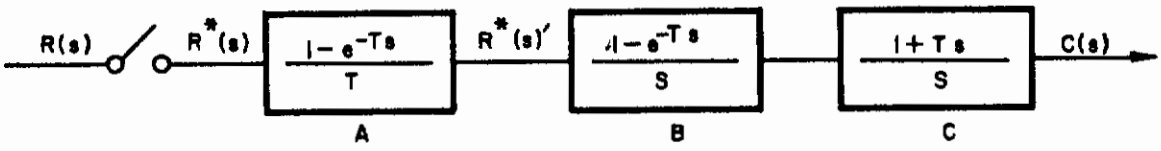


Figure 3.7 Block Diagram of First Order Hold

where block A represents discrete differentiation, block B represents a zero-order hold and block C represents the reconstruction of the continuous signal r(t) from two operations: a ramp given by the term 1/s and a constant amplitude term given by the term T. If the continuous signal r(t) is available a more accurate slope estimate is possible by using continuous differentiation. It is interesting to note that this type of first-order-hold (to be called the "modified F.O.H. circuit") exhibits much less phase shift than the conventional circuit whose frequency behavior was given in Figure 3.6.

The modified F.O.H. circuit is diagrammed in Figure 3.8(a) and its time-domain extrapolation properties shown in Figure 3.8(b).

3.4.1 Frequency characteristics. To evaluate the frequency characteristics of the Modified F.O.H. we again assume the input to be sinusoidal (this is equivalent to considering only the fundamental frequency at the output of the sampler with a sinusoidal input and neglecting all harmonics). The frequency function for the conventional F.O.H. was given in (3.23) as

$$(3.23) \qquad H_1(j\omega) = T\sqrt{1+\omega^2T^2}\left(\frac{\sin \omega T/2}{\omega T/2}\right)^2 \angle -\omega T+\tan^{-1}\omega T$$

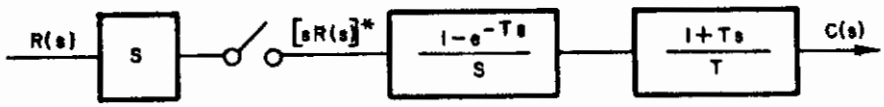
The frequency function of the modified F.O.H. circuit of Figure 3.8(a) is given by inspection by the relation

$$(3.25) \qquad H_{1M}(j\omega) = T\sqrt{1+\omega^2T^2}\left|\frac{\sin \omega T/2}{\omega T/2}\right| \angle -\frac{\omega T}{2}+\tan^{-1}\omega T$$

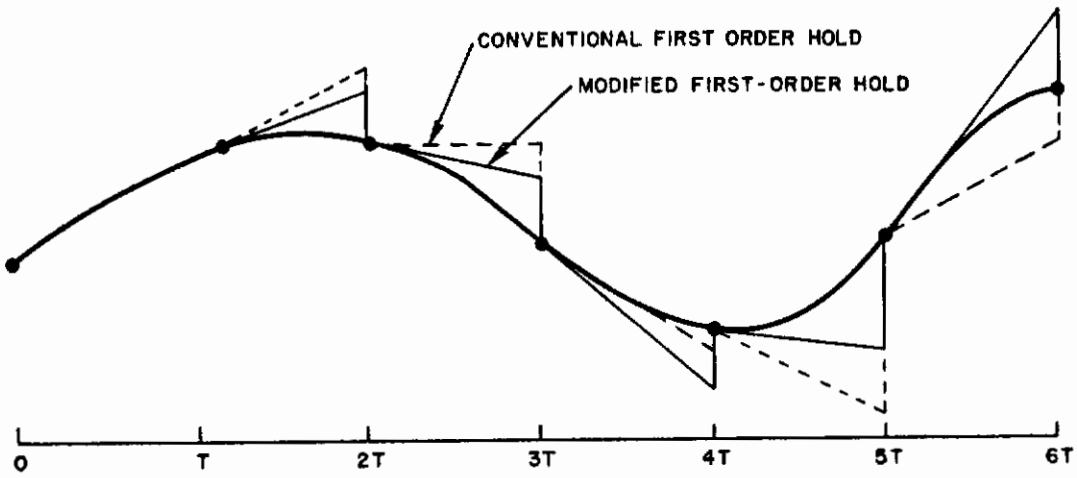
Two facts stand out from a comparison of these expressions:

(1) The modified F.O.H. has $\omega T/2$ radians less phase shift;

(2) The reduction in phase lage is obtained at the expense of more ripple, i.e., less high frequency attenuation. In fact, (3.25) exhibits an amplitude oscillation which tends toward $\sin \omega T/2$ as ωT increases, while the amplitude of (3.23) tends to zero as $\omega T \to \infty$



(a)



(b)

Figure 3.8 Modified First Order Hold Circuit
(a) Block Diagram
(b) Data Reconstruction Between Samples

In a practical situation two factors would tend to mitigate both the amplitude disadvantage and the phase advantage of the modified circuit. First, pure differentiation is not practical due to noise considerations, and consequently high frequency filtering will be required. The practical "differentiator" would have a transfer function

$$G_d(s) = \frac{s}{1+s/\omega_d} \tag{3.26}$$

where ω_d is selected sufficiently large to yield the desired accuracy of differentiation at low frequencies and small enough to limit the resulting high frequency noise. Secondly, a partial velocity correction would often be used (instead of a full velocity correction) thus resulting in less peaking near $\omega T = \pi$.

If the approximate differentiation of (3.26) is used, the frequency function of the modified F.O.H. becomes

$$H_{1M}'(j\omega) = T\frac{\sqrt{1+\omega^2T^2}}{\sqrt{1+\omega^2\tau_d^2}}\left|\frac{\sin \omega T/2}{\omega T/2}\right| \angle -\frac{\omega T}{2} + \tan^{-1}\omega T - \tan^{-1}\omega\tau_d \tag{3.27}$$

where $\tau_d = 1/\omega_d$. Equation (3.27) was actually used to implement the modified F.O.H. in the simulation study described in Chapter 5.

3.4.2 Modified PV hold circuits. Equations (3.25) and (3.27) define the frequency characteristic of a modified first-order hold circuit, but they are readily adapted to the partial velocity correction case. The "conventional" and "modified" PV holds are shown in block diagram form in Figure 3.9. If k = 0 both diagrams reduce to the zero-order hold and if k = 1 they reduce to the F.O.H. and modified F.O.H. circuits respectively.

The frequency characteristics of the modified hold circuits are graphed in Figures 3.10, 11, and 12. Figure 3.10 shows the comparison between the conventional and modified first-order hold circuits and illustrate clearly the decrease in phase shift and

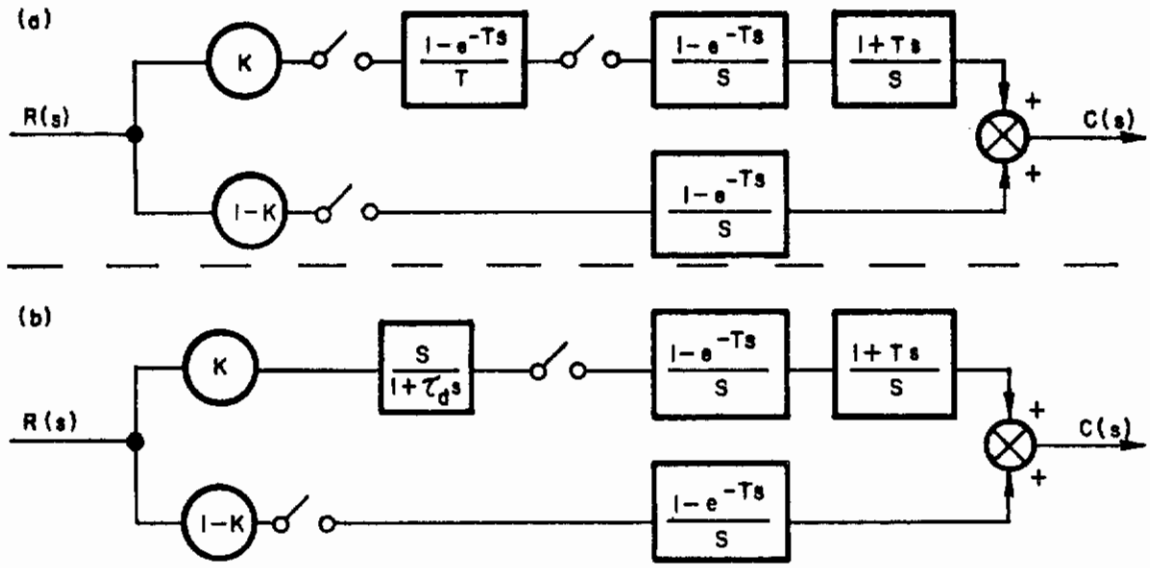


Figure 3.9 Partial Velocity Correction Hold Circuits
(a) Conventional; (b) Modified

decrease in high frequency attenuation which accompany the introduction of continuous differentiation. Figure 3.11 shows the amplitude characteristics of the modified PV hold with several values of k and Figure 3.12 shows the phase characteristic of the modified PV hold for various values of k.

The simulation of the modified hold circuits by means of an analog computer is described in Appendix 4.

3.5 Low Frequency Behavior of Hold Circuits

One of the requirements for the sampling and data-reconstruction portions of the human operator models to be proposed is that the model differe negligibly from the continuous model as the input frequency approaches zero. It is clear that the sampler contributes less and less error as the ratio ω_s/ω_i, sampling frequency to input frequency, increases. This is clear from the sampling theorem. The behavior of the hold circuit is still in doubt.

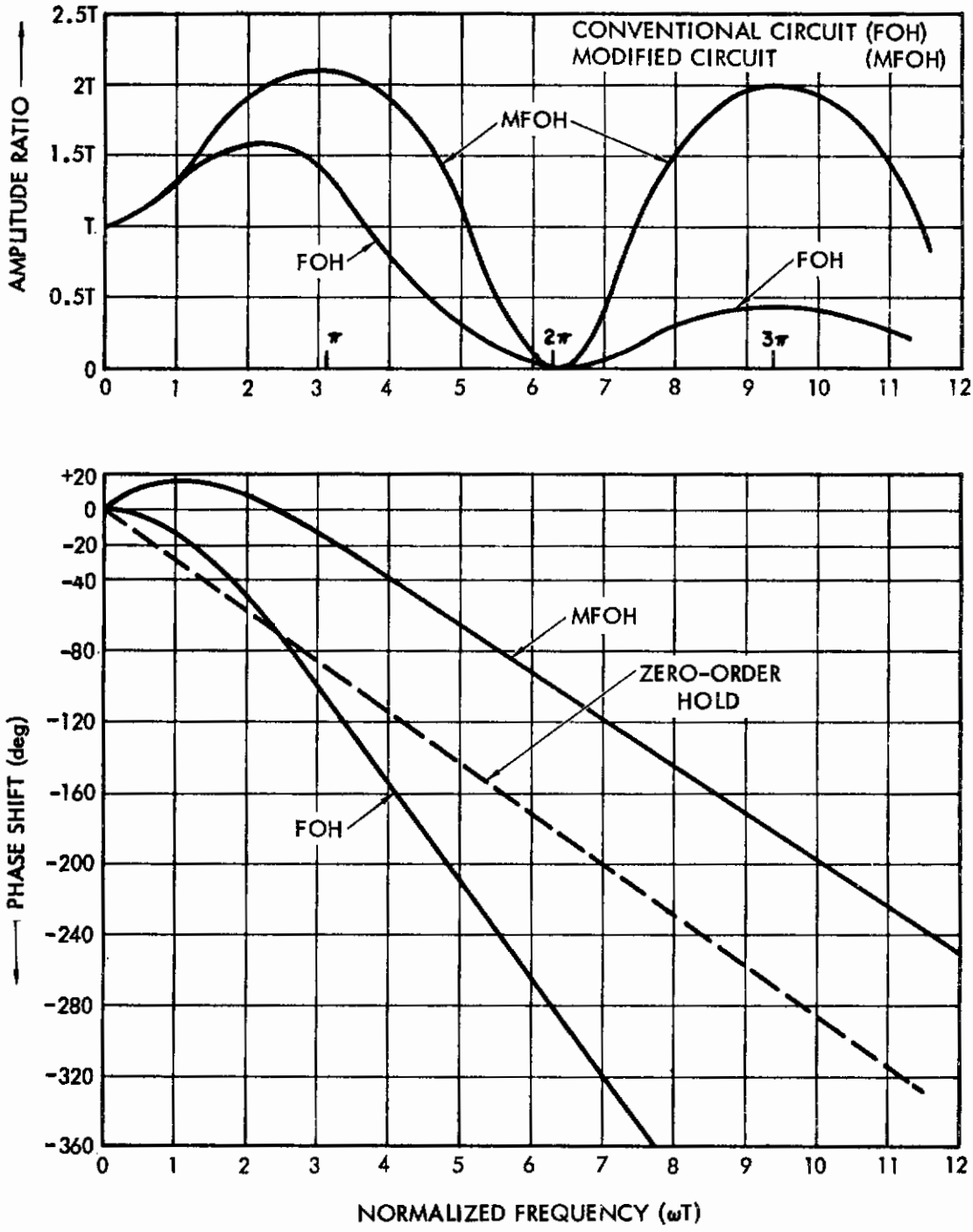


Figure 3.10 Comparison of Frequency Characteristics of First-Order Hold Circuits

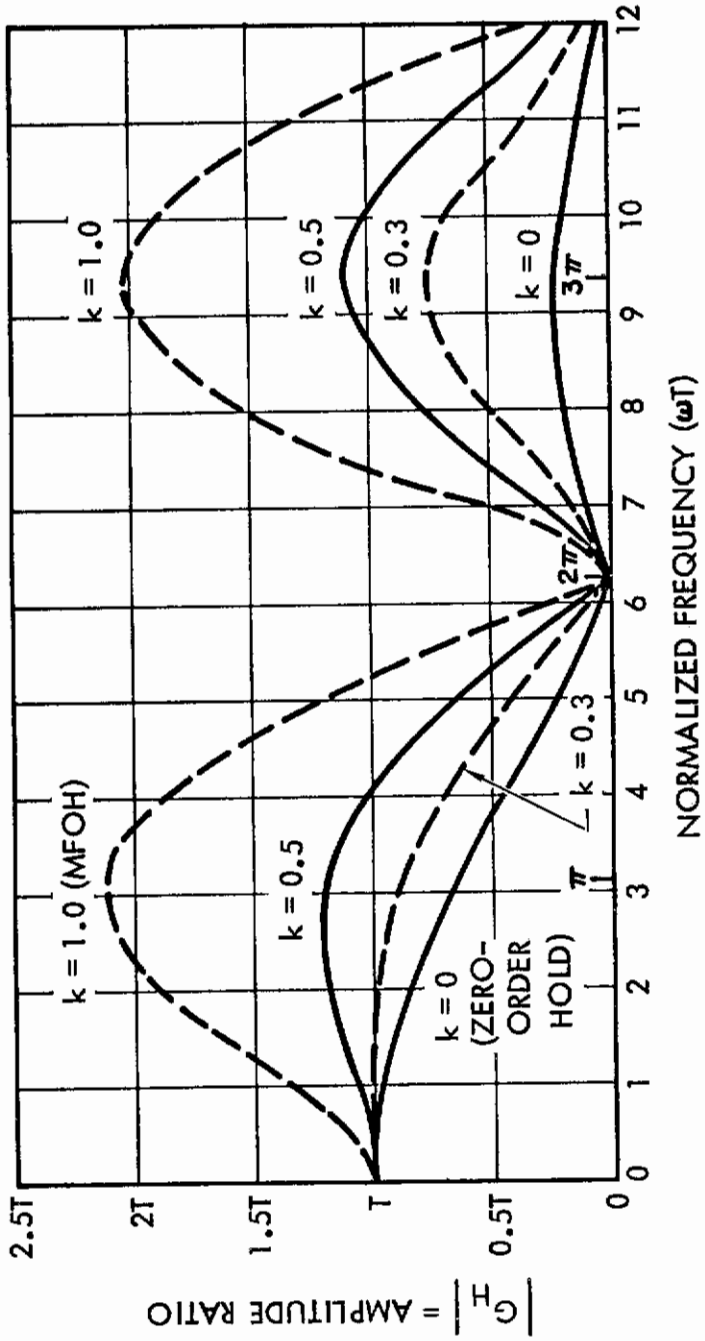


Figure 3.11 Amplitude-Frequency Characteristic of Modified Partial Velocity Holds

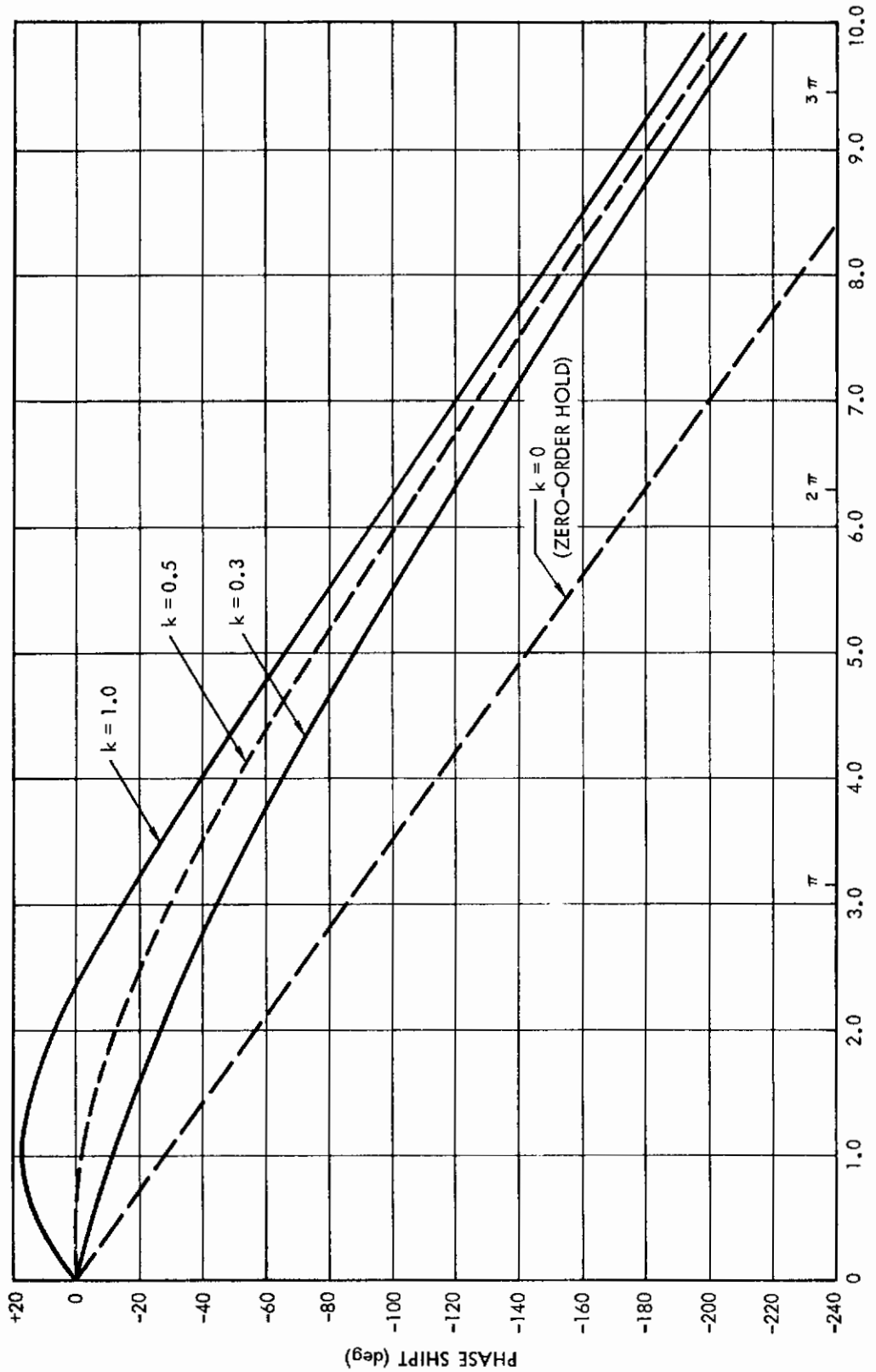


Figure 3.12 Phase-Frequency Curves of Modified Partial Velocity Hold Circuits

The amplitude vs. frequency plots of the hold circuits given in this chapter indicate that

$$\left|H_i(j\omega)\right| \to T \qquad \text{as} \quad \omega \to 0$$

and since the sampler attenuates the input signal by $1/T$ the sample-hold combination approaches unity gain (and 0 phase shift) as $\omega \to 0$. This can also be seen by applying L'Hospital's rule to the frequency functions concerned. For the zero order hold we have, from (3.21):

$$(3.28) \qquad \left|H_o(0)\right| = T \lim_{\omega \to 0} \frac{\sin \omega T/2}{\omega T/2}$$

$$= T \lim_{\omega \to 0} \cos \omega T/2 = T$$

For the first order hold, from (3.23):

$$(3.29) \qquad \left|H_1(0)\right| = T \lim_{\omega \to 0} \sqrt{1+\omega^2 T^2} \left(\frac{\sin \omega T/2}{\omega T/2}\right) = T$$

and similarly for the modified hold.

3.6 Hold Circuits in Human Operator Models

We have seen in the above paragraphs that the data hold or data reconstruction circuit possesses many of the properties required of the intermittent element of the proposed human operator models. In particular, the sample-and-hold elements considered have the following properties.

(1) Their transfer characteristics approach unity as $\omega \to 0$.

(2) They introduce ripple between sampling instants which includes frequencies beyond the range of a band-limited input.

(3) They extrapolate between samples on the basis of the input and/or input rate available.

It thus appears to be reasonable to propose that human operator models be constructed by inserting error-sampling and data-

holds into the continuous models previously utilized. The generic model is of the form indicated in Figure 3.13, where the hold circuit could be of any of the forms discussed in this chapter.

The diagram of Figure 3.13 indicates some of the restrictions on the proposed model: it is designed for tracking with stationary random or random-appearing inputs (thus excluding the single sine-wave inputs) and it is designed for compensatory tracking with the assumption that all other inputs to the operator are negligible by comparison. It is particularly important to note that the "human operator model" block in Figure 3.13 has continuous inputs and outputs, and consequently the existence of sampling cannot be verified by the type of mathematical investigation being conducted. It is possible, however, to examine the implications of sampling and examine the system behavior in view of these implications. In order to do this, in the next chapter, we shall analyze this model with several types of hold circuits and obtain expressions for spectral density functions of the error e(t) and output c(t) which will form the basis for the experimental study reported in Chapters 5 and 6.

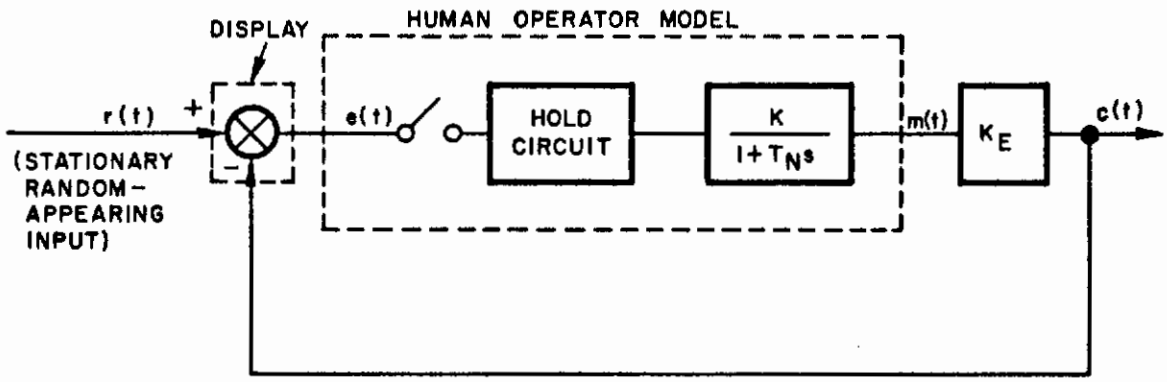


Figure 3.13 Proposed Sampled-Data Model of the Human Operator

Chapter 4

ANALYSIS OF OPERATOR MODELS WITH RANDOM INPUTS

4.1 Introduction

In the previous chapters the background for a sampled-data model of the human operator has been presented. A specific configuration was proposed in Chapter 3, based on the addition of sample-and-hold circuits to "best fit" linear continuous models. In this chapter we propose to examine the power spectra of tracking loops which include human operator models. In particular, closed-form expressions for computation of the power spectral density of the error and output signals in a system of the form of Figure 4.1 will be considered.

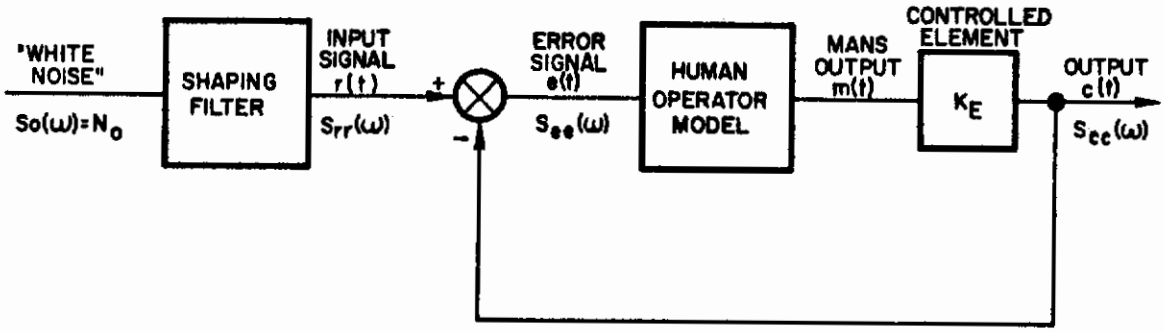


Figure 4.1 Compensatory Tracking System

The following assumptions will be made in the system of Figure 4.1:

(1) The system input will be shaped "white" noise, i.e. it will consist of a stationary random process with constant power spectral density N_o, zero mean, and a Gaussian amplitude distribution;

(2) The controlled system will be approximated by a pure gain K_E;

(3) The operator will be represented by either a continuous model or its equivalent sampled system.

In the concluding portion of the chapter some numerical solutions to the problem will be obtained.

The study of continuous linear systems with random inputs is well established. The relations used as a starting point in this chapter are taken primarily from Davenport and Root (1958), Bendat (1958) and Laning and Batting (1956). The study of linear sampled-data systems with random inputs has been treated in several technical papers. However, since no one paper presents an integrated treatment in a form suitable for this analysis, the development of the required expressions is given in Appendix 2.

4.2 Analysis of System with Continuous Model

If the input to a linear, constant coefficient filter is a stationary process* $x(t)$ characterized by a power spectral density $S_{xx}(j\omega)$ then the power spectral density of the output $y(t)$ is

*Since the processes discussed here are Gaussian stationary random processes and a Gaussian process is completely characterized by its first two moments, these processes are stationary in the strict sense (cf. Davenport and Root, 1958).

determined from the relation

$$S_{yy}(\omega) = \left|F(j\omega)\right|^2 S_{xx}(\omega) \tag{4.1}$$

where $F(j\omega)$ is the frequency response function of the filter and $S_{yy}(j\omega)$ is the output power spectral density. In the case considered here, the noise source is assumed "white" and consequently

$$S_o(\omega) = N_o \tag{4.2}$$

Therefore the input to the tracking system will have a spectral density

$$S_{rr}(\omega) = \left|F(j\omega)\right|^2 N_o \tag{4.3}$$

We are interested in obtaining an expression for the error spectral density and the output spectral density when the continuous operator model is of the form

$$G_c(j\omega) = \frac{K_{HE}\, e^{-j\omega D_c}}{1 + j\,\tau_c\,\omega} \tag{4.4}$$

where K_{HE} is a gain which includes the model gain K_H and the controlled element gain K_E, D_c is the model time-delay in the continuous case and τ_c is the model lag. (The background of eq. (4.4) was given in Chapter 2). The output function $C(j\omega)$ is given by

$$C(j\omega) = \frac{G_c\ (j\omega)}{1 + G_c\ (j\omega)}\ R\ (j\omega) \tag{4.5}$$

since the system has unity feedback, and the error signal transform is given by

$$E(j\omega) = \frac{1}{1 + G_c\ (j\omega)}\ R\ (j\omega) \tag{4.6}$$

Consequently, by application of equation (4.1), the error spectrum is given by

$$S_{ee}(\omega) = \left|\frac{1}{1 + G_c\ (j\omega)}\right|^2\ S_{rr}(\omega) \tag{4.7}$$

and the output spectrum is given by:

$$S_{cc}(\omega) = \left|\frac{G_c\ (j\omega)}{1 + G_c\ (j\omega)}\right|^2\ S_{rr}(\omega) \tag{4.8}$$

By using the model defined in (4.4) we can write

$$\left|\frac{1}{1 + G_c(j\omega)}\right|^2 = \frac{1 + j\tau\omega}{1 + j\tau\omega + \underset{Ke}{\ } - jD\omega}\quad \frac{1 - j\tau\omega}{1 - j\tau\omega + \underset{Ke}{\ } + jD\omega} \tag{4.9}$$

and consequently the error spectrum is given by

$$S_{ee}(\omega) = \frac{(1 + \tau^2\omega^2)}{(1 + K^2) + \tau^2\omega^2 + 2K\ (\cos \omega D - \tau\omega \sin \omega D)}\ S_{rr}(\omega) \tag{4.10}$$

where the subscripts on τ, K and D have been dropped for convenience, and

$$S_{rr}(\omega) = \frac{N_o\ \omega_B{}^2}{\omega^2 + \omega_B{}^2} \tag{4.11}$$

since

$$F(j\omega) = \frac{1}{1 + j\omega/\omega_B} \tag{4.12}$$

Finally, then, the error power spectral density will be given by:

$$S_{ee}(\omega) = \frac{(1 + \tau^2\omega^2)\,\omega_B^{\,2}\,N_o}{(\omega^2 + \omega_B^{\,2})\left[(1+K^2) + \tau^2\omega^2 + 2K(\cos \omega D - \tau\omega \sin \omega D)\right]} \tag{4.13}$$

Similarly, the compute the output power spectral density we first obtain the squared magnitude of the closed-loop system function:

$$\left|\frac{G_c(j\omega)}{1 + G_c(j\omega)}\right|^2 = \left(\frac{K_e^{-j\omega D}}{1 + \tau j\omega + Ke^{-j\omega D}}\right)\left(\frac{K_e^{j\omega D}}{1 - j\tau\omega + Ke^{j\omega D}}\right) \tag{4.14}$$

Substituting (4.14) into (4.8) we get:

$$S_{cc}(\omega) = \frac{K^2\,\omega_B^{\,2}\,N_o}{(\omega^2 + \omega_B^{\,2})\left[(1+K^2) + \tau^2\omega^2 + 2K(\cos \omega D - \tau\omega \sin \omega D)\right.} \tag{4.15}$$

Expressions (4.13) and (4.15) represent power spectra which can be easily measured in experimental situations and compared with their theoretical values. Before performing any numerical computations, however, the corresponding expressions for the sampled case will be developed.

4.3 Analysis of the Zero-Order-Hold Model

The simplest of the proposed sampling models is of the form shown in Figure 4.2, where the hold circuit has the frequency function

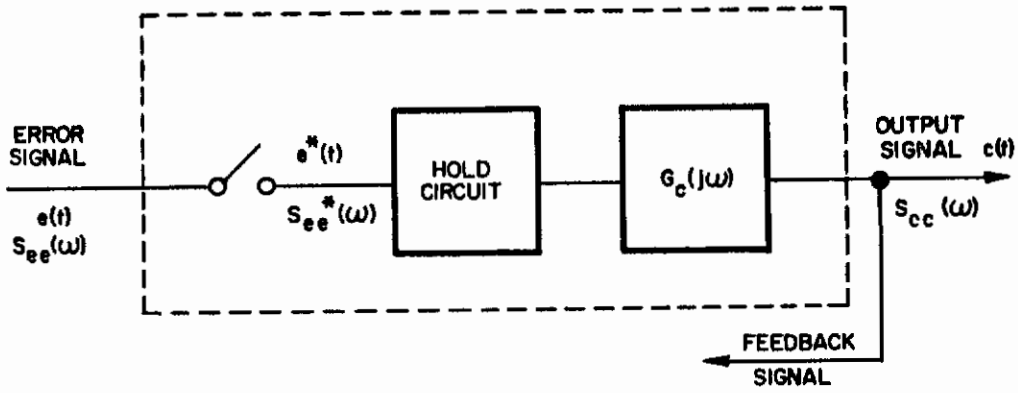


Figure 4.2. Sampled Form of Operator Model

$$H_o(j\omega) = \frac{1 - e^{-jT\omega}}{j\omega} \tag{4.16}$$

The derivation of an expression for the output power spectral density follows a procedure analogous to that for the continuous system, with complications introduced by sampled signals which have repeated spectra along the entire frequency axis. The power spectral density of the continuous error signal e(t) cannot be obtained by a single relation analogous to (4.1) since there is no transfer relationship which explicitly relates $E(j\omega)$ to $R(j\omega)$ in an error sampled system. The theory of the method to be used is given in Appendix 2.

4.3.1 Output power spectrum of zero-order-hold model. To compute the output spectrum we rearrange the circuit of Figure 4.2 into the form of Figure 4.3. The asterisks in the figures are used to denote sampled quantities. To obtain the output spectral density $S_{cc}(j\omega)$ we perform the following operations:

(a) Given the input spectrum $S_{rr}(j\omega)$ compute the corresponding sampled spectrum $S_{rr}^*(j\omega)$. This can be performed using ordinary z transform tables when $S_{rr}(j\omega)$ is a real, rational function of ω^2. Thus, we can write

(4.17) $$S_{rr}^{*}(s) = \frac{1}{T} Z\left[S_{rr}(s)\right]_{z = e^{sT}} = \frac{1}{T^2} \sum_{n=-\infty}^{+\infty} S_{rr}(s + jn\omega_s)$$

where the additional factor of $\frac{1}{T}$ arises due to time-averaging over a sampling period and $S_{rr}(s)$ is the bilateral Laplace transform representation of the power spectral density. The justification for expression (4.17) is given in Appendix 2.

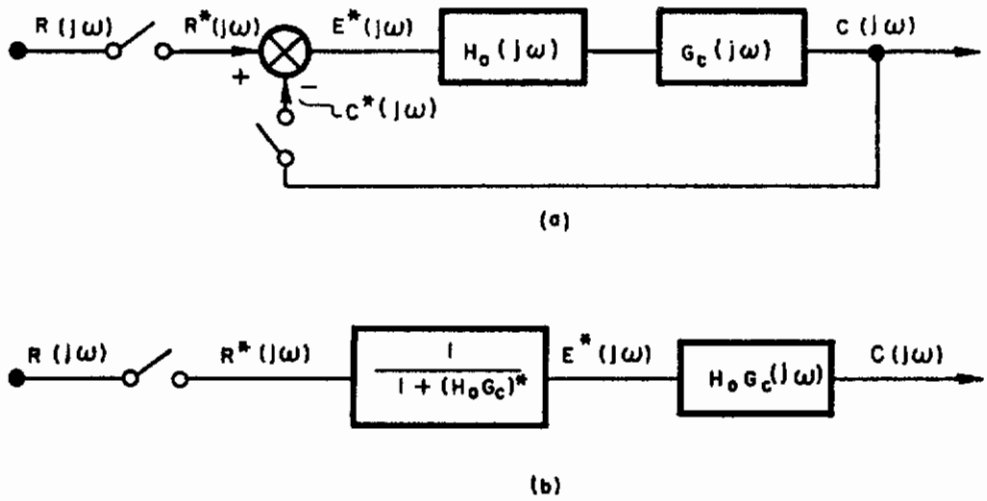


Figure 4.3. Equivalent Representation of Error-Sampled System

(b) Using the sampled input spectrum $S_{rr}^{*}(j\omega)$, and the pulse transfer function of the system, compute the sampled error spectrum $S_{ee}^{*}(j\omega)$ from

(4.18) $$S_{ee}(z) = \left[\frac{1}{1 + G_{so}(z)}\right]\left[\frac{1}{1 + G_{so}(z^{-1})}\right] S_{rr}(z)$$

where

$$(4.19)\quad G_{so}(j\omega) = H_o(j\omega)\, G_c(j\omega) = \left(\frac{1 - e^{-j\omega T}}{j\omega}\right)\left(\frac{Ke^{-Dj\omega}}{1 + j\tau\omega}\right)$$

and $G_{so}(z) = Z\left[G_{so}(s)\right]$. The use of the z-transform in (4.18) is valid since both $e^*(t)$ and $r^*(t)$ are defined only at the sampling instants.

(c) Using the sampled error spectrum $S_{ee}{}^*(j\omega)$ compute the power spectral density of the continuous output $S_{cc}(j\omega)$ from:

$$(4.20)\quad S_{cc}(\omega) = \left|G_{so}(j\omega)\right|^2 S_{ee}{}^*(j\omega)$$

The real frequency equivalent of $S_{ee}(z)$ is obtained simply by the substitution $z = e^{j\omega T}$.

To perform the series of operations outlined above it is recalled from (4.11) that, using the bilateral s-transform,

$$(4.21)\quad S_{rr}(s) = \frac{1}{1 - (s/\omega_B)^2}\quad N_o = \frac{N_o\,\omega_B{}^2}{-s^2 + \omega_B{}^2}$$

Consequently, from (4.17)

$$(4.22)\quad S_{rr}(z) = \frac{1}{T} Z\left[S_{rr}(s)\right] = \frac{-N_o\omega_B\,(\sinh \omega_B T)\, z}{T(z^2 - 2z\cosh\omega_B T + 1)}$$

which exhibits the periodicity characteristic of sampled spectra. The pulse transfer function $G_{so}(z)$ is given by

$$(4.23) \qquad G_{so}(z) = Z\left[\left(\frac{1 - e^{-Ts}}{s}\right)\left(\frac{Ke^{-Ds}}{1+\tau s}\right)\right]$$

$$= K(1 - z^{-1})\, Z_m\left[\frac{e^{-Ds}}{s(1 + \tau s)}\right]$$

where Z_m represents the "modified Z-transform" (cf. Tou, (1959) pp. 184-198). The use of tables of modified transforms makes it possible to obtain the transform of a system with time delay directly. From tables in Tou (p. 588, no. 3.01) we obtain

$$(4.24) \qquad G_{so}(z) = \frac{K\left[z(1 - e^{-a(T-D)}) + (e^{-a(T-D)} - e^{-aT})\right]}{z(z - e^{-aT})} \quad (a \equiv \frac{1}{\tau})$$

If we define new parameters A_1 and A_2 as

$$(4.25) \qquad A_1 = 1 - e^{-a(T-D)} \quad ; \quad A_2 = e^{-a(T-D)} - e^{-aT}$$

then (4.24) can be written as

$$(4.26) \qquad G_{so}(z) = \frac{K(A_1 z + A_2)}{z(z - e^{-aT})}$$

Consequently, since the closed loop transfer function from sampled input to continuous output is given by

$$\frac{C(j\omega)}{R^*(j\omega)} = \frac{G_{so}(j\omega)}{1 + G^*_{so}(j\omega)} \tag{4.27}$$

we have from (4.18) and (4.20)

$$S_{cc}(\omega) = \left| \frac{G_{so}(j\omega)}{1 + G^*_{so}(j\omega)} \right|^2 S_{rr}{}^*(\omega) \tag{4.28}$$

From (4.26) we obtain

$$\frac{1}{1 + G_{so}(z)} = \frac{z(z - e^{-aT})}{K(A_1 z + A_2) + z(z-e^{-aT})} \tag{4.29}$$

which can be written as

$$\frac{1}{1 + G_{so}(z)} = \frac{z(z - e^{-aT})}{z^2 + A_3 z + A_4} \tag{4.30}$$

where

$$A_3 \triangleq KA_1 - e^{-aT} \; ; \quad A_4 \triangleq KA_2 \tag{4.31}$$

Substituting (4.30) and (4.18) we obtain the spectrum of the sampled error:

$$S_{ee}(z) = \frac{1 + e^{-2aT} - e^{-aT}(z+z^{-1})}{(1+A_3^2+A_4^2)+(A_3+A_3A_4)(z+z^{-1})+A_4(z^2+z^{-2})} S_{rr}(z) \tag{4.32}$$

Letting $z = e^{j\omega T}$ (4.32) becomes:

$$(4.33)\quad S^{*}_{ee}(\omega) = \frac{(1+e^{-2aT} - 2e^{-aT}\cos \omega T)(N_o\, \omega_B \sinh \omega_B T)}{2T\left[(1+A_3^2+A_4^2)+2(A_3+A_3A_4)\cos \omega T+2A_4\cos 2\omega T\right](\cosh \omega_B T-\cos \omega T)}$$

Now, since from (4.19) the continuous transfer function is

$$(4.34)\quad G_{so}(j\omega) = \left(\frac{1-e^{-jT\omega}}{j\omega}\right)\left(\frac{Ke^{-jD\omega}}{1+j\tau\omega}\right)$$

we have

$$(4.35)\quad \left|G_{so}(j\omega)\right|^2 = \frac{2K^2(1 - \cos \omega T)}{\omega^2(1 + \tau^2\omega^2)}$$

Substituting the expressions from (4.17), (4.33) and (4.35) into (4.28) we obtain the output spectral density as:

$$(4.36)\quad S_{cc}(\omega) =$$

$$= \frac{(\omega_B K^2 \sinh \omega_B T)(1 + e^{-2aT} - 2e^{-aT}\cos \omega T)(1 - \cos \omega T)}{T\omega^2(1+\tau^2\omega^2)\left[(1+A_3^2+A_4^2)+2(A_3+A_3A_4)\cos \omega T+2A_4 \cos 2\omega T\right](\cosh \omega_B T-\cos \omega T)}$$

4.3.2 Error power spectrum of zero-order hold model. From Appendix 2, the expression for the power spectral density of the continuous error e(t) is given by

$$(4.37)\quad S_{ee}(\omega) = S_{rr}(\omega) - \frac{2}{T} S_{rr}(\omega)\, R_e\left[\frac{G_{so}(j\omega)}{1 + G^{*}_{so}(j\omega)}\right] + S^{*}_{rr}(\omega)\left|\frac{G_{so}(j\omega)}{1 + G^{*}_{so}(j\omega)}\right|^2$$

where, as before, $S_{rr}(j\omega)$ is the power spectral density (PSD) of the continuous input, $S_{rr}^{*}(j\omega)$ is the sampled PSD and $G_{so}(j\omega)$ has been defined above in (4.19). The third term in (4.37) will be recognized from (4.28) as the PSD of the output signal c(t). Consequently, the only unknown in (4.37) is the term

$$S_{rc}(\omega) = \frac{2}{T} S_{rr} \mathrm{Re}\left[\frac{G_{so}(j\omega)}{1 + G_{so}^{*}(j\omega)}\right] \tag{4.38}$$

which arises from the cross-correlation function between r(t) and c(t).

The open loop transfer function $G_{so}(j\omega)$ is given in equation (4.34), and the corresponding pulse transfer function is

$$G_{so}(z) = \frac{K(A_1 z + A_2)}{z(z - e^{-aT})} \tag{4.39}$$

where the coefficients A_1 and A_2 have been defined in (4.25). Therefore, one obtains from (4.30) and (4.31) that

$$\frac{1}{1 + G_{so}^{*}(j\omega)} = \left.\frac{z(z - e^{-aT})}{z^2 + A_3 z + A_4}\right]_{z=e^{j\omega T}} \tag{4.40}$$

As before, the A_i represent functions of K, T, D, and τ. The complete closed loop frequency function is then given by

$$\frac{G_{so}(j\omega)}{1+G_{so}^{*}(j\omega)} = \frac{K(1-e^{-j\omega T})(e^{-j\omega D_{so}})(e^{j\omega T})(e^{j\omega T}-e^{-aT})}{(j\omega)(1+j\omega\tau)(e^{2j\omega T}+A_3 e^{j\omega T}+A_4)} = \frac{C}{R^{*}}(j\omega) \tag{4.41}$$

Multiplying by the complex conjugate of the denominator we obtain

$$(4.42)\quad \frac{C}{R^*}(j\omega) =$$

$$= \frac{-jK(1-e^{-j\omega T})(e^{-j\omega D})(e^{j\omega T}-e^{-aT})(e^{-j2\omega T}+A_3e^{-j\omega T}+A_4)(1-j\tau\omega)}{\omega(1+\tau^2\omega^2)(A_{10}+A_{11}\cos\omega T+A_{12}\cos 2\omega T)}\,(e^{j\omega T})$$

where

$$A_{10} \triangleq 1 + A_3^2 + A_4^2$$

$$A_{11} \triangleq 2(A_3 + A_3 A_4)$$

$$A_{12} \triangleq 2A_4$$

After multiplying out and collecting terms the numerator can be written as

$$(4.43)\quad \frac{C}{R^*}(j\omega) = -\frac{K\left[(\omega\tau\cos\omega D+\sin\omega D)+j(\cos\omega D-\omega\tau\sin\omega D\right]F_o(A,\omega)}{\omega(1+\tau^2\omega^2)(A_{10}+A_{11}\cos\omega T+A_{12}\cos 2\omega T)}$$

where $F_o(A,\omega)$ is given by

$$(4.44)\quad F_o(A,\omega) = A_5 + A_4e^{2j\omega T}+A_6e^{j\omega T}+A_7e^{-j\omega T}+A_8e^{-2j\omega T}$$

The coefficients in (4.44) are defined by:

$$(4.45)\qquad \begin{aligned} A_5 &\triangleq 1 - (1 + e^{-aT})A_3 + e^{-aT}A_4 \\ A_6 &\triangleq A_3 - (1 + e^{-aT})A_4 \\ A_7 &\triangleq -(1 - e^{-aT}) + e^{-aT}A_3 \\ A_8 &\triangleq e^{-aT} \end{aligned}$$

Expression (4.44), separated into real and imaginary parts, becomes

$$(4.46)\qquad F_o(A,\omega) = \left[A_5 + (A_6 + A_7)\cos \omega T + (A_4 + A_8)\cos 2\omega T\right] + j\left[(A_6 - A_7)\sin \omega T + (A_4 - A_8)\sin 2\omega T\right] \triangleq F_{OR}(A,\omega) + j\, F_{OI}(A,\omega)$$

With the aid of (4.46) and (4.43), the real part of the desired closed-loop frequency function becomes

$$(4.47)\qquad \mathrm{Re}\ \frac{C}{R^*}(j\omega) = \frac{-K\left\{F_{OR}(A,\omega)\left[\omega\tau \cos \omega D + \sin \omega D\right] - F_{OI}(A,\omega)\left[\cos \omega D - \omega\tau \sin \omega D\right]\right\}}{\omega(1+\tau^2\omega^2)\,(A_{10} + A_{11}\cos \omega T + A_{12}\cos 2\omega T)}$$

and the power spectral density of the continuous error for the zero order hold model can be obtained using (4.37) as

$$(4.48)\qquad S_{ee}(\omega) = S_{rr}(\omega) + S_{cc}(\omega) - \frac{2}{T}\ \mathrm{Re}\ \frac{G_{so}(j\omega)}{1 + G^*_{so}(j\omega)}$$

Clearly, expressions such as (4.47) are too complex for much manual calculation. They can, however, be programmed for digital computation rather easily, and some results are presented later in this chapter.

4.4 Analysis of the First-Order Hold Model

The zero-order-hold has been presented in the preceding section largely for completeness and because it illustrates the basic ideas of the method. From the considerations of Chapters 2 and 3 it is clear that velocity extrapolation is a more reasonable

assumption. In this section the expressions for the output and error PSD for the first order hold model will be developed in an analogous manner to the preceding sections.

The model considered is that shown previously in Figure 4.2, but now the hold circuit is described by the transfer function

$$H_1(s) = \left(\frac{1 + Ts}{T}\right)\left(\frac{1 + e^{-Ts}}{s}\right)^2 \tag{4.49}$$

The hold circuit is in series with the continuous model of Section 4.2, where only the time delay D_1 may require adjustment to take into account the phase shift introduced by the hold circuit.

4.4.1 Output PSD of first order hold model. As previously, we begin with the open-loop transfer function

$$G_{s1}(s) = \left(\frac{1 + Ts}{T}\right)\left(\frac{1 - e^{-Ts}}{s}\right)^2\left(\frac{K_e^{-D_1 s}}{1 + \tau s}\right) \tag{4.50}$$

The corresponding z-transform is given by

$$G_1(z) = \frac{K}{T}(1 - z^{-1})^2 \, Z\left\{\frac{(1 + Ts)\, e^{-D_1 s}}{s^2(1 + \tau s)}\right\} \tag{4.51}$$

which can be obtained from a table of "modified" transforms with $m = 1 - \frac{D_1}{T}$. The result, after some manipulation, reduces to

$$G_1(z) = \frac{K(Q_1 z^2 + Q_2 z + Q_3)}{Tz^2(z - e^{-aT})} \tag{4.52}$$

where

$$Q_1 = (2T - D_1 - \tau) + (\tau - T)e^{-a(T-D_1)}$$

(4.53)
$$Q_2 = T - (1 + e^{-aT})(2T - D_1 - \tau) - 2(\tau - T)e^{-a(T-D)}$$

$$Q_3 = - Te^{-aT} + (2T - D_1 - \tau)e^{-aT} + (\tau-T)e^{-a(T-D)}$$

The pulse transfer function relating E(z) to R(z) then becomes

(4.54)
$$\frac{E(z)}{R(z)} = \frac{1}{1 + G_{s1}(z)} = \frac{z^2(z - e^{-aT})}{z^3 + P_1z^2 + P_2z + P_3}$$

where the coefficients are defined in terms of the Q_i of (4.53) as:

$$P_1 = (KQ_1 - Te^{-aT})/T$$

(4.55)
$$P_2 = KQ_2/T$$

$$P_3 = KQ_3/T$$

The sampled error spectrum then becomes

(4.56)
$$S_{ee}(z) = \left[\frac{1}{1 + G_{s1}(z)}\right]\left[\frac{1}{1 + G_{s1}(z^{-1})}\right] S_{rr}(z) =$$

$$= \frac{(1 + e^{-2aT}) - e^{-aT}(z + z^{-1})}{(1+P_1^2 + P_2^2 + P_3^2)+(P_1+P_1P_2+P_2P_3)(z+z^{-1})+(P_2+P_1P_3)(z^2+z^{-2})+P_3(z^3+z^{-3})}$$

Finally, since

(4.57)
$$S_{cc}(\omega) = G_{s1}(j\omega)^2 S_{ee}^*(j\omega)$$

we substitute (4.56) and (4.50) in (4.57) to obtain

(4.58) $S_{cc}(\omega) =$

$$= \frac{\left\{4K^2(1+T^2\omega^2)(1+e^{-2aT}-2e^{-aT}\cos\omega T)(1-\cos\omega T)^2\right\}S_{rr}^*(\omega)}{T^2\omega^2(1+\tau^2\omega^2)\left[(1+P_1^2+P_2^2+P_3^2)+2(P_1+P_1P_2+P_2P_3)\cos\omega T+2(P_2+P_1P_3)\cos 2\omega T+2P_3\cos 3\omega T\right]}$$

where, as before, for the filtered noise input, $S_{rr}^*(\omega)$ is given by

$$(4.59) \qquad S_{rr}^*(\omega) = \frac{N_o\,\omega_B \sinh \omega_B T}{2T(\cosh \omega_B T - \cos \omega T)}$$

4.4.2 Error spectrum for first-order hold model. In an analogous manner to the derivation of the PSD expression for the zero-order-hold system in Section 4.3.2, we begin with the open-loop transfer function $G_{sl}(j\omega)$ gain in (4.50) and the pulse transfer function $E(z)/R(z)$ from (4.54). The closed-loop relation $C(j\omega)/R^*(j\omega)$ then becomes

$$(4.60) \qquad \frac{C(j\omega)}{R^*(j\omega)} =$$

$$= \frac{G_{sl}(j\omega)}{1+G_{sl}^*(j\omega)} = \frac{K(1+j\omega T)(1-e^{-j\omega T})^2(e^{-j\omega D_1})(e^{j2\omega T})(e^{j\omega T}-e^{-aT})}{T(j\omega)^2(1+j\omega\tau)(e^{3j\omega T}+P_1e^{2j\omega T}+P_2e^{j\omega T}+P_3)}$$

where the coefficients P_i have been defined in (4.55). The real part of (4.60) is obtained by multiplying by the complex conjugate of the denominator and collecting terms to obtain, after some algebra:

$$(4.61)\quad \mathrm{Re}\,\frac{C(j\omega)}{R^{*}(j\omega)} =$$

$$= \frac{-K\left\{F_{1R}(P,\omega)\left[(1+\tau T\omega^2)\cos\omega D+\omega(T-\tau)\sin\omega D\right]-\left[\omega(T-\tau)\cos\omega D-(1+\tau T\omega^2)\sin\omega D\right]F_{1I}(P,\omega)\right\}}{T\omega^2(1+\tau^2\omega^2)(P_{10}+P_{11}\cos\omega T+P_{12}\cos 2\omega T+P_{13}\cos 3\omega T)}$$

where the various coefficients are defined as follows:

$$(4.62)\begin{cases} F_{1R}(P,\omega) = P_4+(P_5+P_7)\cos\omega T+(P_6+P_8)\cos 2\omega T+(P_3+P_9)\cos 3\omega T \\ F_{1I}(P,\omega) =(P_5-P_7)\sin\omega T+(P_6-P_8)\sin 2\omega T+(P_3-P_9)\sin 3\omega T \end{cases}$$

and

$$(4.63)\quad \begin{aligned} P_4 &\triangleq 1-(2+e^{-aT})P_1+(1+2e^{-aT})P_2-e^{-aT}P_3 \\ P_5 &\triangleq P_1-(2+e^{-aT})P_2+(1+2e^{-aT})P_3 \\ P_6 &\triangleq P_2-(2+e^{-aT})P_3 \\ P_7 &\triangleq -(2+e^{-aT})+(1+2e^{-aT})P_1-e^{-aT}P_2 \\ P_8 &\triangleq (1+2e^{-aT})-e^{-aT}P_1 \\ P_9 &\triangleq -e^{-aT} \\ P_{10} &\triangleq 1+P_1^2+P_2^2+P_3^2 \\ P_{11} &\triangleq 2(P_1+P_1P_2+P_2P_3) \\ P_{12} &\triangleq 2(P_2+P_1P_3) \\ P_{13} &\triangleq 2P_3 \end{aligned}$$

Therefore, the power spectral density of the error signal in the first-order hold model is determined by substituting (4.61) into the relation

$$(4.64) \qquad S_{eel}(\omega) = S_{rr}(\omega) + S_{ccl}(\omega) - \frac{2}{T} S_{rr} \; \mathrm{Re} \frac{G_{sl}(j\omega)}{1 + G_{sl}^{*}(j\omega)}.$$

4.5 Analysis of the "Modified" Hold Model

The modified first order hold (MFOH) representation was introduced in Chapter 3 and compared with conventional zero-order and first-order-hold circuits. In the present section we shall derive expressions for computing the output and error PSD of the MFOH circuit. In a later section of this chapter these spectral densities will be compared with those of other models for specific values of parameters.

Consider the circuit of Figure 4.4(a) which shows the MFOH circuit. In order to compute the output spectrum we redraw the circuit into the form shown in Figure 4.4(b) and (c) where $Y_F(j\omega)$ denotes the frequency function of the approximate differentiator. The error spectrum can be obtained from Figure 4.4(b) by noting that

$$(4.65) \qquad S_{e'e'}(\omega) = S_{r'r'}(\omega) + S_{c'c'}(\omega) - \frac{2}{T} S_{r'r'}(\omega) \, \mathrm{Re} \, \frac{G_{lm}(j\omega)}{1+G_{lm}Y_F^{*}(j\omega)}$$

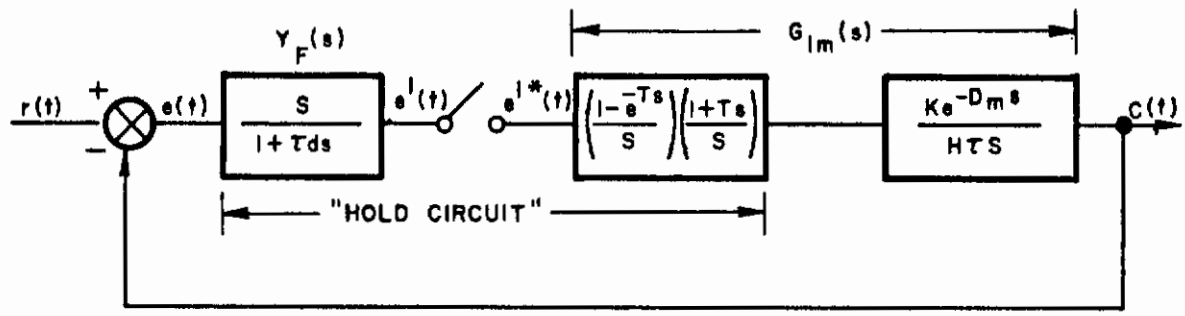
where

$$(4.66) \qquad \begin{aligned} S_{r'r'}(\omega) &= \left|Y_F(j\omega)\right|^2 S_{rr}(\omega) \\ S_{c'c'}(\omega) &= \left|Y_F(j\omega)\right|^2 S_{cc}(\omega) \end{aligned}$$

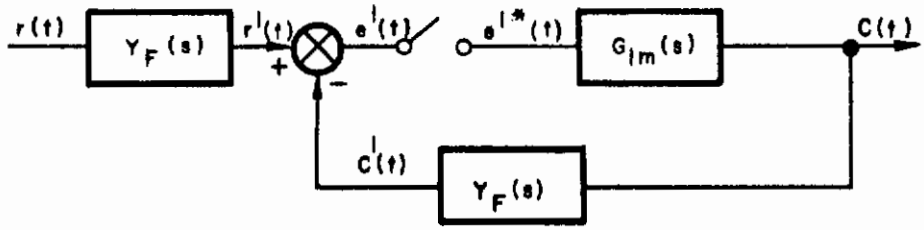
and

$$(4.67) \qquad E'(j\omega) = E(j\omega)\, Y_F(j\omega)$$

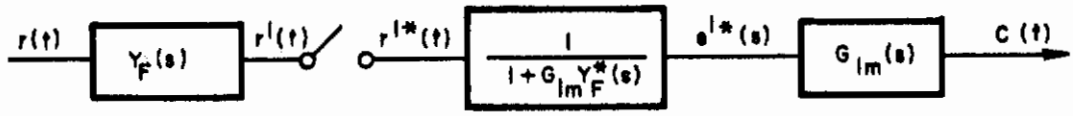
and consequently the desired continuous error PSD is given by:



(a)
ORIGINAL MFOH SYSTEM



(b)
EQUIVALENT SYSTEM WITH "ERROR" SAMPLING



(c)
EQUIVALENT SERIES FORM

Figure 4.4 Equivalent Forms of MFOH Model

$$(4.68)\quad S_{ee}(\omega) = S_{rr}(\omega) + S_{cc}(\omega) - \frac{2}{T} S_{rr}(\omega)\ \mathrm{Re}\ \frac{G_{lm}(j\omega)}{1 + G_{lm}\, Y_F^{*}(j\omega)}$$

4.5.1 Output PSD of MFOH operator model. We begin by deriving an expression for the output power spectral density $S_{cc}(\omega)$. As before, the input PSD is given by

$$(4.69)\qquad S_{rr}(\omega) = \frac{N_o\, \omega_B^2}{\omega^2 + \omega_B^2}$$

which represents "white noise" filtered by a first-order lag. The "differentiated" spectrum $S_{r'r'}(\omega)$ is then given by

$$(4.70)\quad S_{r'r'}(\omega) = \left|Y_F(j\omega)\right|^2 S_{rr}(\omega) = \frac{(N_o b^2\, \omega_B^2)\, \omega^2}{(\omega^2 + \omega_B^2)(\omega^2 + b^2)};\ (b \equiv \frac{1}{\tau_d})$$

The equivalent sampled spectrum is obtained by first expanding (4.70) in partial fractions in order to obtain functions which appear in transform tables. In terms of the complex frequency s the result is:

$$(4.71)\qquad S_{r'r'}(s) = \frac{-N_o\, b^2\, \omega_B^2}{b^2 - \omega_B^2}\left[\frac{\omega_B^2}{\omega_B^2 - s^2} - \frac{b^2}{b^2 - s^2}\right]$$

and corresponding sampled spectrum is

$$(4.72)\quad S_{r'r'}{}^{*}(\omega) = \frac{C_1}{2T}\left[\frac{b \sinh bT}{\cosh bT - \cos \omega T} - \frac{\omega_B \sinh \omega_B T}{\cosh \omega_B T - \cos \omega T}\right]$$

where

$$C_1 = \frac{N_o\, \omega_B^2\, b^2}{b^2 - \omega_B^2}$$

Equation (4.72) can be used to compute the output PSD from the relation

$$S_{cc}(\omega) = \left| \frac{G_{1M}(j\omega)}{1 + G_{1M} Y_F^*(j\omega)} \right|^2 S^*_{r'r'}(j\omega) \tag{4.73}$$

The transfer functions $G_{1M}(j\omega)$ and $Y_F(j\omega)$ are identified in the block diagrams of Figure 4.4, i.e.

$$G_{1M}(j\omega) = \frac{Ke^{-D_m s}(1 + Ts)(1 - e^{-Ts})}{s^2(1 + \tau s)} \tag{4.74}$$

$$Y_F(s) = \frac{bs}{s + b} \tag{4.75}$$

(For purposes of computing the loop transfer function $G_{1M}Y_F(s)$ we shall assume $Y_F(s) \cong s$, since $G_{1M}(s)$ is sufficiently low pass to make the errors of approximations negligible). Consequently we have

$$G_{1M}Y_F(s) \cong \frac{Kae^{-Ds}(1 - e^{-Ts})(1 + Ts)}{s(s + a)} \tag{4.76}$$

from which the modified Z transform can be obtained and written in the form

$$G_{1M}Y_F(z) = \frac{K(B_1 z + B_2)}{z(z - e^{-aT})} \tag{4.77}$$

where the coefficients are given by

$$\begin{aligned} B_1 &= 1 - (1 - Ta)e^{-a(T-D)} \\ B_2 &= (1 - Ta)e^{-a(T-D)} - e^{-aT} \end{aligned} \tag{4.78}$$

The desired closed-loop transfer function for substitution into (4.37) can then be written in the form

$$(4.79)\quad \frac{G_{1M}(s)}{1+G_{1M}Y_F^*(s)} = \left[\frac{Ke^{-Ds}(1+Ts)(1-e^{-Ts})}{s^2(1+\tau s)}\right]\left[\frac{z(z-e^{-aT})}{z^2 + B_3 z + B_4}\right]_{z=e^{sT}}$$

where the coefficients are defined by

$$(4.80)\qquad B_3 = K B_1 - e^{-aT};\quad B_4 = K B_2$$

Substituting in (4.73) we obtain the output spectrum:

$$(4.81)\quad S_{cc}(\omega) = \frac{\left[2K^2(1+T^2\omega^2)(1-\cos\omega T)(1+e^{-aT}-2e^{-aT}\cos\omega T)\right]S^*_{r',r'}(\omega)}{\omega^4(1+\tau^2\omega^2)\left[(1+B_3^2+B_4^2)+2(B_3+B_3B_4)\cos\omega T+2B_4\cos\omega T\right]}$$

4.5.2 Error PSD of MFOH operator model. To obtain the error spectrum we consider the real part of (4.79) for $s = j\omega$, i.e. of the function

$$(4.79a)\quad \frac{G_{1M}(j\omega)}{1 + G_{1M}\,Y_F^*(j\omega)} =$$

$$= \frac{Ke^{-jD\omega}(1+jT\omega)(1-e^{-jT\omega})(e^{jT\omega})(e^{jT\omega} - e^{-aT})}{(j\omega)^2(1 + j\tau\omega)(e^{2j\omega T} + B_3e^{j\omega T} + B_4)} \triangleq X_{1M}(j\omega)$$

Multiplying the numerator and denominator of (4.81) by the complex conjugate of the denominator we obtain

$$(4.82)\quad X_{1M(j\omega)} =$$

$$= -\frac{K(1+jT\omega)(1-e^{-jT\omega})(e^{jD\omega})(e^{2jT\omega}-e^{-aT}e^{jT\omega})(1-j\tau\omega)(e^{-j2\omega T}+B_3e^{-j\omega T}+B_4)}{(1+\tau^2\omega^2)\left[(1+B_3^2+B_4^2)+2(B_3+B_3B_4)\cos\omega T+2B_4\cos 2\omega T\right]}$$

the real part of which can be written as

$$(4.83)\quad R_eX_{1M}(j\omega) =$$

$$= \frac{-K\left\{F_{MR}\left[(1+\tau T\omega^2)\cos\omega D+\omega(T-\tau)\sin\omega D\right]-F_{MI}\left[\omega(T-\tau)\cos\omega D-(1+\tau T\omega^2)\sin\omega D\right]\right\}}{\omega^2(1+\tau^2\omega^2)\left[(1+B_3^2+B_4^2)+2(B_3+B_3B_4)\cos\omega T+2B_4\cos 2\omega T\right]}$$

where the coefficients are defined as follows:

$$F_{MR}(B,\omega) \triangleq B_5 + (B_6+B_8)\cos\omega T + (B_4+B_9)\cos 2\omega T$$

$$F_{MI}(B,\omega) \triangleq (B_6-B_8)\sin\omega T + (B_4-B_9)\sin 2\omega T$$

$$(4.84)$$

$$B_5 \triangleq (1-B_3) - e^{-aT}(B_3-B_4) \quad ; \quad B_8 = -\left[1+e^{-aT}(1-B_3)\right]$$

$$B_6 \triangleq (B_3-B_4) - e^{-aT}B_4 \quad ; \quad B_9 = e^{-aT}$$

Thus, the error spectral density for the modified first order hold is given by substituting (4.83) into:

$$(4.85)\quad S_{ee_M}(\omega) = S_{rr}(\omega) + S_{cc_M}(\omega) - \frac{2}{T}S_{rr}(\omega)\ \mathrm{Re}\left[\frac{G_{1M}Y_F(j\omega)}{1+G_{1M}Y_F^*(j\omega)}\right]$$

4.6 Numerical Evaluation and Comparison of Spectra

In order to evaluate the three sets of expressions for the power spectral density of the proposed operator models it will be necessary to select α numerical values for the various parameters, i.e. to select a particular situation in which a quasi-linear model is appropriate. In addition to the parameters which characterize the continuous model (i.e. K, D_c and τ) it is also necessary to select a value of sampling period T. With these values fixed, it is possible to evaluate and plot the power spectral density of the output and error of the various models. In order to select appropriate values we shall use an experiment from the work of J. Elkind (1957), which was described in Chapter 2.

4.6.1 Elkind's experiment F1. The forcing function used by Elkind consisted of a sum of up to 144 sinewaves separated by equal frequency increments but of random phase relation to each other. In his experiment F1 the frequency separation was 0.02 cps (therefore a range of 0.02 to 2.88 cps) and the resulting signal was filtered by a lag filter with a break frequency of 0.24 cps. The resulting spectrum can be considered an approximation to the continuous spectrum used in the derivation above. The resulting data were fitted with the quasi-linear transfer function:

$$G_H(j\omega) = \frac{3.35\, e^{-0.139 j\omega}}{1 + j\omega/1.13} \tag{4.86}$$

Since the experimental data from which this transfer relationship was derived are available in Elkind's dissertation, it seems logical to use them for a preliminary investigation of the sampled model. In particular, we shall be interested in comparing the measured values of power spectral density $S_{cc}(\omega)$ and $S_{ee}(\omega)$ with those computed from the continuous model (using the data of eq. 4.86) and those computed using the sampled-data models.

Two limitations of this comparison should be noted immediately.

(1) The quasi-linear continuous model does not account for all of the operator's output, and an uncorrelated noise with spectrum $S_{nn}(\omega)$ must be added for a complete representation. However, our intention is to see whether a portion of the "remnant" is accounted for with the new models.

(2) Elkind's data for experiment F1 are based on the average of 4 runs conducted by a single well-trained subject in a particular experimental situation. Neither the subject nor the particular tracking equipment may be "typical" or "average".

In spite of these limitations, it is extremely useful to use data available in the literature as a preliminary feasibility test.

4.6.2 Selection of the sampling interval T for the sampled model. From the continuous model of (4.86) we have three of the parameters of the models, namely:

$$K = 3.35$$
$$a = 1/\tau_c = 1.13 \text{ rad/sec.}$$
$$D_c = 0.139 \text{ sec.}$$

Furthermore, the shaping filter break frequency used in Elkind's experiment was 0.24 cps, and thus:

$$\omega_B = 0.24\ (2\pi) = 1.5 \text{ rad/sec.}$$

These four parameters are sufficient for computing the error and output PSD of the continuous model. For the sampled-data models it is necessary to determine a value of sampling period T. When T is fixed, the delay D_s will have to be adjusted to take into account the phase shift introduced by the hold circuit.

In order to set the value of $T = 1/f_s$ for the model let us examine the experimentally measured values of $S_{cc}(\omega)$ and $S_{ee}(\omega)$ from Elkind's report for this particular experiment. The two curves are given in Figure 4.5 (both curves are normalized by having the ordinates divided by the total input power (since $P_{in}=\frac{1}{2\pi}\int_0^\infty S_{rr}(\omega)d\omega$). Both of the curves of Figure 4.5 exhibit significant peaks in the vicinity of f = 1.5 cps. If we now recall that the first-order hold circuits discussed in Section 3.3 also exhibit characteristic peaks near $f_s/2$, we can make the following assumption:

> It will be assumed that peaks (if they exist) in power spectra of tracking data, correspond to one-half the sampling frequency of the equivalent sampled-data model.

Clearly, this assumption will require re-examination when additional experimental data are available. Such a re-examination will be made in Chapters 5 and 6, in connection with a discussion of the experimental plan of this study. However, on the basis of the spectral density plots given in Elkind's work, it appears that all experiments with a forcing function where significant energy appears above approximately 1 cps do in fact exhibit peaks in a range of 1 to 2 cps. On the basis of Appendix 1, the psychological evidence for intermittency points to a sampling frequency between 2 and 4 cps; consequently it can be assumed that the location of the peak is a reasonable guide to the model sampling frequency. At this stage of the study, this is merely an assumption which appears to be reasonable. For the experiment at hand, we shall assume T = 1/(2x1.5 cps) = 0.33 seconds.

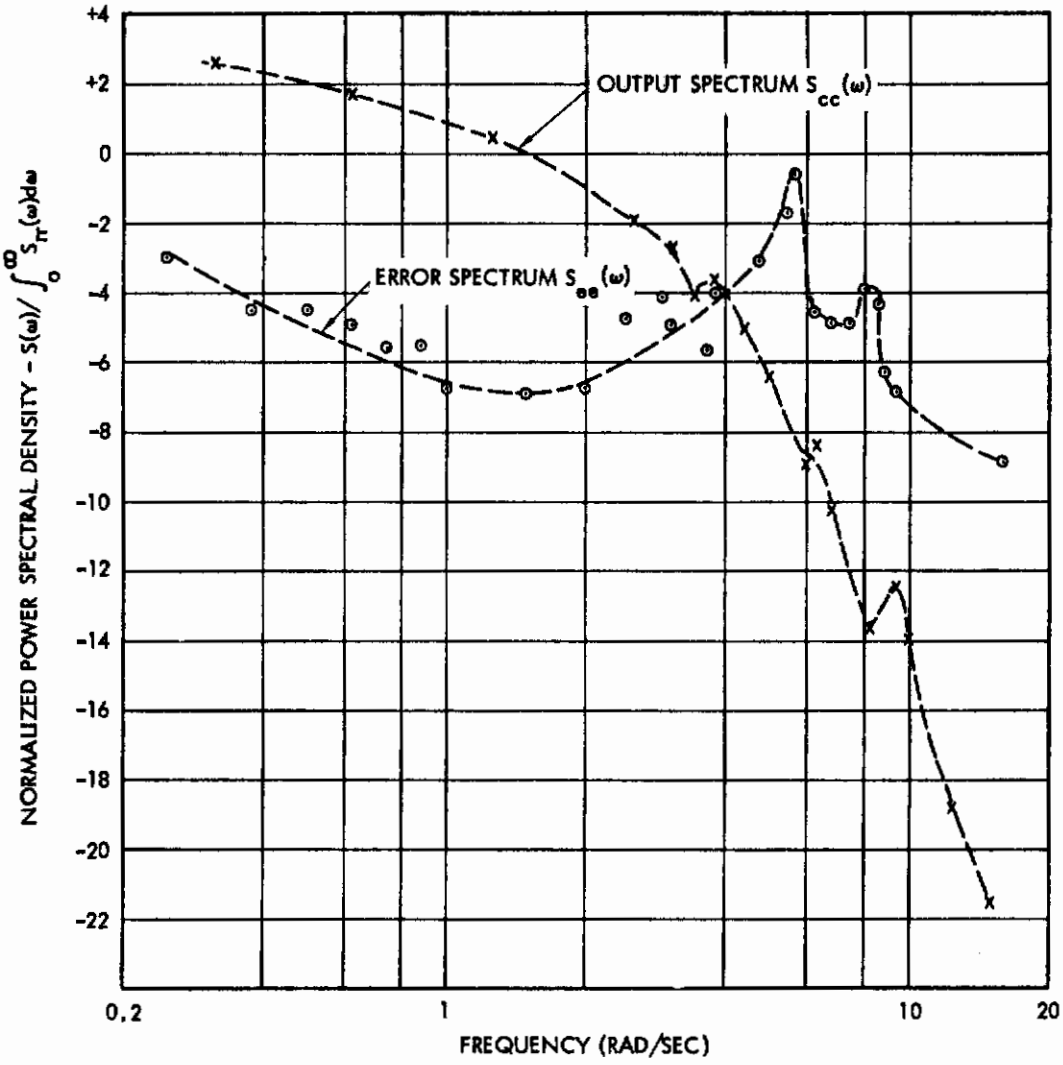


Figure 4.5 Experimental Data from Elkind (1956) (Experiment F1)

4.6.3 Selection of time delay D_s in the sampled model. Once the sampling period T is given, we can turn to the characteristics of the hold circuits given in Chapter 3 to determine values of D_s. Since the hold circuits introduce phase-shift and consequently an "effective time delay", we shall set the model delay such that the total open loop delay equals the continuous delay D_c, i.e.

$$D_s = D_c - D_H \tag{4.87}$$

where

D_c = continuous model delay (from fitted data)

D_H = effective or average delay due to hold circuit in frequency range of interest

D_s = sampled model delay

The significance of the values of D_s obtained in this way (which may be negative or positive) will be discussed in Chapter 6. Here, our objective is simply to obtain a set of data for comparing theoretical spectra.

For the zero-order hold model it can be recalled that the time delay D_{HO} is given by

$$D_{HO} = \frac{1}{2f_s} = \frac{T}{2} \tag{4.88}$$

i.e., the zero-order hold introduces a time delay of one-half of the sampling period. Consequently, we can set

$$D_{so} = D_c - \frac{T}{2} \tag{4.89}$$

For the first-order hold the situation is more complex, since this model does not introduce a linear variation of phase shift with frequency. However, for the range of frequencies we are considering, and from an examination of the phase-shift characteristics of the hold circuits (Figure 3.10) we can assume that

(4.90)
$$D_{H1} \cong 0.5\ T \quad \text{for} \quad 0 < \omega T < 4$$
$$D_{H1} \cong 0.8\ T \quad \text{for} \quad 0 < \omega T < 8$$

and consequently, since we expect T to be of the order of 1/3 sec and a maximum frequency of interest to be about 12 rad/sec, we can use:

(4.91)
$$D_{s1} \cong D_c - 0.5\ T$$

Similarly, the modified hold circuit introduces an effective delay of approximately zero for the range of frequencies of interest. Consequently, we shall take

(4.92)
$$D_{sM} \cong D_c$$

The values of the parameters chosen for all the models for this case are summarized in Table 4.1 below.

Table 4.1

Parameter Values for Demonstration Case

Model	Gain K	Time Constant (sec)	Time Delay (sec)	Sampling Period T (sec)
Continuous	3.35	.886	.14	-
Zero-order hold	3.35	.886	0.0	0.33
First-order hold	3.35	.886	0.0	0.33
Modified first-order	3.35	.886	0.14	0.33

With the sampling period set at 0.33 sec, the effective delay is approximately $\frac{1}{2}$ T or 0.16 sec, rather than 0.14 as in Elkind's continuous model. However, for the purpose of studying the general shapes of the spectral density curves, this is considered adequate, especially since it simplifies the computation.

The output power spectral density curves for the four models are plotted in Figure 4.6 and the error PSD curves in Figure 4.7. Both figures include experimental points from Elkind, not to indicate the "fit" of the various models, but to give a general impression of the relation among the various models and the scatter of a typical set of experimental points. (The points are taken from Elkind's dissertation where they appear as approximately 3"x4" plots which are quite difficult to read. Furthermore, there appears to be some ambiguity in his selection of normalizing factors for the curves).

4.7 Evaluation of the Sampled Models

On the basis of an examination of the curves in Figures 4.6 and 4.7 the following observations can be made:

(1) The ZOH model produces an error spectrum which deviates extremely from both continuous model and experimental data in the middle frequency region.

(2) The MFOH model, while behaving in a reasonable way at high frequencies, exhibits excessive gain at very low frequencies.

(3) The first-order hold model, while appearing quite reasonable in its error spectrum, exhibits somewhat excessive peaking in its output spectrum. In spite of this defect, however, it appears to give the most promise of yielding satisfactory results.

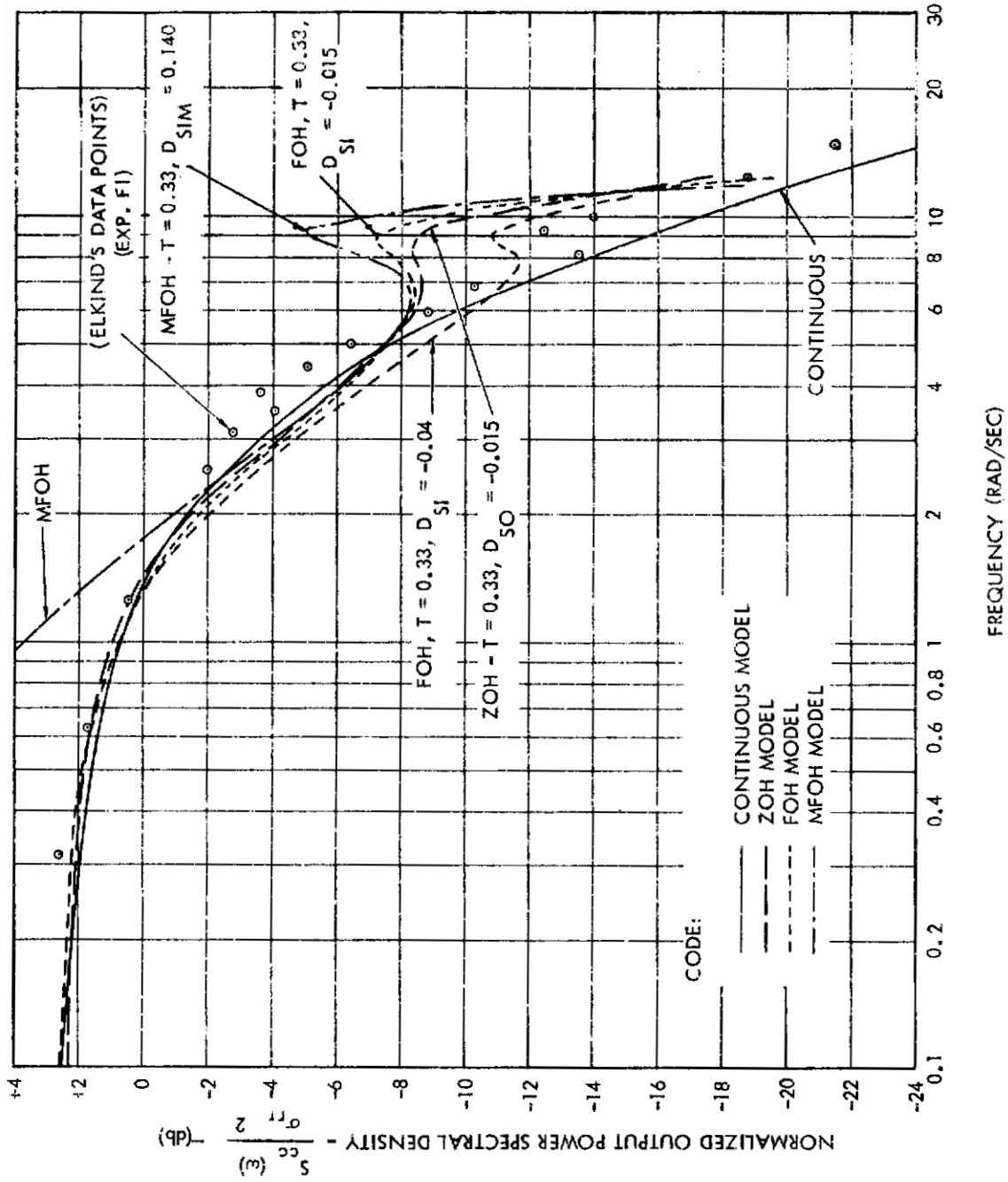


Figure 4.6 Output Spectral Density for Various Human Operator Models

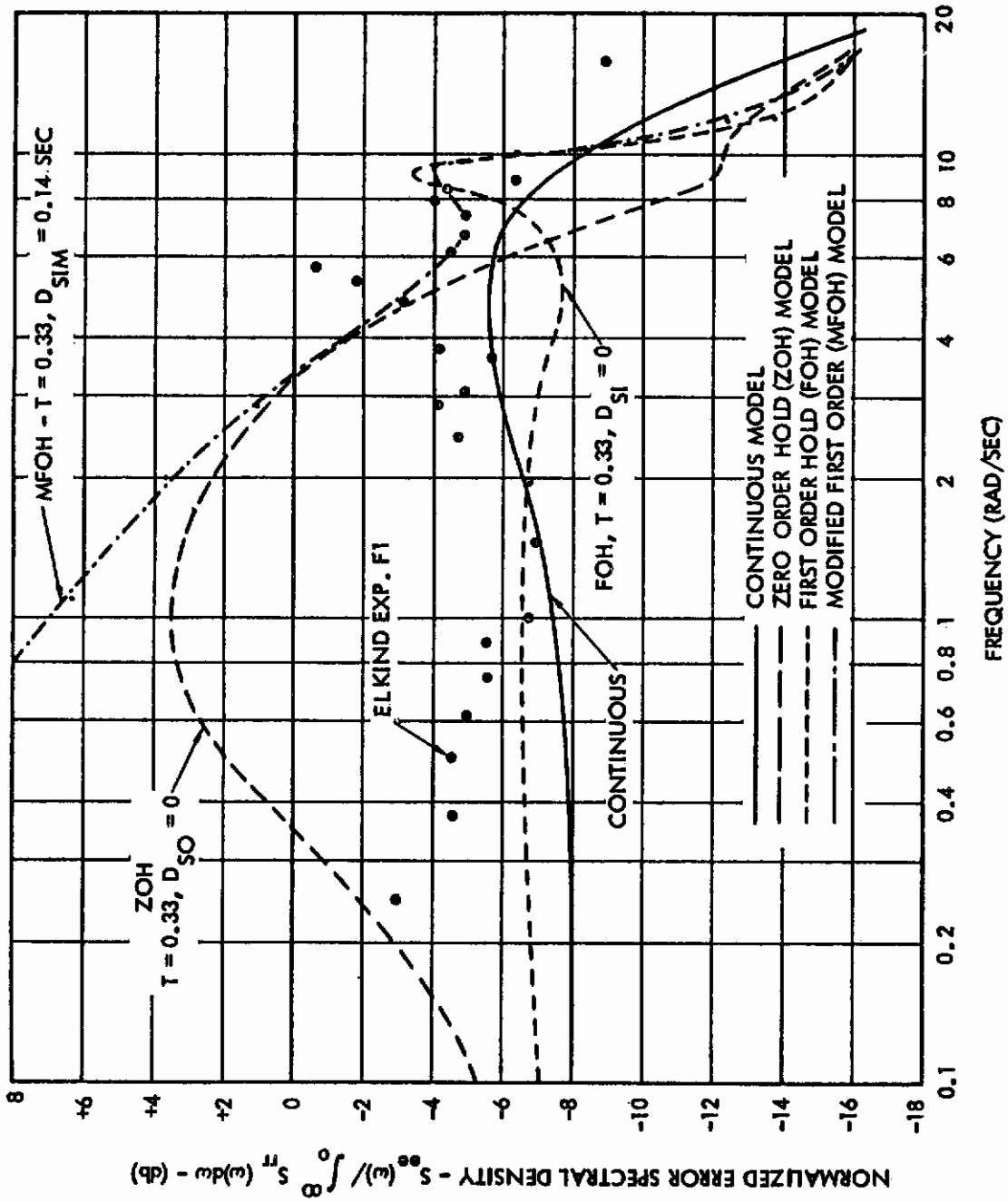


Figure 4.7 Error Spectral Density for Various Human Operator Models

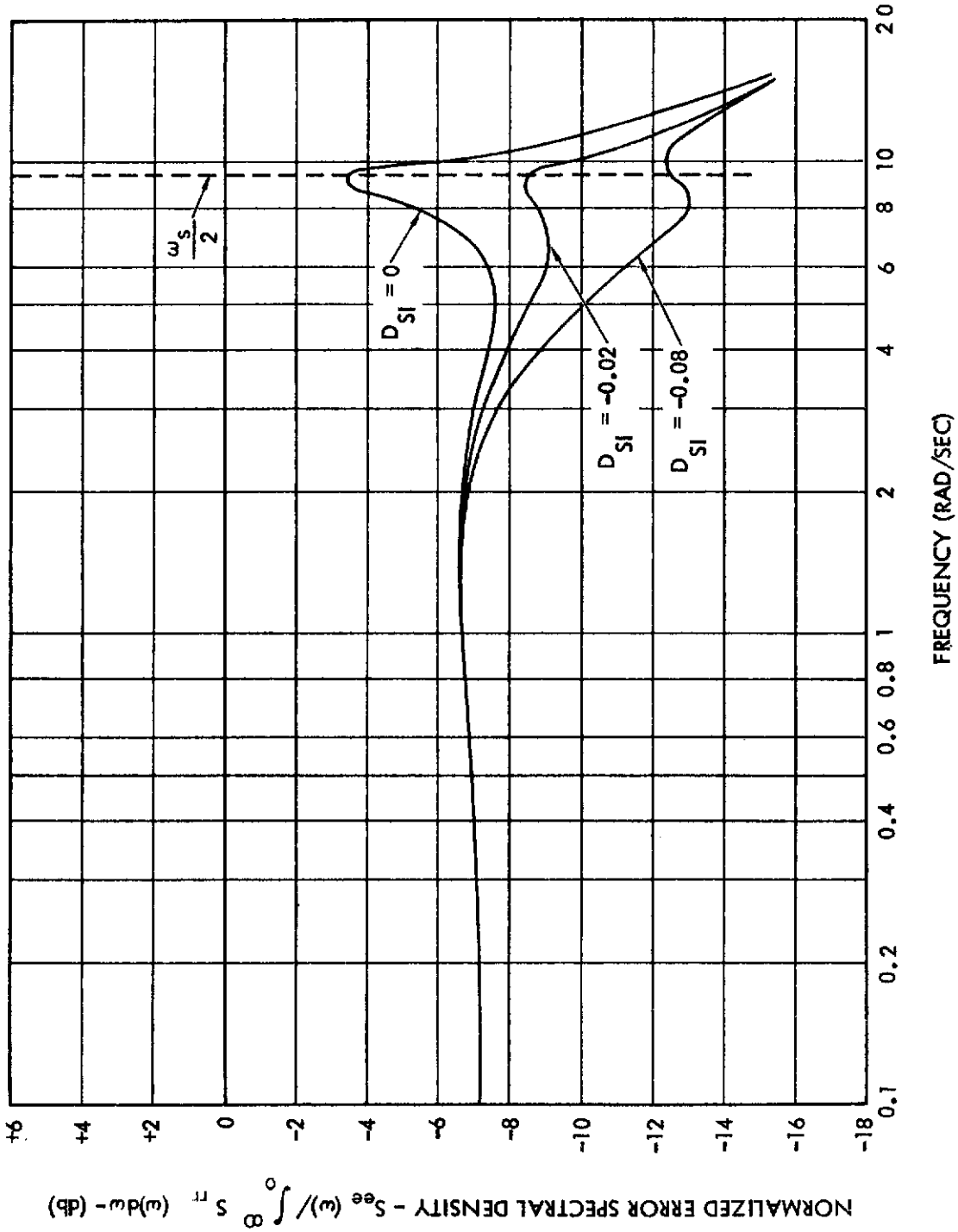


Figure 4.8 Sensitivity of Error Spectrum of FOH Model to Time Delay Variation for T = 0.33 sec

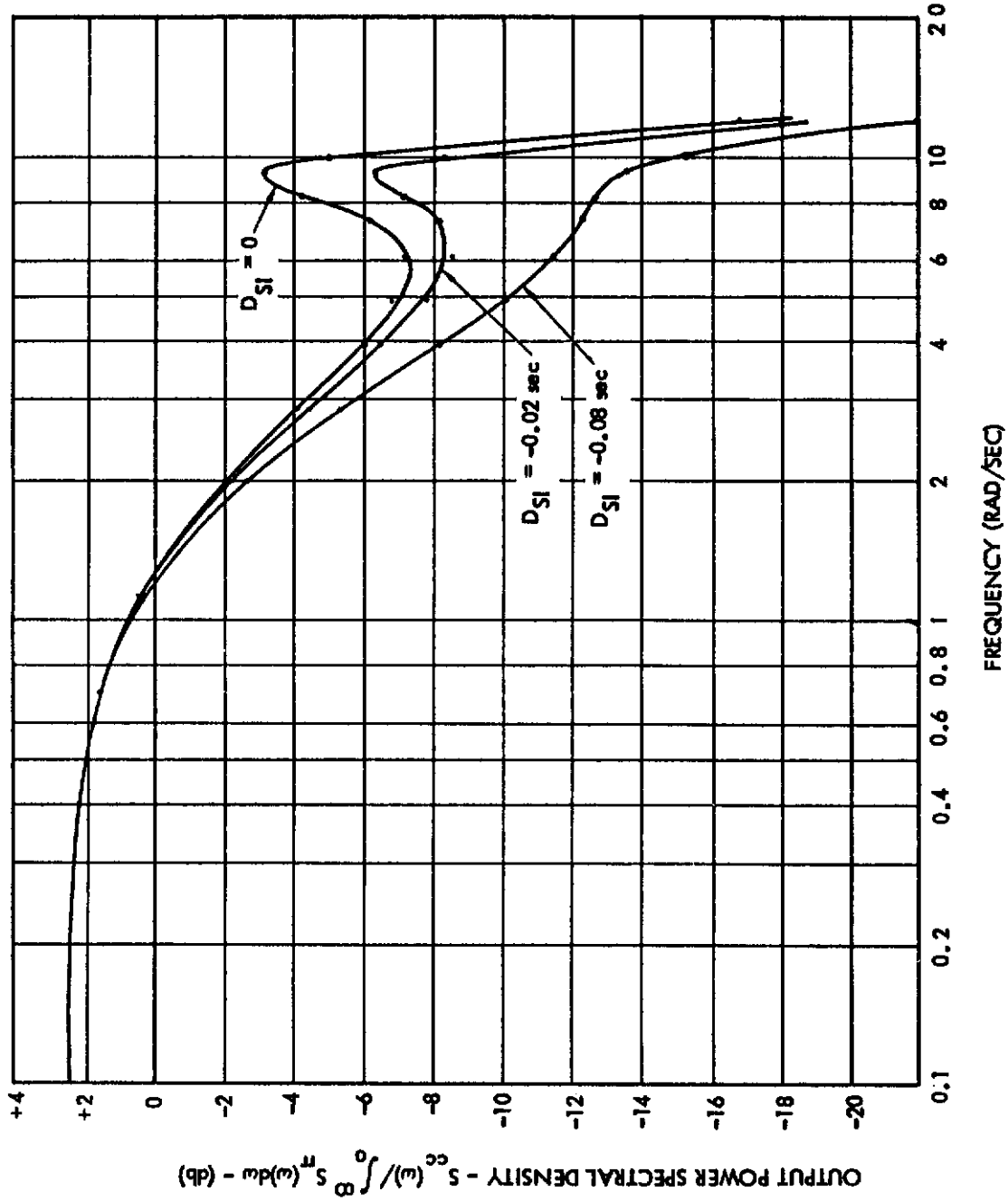


Figure 4.9 Sensitivity of Output Spectrum of FOH Model to Time Delay Variation (T = 0.33 sec)

(4) It is clear that more experimental data are needed to evaluate any proposed sampled models. However, since the zero-order hold does not possess the velocity-extrapolation properties which are desired, in addition to having the behavior illustrated in Figure 4.7, it will be abandoned here. The modified hold circuit likewise will be considered only in a cursory fashion in the remainder of the report. However, since it does offer some interesting possibilities for future work, it will not be abandoned entirely.

In order to examine the significance of the peak in the first-order hold sampled data model, a number of runs were made for small variations in the time delay, D_{s1}. The resulting spectra are shown in Figures 4.8 and 4.9. All three of these curves reveal the characteristic peaks near $\frac{1}{2}$ the sampling frequency. It can be noted that the most "reasonable" of these curves have negative values of delay, i.e., the model incorporates some prediction. The net phase shift of the model, however, is leading only for very low frequencies (where prediction by the human operator is also known to occur) and lagging at higher frequencies.

In summary, the sampled-data model with a first-order hold exhibits a power spectrum which strongly peaks near one-half the sampling frequency, and generally resembles tracking spectra found elsewhere in the literature. However, to evaluate this model more carefully additional experimental data were needed. These are the subject of the next chapter.

Chapter 5

EXPERIMENTAL PROGRAM

In this chapter the purpose and general aspects of the experimental portion of the research program are described. The construction of equipment and the selection of experiments and subjects are discussed. The results are presented and discussed in detail in Chapter 6.

5.1 Purpose of the Experimental Study

In the last chapter several models of the human operator based on sample-and-hold operations were analyzed. The models were proposed on an a priori basis, that is, they represented a theoretical hypothesis based on general features of human tracking and on the properties of sampled-data systems, rather than being attempts to "match" experimental data. In order to compare the behavior of the models with experiments a number of assumptions were required in connection with one set of data from the literature. These assumptions were related to the validity of Elkind's data, to the universality of certain spectral characteristics and to the method of incorporating them into a model. We are therefore left with the following set of questions:

(1) Are the characteristics of the measured error spectra in Elkind's experiment typical features of human tracking or are they an artifact of his particular experiment?

(2) Will such characteristics appear with different operators and with inputs of varying difficulty in a setting different from Elkind's?

(3) Can the assumption that spectral peaks are related to model sampling frequencies be justified on the basis of more data? If so, can a systematic procedure for constructions of the sampled-data model for a particular task be developed?

(4) Are there other implications of the sampling hypothesis which can be tested experimentally?

The experimental program was designed to answer the above questions. Its primary purpose was to get more data and to use this data for developing and justifying a synthesis procedure. The secondary purpose was to perform some qualitative, preliminary experiments concerned with the implications of sampling.

To answer Questions (1) and (2), error spectra of ten operators performing compensatory tracking were recorded using analog computer techniques. The forcing functions were random-appearing, consisting of 10 sinusoids ranging in frequency from 0.16 to 1.6 cps.

To answer Question (3) data for two representative subjects from the previous sample, the tracking data were recorded on magnetic tape and analyzed using a digital computer in order to obtain continuous quasi-linear models and equivalent sampled-data models. The power spectra resulting from these models were then compared with the experimentally recorded data.

Finally, Question (4) was investigated by exposing several operators to sampled input signals. This procedure made it possible to determine whether the location of the "sampling peak" in the operator's error spectral density can be related to an intermittency in the presented data.

In addition to the measurement of power spectral density in the tracking experiments, some time domain recordings of typical tracking responses were made and compared with recordings of the output and error signals of the human operator models discussed previously.

5.2 Analog Measurement of Power Spectral Density

In Chapter 4 we have obtained expressions and computed output and error spectra of various human operator models inputs which consisted of continuous RC-filtered noise. That is, the forcing functions were considered to be sample functions of strictly stationary random processes with a flat continuous spectra ("white" gaussian noise). For the experimental work the input signal consisted of a sum of sine waves. Consequently, in this section the concepts needed for an interpretation of the spectral density of periodic functions are presented briefly and the method of measurement is indicated. A more detailed treatment of this topic is given in Appendix 3.

5.2.1 Power Spectra of Periodic Functions. Consider the function f(t) given by a finite number of sinusoids:

$$f(t) = \sum_{n=1}^{N} C_i \cos(\omega_i t - \theta_i) \tag{5.1}$$

If $N \to \infty$ eq. (5.1) can be considered the Fourier series representation of f(t) in an interval (-T,T), (provided that f(t) satisfies the Dirichlet conditions; see, for example, Pipes (1960)). In Appendix 3 the power spectral density of such a Fourier series representation is indicated. However, we are more concerned with f(t) as a finite sum of sinusoids defined only over an interval (-T, T), i.e. we define a new function $f_T(t)$ such that:

$$(5.2) \qquad \begin{aligned} f_T(t) &= f(t) && |t| \leq T \\ f_T(t) &= 0 && |t| > T \end{aligned}$$

We can then define the power spectral density of $f_T(t)$ as being

$$(5.3) \qquad S_{ff_T}(\omega) = \frac{\left| F_T(\omega, f) \right|^2}{2\pi T}$$

where

$$(5.4) \qquad F_T(\omega, f) = F_T\left[f(t)\right] = \int_{-\infty}^{+\infty} f_T(t) e^{-j\omega t} dt = \int_{-T}^{T} f(t) e^{-j\omega t} dt$$

is called the truncated Fourier Transform of f(t). Since physical measurements are necessarily finite in duration, the convergence of (5.3) to $S_{ff}(\omega)$ as $T \to \infty$ is not of importance here (see, for example, Laning and Battin (1956) or Bennett (1956) and Appendix 3).

Now it is shown in Appendix 3 that substitution of (5.1) into expressions (5.4) and (5.3) leads to the relation

$$(5.5) \qquad S_{ff_T}(\omega) = \frac{T}{2\pi} \sum_{n=1}^{N} \frac{C_n^2 \sin^2\left[(\omega_n - \omega)T\right]}{(\omega_n - \omega)^2 T^2}$$

for the rectangular "time window" of Equation 5.2.

For a particular frequency ω_n contained in the input series, Equation (5.5) becomes

$$(5.6) \qquad S_{ff_T}(\omega_n) = \frac{T}{2\pi} C_n^2$$

This latter result is obtained by applying L'Hospital's rule to the right hand side of (5.5).

It should be noted that the expression of (4.1-36) indicates that even with a single sine wave, say C_i, there is power present at all frequencies, having the characteristic spectrum shown in Figure 5.1.

Alternatively, we can define the power spectrum by measuring the average power in a band of frequencies of width ($\triangle \omega$) centered on ω_i. Now, since from (5.1) the average power for all frequencies in f(t) is clearly given by:

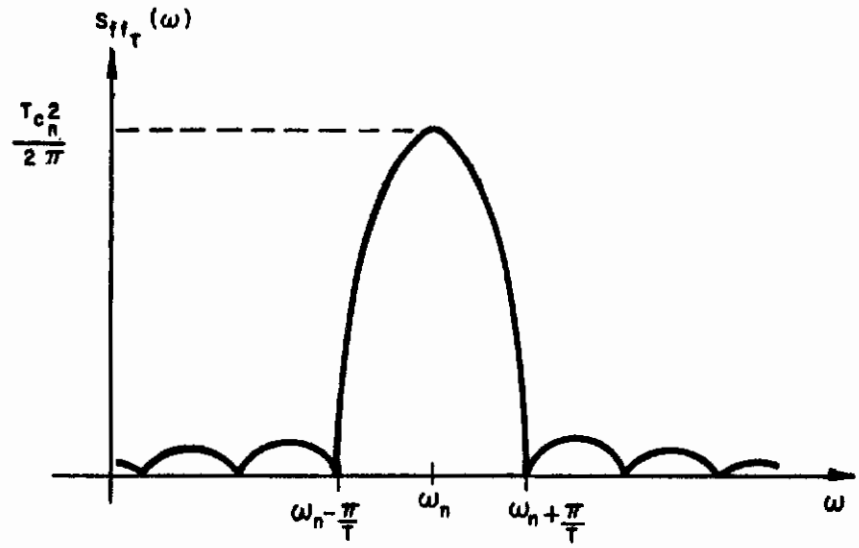


Figure 5.1 Power Spectrum for Finite Duration Sine Wave of Frequency ω_i

$$(5.7) \qquad P_{av} = C_o^2 + \sum_{n=1}^{N} \frac{C_n^2}{2}$$

then the power spectral density can be defined as

$$(5.8) \qquad S_{ff}(\omega_n) = \frac{\text{Ave. power in freq. bandwidth } \Delta\omega_n}{\Delta\omega_n} = \frac{\Delta P_n}{\Delta \omega_n}$$

Since the elementary frequency bandwidth is determined by the period for a Fourier series, we have

$$(5.9) \qquad \Delta \omega_n = \omega_n - \omega_{n-1} = 2\pi \left(\frac{n}{2T} - \frac{n-1}{2T} \right) = \frac{\pi}{T}$$

Then, substituting in (5.8):

$$(5.10) \qquad S_{ffT}(\omega_n) = \frac{\Delta P_n}{\Delta \omega_n} = \frac{C_{n/2}^2}{\pi/T} = \frac{T\, C_n^2}{2\pi}$$

Equation (5.10) is identical with (5.5) given previously. In both of these equations, as expected, the spectral density tends to infinity as $T \to \infty$.

Therefore, we shall define the power spectral density as being proportional to the average power in the corresponding term in the sine (or Fourier) series, where the coefficient of proportionality depends on the run length. Thus, by definition:

$$(5.11) \qquad S_{ffT}(\omega_n) = K_e C_n^2 = \left(\frac{T}{2\pi}\right) C_n^2$$

It should be noted, however, that the problem at hand is not a theoretical one, but the practical one of measurement of a com-

plex signal for a period T_R=2T seconds and the determination, from this measurement, of an estimate of $S_{ff_T}(\omega)$. Due to the "lobing" of the frequency characteristics (as shown in Figure 5.1), any attempt to extract c_n^2 with finite-width, realizable filters will be subject to error, as discussed in Appendix 3. We have chosen a run length (2T) equal to approximately 32 cycles of the lowest frequency present in the input. In this case the separation between zeros of the frequency characteristic is π/T=0.006 cps, or 1/32 of f_1. Since for our experiment we had f_1= 0.18 cps, the corresponding run length 2T for 32 cycles is T_R= 180 sec. Therefore a filter which would measure the power in the primary lobe must be at least .006 cps wide on each side of the center frequency. Our filters are about .01 cps wide, as sketched below in Figure 5.2.

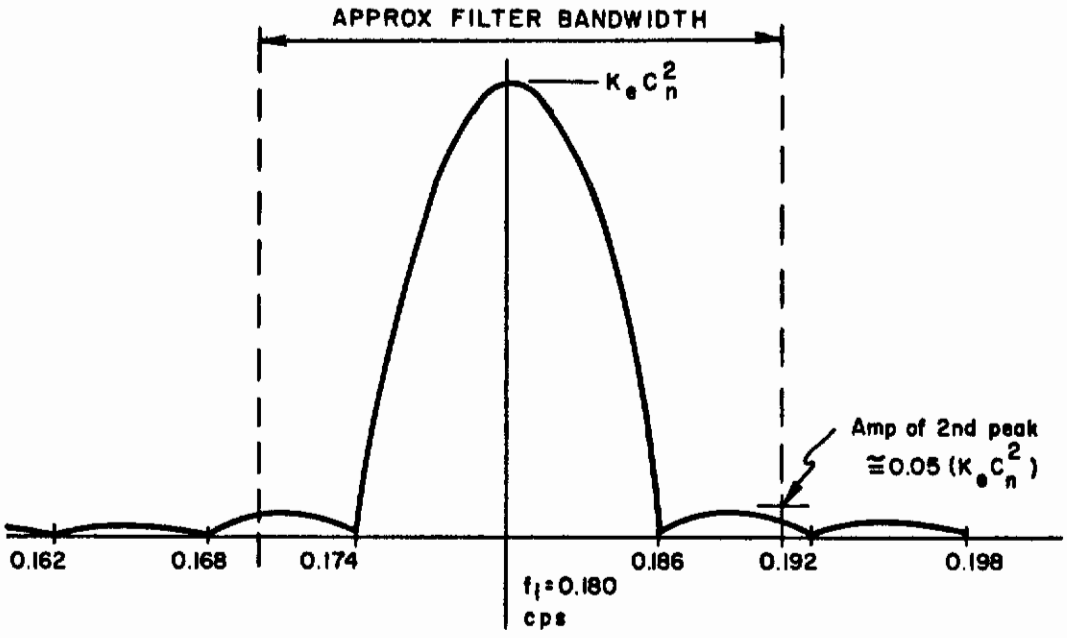


Figure 5.2 Filter Bandwidth for Extraction of Fundamental Frequency

5.2.2 Measurement of the Power Spectral Density. The method selected for measurement of the spectrum $S_{ffT}(\omega)$, as defined above, is based on modulation and low pass-filtering. The method will be indicated briefly and heuristically here; more details are given in Appendix 3.

We consider the function f(t) given in Equation (5.1) in its equivalent form:

$$f(t) = \sum_{i=1}^{N} (a_i \sin \omega_i t + b_i \cos \omega_i t) \tag{5.12}$$

where the ω_i are not necessarily multiples of some basic frequency. Multiplication by $\sin \omega_j t$ and $\cos \omega_j t$ respectively produces:

$$f(t) \sin \omega_j t = \sin \omega_j t \sum_{i=1}^{N} (a_i \sin \omega_i t + b_i \cos \omega_i t)$$

$$f(t) \cos \omega_j t = \cos \omega_j t \sum_{i=1}^{N} (a_i \sin \omega_i t + b_i \cos \omega_i t) \tag{5.13}$$

Now, if there is a frequency component ω_j present in f(t), then the product results in the following:

$$f(t) \sin \omega_j t = a_j \sin^2 \omega_j t + \sin \omega_j t \sum_{i} \text{(N-1) terms} \tag{5.14}$$

$$= \frac{a_j}{2} - \frac{a_j \cos 2\omega_j t}{2} + \sin \omega_j t \sum \text{(N-1) terms}$$

and similarly

$$(5.15) \qquad f(t)\cos\omega_j t = \frac{b_j}{2} + \frac{b_j \cos 2\omega_j t}{2} + \cos\omega_j t \sum (N-1)\text{terms}$$

We now note that (5.14) and (5.15) consist of a d.c. term and an oscillatory component and we use low pass filters to extract the d.c. term. If the filters were perfect, we would extract the d.c. components $a_j/2$ and $b_j/2$ exactly. Since filters of zero-width and rectangular cutoff are not physically realizable, we obtain instead a ripple component, the magnitude of which depends on the separation of the components ω_i in the original signal i.e., on the run length. Averaging the resulting value decreases the effect of the ripple. If we denote the filter outputs as

$$(5.16) \qquad \begin{aligned} q_s &= \frac{a_j}{2} + \epsilon_a(t) \\ q_c &= \frac{b_j}{2} + \epsilon_b(t) \end{aligned}$$

Then we can square these terms to obtain, after adding:

$$(5.17) \qquad q_s^2 + q_c^2 = \frac{1}{4}(a_j^2 + b_j^2) + \left[(a_j\epsilon_a + b_j\epsilon_b) + (\epsilon_a^2 + \epsilon_b^2)\right],$$

Filtering the result tends to reduce the effect of the oscillatory terms due to the errors ϵ ; as noted in Appendix 3, the effectiveness depends on the filter bandwidth and averaging time as well as on the separation of frequencies ω_i. Thus we obtain an estimate of the spectral density as

$$(5.18) \qquad S_{ffT}(\omega_j) \cong \frac{k_e}{T} \int_{-T}^{T} (g_s^2 + g_c^2)dt = k_e c_j^2$$

In block diagram form this method of spectral analysis is given in Figure 5.3. Since the driving function used in the experiments consisted of 10 sine waves, 10 identical circuits to Figure 5.3 were required. The design of similar analyzers is discussed by Seltzer and McRuer (1959).

Appendix 3 also gives details on the construction of the low-pass filters indicated in Figure 5.3. Basically, second-order filters with a damping ratio of 0.7 and a resonant frequency of 0.01 cps were selected. The filters were mechanized by using RC networks and operational amplifiers. The squaring operations were performed using analog multipliers and the averaging by means of integrators.

5.3 Conditions of the Experiment

The general conditions of the experiment were the following:

(a) The tracking was compensatory and one-dimensional with negligible controlled element dynamics.

(b) The input forcing function was a "random-appearing" sum of 10 non-harmonic sine waves which approximated a gaussian process.

(c) The generation of the forcing function as well as the real-time measurement of power spectra were performed using an electronic analog computer.

(d) The objective of each run was to record the power spectral density of the loop error as well as the time history of forcing function, operator error and output, and sampled model error and output.

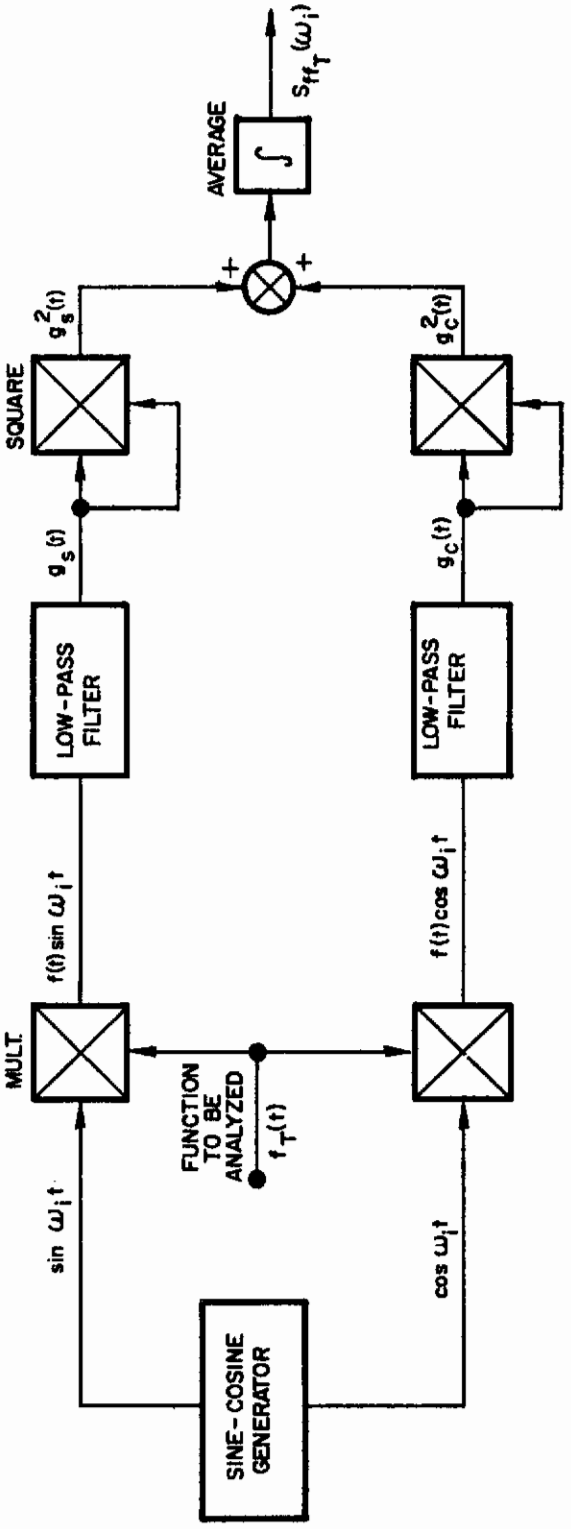


Figure 5.3 Block Diagram of One Channel of Analog Spectral Analyzer

(e) The range of input function frequency was such that the task was "difficult" for all operators, with the effect that the tracking error was of the same order of magnitude as the input.

5.3.1 Equipment. The tracking apparatus is illustrated in the photograph of Figure 5.4 and in the block diagram of Figure 5.5 A DuMont Type 304 oscilloscope was used for the display, with the dot being adjusted to approximately 1/8" diameter. Horizontal motion away from the center represented the system error; zero error corresponded to a position between two black vertical lines painted on the

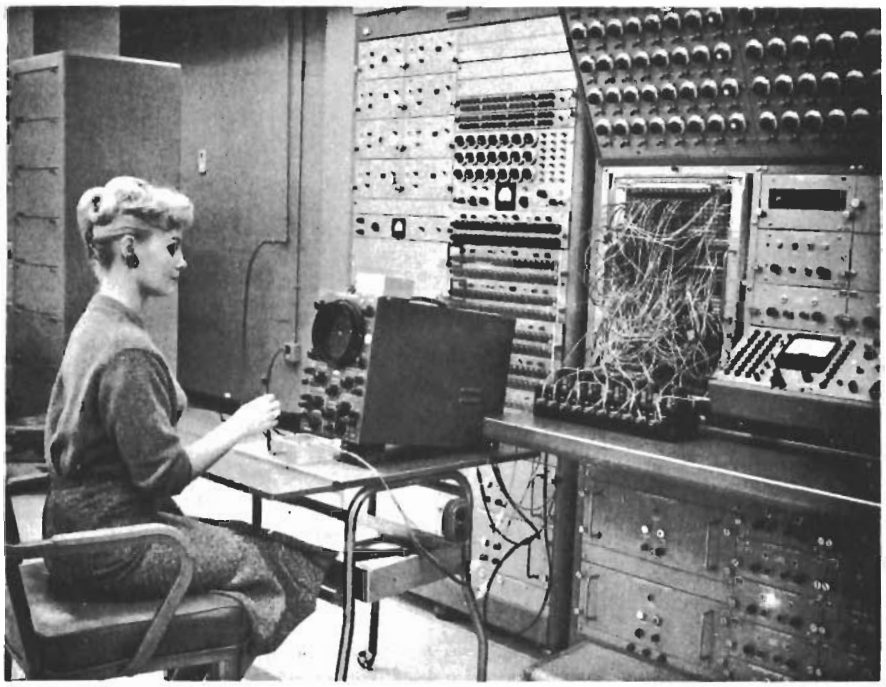


Figure 5.4 Experimental Arrangement

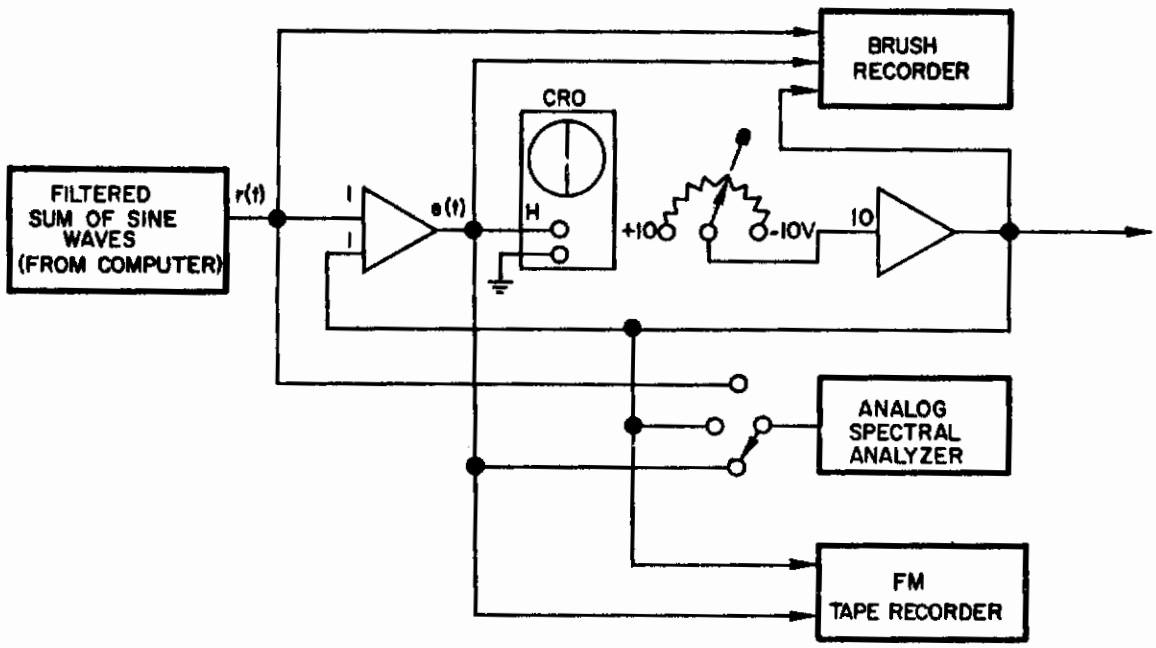


Figure 5.5 Block Diagram of Tracking Equipment

oscilloscope face. The maximum excursion was adjusted to approximately 2 inches each side of zero. The control lever was constrained to move in one plane (to the left and right, in the same plane as the CRO face) thus positioning a 10K wirewound potentiometer and providing feedback voltage as indicated in the block diagram.

For the intermittent display situation a power relay was used to add a series resistor to the "INTENSITY" circuit of the CRO, of sufficient magnitude to blank the spot completely. The pulses used to turn the spot on and off were provided by the sampled-data portion of the analog computer and are discussed below in Appendix 7.

The subject was seated directly in front of the oscilloscope and was allowed to adjust his viewing distance and select his arm and hand position as he desired. Since all the subjects were right handed, the control lever was located slightly to the right of the oscilloscope for the most comfortable arm position.

Input and output signals for all rms were recorded on strip-chart recorders (and for the last set of runs were also recorded on a Precision Inst. Type PS 207 FM tape recorder. A switch provided for the selection of any signal to be fed to the analog spectral analyzer. In nearly all the runs (except for calibration purposes) this signal was the tracking error.

The analog computer utilized consisted of 4 consoles of assorted equipment including some 110 operational amplifiers, 4 sample-and-hold circuits, 18 electronic multipliers (dual channel) and 12 servo multipliers.

The entire experiment was performed in the Analog Computation Center at Space Technology Laboratories, Incorporated in El Segundo, California.

5.3.2 Input Signals. The selection of a sum of sine waves was based on the desire to obtain an input which was random-appearing and yet completely predictable mathematically. The periods were selected such that they were non-harmonic within 0.02 seconds. The frequencies selected have an approximately linear spacing between 1 and 12 rad/sec., this range is close to the upper limit of the tracking ability of human operators and consequently should be adequate for testing models in the range where "remnants" are high. The frequencies used appear in Table 5.1.

The sine waves were started in phase. However, the measurement of power spectral density began about 120 seconds after t = 0 (in order to allow for transients to decay), at which time the

phase relationships of the sinusoids were arbitrary. It can be shown by the Central Limit Theorem (Bendat, 1958, for example) that the probability density function for the sum of n independent sinusoids

Table 5.1

Frequencies of Sine Waves for Driving Function f(t)

Sine Wave No.	Frequency (rad/sec)
1	1.122
2	1.995
3	2.860
4	3.940
5	4.910
6	6.10
7	7.32
8	8.25
9	9.31
10	12.00

approaches a Gaussian distribution as $n \to \infty$. The forcing function used here was of the form

$$f(t) = \sum_{n=1}^{10} a_n(\cos \omega_n t + \phi_n) \tag{5.19}$$

Siskind (1961) shows that for $N > 8$ the distribution is indistinguishable from a Gaussian one for most practical purposes. A plot of the amplitude probability density function for a 100 second sample of f(t)*, the amplitude quantized into 32 equal levels on

*Actually the Figure shows the probability density for f(t) after filtering by a first order lag. However, if f(t) is Gaussian, f(t) after filtering by a linear filter, will also be Gaussian.

each side of zero (arbitrary units), is shown in Figure 5.6. While a statistical normality test has not been applied to this sample distribution, the measures of skewness and excess are not large, and slight departures from normality are not considered to be too significant in this study.

The actual forcing function consisted of f(t) after it was filtered by a first-order lag with a break frequency at either 1.5 or 3.0 rad/sec. The measured spectra of these forcing functions are given in Appendix 3, in connection with a discussion of the calibration of the analog spectral analyzer. The resulting functions are quite random in appearance, as can be seen by the two samples shown in Figure 5.7.

5.3.3 Specific Experiments and Procedure. A total of six (6) cases were considered, the first two being used primarily for training purposes. The basic input spectra were those discussed above, i.e., the sum of 10 sine waves filtered by a low-pass filter with a break frequency of either 1.5 or 3.0 rad/sec. In addition, the visual display was either continuous or sampled. The training cases were run either with only the first 4 sinusoids and filter at 1.5 rad/sec or with all 10 sinusoids and an attenuated input obtained by moving the filter break frequency to 0.75 rad/sec. The six cases are shown below in tabular form.

Table 5.2
Conditions of Tracking Experiment

Case No.	Number of Sinewaves	Input Filter Break Freq. $\left(\frac{\text{rad}}{\text{sec}}\right)$	Display Condition
1	4	1.50	Continuous
2	10	0.75	Continuous
3	10	1.50	Continuous
4	10	3.00	Continuous
5	10	1.50	Intermittent
6	10	3.00	Intermittent

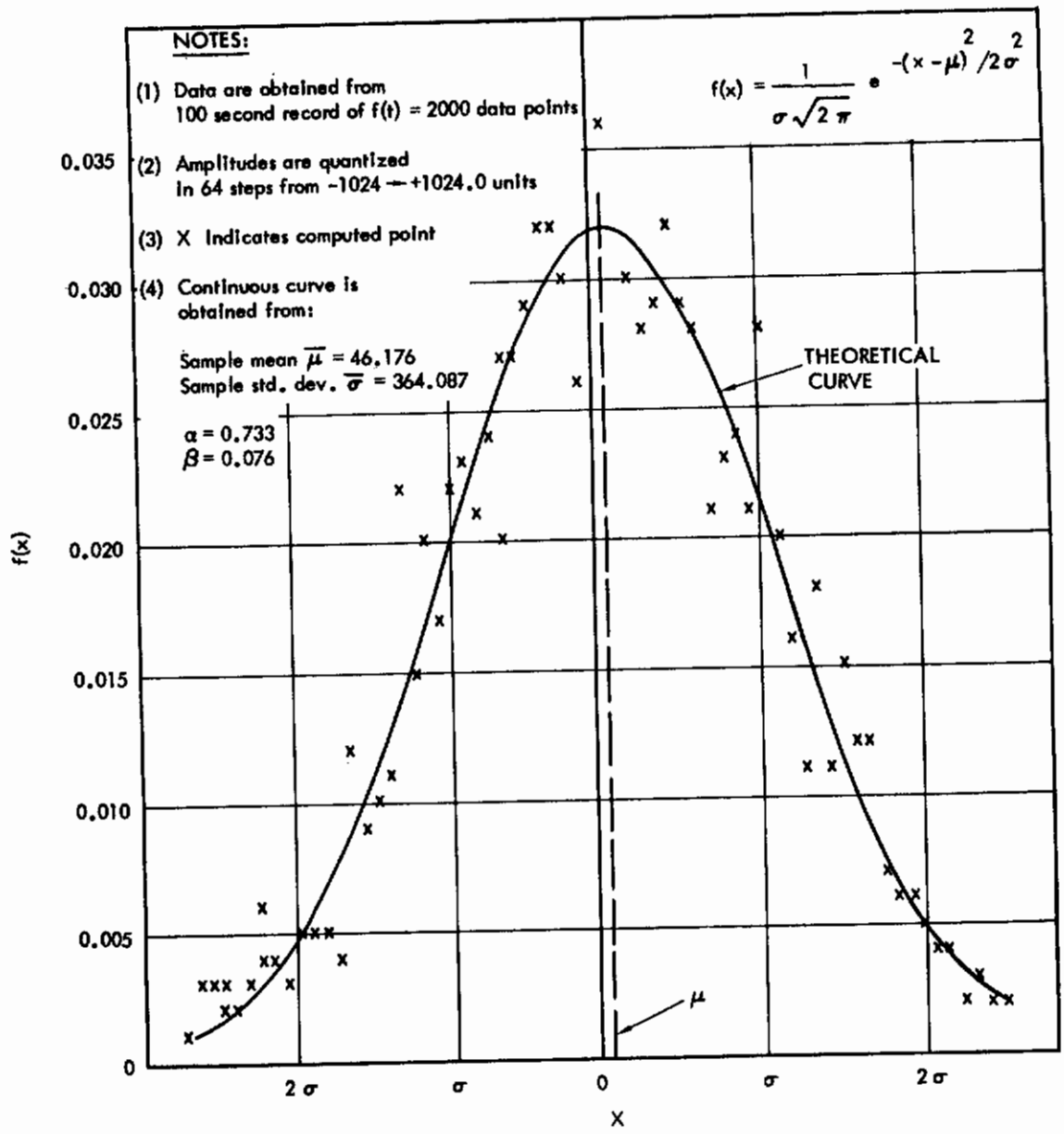
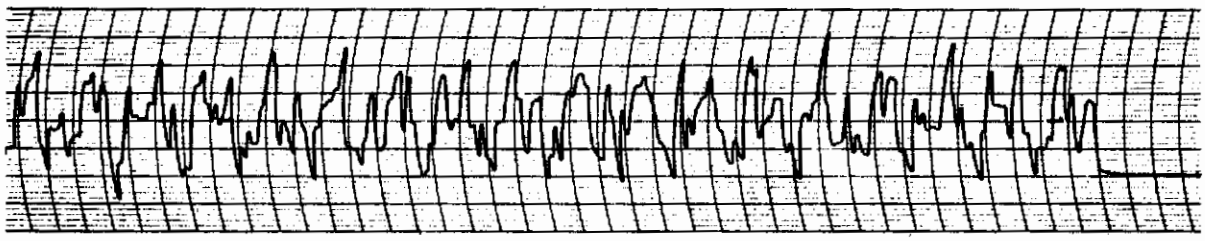
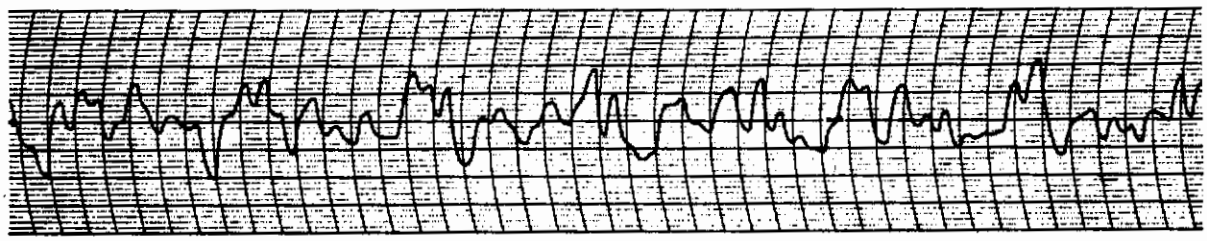


Figure 5.6 Amplitude Probability Density Function for Input Function f(t)



(a) "Easy" Forcing Function: -
Filter Break Frequency = 1.50 rad/sec
Paper Speed = 2 mm/sec



(b) "Difficult" Forcing Function:
Filter Break Frequency = 3.00 rad/sec
Paper Speed = 5 mm/sec

Figure 5.7 Sample Waveforms of Forcing Functions

The following comments can be made on the experimental procedure:

(1) Each run lasted 5 minutes. Of this time, 120 seconds was required for filter transients to become negligible. At $t = t_o + 120$ the output averaging circuits were connected. Averaging took place over the final 180 seconds of each run.

(2) Subjects unfamiliar with tracking were asked to track between 6 and 12 trial runs using Cases 1 and 2 or Case 3, until the results of successive runs were consistent within approximately $\pm$20% in the recorded values of power spectral density obtained from the analog computer.

(3) Approximately 100 good runs were made during a 1 week period.

(4) The last 8 runs using two well-trained subjects were recorded on magnetic tape for digitizing and processing by the IBM 7090 correlation and spectral program.

5.3.4 Subjects. A total of 10 subjects was used in various portions of the experiment. Good data is available from 8 of them who were able to participate in most cases of the experimental program. Various backgrounds were used in order to obtain as wide a range of results as possible. The subjects can be identified as follows:

Subject	Background	Previous Tracking Experience
AAB	Grad. student UCLA	None
GAB	Grad. student UCLA	Some
WWB	Administrative aide, STL	None
MJF	Engineer, STL	None
JJP	Engineer, STL	None
FCR	Engineer, STL and ex-pilot	Extensive
GAR	Grad. student UCLA	None
TCR	Technician, STL	None
GNW	Grad. student UCLA	Extensive
KEZ	"	Some

All the subjects participated as volunteers. Of the 4 subjects used most extensively, GNW and KEZ were aware of the purpose of the experiment, while WWB and JJP were not and had no background in the field.

5.4 Simulation of the Sampled-Data Models

It has been mentioned above that some of the proposed models of the human operator were simulated on the analog computer and recordings made during a number of runs. While the study of sampled-data systems by means of analog equipment has been performed for several years, a number of innovations were made in the course of this study. Consequently some material on sampled-data simulation techniques is collected in Appendix 4. In particular, the techniques utilized for the simulation of first-order and modified first order hold circuits are discussed in detail.

The key to the simulation method utilized is the analog sample-and-hold channel consisting of two amplifiers, as shown in Figure 5.8. Basically, this channel is a simulation of a sample-and-zero-hold operation. Its operation is based on the presence of two pulse trains:

(a) The S-pulse (or "sample" pulse) which actuates the relay indicated by the letter "S", and
(b) The P-pulse (or "present" pulse) which activates the second relay a short (but adjustable) time later.

The first amplifier samples and holds the input function upon arrival of the S-pulse. The sampled signal is held for a delay time of D seconds (where $D < T$, the sampling interval) and then, upon arrival of the P-pulse, it is transferred to the second amplifier. This arrangement offered two great advantages for our experimental study:

(1) An adjustable time-delay was available, so that with any sampled model it was easy to insert a true time delay into the system as required by the model; and

(2) In order to provide a "blanking" signal for the sampled CRO display, the sample-hold channel provided relay contacts which closed at the sampling frequency and remained closed for prescribed lengths of time; these were used to keep the spot "on" for approximately one half the sampling period regardless of sampling frequency.

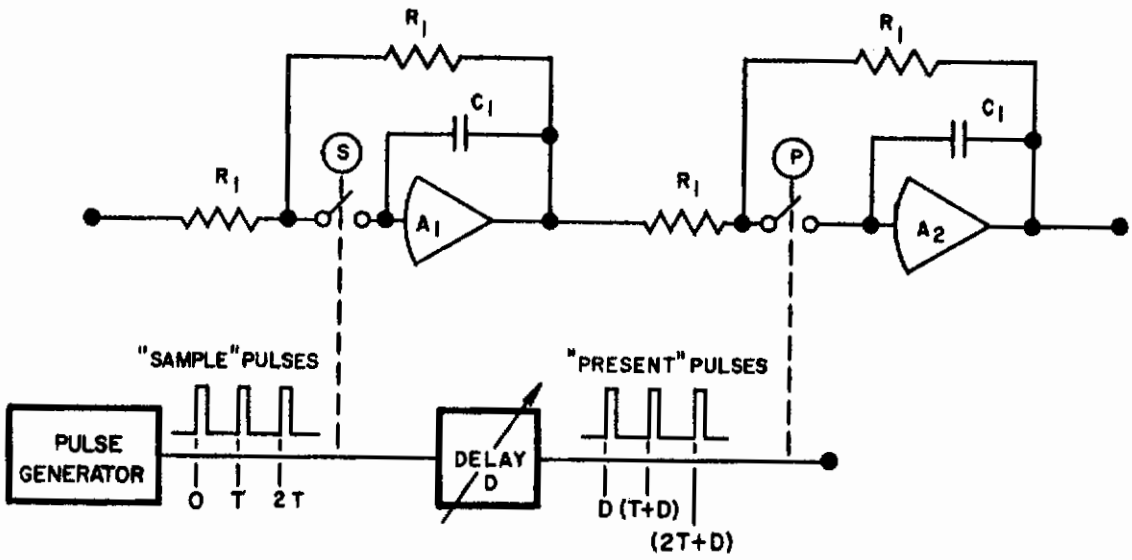


Figure 5.8 Analog Sample and Hold Channel

Chapter 6

DISCUSSION AND EVALUATION OF EXPERIMENTAL RESULTS

6.1 General Comments

The major purpose of the experimental study, as outlined in Chapter 5, has been to obtain power spectral density measurements of the tracking error in compensatory tracking, in order to have data which could be used for comparison with spectra of "a priori" sampled-data models of the operator. The major effort was centered on obtaining error spectra since the variation of the error spectral density is relatively small within the frequency ranges of interest, while the output spectrum attenuates rapidly. Consequently, the error spectrum is a more sensitive indicator of the system properties. In addition to a discussion of the error spectra, this chapter will also present some comments concerning output spectra and time domain response, as well as presenting comparisons of measured and computed data.

The chapter is divided into four major sections which discuss the following features of the results:

(1) The characteristics of the error spectrum (at the input frequencies) for continuous inputs

(2) The effects of intermittent displays on the spectra

(3) The characteristics of the spectra between and beyond the input frequencies

(4) Comparisons of computed and measured responses, both in the time and frequency domains.

Items (1) to (3) are primarily qualitative, while Item (4) concerns the quantitative comparison of experimental with analytical results.

6.2 General Characteristics of the Measured Error Spectra

As outlined in Chapter 5, error spectra were measured for eight operators under various experimental conditions.

The error spectral density plots from the analog measurements described previously are given in Appendix 6. A typical set of runs for Case 3, Subject JP, is shown in Figure 6.1. Before considering the significance of these curves, it should be noted that:

(1) For all the runs of cases 3 and 4 the input consisted of a filtered sum of 10 sinusoids, and

(2) The points plotted on these curves represent values of average power in a narrow frequency band, as measured by the analog method described in Appendix 3. While these points have been connected with dotted lines in the figures, the data from which they were obtained give us no information on the presence of energy at frequencies intermediate between the 10 measured points. (The digital computation of spectral density, to be discussed below shows that the power between the 10 frequency peaks is lower by 10 to 20 db than the peak level).

6.2.1 Relation to Elkind's Work. In Chapter 3 it was noted that the error spectra obtained by Elkind (1956) were characterized by dual peaks, at least for "difficult" inputs. The general pattern was that of a maximum error at some intermediate frequency followed by a second peak after which the error attenuates rapidly with increasing frequency. The same general pattern appears in most of the spectra recorded here. This resemblance is significant since the experiment was conducted under somewhat different physical conditions (e.g. a "stick" controller vs. Elkind's conducting stylus) and the input signal consisted of only 10 sinusoids as contrasted with Elkind's 144 sinusoids.

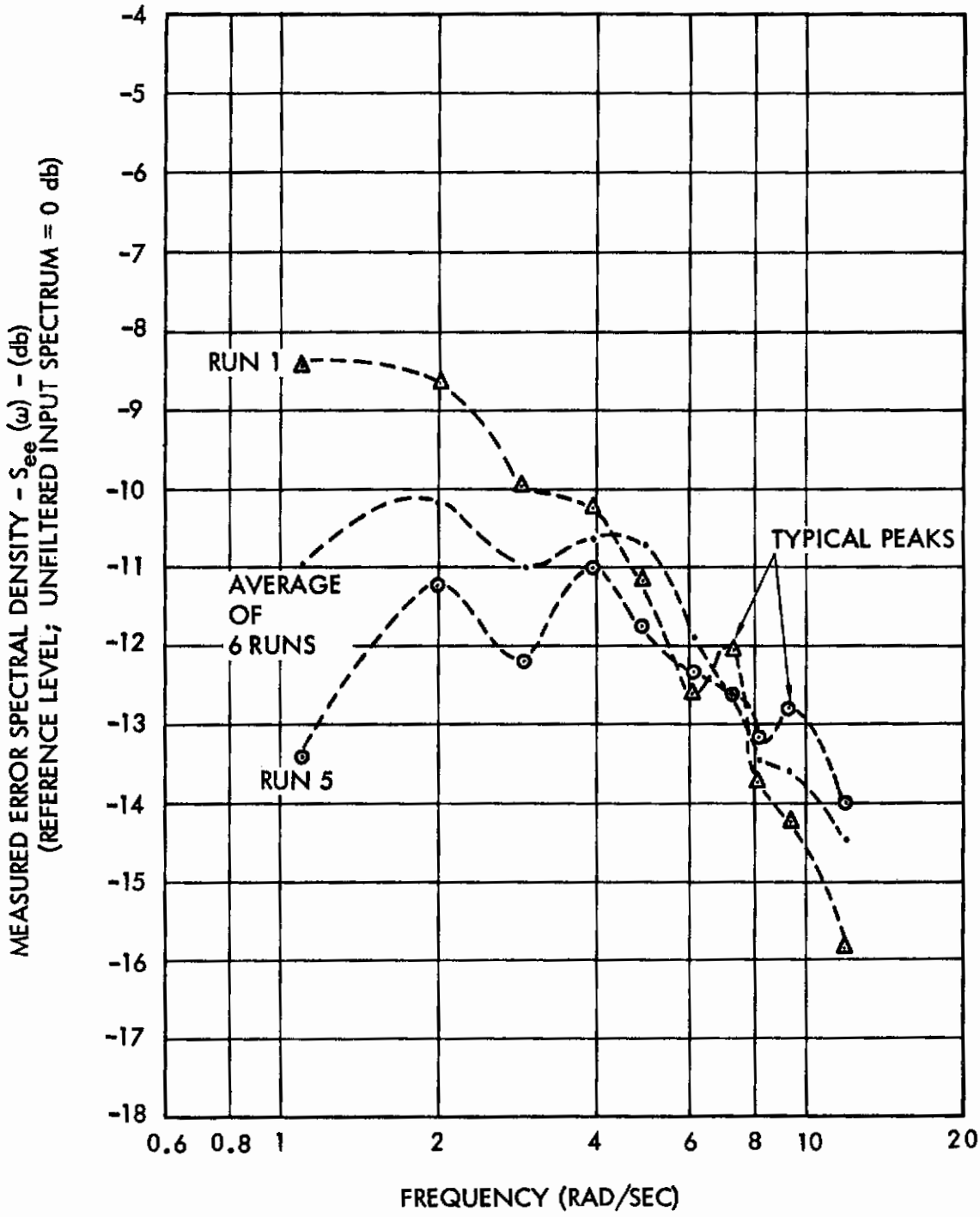


Figure 6.1 Typical Runs for Subject JP and Averaged Spectrum (Case 3)

6.2.2 Comparison with "a priori" sampled models. The spectral characteristics of several sampled-data models were computed in Chapter 4. It was shown that the error spectra obtained by using first-order hold (FOH) models (both conventional and "modified") produced a general spectral characteristic of similar shape to that of Elkind's experiment. Furthermore, the second peak was shown to occur at or near $\omega_s/2$ (where ω_s is the model sampling frequency.) Consequently, the presence of similar features in this experiment gives us a clue to relationships with sampled models, which will be explored quantitatively in Section 6.5 below.

6.2.3 Occurrence and location of the "sampling peak". For convenience and ease of identification we shall designate the 1 to 3 db peak which follows the maximum in the operator's error spectrum, as the "sampling peak". An examination of the error spectra in Appendix 8 shows that the maxima occur in the range of 3-6 rad/sec and the "sampling peaks" in the range of 6-10 rad/sec. The occurrence and frequency distribution of the "sampling peaks" in the various experiments is summarized in Table 6.1.

The variation in location of the "sampling peak" is probably due not only to the inherent differences and range of human capability, but to factors such as learning and task difficulty which will be considered later. The hypothesis we shall explore further in later chapters is that if the human operator does indeed operate as a sampling system, he does so as a system with variable sampling frequency. Over a given run however, the frequency may be approximately constant. Two consequences of this hypothesis would be: (1) That some runs would show no clearly evident "sampling peak" if the frequency varies greatly during the run, and (2) That the peak location could vary from run to run to some extent depending upon motivational factors as well as learning, experience, instructions, etc.

6.2.4 On averaging of spectra. The variation in location of the "sampling peak" has a very interesting consequence. If a number of runs with different "sampling peaks" are averaged, the peak may be

Table 6.1

Occurrence and Location of "Sampling Peak"
in Experimental Data

Case No. and Description	Approximate Freq. of Peak (rad/sec)	No. of Runs	% of Total Peaks for This Case
Case 3:	4.9	1	3.3
Sum of 10 sine waves; filtered by lag filter with break freq. ω_B = 1.50 rad/sec	7.3	7	23.4
	8.2	4	13.3
	9.3	15	50.0
	10.0 or above	3	10.0
(33 runs total)	no peak evident	3	-
	TOTAL	33	100.0
Case 4:	7.3	3	20.0
Sum of 10 sine waves; filtered by lag filter with break freq. ω_B = 3.00 rad/sec	8.2	1	6.7
	9.3	5	33.3
	10.0 or above	6	40.0
(16 runs total)	no peak evident	1	-
	TOTAL	16	100.0

obscured by the averaging process. In Figure 6.1 the average of the six runs is also indicated and it can be seen that this "average spectrum" does not indicate a clear peak in the higher frequencies. Thus, it is possible that ordinary ensemble averaging methods, which are used due to the variability of tracking data, will have to be used carefully in future studies of the sampling hypothesis.

6.2.5 The effects of learning. While the study of changes in spectral density due to learning was not among the primary objectives of this research, some qualitative observations can be made on the basis of the data in Appendix 6. It should be noted that these are no more than observations which cannot be substantiated with the data at hand.

The learning period for the inexperienced subjects included approximately 10 five-munute runs and 2-4 runs for experienced trackers. While 50 minutes of tracking experience over a period of a a week is probably sufficient to adapt to the equipment at hand, the training problem is aggravated by the fact that the task was extremely difficult for all the operators. The time-domain traces to be discussed later are an indication of the difficulty of the task, since the magnitude of the error signal will be seen to be approximately equal to the input magnitude. The following comments on learning can be made:

(1) The "sampling peak" tends to move to higher frequencies with more practice.

(2) The low frequency error tends to decrease with practice, but at the expense of increases in high frequency error. The subjects were simply instructed to "keep the dot as close to the center of the screen as possible" with no directions begin given on high or low frequencies.

(3) The mean square of the error signal tends to decrease with training (values of mean squared error are in Table 6.2 below).

(4) With additional experience, the "sampling peak" tends to become more clearly defined for each subject.

None of the above observations can be regarded as definitive statements of fact; but they are interesting suggestions for future experiments. The reduction of mean squared error is a reasonable effect of training. The reduction of low-frequency error can be attributed to an increase in d-c gain by the operator.

6.2.6 The effect of task difficulty. Consider Figure 6.2 where the error spectra of Subject JP are plotted for several runs of Case 4. This latter case differs from that of Figure 6.1 only in the break frequency of the input filter, which is now 3.0 rad/sec rather than 1.5 rad/sec. The curves of Figure 6.2 are typical of the complete set of data in Appendix 6 and lead to the following observations:

(1) As a result of this particular increase in task difficulty, the error spectrum is considerably more "flat." The range of maximum to minimum error power in the frequency bands considered is about 3 db, while it is about 6 db in Figure 6.1.

(2) The "sampling peak" is less clearly defined, appearing more as a leveling or slight increase in error spectral density at the highest frequencies considered. Assuming that the measured peaks are valid, it can be noted that they have shifted to a higher frequency, by inspection of the distribution of peak locations, as given in Table 6.1.

(3) The mean squared error as given by the sum of the 10 measured error spectral densities (an approximation to the integral $\int S_{ee}(\omega)\, d\omega$) is significantly greater than for case 3. For Subject JP, the MSE is nearly twice the value for case 3, for Subj. WB it about 2-1/2 times as large. However, if these error spectral densities are normalized by dividing by the

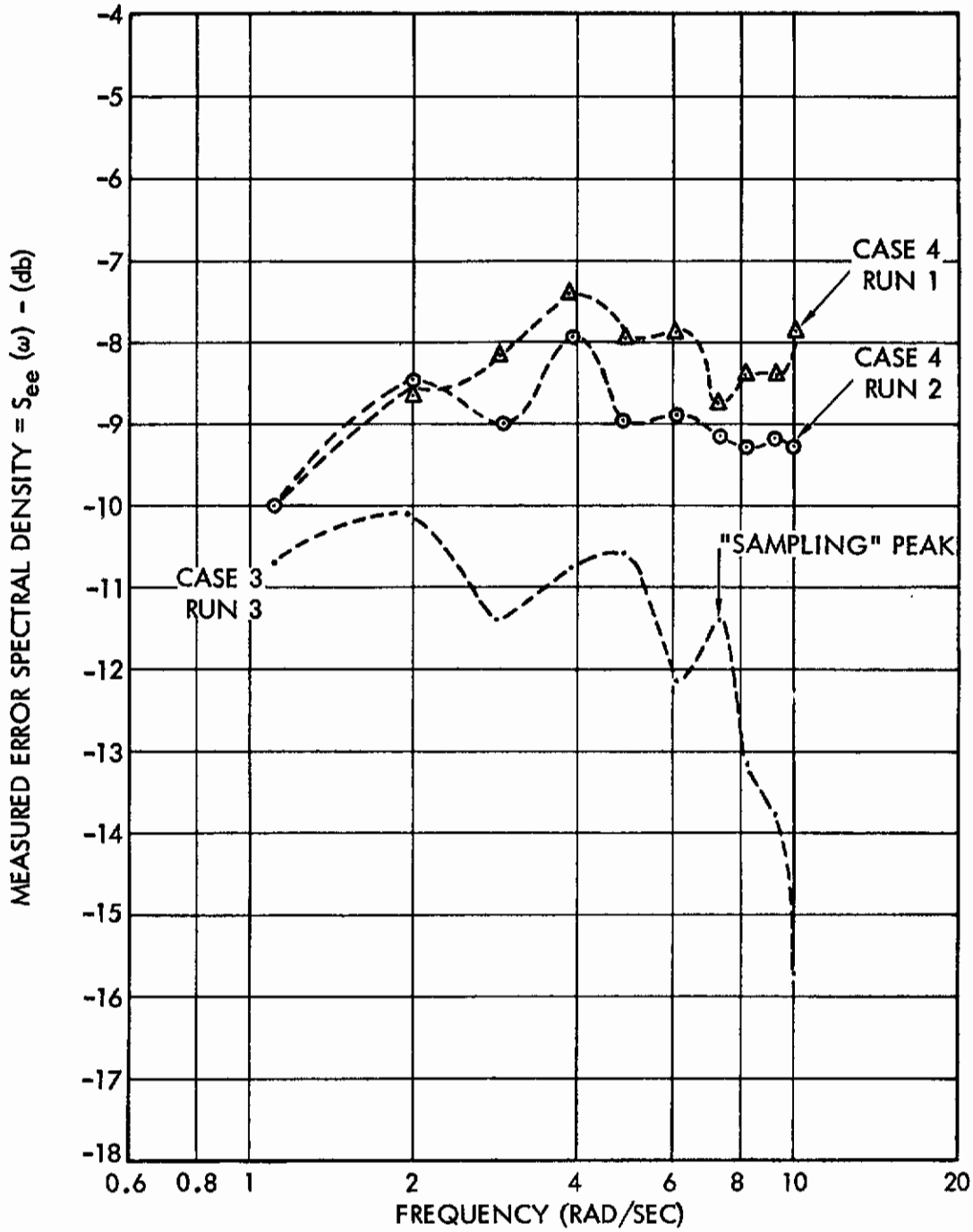


Figure 6.2 Comparison of Cases 3 and 4 for Subject JP

total input power, it can be seen from Table 6.2 that one operator's performance has improved slightly. The great difficulty of this specific tracking task (both cases 3 and 4) can be seen by noting that the error power is approximately 1/2 the input power over the range of frequencies considered.

(4) The variation among subjects is somewhat greater, possibly because for some operators (such as WB), Case 4 represents essentially a limit of performance (see Table 6.2). As with Case 3 the variation of MSE is due in large part to the ability of some operators to reduce the low frequency errors, where the amount of input power is largest.

6.2.7 Differences between subjects. Table 6.2 gives a good indication of the range of variability for this experiment. Some past experimenters (such as Elkind) have tended to dismiss variation among subjects as insignificant, and this is probably true for sufficiently "easy" inputs. However, it appears reasonable that the variation of mean squared error in the present experiment is due to differences in individual ability, which could not be completely eliminated by training.

However, while error "scores" differed quite widely, the shapes of the spectral density curves are very similar. It will be shown below when the spectra of the models are introduced, that much of the variation in experimental data can be accounted for by gain changes or "reaction time" changes in the model.

6.3 Response to An Intermittent Display

All the results discussed above were based on Cases 3 and 4, where the operator attempted to keep a dot in the center of the oscilloscope screen. Cases 5 and 6 were concerned with an attempt to study another implication of the sampling hypothesis.

Examination of all the data for sampled inputs in Appendix 6 makes it clear that: (1) the variation between subjects is considerably less than it was for the continuous case, and (2) the variation in location of the "sampling peak" is considerably reduced.

The fact that in all the sampled runs made (a total of 21) the "sampling peak" occurs at or near 1/2 the display sampling frequency is highly significant and provides support for the assumptions previously made in the construction of the sampled models. (The qualification "at or near $\omega_s/2$" must be made since the input included only 10 discrete frequencies, none of which was exactly equal to one-half the sampling frequency.)

The effect of changing from a continuous to a sampled display can be summarized by noting that:

(1) The basic shape of the error spectra was unchanged.

(2) The location of the "sampling peak" is clearly related to the display sampling frequency.

6.4 Energy Beyond the Input Bandwidth

One of the reasons for selecting the sampled-data approach to human operator models was that the sampling operation generates frequencies not present in the input signal. When the input signal is band-limited, as was the case in this experiment, the effect of the sampler is to repeat the input frequencies at harmonics of the sampling frequency ω_s. Thus, if the input spectral density consists of discrete spectral "lines," i.e.,

$$(6.1) \qquad S_{rr}(\omega) = \frac{1}{2} \sum_{i=1}^{N} a_i \left[\delta(\omega - \omega_i) + \delta(\omega + \omega_i) \right]$$

then the corresponding sampled spectrum is given by

$$(6.2) \qquad S_{rr}^*(\omega) = \frac{1}{T^2} \sum_{n=-\infty}^{+\infty} S_{rr}(\omega + n\omega_s)$$

TABLE 6.2

Mean Square Error* For All Runs

Case No.	Trial No.	SUBJECTS KZ	JP	GB	AB	WB	GW	FR	GR
3	1	.778	.755	.791	.805	.974	.500	.620	1.06
	2	.632	.979	.876	.841	1.02	.429	.622	.910
	3	.644	.663	.885	.941	.949		.546	
	4	.630	.689	.850		1.10			
	5	.520	.562	.700		1.01			
	6		.744	.886		.936			
	7					.845			
Average		.640	.732	.815	.861	.976	.464	.596	.980
Normalized Average*		.398	.456	.507	.536	.609	.289	.372	.610
4	1	1.45	1.43	2.10		2.34	1.15		
	2	1.40	1.23	1.79		2.57	1.03		
	3	1.45					1.02		
	4	1.59					1.07		
	5	1.29							
	6	1.56							
Average		1.45	1.33	1.95		2.45	1.07		
Normalized Average*		.431	.396	.58		.729	.318		
5 (f=2)	1		1.11	.875	.982	1.02			
	2		1.12	.885	1.02	1.15			
Average			1.11	.880	1.00	1.18			
Normalized Average*			.690	.548	.623	.735			
5 (f=3)	1	.950	.925			.988	.815		
	2	.843	.768			.895	.723		
		.810	.826						
		.768							
Average		.843	.840			.942	.769		
Normalized Average*		.525	.523			.586	.479		

TABLE 6.2 (Continued)

Case 6	SUBJECT		
	GW	KZ (f_s=3)	KZ (f_s=6)
MSE:	1.81	2.13	2.39
Normalized Value*:	.539	.634	.710

*Definitions

Since the spectral density of the error signal e(t) is only measured at 10 discrete frequencies, we defined the mean-square of the error as the area of a series of rectangles 1 rps wide, and with an amplitude equal to the respective measured value of error spectral density, $S_{eeT}(\omega)$. That is,

$$\overline{e^2} = MSE \triangleq k_1 \sum_{i=1}^{10} \left[S_{ee_T}(\omega_i)\right]\Delta\omega, \quad \Delta\omega = 1$$

The coefficient k_1 is selected in such a way that the measured power spectral density for each input sine wave, before filtering is equal to 1. Then, the MS value of the unfiltered input, by the above definition, is

$$\overline{f(t)^2} = k_1 \sum_{i=1}^{10} S_{ff_T}(\omega_i) = 10\, k_1\, S_{ff_T}(\omega_1) = 10$$

Normalization.

The averaged values of MSE are normalized with respect to the total input power (i.e. the MS value of the input signal) for the particular run under consideration. The two values, for the two break frequencies of input filter, then become:

$$\overline{r(t)^2} = \sum_{i=1}^{10} \left(\frac{\omega_B^2}{\omega_i^2 + \omega_B^2}\right) K_1 S_{ff_T}(\omega_i) = \begin{cases} 1.606 \text{ for } \omega_B = 1.50 \text{ rad/sec} \\ 3.362 \text{ for } \omega_B = 3.00 \text{ rad/sec} \end{cases}$$

The mean-square of the signal r(t) is of course also given by the value of the autocorrelation at zero, and for a Gaussian process corresponds to the variance of the signal:

$$\overline{r(t)^2} = R_{rr}(0) = \sigma_r^2$$

Consider the block diagram of Figure 6.3 which includes the proposed human operator model. In Chapters 2 and 4 some of the consequences of the presence of the sampler in the loop were discussed. It was pointed out that the sample-and-hold operations should result in: (1) certain characteristics of the spectral densities $S_{ee}(\omega)$ and $S_{cc}(\omega)$ of the error and output respectively; and (2) the presence of energy beyond the bandwidth of the input signal. The experimental study of implication (1) was discussed in Section 6.2 above, while that of implication (2) will be mentioned in the next section, (6.4). However, neither of these deductions were considered sufficiently crucial to the sampling hypothesis, since the input and outputs of the operator are continuous. It was therefore decided to examine the implications of presenting the operator with a sampled or intermittent input.

If the block diagram of Figure 6.3 is a valid representation and the visual input is made intermittent by being turned off for T/2 seconds every T seconds, we would expect the following consequences:

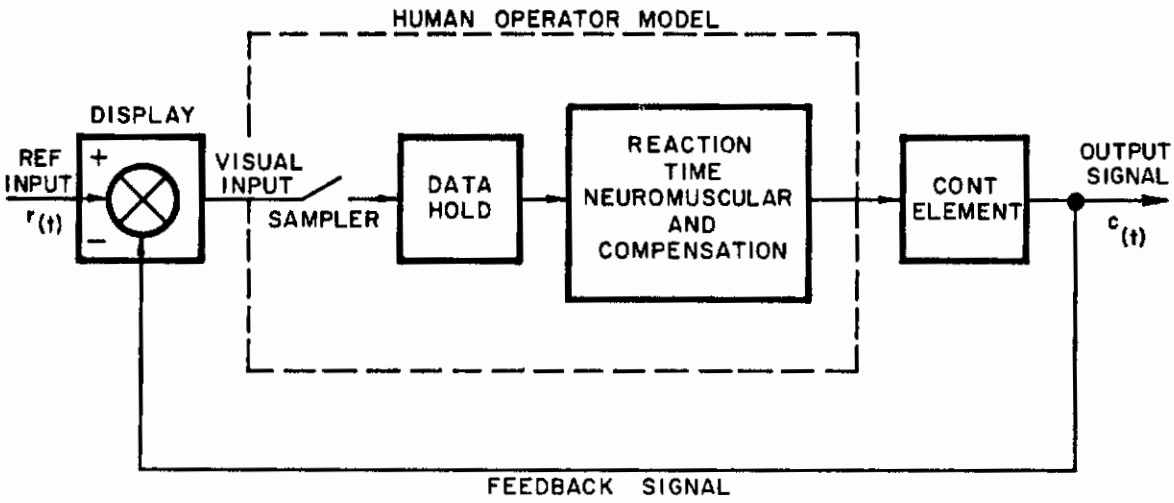


Figure 6.3 Block Diagram of Tracking Loop with Model

(1) If the display sampling rate is equal to the operator's sampling rate and the sampling times are synchronized, his tracking characteristics should not change significantly from the equivalent continuous display case. Since the sampler is simply an open-circuit between sampling instants, the presence of information between sampling times is not important to its operation.*

(2) If the operator is able to track a sampled display, he is forced to operate in a sampled-data manner and at a constant sampling frequency. This implies that the "sampling peak" in the operator's spectrum should correspond to the input sampling frequency and it should be more clearly defined than in the continuous input case.

(3) If the input sampling frequency is changed to a new rate with which the operator can still synchronize this should be evidenced as a shift in the location of the "sampling peak" in his response.

Not enough data were obtained to give definitive verification of these hypotheses. However, the limited data not only do not contradict them, but suggest that they may be valid. Consider the comparison of mean-squared values of the error signal $e(t)$ in Table 6.3. Two interesting observations can be made from this data: First, the range of the MSE for the sampled display is smaller than for the continuous display, and secondly, it represents less of a change for the "average" tracker than for the excellent tracker. One can only speculate about the meaning of these observations. However, at least for some of the subjects, it is clear that the sampled display did not represent a dramatic change.

*This statement, while valid for a mechanical system, has physiological implications which may not be valid. These will be discussed briefly in Chapter 8.

Table 6.3. Mean Squared Error (MSE for Samples and Continuous Displays.

Subject	Continuous Input	Sampled Input	
	MSE for Case 2	Sampling frequency (Samples/Sec)	MSE for Case 5
	Average Values		Average Values
KZ	0.640	3	0.843
JP	0.732	2	1.11
		3	0.840
GB	0.815	2	0.880
AB	0.861	2	1.00
WB	0.976	2	1.18
		3	0.942
GW	0.464	3	0.769

Examination of the shapes of the spectral density curves gives considerable credence to implications (2) and (3) made above. As before, the curves are included in Appendix 6, except for sample curves for Subject JP which are given in Figure 6.4. The average curve for the comparable continuous input (Case 3) is also included for comparison. It can be seen clearly that the change in sampling frequency results in a change in location of the "sampling peak." Not enough data are available to perform statistical significance tests on these observations, but they are included here as qualitative support for the basic hypothesis.

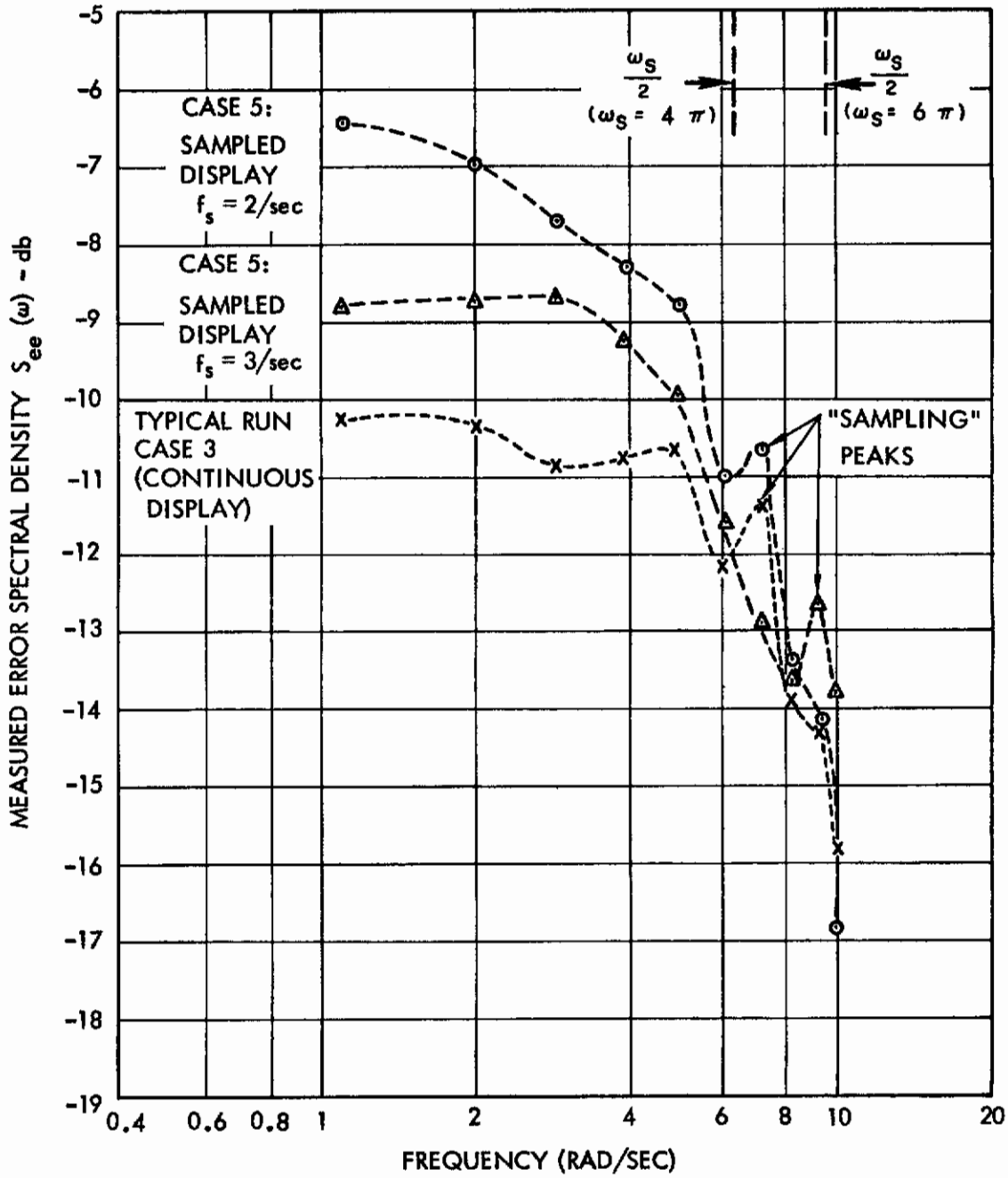


Figure 6.4 Typical Runs for Subject JP with an Intermittent Display
f_s = 2/sec and 3/sec
ω_B = 1.50 rad/sec

which becomes

$$(6.3)\quad S_{rr}^{*}(\omega) = \frac{1}{2T^2} \sum_{n=-\infty}^{+\infty} \sum_{i=1}^{N} a_i \left[\delta(\omega + n\omega_s - \omega_i) + \delta(\omega + n\omega_s - \omega_i) \right]$$

An analogous situation applies if the energy is concentrated in finite spectral peaks (as in the present case) rather than impulses. From Appendix 2 or Chapter 4, the output spectral density of an error-sampled system can be written in the form

$$S_{cc}(\omega) = \left| \frac{G(j\omega)}{1 + G^{*}(j\omega)} \right| S_{rr}^{*}(\omega)$$

where $G(j\omega)$ is the frequency function of the open loop system, and $S_{rr}^{*}(\omega)$ is given by (6.3). Consequently, the system output will contain frequencies beyond the bandwidth of the input.

The human operator's output is known to contain all frequencies. When quasi-linear models are used, these additional frequencies are included as "noise," and reports such as Elkind's (1956) give extensive plots of the spectral density of the additive noise.

In the present experiment the analog method of spectral analysis provided no information about the spectrum except at the 10 input frequencies. However, as noted previously, a number of runs were recorded on magnetic tape and analyzed by a digital computer method due to Blackman and Tukey (1958) which is described in Appendix 5.

The digitally computed output spectral density for subject KZ (run R-1, Case 3) is shown in Figure 6.5 (a). Figure 6.5 (b) shows the corresponding output spectral density for the MFOH model, with model parameters adjusted to give good visual agreement during the run. (Final model parameter values were not available during the experiment and could not be used to obtain the data.) While the fit

of the peaks could have been improved somewhat by using the procedure described later (See Figure 6.9 (a)), the spectra are included here in order to compare amplitudes between and beyond the 10 input frequencies. The following conclusions can be made from an inspection of this figure:

(1) The magnitude of the power spectrum obtained from the model between the 10 input frequencies is comparable to the experimental data.

(2) The magnitude of the power spectrum obtained from the model beyond the 10 input frequencies of the same order of magnitude as the "noise" generated by the man, but attenuates somewhat more rapidly with frequency.

(3) Both the man and model spectra beyond the 10 input sinewaves are characterized by a series of peaks, which would be expected from sampling.

These conclusions are of crucial importance for the sampling hypothesis. Ward (1958) based much of his work on a qualitative inspection of the "noise" in his sampling model, but he made no attempt to measure the power spectra. The importance of these conclusions stems from the fact that the output of any continuous linear model will contain only the input frequencies. Other frequencies can be generated only by sampling or by nonlinear behavior. If a noise generator is added to the linear, continuous model, the spectral peaks still remain unexplained.

The ouput "noise" spectra for runs R-3 through R-7 are considerably larger in magnitude than the one shown in Figure 6.5 (a), because the input bandwidth is higher and the operators behavior is more erratic. The "noise" generated by the model in these runs is of the same order of magnitude as that of Figure 6.5 (b), and therefore represents a smaller proportion of the total power than for R-1. (The spectra for Runs R-1 through R-7 are given in Appendix 5.)

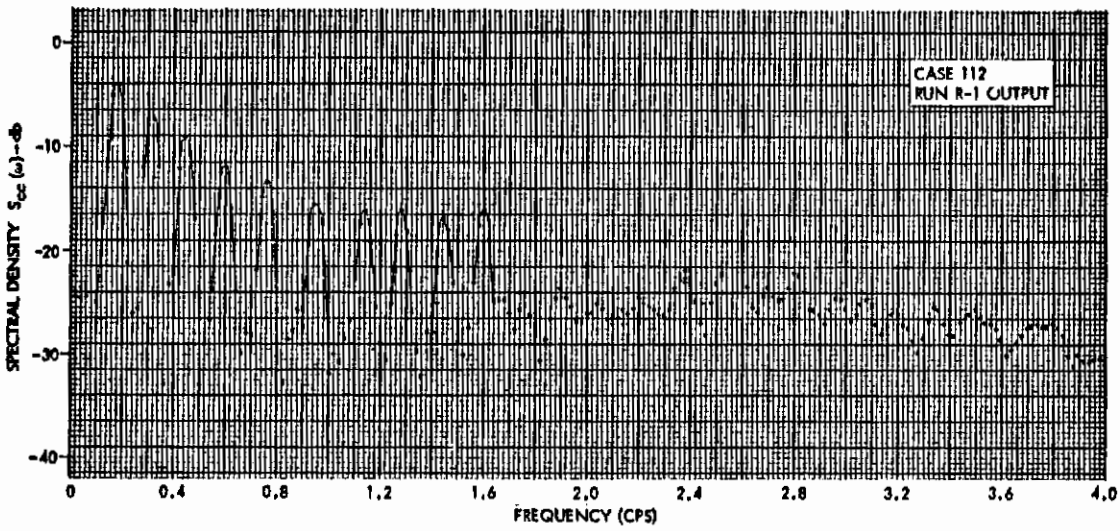


Figure 6.5 (a) Output Spectral Density from Run R-1

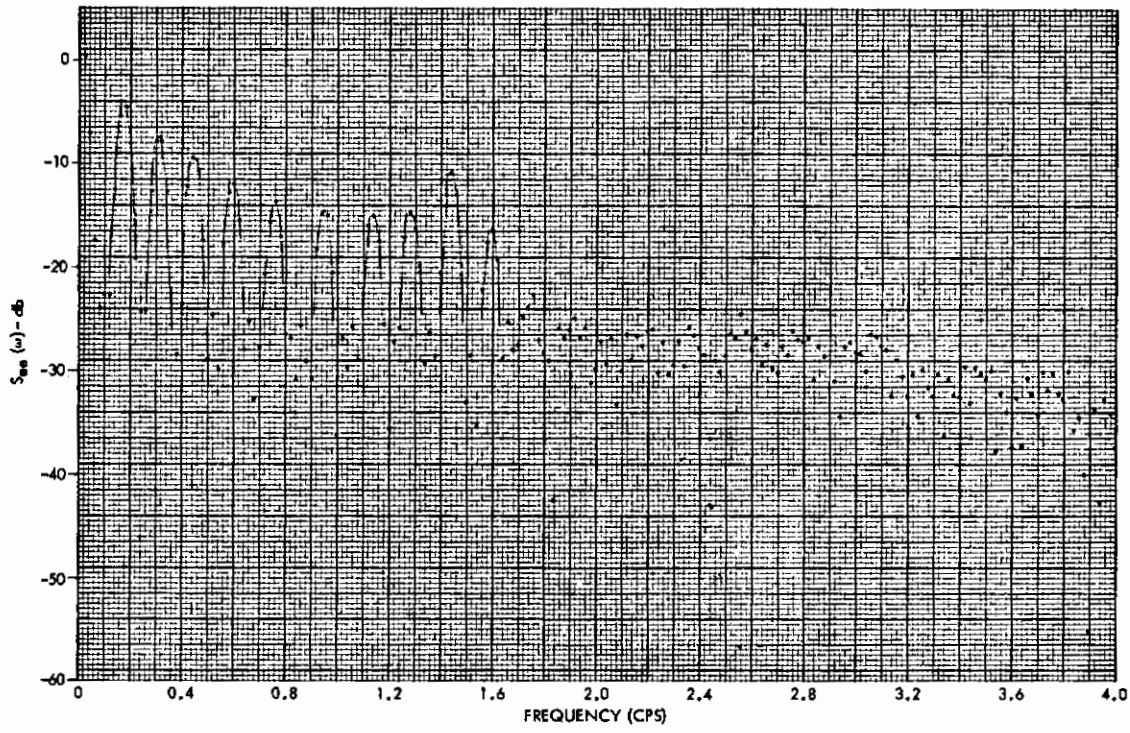


Figure 6.5 (b) Output Spectral Density of Approximate MFOH Model for Run R-1

Thus, the effect of the sampler in the model is to generate spectral peaks beyond the input bandwidth, but these peaks do not account for all the "noise" power evident in Figure 6.5 for example. Consequently, the human operators' behavior cannot be completely specified with continuous or sampled linear models, but requires other phenomena for its explanation. The sampled model does, however, "explain" a part of the spectrum beyond the range of frequencies present in the input. The continuous linear model, of course, predicts zero energy at these frequencies.

6.5 Comparison with Predicted Error Spectral Densities

In this section we shall consider an essential part of the "proof of the pudding" for the sampling hypothesis, as presented in this report. Based on the analysis of Chapter 4, we shall compute the error and output spectral densities of the human operator models considered, both continuous and sampled, and compare them with experimental results.

The evaluation will be applied to the 7 runs for two subjects, which were recorded on magnetic tape and thus accessible to the spectral analysis techniques required for construction of the mathematical models. The seven runs to be analyzed are listed in Table 6.4

Table 6.4. Experimental Data Selected for Comparison with Analysis.

Run No.	Subject	Case No.	No. of Input Sine Waves	Break Freq. of Input Filter	Type of Display	Sampling Freq.
R-1	KZ	3	10	1.5	Cont.	-
R-2	KZ	5	10	1.5	Samp.	3/sec
R-3	KZ	4	10	3.0	Cont.	-
R-4	KZ	6	10	3.0	Samp.	3/sec
R-5	KZ	6	10	3.0	Samp.	6/sec
R-6	GW	4	10	3.0	Cont.	-
R-7	GW	6	10	3.0	Samp.	3/sec

6.5.1 Construction of continuous quasi-linear model. The construction of the continuous model is based on a method due to Booton (1953) presented in earlier chapters. It will be summarized here only briefly. If in the block diagram of Figure 6.6 the input r (t) is a stationary Gaussian process, then the best linear fit (in the mean-square sense) to the characteristics of the unknown system is obtained from the relation

$$G(j\omega) = \frac{S_{rc}(j\omega)}{S_{re}(j\omega)} \tag{6.5}$$

where $S_{rc}(j\omega)$ is the cross-spectral density from input to output and $S_{re}(j\omega)$ is the cross-power spectral density from input to error. Since both $S_{rc}(j\omega)$ and $S_{re}(j\omega)$ are obtained from experimental data, their ratio is the ratio of two complex numbers for each frequency considered. These plots of amplitude and phase versus frequency are fitted with continuous models of the form

$$G_c(j\omega) = \frac{K e^{-jD_c \omega}}{1 + j\tau_c \omega} \tag{6.6}$$

since it has been shown previously by McRuer and Krendel (1957) that (6.6) is the best linear fit possible to the data for systems with no controlled element dynamics.

The procedure outlined in the preceding paragraph was followed for the six runs listed in Table 6.4. The fits to the plotted data were made "by eye" and checked independently by another engineer. The fitted frequency functions and the original data are given in Appendix 8. From an inspection of the fitted curves it can be seen that they agree with the experimental data within approximately 0.5 db in amplitude and 10° in phase, except at the 2 or 3 highest frequency points (where the disagreement in amplitude may be larger) and the lowest frequency point (where the phase error may be larger).

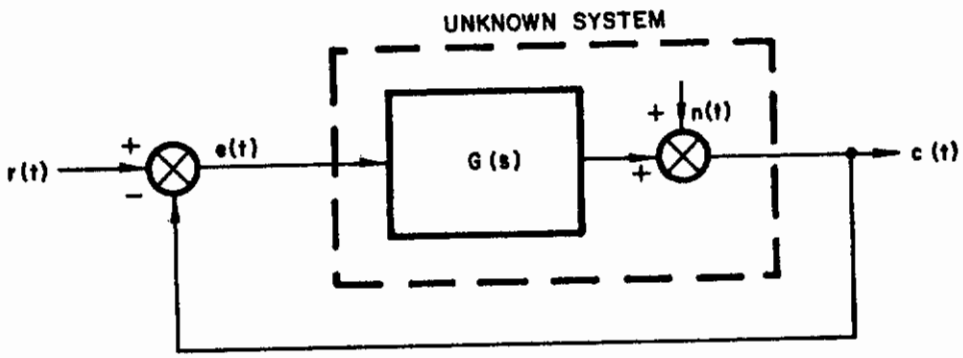


Figure 6.6 Block Diagram Representation of Unknown System as a Conbination of a Linear Element and a Noise Source

The selection of an analytic function of frequency which will provide a "best fit" in some sense to a set of points in the frequency domain is usually called the approximation problem of network synthesis, and is discussed, for example, by Guillemin (1957), (Chapter 14) and Balabanian (1958) (Chapter 9). However, the approximation problem, when formulated analytically, usually leads to rational fractions in the frequency domain, thus excluding the pure delay $e^{-jD\omega}$ Least-squares criteria for approximation can be used and instrumented on a digital computer. For the purposes of this report, a visual approximation was considered adequate.

It has been stated above that the input signal r (t) must have a Gaussian probability density function. The signal used in the present experiment was approximately Gaussian. Its statistical properties are indicated in Figure 5.6 which shows the probability density function of r (t).

6.5.2 Construction of the sampled models. The construction of the sampled-models has been discussed in Chapter 3. Basically, the method involves the addition of sample and hold circuits to the continuous models of the previous section, bearing in mind that "hold" circuits exhibit phase shift and consequently introduce an effective "time delay" into the model. Consequently, the synthesis of the sampled model consists of three steps:

(1) Select the parameters K, $a = 1/\tau_c$, and D_c from the continuous models.

(2) Select the sampling frequency f_s by observing the location of "sampling peaks" in the operator's error spectrum, and letting the frequency location of the peak, $f_p = f_s/2$.

(3) Select the time delay for the sampled model by subtracting the effective delay of the hold (D_H) from the delay obtained in the continuous model, i.e., let

(6.7) $$D_s = D_c - D_H$$

As we have mentioned in Chapter 3, the first-order and modified first order held circuits exhibit different amounts of phase shift, consequently the values D_{s1} and D_{sM1} will be different. The effective delay due to the first-order hold circuit has been found to be

$$D_{H1} \cong \frac{T}{2}; \qquad T = \frac{1}{f_s} \tag{6.8}$$

and that for the modified first order hold:

$$D_{HM1} \cong 0 \tag{6.9}$$

Based on these considerations, the parameters of the various models take on the values given in Table 6.5 for the seven runs described in Table 6.4.

Table 6.5. Parameters for Continuous and Sampled Models.

Run No.	Continuous Model Parameters			Additional Sampled Model Parameters		
	K	a(sec^{-1})	D_c(sec)	T = $1/f_s$ (sec)	D_{s1}(sec)	D_{sM1}(sec)
R1	2.30	1.84	0.102	0.33	-0.06	0.102
R2	2.24	1.25	0.114	0.33	-0.05	0.114
R3	1.52	2.51	0.092	0.33	-0.08	0.09
R4	0.85	3.76	0.142	0.38	-0.05	0.14
				0.29*	-0.01	0.14
R5	1.00	1.84	0.133	0.38	-0.05	0.14
				0.29*	-0.01	0.14
R6	2.20	3.76	0.107	0.27	0.01	0.11
R7	1.26	4.40	0.131	0.33	-0.04	0.13
					+0.02*	

*Alternate values also plotted in the Figures.

6.5.3 Computation of spectral density of models. The relations developed in Chapter 4 for the computation of power spectral density were based on continuous input spectra. That is, the input spectrum $S_{rr}(j\omega)$ was assumed to be produced by the action of a shaping filter on white Gaussian noise. In the experimental situation reported here the input spectrum was by no means "white" and thus we must examine whether the relations of Chapter 4 are valid.

In the case of the continuous model, the equations of Chapter 4 apply directly, if we are careful about the definition of the input spectrum. In Appendix 3 it is shown that the power spectral density of a sum of 10 sine waves:

$$r(t) = \sum_{i=1}^{10} c_i \cos(\omega_i t - \theta_n) \tag{6.10}$$

measured for a finite length of run T_R can be expressed (under certain conditions) as

$$S_{rr_T}(\omega) = \frac{T}{2\pi} \sum_{i=1}^{10} \frac{c_i^2 \sin^2(\omega_i - \omega) T}{\left[(\omega_i - \omega) T\right]^2}, (T_R = 2T) \tag{6.11}$$

Expression (6.11) shows that the finite-length spectrum contains power at all frequencies. At the input frequencies only, the spectral estimate is obtained by taking the limit of (6.11) as $\omega \to \omega_i$, to result in

$$S_{rr_T}(\omega_i) = \frac{T}{2\pi} c_i^2 \tag{6.12}$$

for each frequency. At each of the input frequencies one can obtain the output spectral density from

$$(6.13) \qquad S_{cc}(\omega_i) = \left| \frac{G(j\omega_i)}{1 + G(j\omega_i)} \right|^2 S_{rr}(\omega_i)$$

in an analogous manner to the continuous spectrum case.

The problem is more complicated in the case of the sampled models because of the effects of sampling. In equations (4.58) and (4.64) for example, which give the output and error spectral density of the first-order hold sampled model, the sampled spectrum $S_{rr}^*(\omega)$ appears. This quantity has a very different meaning for the continuous than for the discrete spectrum. For the continuous spectrum case, we have shown in Appendix 2, the sampled spectrum can be obtained simply from Z-transform tables as

$$(6.14) \qquad S_{rr}^*(\omega) = \frac{1}{T} Z\left\{ S_{rr}(s) \right\}_{z = e^{j\omega T}}$$

$$= \frac{1}{T^2} \sum_{n=-\infty}^{+\infty} S_{rr}(\omega + n\omega_s)$$

Thus, for a continuous spectrum which is not band-limited, the resulting sampled spectrum is periodic and includes the effects of "folding" of the higher harmonics. This process is illustrated in Figure 6.7.

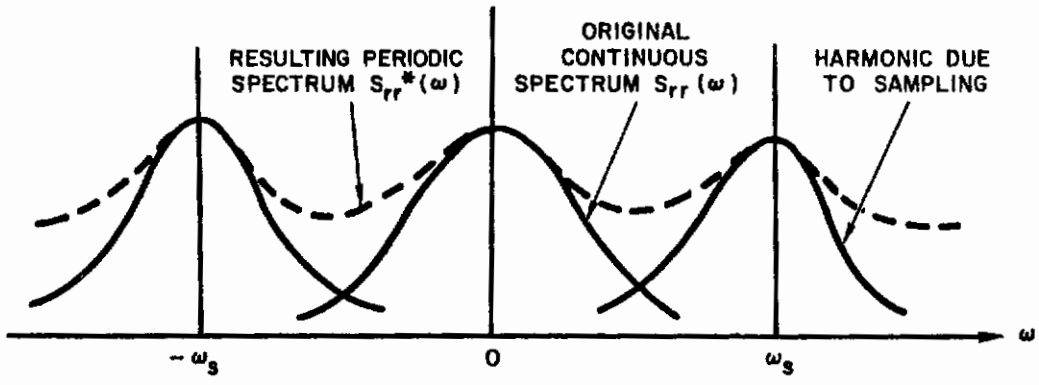


Figure 6.7 Sampled Power Spectral Density

When the spectrum consists of narrow peaks in the frequency domain (neglecting the side-lobes due to equation (6.11) for the moment), the folding may introduce energy where none was present before sampling, or it may add to the peaks if, for example, $\omega_j = \omega_s = \omega_i$. If, however, the input spectrum does not exceed $\omega_s/2$, no folding takes place, and we can write

$$S_{rr}^{*}(\omega) = \frac{1}{T^2} \sum_{n=-\infty}^{+\infty} S_{rr}(\omega + n\,\omega_s) \tag{6.15}$$

and the resulting situation is approximately as depicted in Figure 6.8.

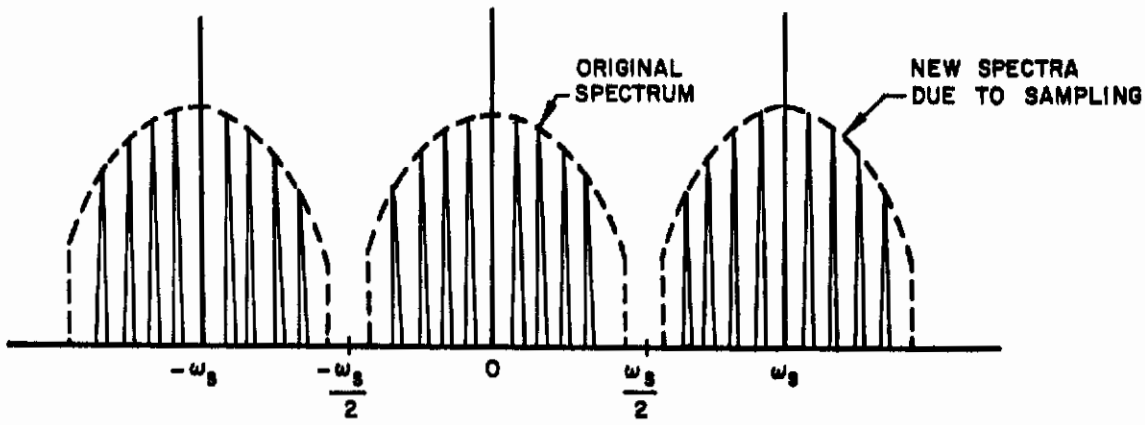


Figure 6.8 Sampled Spectrum Due to Sum of Sinusoids with No Folding

The location of the peaks in the measured spectra of Appendix 6 makes it reasonable to assume that $f_s \cong 3$/sec and therefore that most probably some folding does take place (since our highest input sinusoid is at 1.6 cps). Furthermore, since the operator certainly does generate output signals at frequencies both between and beyond the 10 input sine waves, it is important to at least attempt to take them into account. Since the input spectrum is given by (6.11), it is possible to substitute this relation into (6.15) and solve for the spectra by taking one term in the series at a time. To avoid this extremely time-consuming procedure, two alternate methods were

employed: (1) the input signal was considered to be filtered "white" noise, or (2) the input signal was assumed to consist only of spectral "lines" at the 10 input frequencies and folding neglected completely. It is clear that neither assumption is correct: (1) will result in excessive energy between and beyond the 10 input "peaks" in the spectrum, while (2) will give no energy there. It is therefore perhaps surprising that remarkably good results have been obtained with the use of assumption (1).

With the above interpretation in mind, digital computer programs were prepared for the solution of equations (4.58) and (4.64) for the output and error spectra of the first-order hold model, and equations (4.81) and (4.85) for the "modified" first-order hold model spectra. Two versions of the programs made possible the use of continuous or discrete input spectra. An additional program was based on equations (4.13) and (4.15) for the continuous model. The values of parameters from Table 6.5 were substituted into these programs. The resulting plots of computed versus measured spectral density are given in Figures 6.9 through 6.15 for the seven runs under consideration.

6.5.4 Evaluation of computed versus measured spectra. An examination of the curves in Figures 6.9 to 6.15 leads to the following conclusions:

(1) The "fit" of the model is much better for output spectra than for error spectra. In general, the error spectra, while having the right shape, attenuate more rapidly than the experimental points. (All the computed spectra are indicated by dotted lines, or solid lines for continuous models to show the behavior of the equations involved. Of course, the experimental data are valid only at the points indicated.)

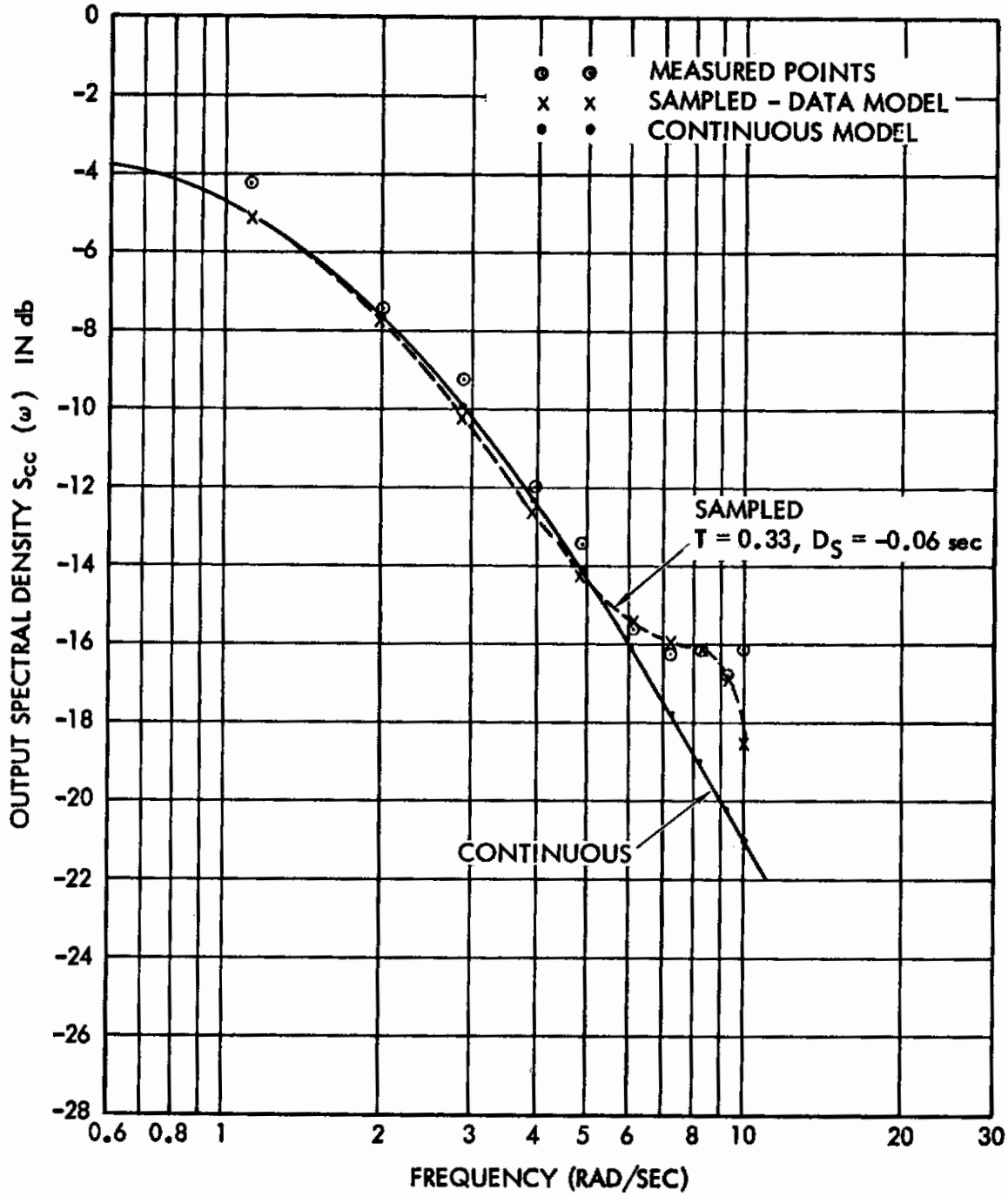


Figure 6.9 (a) Output Power Spectral Density (Run R-1) Comparison of Experimental and Analytical Data

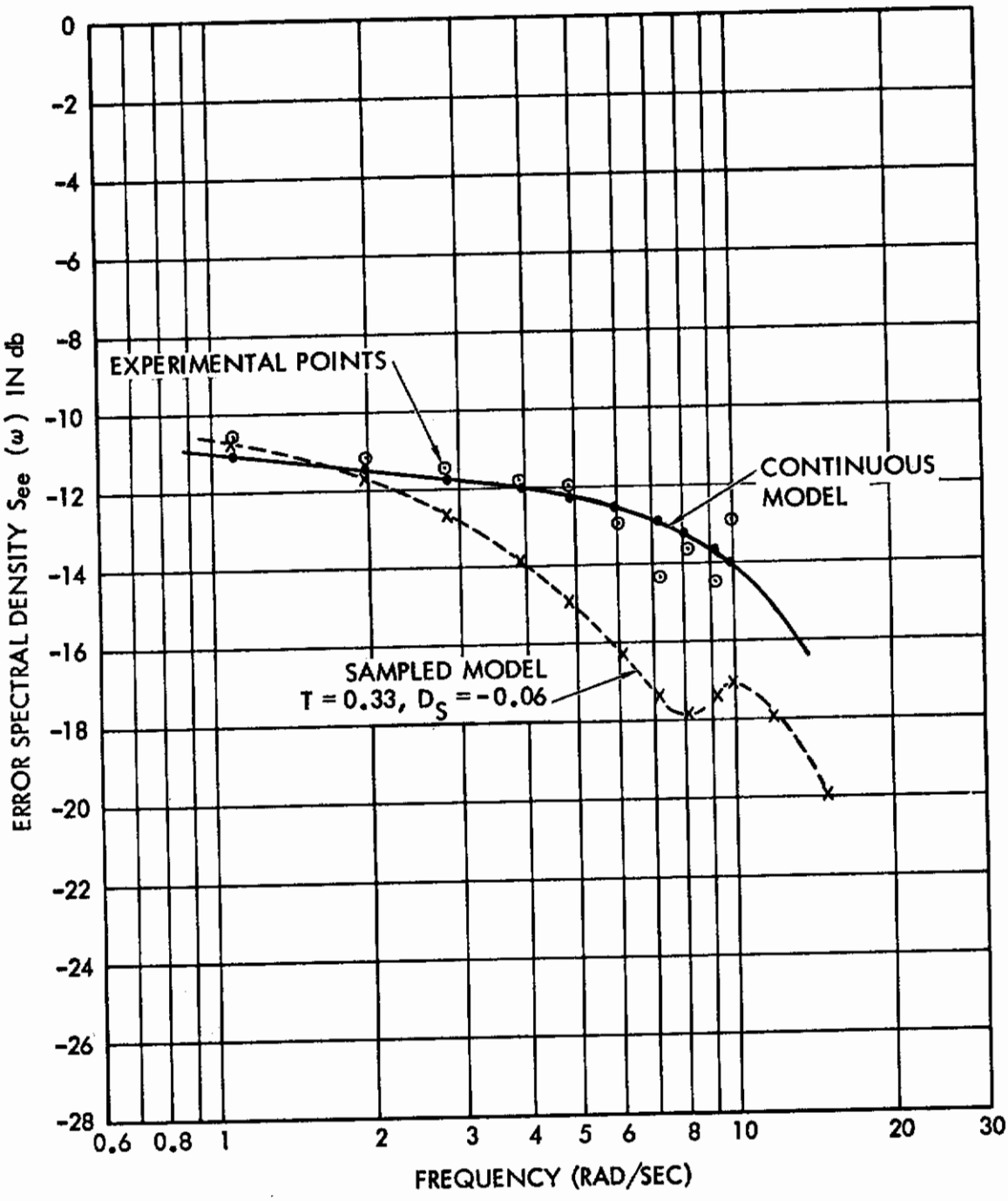


Figure 6.9 (b) Error Power Spectral Density for Run R-1 Comparison of Experimental and Analytical Values

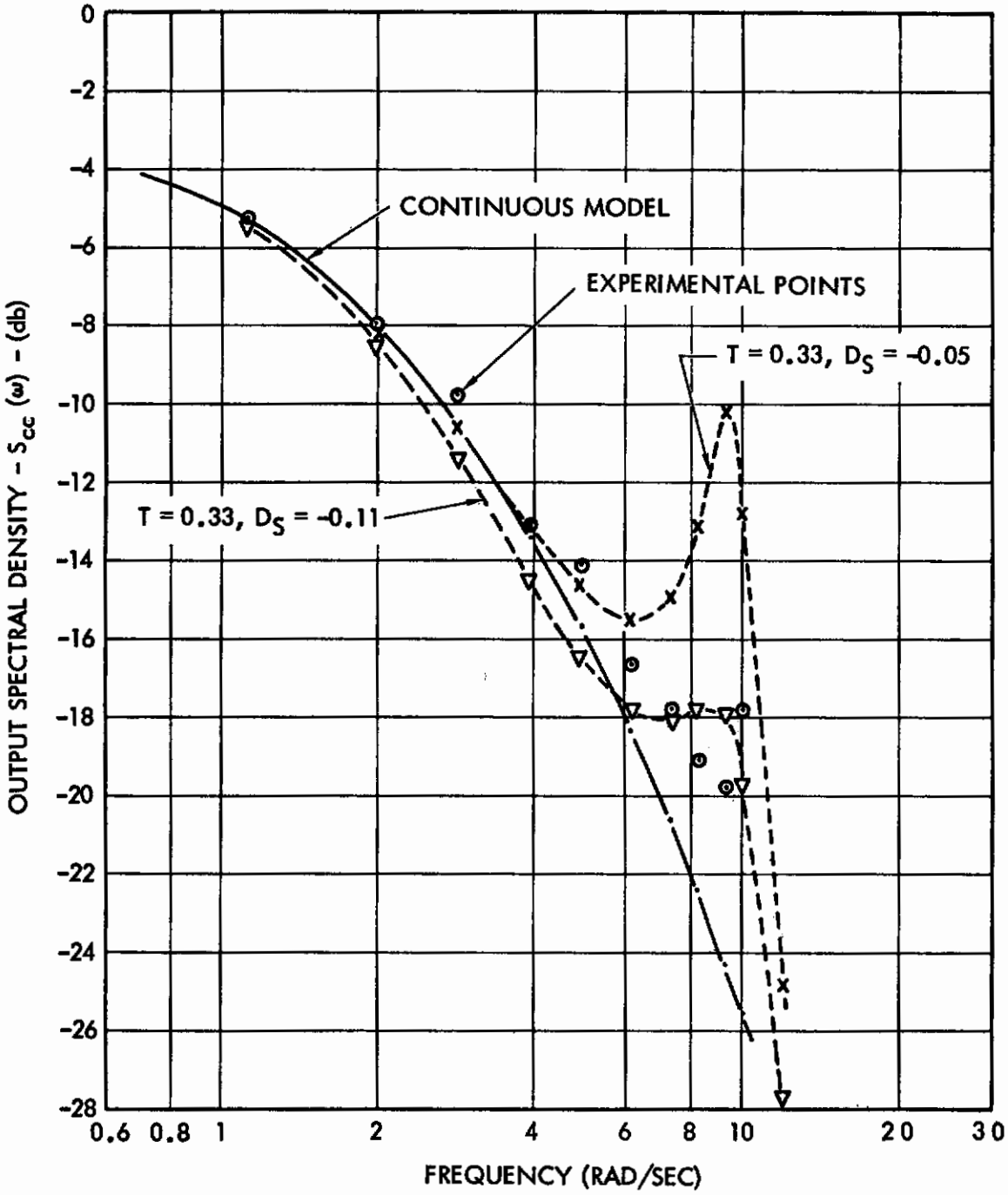


Figure 6.10 (a) Comparison of Experimental and Analytical Values of Output Spectral Density for Run R-2

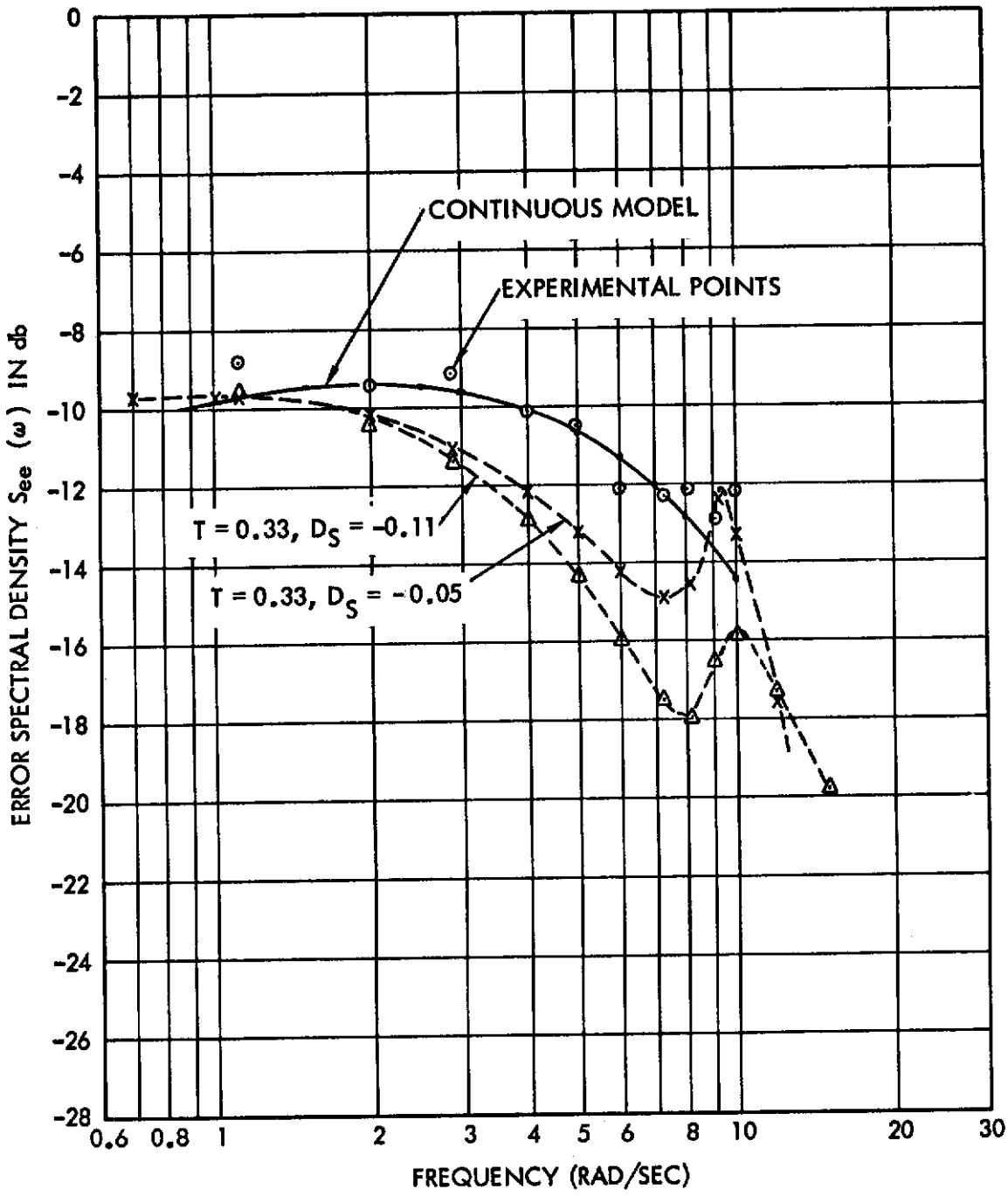


Figure 6.10 (b) Comparison of Experimental and Analytical Values of Error Spectral Density for Run R-2

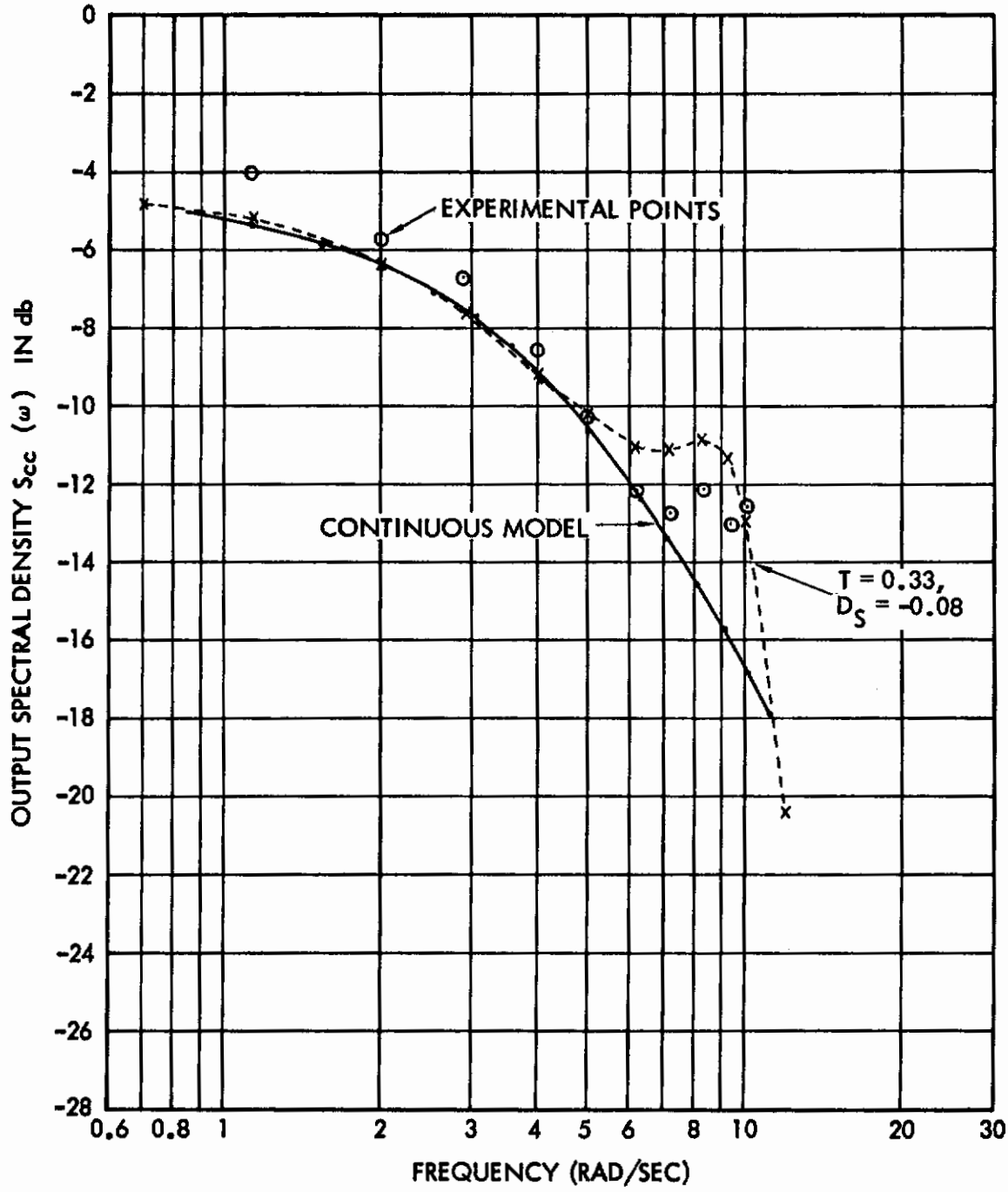


Figure 6.11 (a) Comparison of Experimental and Analytical Values of the Output Spectral Density for Run R-3

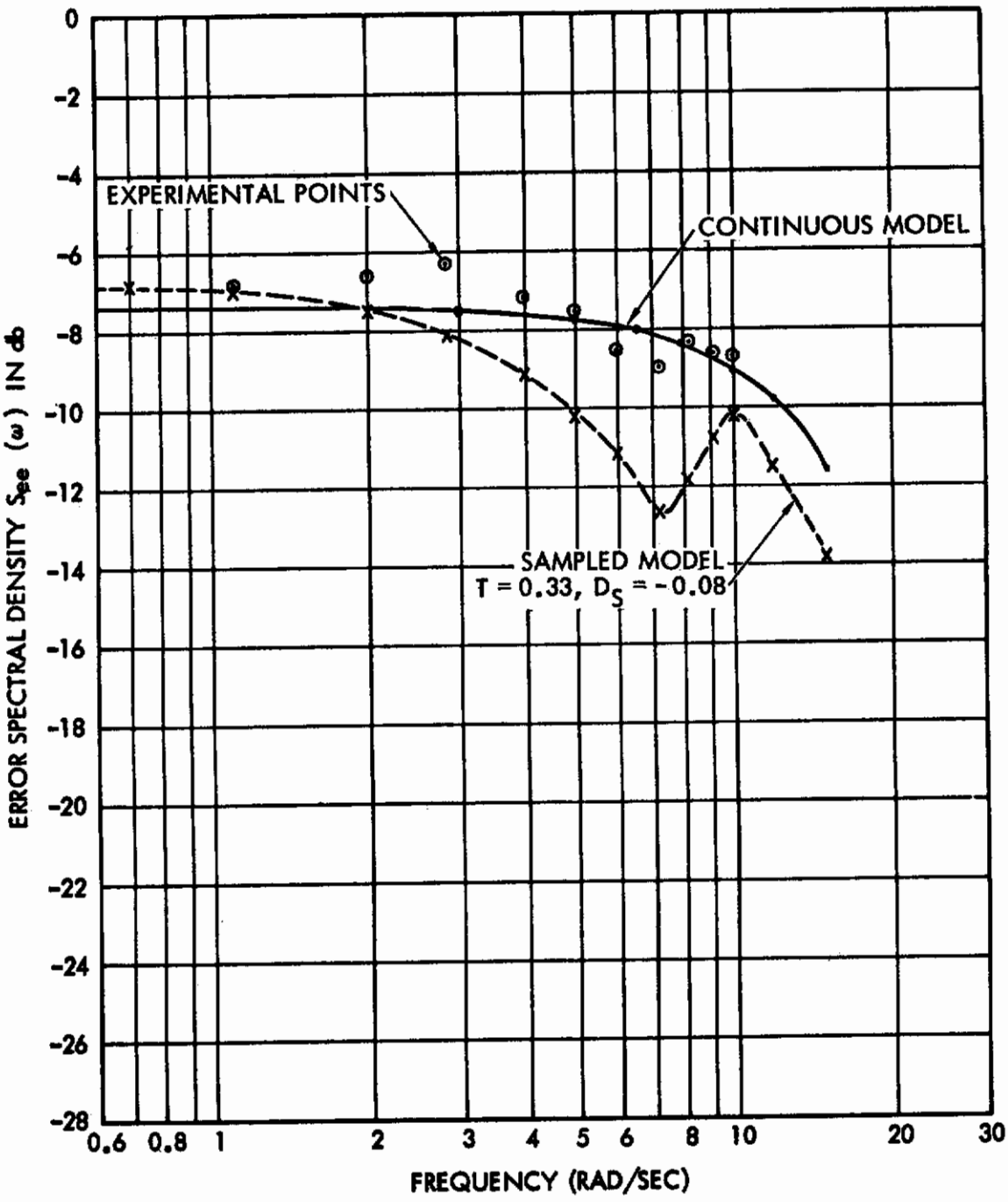


Figure 6.11 (b) Comparison of Experimental and Analytical Values of the Error Spectral Density for Run R-3

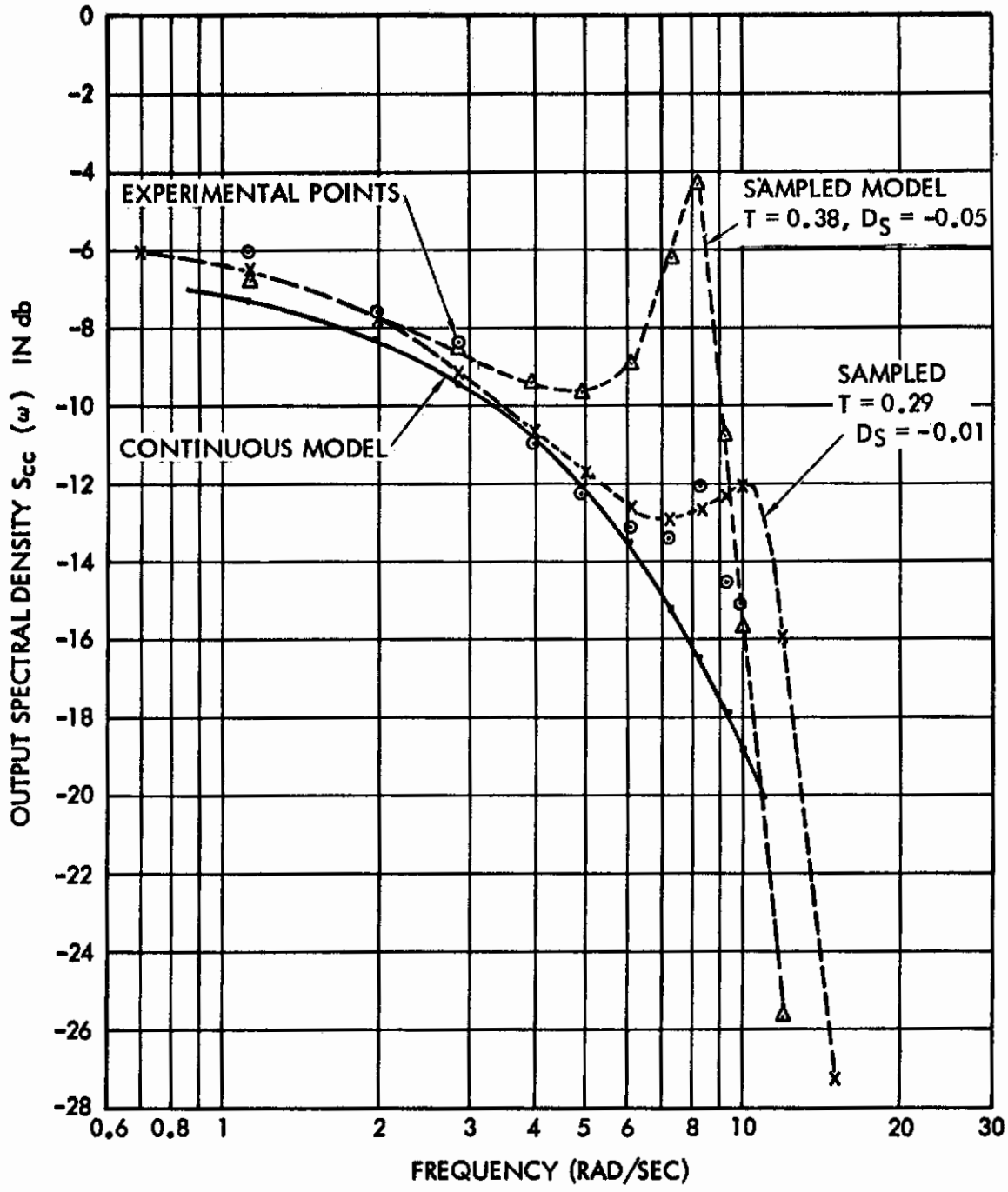


Figure 6.12 (a) Comparison of Experimental and Analytical Values of the Output Spectral Density for Run R-4

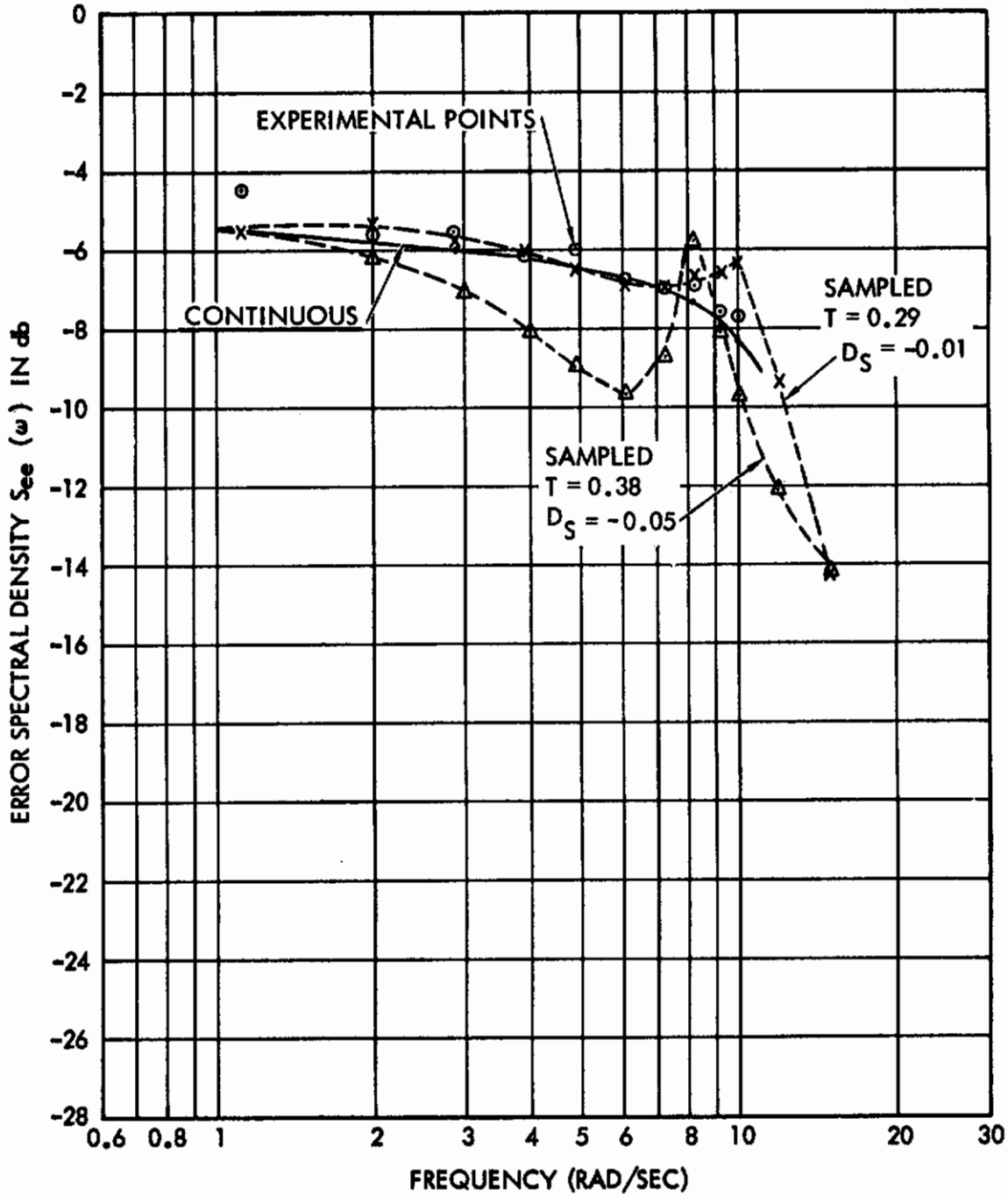


Figure 6.12 (b) Comparison of Experimental and Analytical Values of the Error Spectral Density for Run R-4

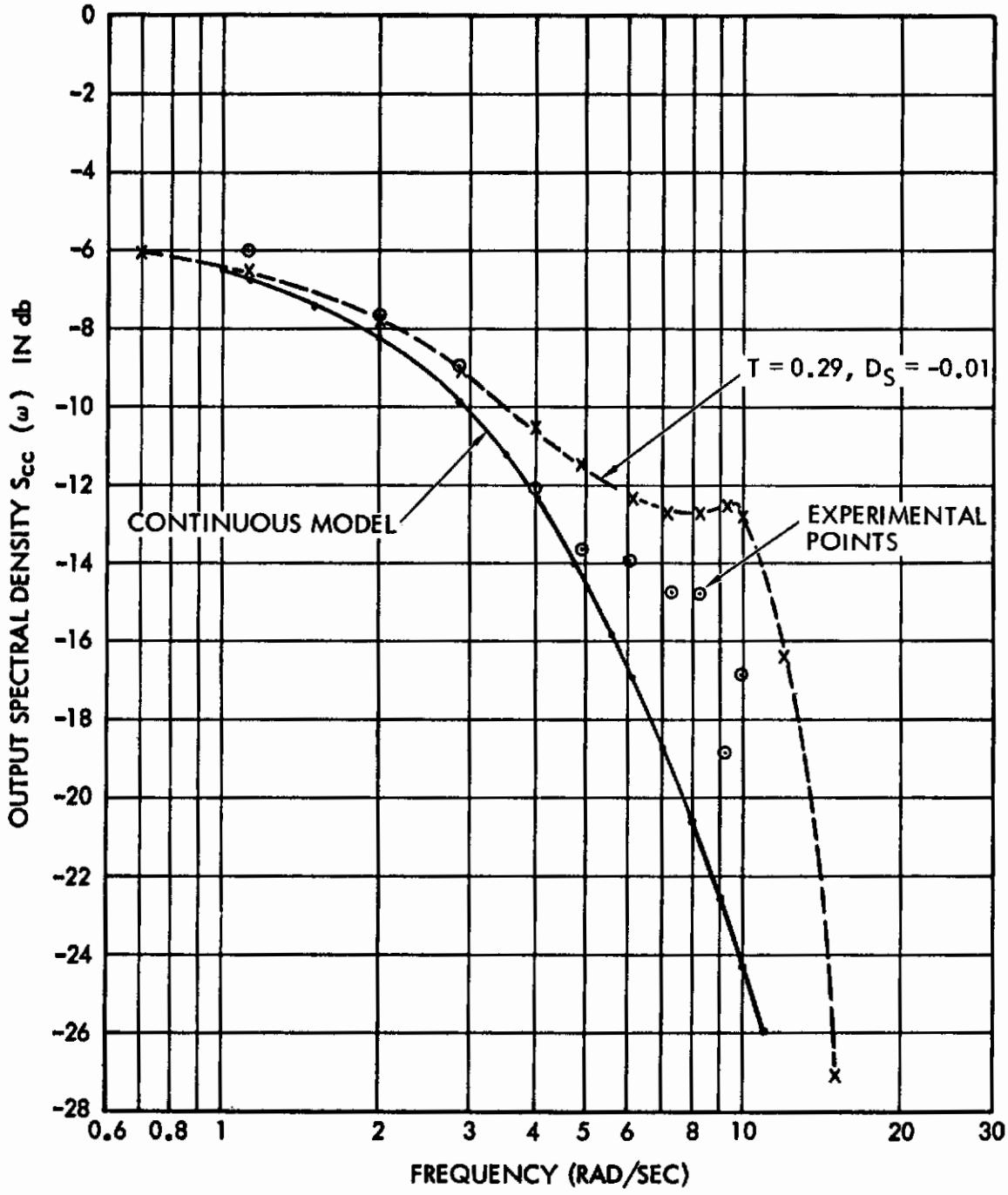


Figure 6.13 (a) Comparison of Experimental and Analytical Values of the Output Spectral Density for Run R-5

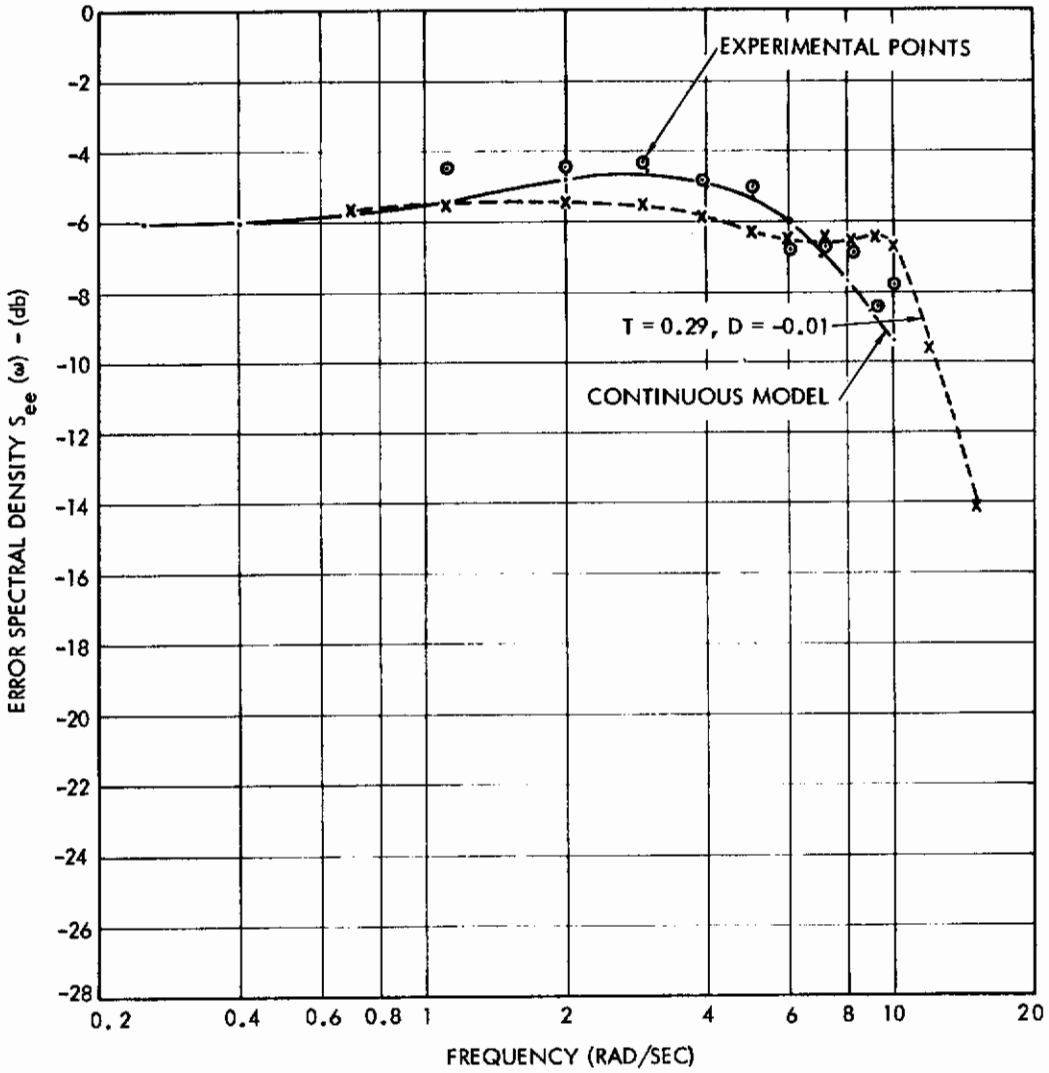


Figure 6.13 (b) Comparison of Experimental and Theoretical Values of the Error Spectral Density for Run R-5

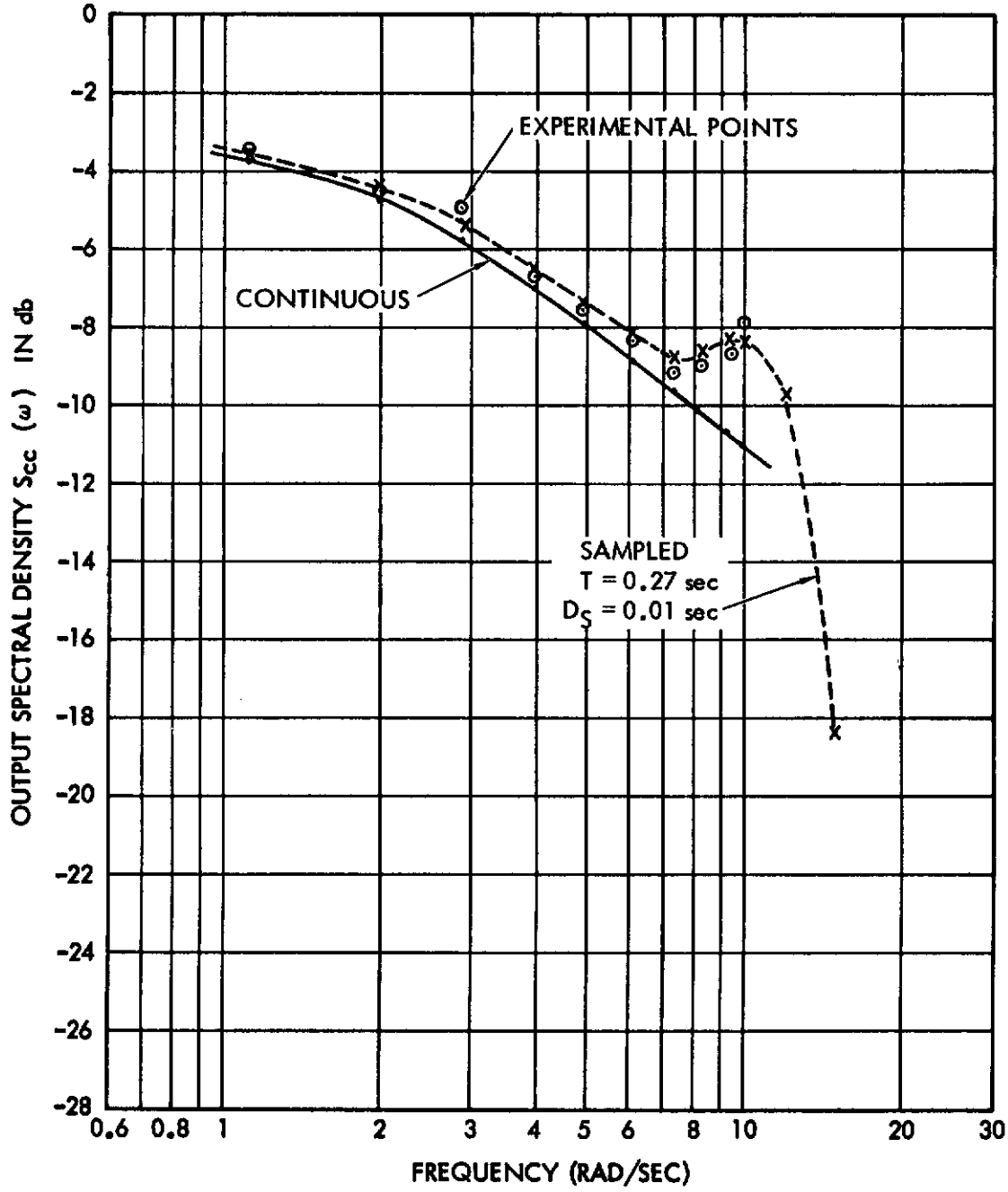


Figure 6.14 (a) Comparison of Experimental and Theoretical Values of the Output Spectral Density for Run R-6

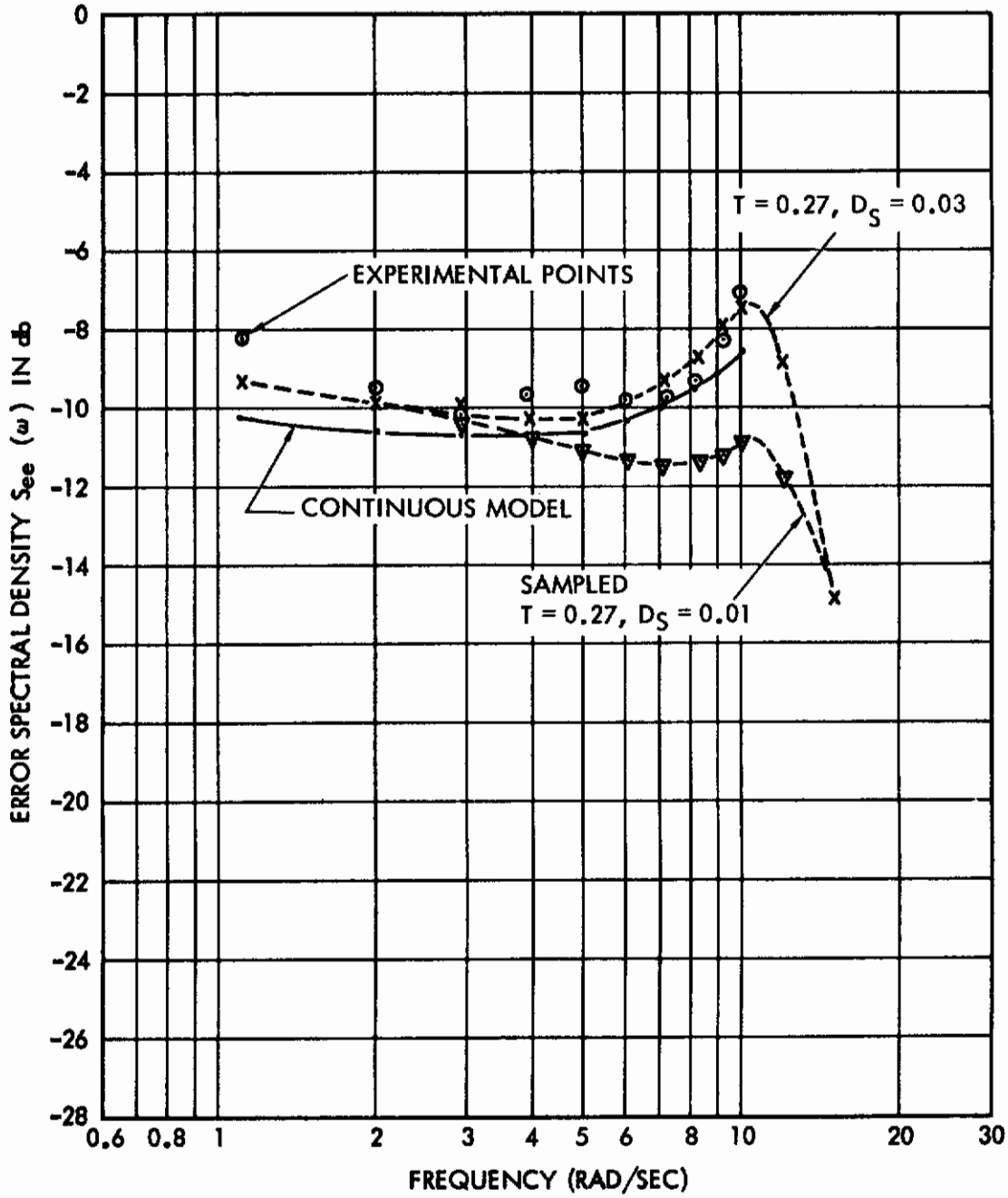


Figure 6.14 (b) Comparison of Experimental and Theoretical Values of the Error Spectral Density for Run R-6

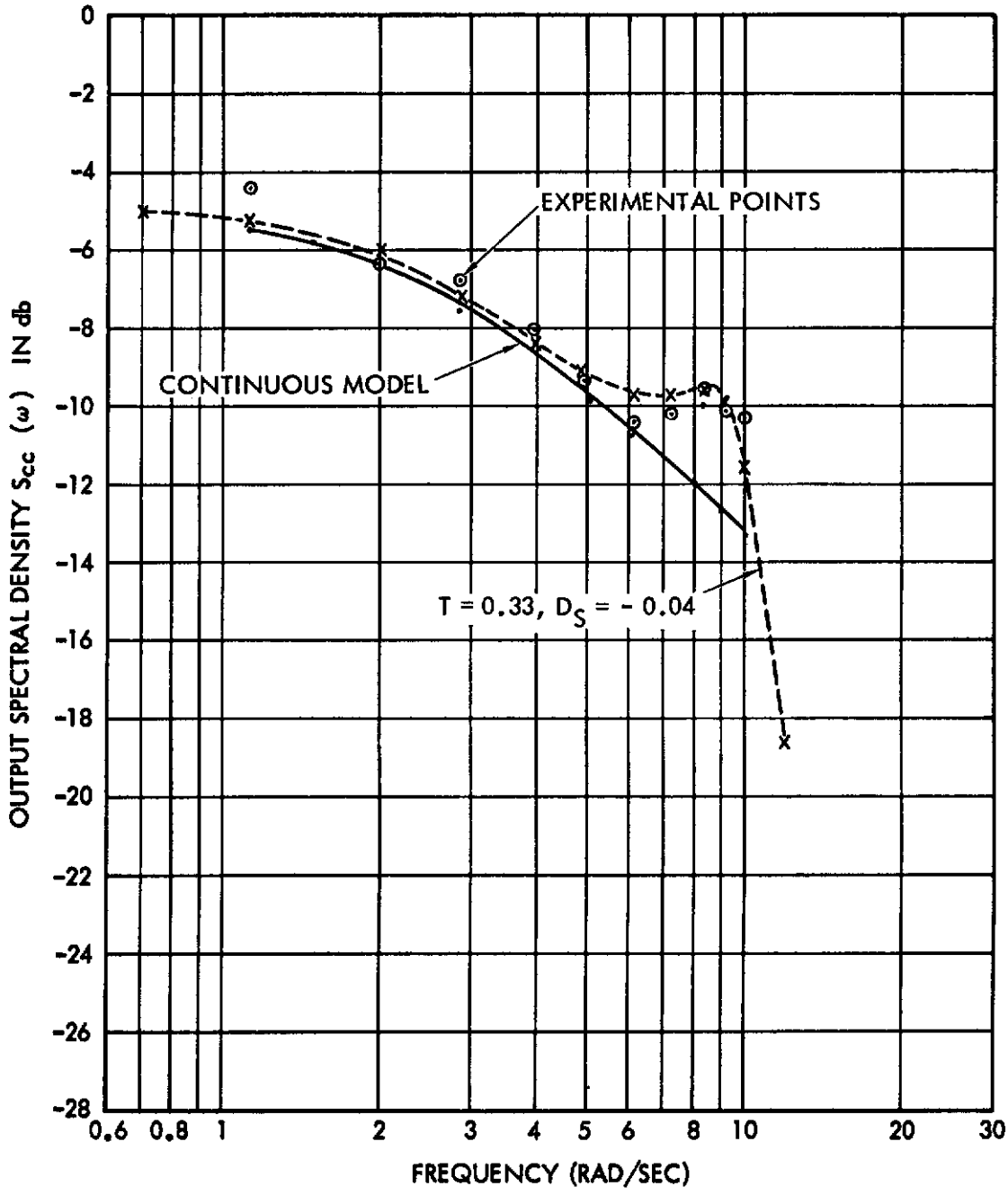


Figure 6.15 (a) Comparison of Experimental and Theoretical Values of the Output Spectral Density for Run R-7

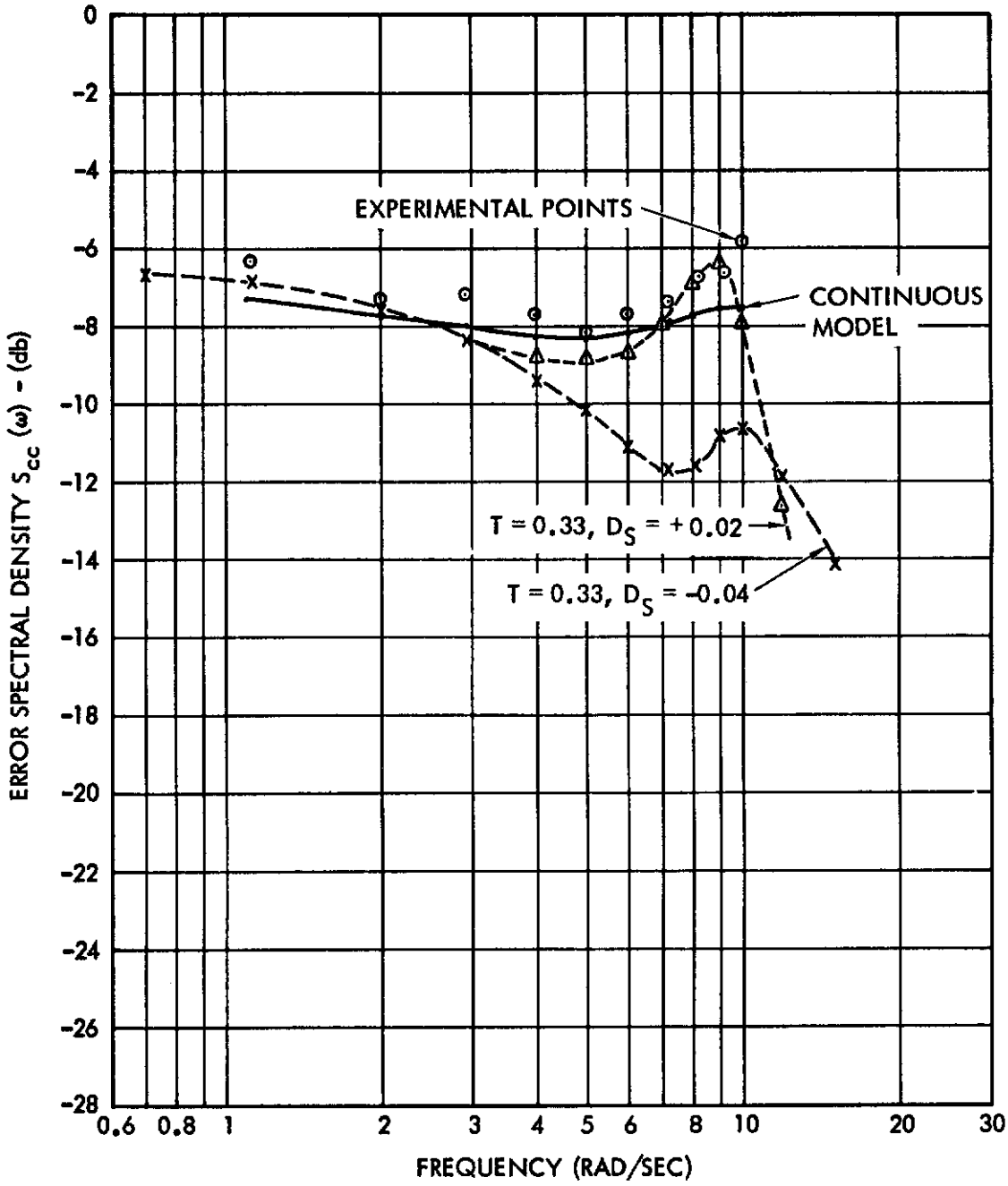


Figure 6.15 (b) Comparison of Experimental and Theoretical Values of the Error Spectral Density for Run R-7

(2) It is apparent that the runs recorded for analysis were selected in an unfortunate way. Only runs R-1 and R-2 made use of the "easier" input (filter break frequency at 1.5 rps) with the remainder being taken from cases 4 and 6. The selection of sampling frequency for the model is based on finding a clear "sampling peak" in the recorded spectra. In none of the runs is it really clear that the peak had been included in the ten frequencies under consideration. The two subjects (KZ and GW) were both excellent trackers. But since the available frequencies were limited, trackers such as JP or WB who exhibited peaks at lower frequencies may have given a clear indication of the sampling frequency to use in the sampled-data model. Consequently, the fits shown in some of figures are based on values of $f_s/2$ which are greater than 10 rad/sec, and thus beyond the range of the data.

(3) All the first-order hold sampled-data systems used for operator models included predictors, i.e., the value of time delay D_{si} became negative for the sampling frequencies in question. While such a result is not inconceivable from a theoretical point of view (i.e., the effective delay from the model is clearly positive, but an advance is required to correct for the excessive delay from the hold circuit), it is clearly inconvenient. For the digital computation program, the sign of the delay time is immaterial. For simulation purposes, however, it is clearly unrealizable.

(4) For the single low frequency run (R-1) the accurate selection of the sampling frequency for the model (i.e., location of the "sampling peak" in the

spectrum) is not too significant. Figure 6.16 shows the effect of changing the sampling period from 0.33 to 0.28 and 0.38 sec, while maintaining the correct overall time delay. It can be seen that even with this relatively large parameter change, the models behave in a very similar way. Again, more experimental points beyond 10 rad/sec would have been helpful in selecting the frequency.

(5) The use of discrete input spectra ("spectral lines") instead of continuous spectra for the input resulted in invariably worse fits to the data, generally by 3 to 6 db or more. A typical result is shown in Figure 6.17 for the error spectrum of Run R-1. Apparently the inclusion of additional energy between the input sinusoids did not hurt the model as badly as complete omission of it as a result of a complete omission of the "folded spectra" due to sampling.

(6) The match to the spectra obtained with sampled displays was variable. For Run R-2, where the best results should be expected since the input was easier, the value of time delay obtained by the method outlined previously resulted in an excessive peak in the computed output spectrum. Much better results are obtained when an additional 0.06 sec of prediction are included in the model, but the results are still not outstanding. It is possible that a higher sampling frequency would give a better fit, but it is difficult to justify such a model when the display was sampled at 3 cps. A similar situation applies to Run R-4, also with a sampled display at 3 cps, where a peak location would

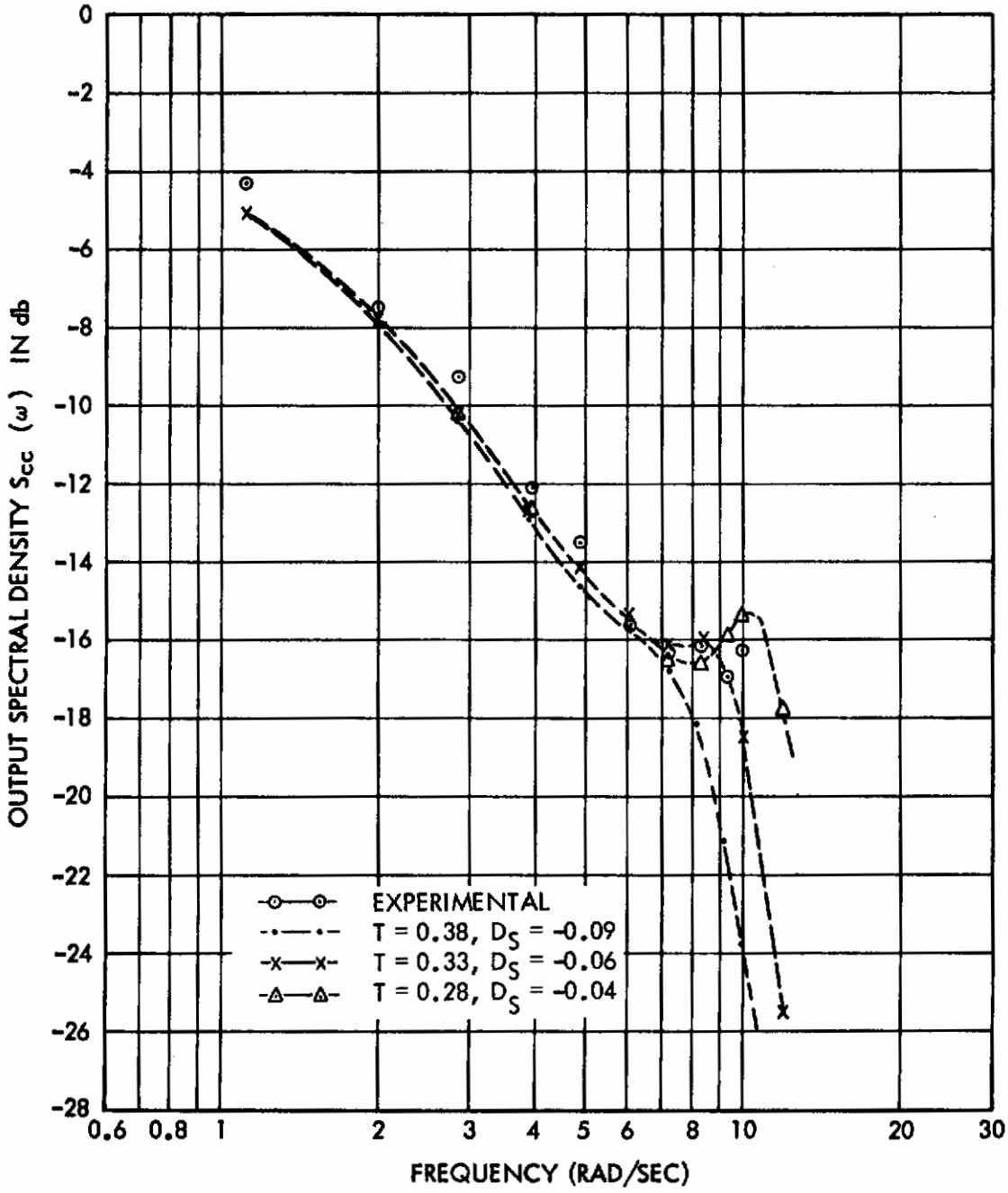


Figure 6.16 (a) Effect of Sampling Frequency on Output Spectra for Run R-1

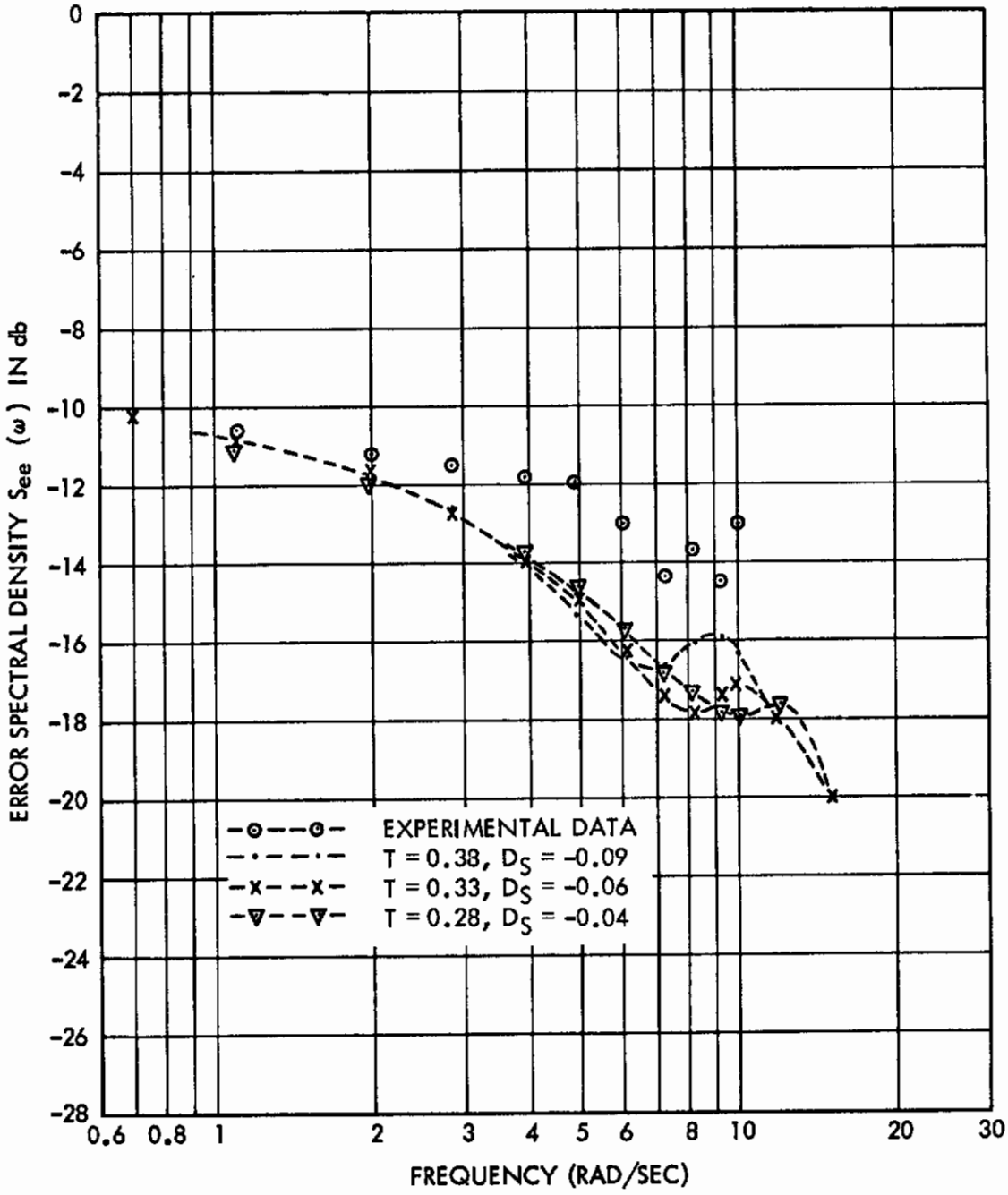


Figure 6.16 (b) Effect of Sampling Frequency on Error Spectra for Run R-1

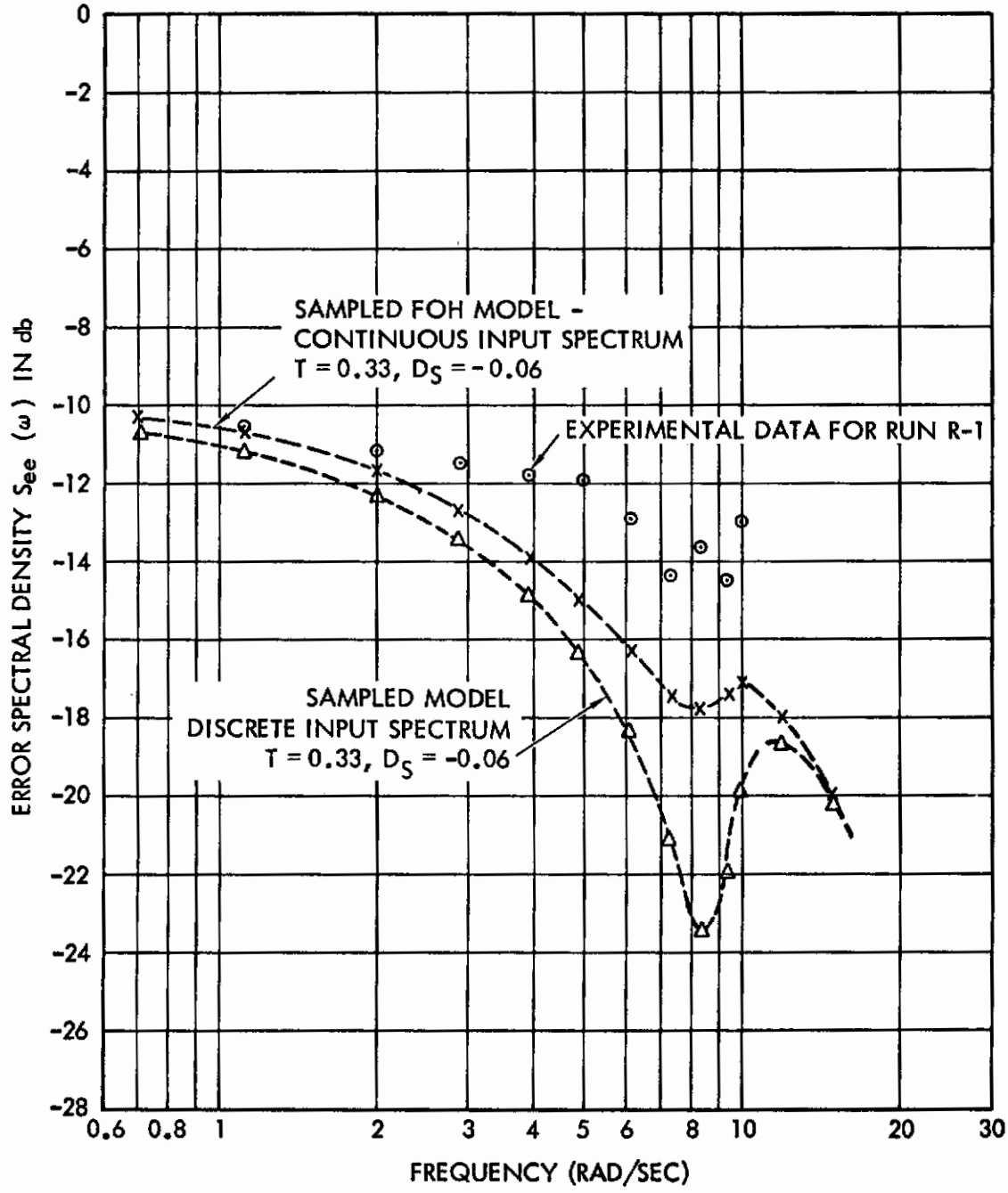


Figure 6.17 Error Spectral Density Pilots for Run R-1 Effect of Discrete Verus Continuous Input Spectrum

suggest T = 0.38 sec, but a much better fit is obtained with T = 0.29. On the other hand, Run R-7 results in an output spectrum which is fitted very well by the expected values of T and D.

In Run R-5 the display was sampled at 6 cps, and the subject found this run extremely difficult. The sampled model gives a reasonable fit with T = 0.29 sec.

(7) The MFOH models, in general, yielded fairly good fits at frequencies above 3 or 4 rad/sec, and exhibit too much gain at lower frequencies, as has been noted previously.

6.6 Time domain results.

In addition to the frequency domain results presented above as plots of power spectral density, time recordings were also made during most of the runs. The model parameters for the subjects were of course yet undetermined, since a spectral analysis was required to obtain them. Therefore, the values of a and D were set to the Elkind values mentioned previously, and the model gain K_s adjusted to obtain the best visual agreement. For most runs, the model sampling frequency was set to 3.0 cps. The modified hold was used in order to make time delay simulation convenient.

In general, in spite of the lack of detailed and accurate model information, the time traces of man and sampled model show remarkable resemblance. No attempt was made to obtain results for the continuous model, since the analog computer simulation of this model would require simulation of a continuous time delay of approximately 0.14 sec. The simulation of continuous systems with pure delay requires either expensive equipment (such as tape recorders with movable heads) or approximate methods such as those using Pade approximants to the operator e^{-Ds}.

Portions of tracking runs for various subjects are shown in Figures 6.18 to 6.20. In all cases it can be noted that:

(1) The magnitude of the error signal may exceed that of the input signal at high frequencies.

(2) Both man and model show a tendency to magnify input disturbances when these occur near the "sampling frequency" of the model.

6.7 Summary of results

The major results discussed in this chapter can be summarized as follows:

(1) The measured error spectral densities have characteristics which agree both with previous experimental data and with those of the sampled-data models proposed in this report.

(2) The distinguishing characteristics of the sampled models are: (a) the presence of a spectral peak in the range of 6 to 10 rad/sec, and (b) the presence of energy beyond the range of the input function bandwidth. Both of these characteristics agree with experimental data.

(3) The response of the subjects to intermittent displays provides considerable support to the hypothesis that the operator's behavior can indeed be represented by sampling operations followed by extrapolating circuits.

(4) Power spectral densities of operator output, computed from models synthesized by procedures outlined in this chapter, result in closer agreement with experimental data in the range of 6 to 10 rad/sec than those from conventional continuous operator models. The error spectra, while in general having a similar shape to measured data, result in a better fit to experimental data only in certain of the experiments, when computed according to the assumptions made in this chapter.

(5) Time responses of sampled-data models are very similar to those of human operators, even for the tracking signals used in this report, which represent a very difficult task.

Notes on Figures 6.18, 6.19, and 6.20
(on following pages)

For all three figures:

Channel 1: Input function $f(t)$
Channel 2: Man's output $c_h(t)$
Channel 3: Man's error $e_h(t)$
Channel 4: Error in MFOH model $e_m(t)$
Channel 5: Output of model $c_m(t)$
Channel 6: Output of hold circuit in model

All vertical scales are equal, showing the attenuation which takes place from input to output, and the magnification of high frequencies in the error signal.

Figure 6.18

Subject GW: Paper speed = 2 mm/sec
Input filter break frequency = 1.50 rad/sec (Case 3)
Model sampling frequency = 3/sec; K = 2

Figure 6.19

Subject WB: Paper speed = 5 mm/sec
Input filter break frequency = 3.00 rad/sec (Case 4)
Model sampling frequency = 3/sec; K = 3

Figure 6.20

Subject GW: Paper speed = 5 mm/sec
Input filter break frequency = 3.00 rad/sec (Case 4)
Model sampling frequency = 6/sec; K = 2

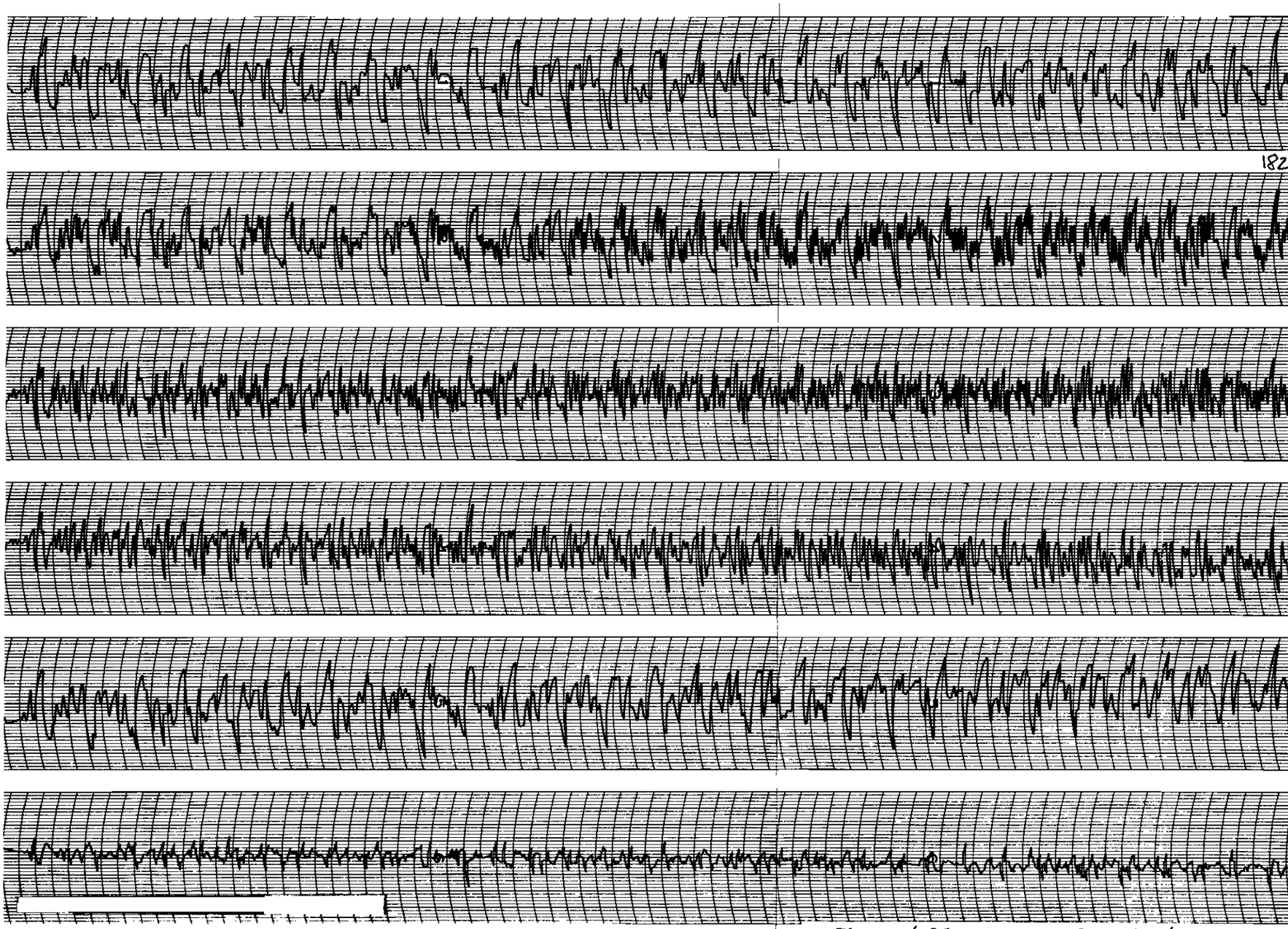


Figure 6.18 GW ω_B = 1.5 2mm/sec f_s = 3

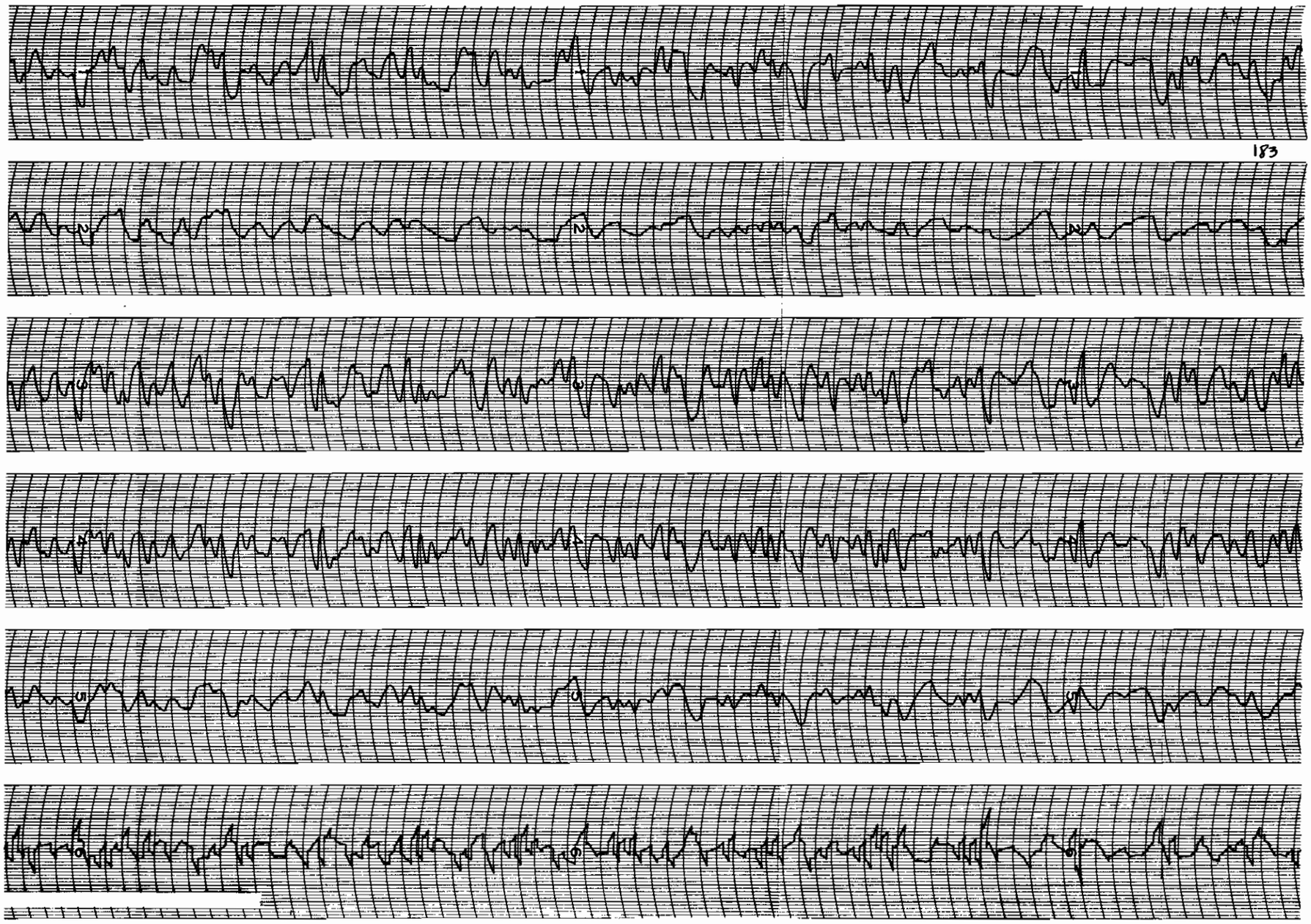


Figure 6.19 WB ω_B = 3.0 5mm/sec f_s = 3

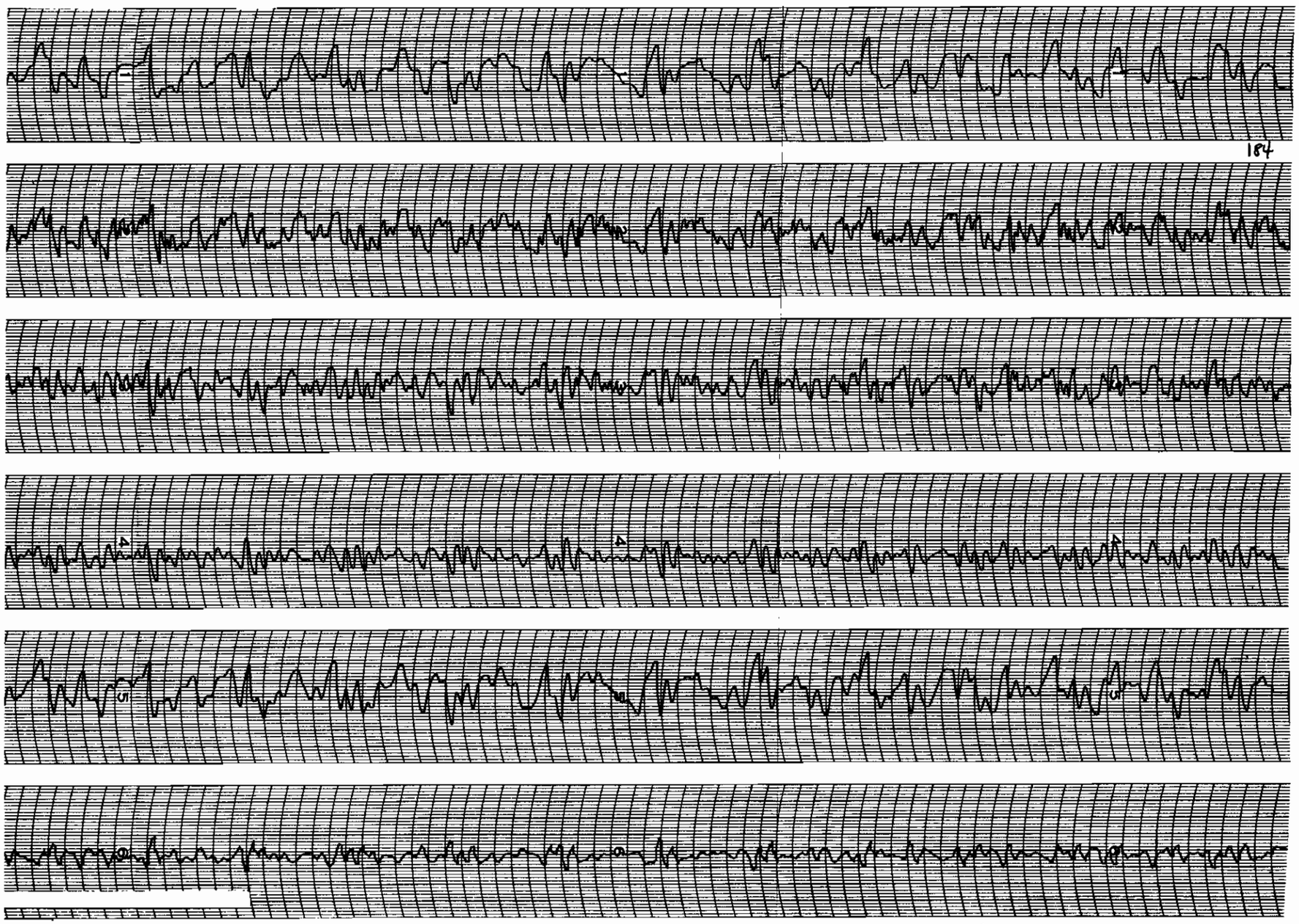


Figure 6.20 GW ω_B = 3 5mm/sec f_x

Chapter 7

SAMPLED-DATA MODELS WITH VARIABLE SAMPLING RATES

7.1 Introduction

The various sampled-data models of the human operator presented so far have been linear, with a constant sampling rate and time invariant parameters. However, it has been noted in the discussion of the previous chapter that the variation in location of certain peaks in the power spectral density of the operator's error may be due to a variation of the effective "sampling frequency". It would be reasonable to expect that if the sampling in the models considered here actually corresponds to a physiological process, that some kind of optimizing activity takes place by continuous adjustment of the sampling frequency on the basis of an appropriate performance criterion. If such an adjustment is a function of system error, for example, the operation would be non-linear. However, since the measurement of operator responses must take place over relatively long periods, the "average" or "effective" sampling rate may be approximately constant; thus resulting in the essentially linear behavior previously discussed. The purpose of this chapter is to outline some analytical approaches for the study of sampled-data systems with variable sampling rates. While no experimental data are yet available, and the variable-rate models have not been fully developed, the analysis methods are presented here since they represent a step in the direction of further research, both in human operator models and in sampled-data theory.

The basic system considered in this chapter is shown in Figure 7.1. Three techniques for the analysis of such a system are presented.

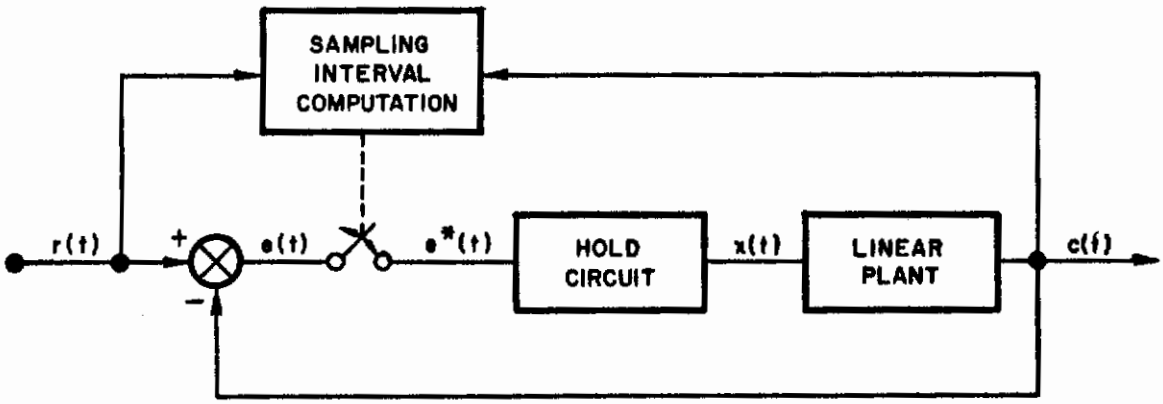


Figure 7.1 Prototype Adaptive Sampled-Data System

(1) Transient response evaluation using difference equations with variable coefficients

(2) Stability analysis using the T-locus

(3) Investigation of asymptotic stability using the "direct method" of Lyapunov.

The methods will be illustrated with examples which will not necessarily represent a human operator, in order to keep the basic points as simple as possible.

7.2 Status of Variable-Rate Sampled Data Theory

Very little work is available in the literature on sampled-data systems in which the sampling interval is variable. Time-

dependent sampling schemes (in which the sampling periods are given by a known function of time) have been treated for a limited class of such functions. Thus, cyclically-varying sampling rates have been treated by Hufnagel (1958), Jury and Mullin (1959) and Friedland (1960). A summary of various techniques is given by Hufnagel (1959) and Bekey (1961). The work of Kalman and Bertram (1959) is applicable to a fairly broad class of sampled data systems, but the methods are not readily applicable to general variable-rate systems. The problem of stability, in particular, has been neglected in the literature. It is discussed briefly by Hufnagel (1959) and by Kalman and Bertram (1959), in connection with periodically time varying systems. Tartakovskii (1957) has presented a method of studying stability of sampled-data systems with variable parameters based on the variable-transfer function concept developed by Zadeh. No stability studies of systems of the type of Figure 7.1 have appeared in the literature yet, to the author's knowledge.

A sampled-data system of the type presented here can be considered adaptive if the sampling frequency is adjusted automatically in accordance with an appropriate performance criterion. The systems to be analyzed here are examples of a class of systems which will minimize the sampling frequency consistent with stability and performance requirements.

The importance of such systems stems from the utilization of digital computers as elements of the control loop in an increasingly large number of applications. Since such a computer will generally be time-shared among a number of subsystems, it is desirable to utilize it only as frequently as necessary to achieve the desired stability and performance characteristics. It should be noted that such utilization may

also result in lower power utilization; an important issue in space vehicle control systems.

7.3 Evaluation of Transient Response

Since the sampling rate in Figure 7.1 is considered to be dependent on the state of the system, the system is nonlinear and ordinary Z transform analysis is not applicable. However, the time domain behavior of a sampled data system can be evaluated using difference equations even if the sampling interval is of variable width, since, regardless of the sampling period, sampled-data systems are open-loop between sampling instants. The response can be evaluated sample-by-sample. Such a method is discussed by Jury and Mullin (1959) for the periodically varying sampling rate case.

If we let the linear plant transfer function be G(s), corresponding to an impulse response h(t), then we can write (for $t_n \leq t \leq t_{n+1}$)

$$(7.1) \qquad c(t) = x(t_n)\, g(t-t_n) + \sum_{p=o}^{q-1} f_p(t-t_n) \left[\frac{d^{(p)} c(t)}{dt^{(p)}} \right]_{t=t_n}$$

where $x(t_n)$ is the magnitude of the hold circuit output at time t_n

$g(t)$ is the step function response of the system G(s)

and $f_p(t)$ are time functions resulting from the influence of past samples on the output signal c(t).

To completely describe a q-th order system, q first-order equations (or one q-th order equation) will be required. The additional equations can be obtained from (1) by differentiation. The application of this technique requires no specific knowledge of the sampling times t_n in advance, and they can be computed as a function of the system variables. These concepts are best

illustrated by means of an example.

Consider the system of Figure 7.1 with

$$G(s) = \frac{K}{s(s+a)} \tag{7.2}$$

and a zero-order hold circuit. The output can be evaluated (using Laplace transforms, for example). During the n-th sampling interval, the output satisfies the equation.

$$c(t) = KX_n\left[(t-t_n) - 1+e^{-(t-t_n)}\right] + C_n+\dot{C}_n\left[1-e^{-(t-t_n)}\right] \tag{7.3}$$

$$(t_n \leq t \leq t_{n+1})$$

which correspond to the form of Equation (7.1) and where

X_n = output of zero order hold at time $t_n = x(t_n) = e(t_n)$

$c(t)$ = continuous system output

C_n = value of output at n-th sampling instant = $C(t_n)$

$\dot{C}_n$ = value of output rate at n-th sampling instant = $(dc/dt)_{t = t_n}$

Differentiating (7.3) once we obtain

$$\dot{c}(t) = KX_n\left[1 - e^{-(t-t_n)}\right] + \dot{C}_n e^{-(t-t_n)} \tag{7.4}$$

Setting $t = t_{n+1}$ in Equations (7.3) and (7.4) we obtain

$$C_{n+1} = KE_n\left[T_n - 1 + e^{-T_n}\right] + C_n + \dot{C}_n\left[1 - e^{-T_n}\right] \tag{7.5}$$

$$\dot{C}_{n+1} = KE_{n-1}\left[1 - e^{-T_n-1}\right] + \dot{C}_n e^{-T_n} \tag{7.6}$$

where $T_n \triangleq t_{n+1} - t_n$ is the length of the n-th sampling interval.

Corresponding equations for C_n and $\dot{C}_n$ are:

$$(7.7) \quad C_n = KE_{n-1}\left[T_{n-1} - 1 + e^{-T_{n-1}}\right] + C_{n-1} + \dot{C}_{n-1}\left[1 - e^{-T_{n-1}}\right]$$

$$(7.8) \quad \dot{C}_n = KE_{n-1}\left[1 - e^{-T_{n-1}}\right] + \dot{C}_{n-1}\left[e^{-T_{n-1}}\right]$$

Terms containing $\dot{C}_n$ and $\dot{C}_{n-1}$ can be eliminated from Equations (7.5) - (7.8). Then substituting $X_n = R_n - C_n$ for the error at the n-th sampling instant, we obtain the relation

$$(7.9) \quad C_{n+1} + \left[K(T_n - 1 + e^{-T_n}) - \frac{1 - e^{-(T_n + T_{n-1})}}{1 - e^{-T_{n-1}}}\right] C_n$$
$$+ \left[K(1 - e^{-T_{n-1}} - T_{n-1}e^{-T_{n-1}}) + e^{-T_{n-1}}\right] \left(\frac{1 - e^{-T_n}}{1 - e^{-T_{n-1}}}\right) C_{n-1}$$
$$= K(T_n - 1 + e^{-T_n})R_n + K(1 - e^{-T_{n-1}} - T_{n-1}e^{-T_{n-1}})\left(\frac{1 - e^{-T_n}}{1 - e^{-T_{n-1}}}\right) R_{n-1}$$

This equation, a difference equation with time-varying coefficients, yields the output at any sampling instant as a function of the values of the input and output at the two previous intervals and the lengths of the two previous sampling intervals, T_n and T_{n-1}.

Let us now assume that the sampling period is governed by the relation

$$(7.10) \quad T_n = \frac{\alpha}{1 + \beta |E_n|}$$

where α and β are constants. Then, if the error E_n increases,

the sampling period will decrease, and the system will tend to become more stable (as will be noted in the next Section).

Given the "control law" of Equation (7.10) and the difference equations (7.9) the output response can be computed sample-by-sample. While this is clearly a laborious procedure, it can be easily instrumented on a digital computer. The procedure above can be applied directly to more complex systems.

7.4 Stability Determination Using the T-locus

The stability of linear sampled data systems is often studied by using the root-locus method (see for example, Ragazzini and Franklin (1958) p. 105). The root locus is a plot of the roots of the characteristic equation of the system as function of the open-loop gain. For linear time-invariant sampled data systems, if the characteristics equation is

$$1 + KG(z) = 0 \tag{7.11}$$

where $G(z)$ is the open-loop pulse transfer function, then the root locus is plot of those values of z which satisfy the condition

$$KG(z) = 1 \angle \pi \pm 2\pi n \qquad n = 1, 2, 3, \ldots \tag{7.12}$$

If the roots of (7.11) lie inside of the unit circle in the complex z-plane, they represent a stable system. If they lie outside, the system is unstable and if they lie on the unit circle the system will be neutrally stable.

Consider now a system in which the sampling rate is not constant, but varies slowly in comparison to the system dyanmics. This means that at any particular time, the sampling period T can be considered approximately constant. The stability of the system can be examined by plotting the root locus with the sampling period T as the parameter. This locus, which we shall term the "T-locus", will give an indication of the dependence of the system's stability on the sampling rate, provided that the initial assumption is not violated, i.e. that rate of change of sampling period T is slow compared to the system time constants.

Several classes of T-loci can be readily identified. In some systems the T-locus behaves like the ordinary gain or K-locus. In other systems, when the position of the open-loop poles and zeros is affected by the value of T, the locus behavior is quite different. Consider the following examples of various types of loci:

(1) T-locus behaves like K-locus

In the system of Figure 7.1, let the hold be zero order, and the plant be

$$G(s) = \frac{K}{s} \tag{7.13}$$

Then the open loop transfer function is

$$GH(s) = \frac{K(1-e^{-Ts})}{s^2} \tag{7.14}$$

and the corresponding pulse transfer function is

$$GH(z) = \frac{KT}{z-1} \tag{7.15}$$

Consequently, the characteristic equation is given by

$$1 + GH(z) = 0 \tag{7.16}$$

or

$$z + (KT-1) = 0 \tag{7.17}$$

Clearly, the locus behaves in an identical way to the K-locus. For $T = 0$, $z = 1$, and $|z| > 1$ for $T > 2/K$, i.e. the system is stable only for $T < 2/K$. The locus is always real as indicated in Figure 7.2(a). As usual, the open loop pole is indicated by an "x".

(2) T-locus resembles K-locus, but open-loop roots move

Now let

$$G(s) = \frac{Ka}{s+a} \tag{7.18}$$

Then the open loop pulse transfer function is

$$GH(z) = \frac{K(1-e^{-aT})}{z - e^{-aT}} \tag{7.19}$$

Note that the position of the open loop pole (located at $z = e^{-aT}$) now depends on the sampling period. The characteristic equation is

$$z + K - e^{-aT}(1 + K) = 0 \tag{7.20}$$

and the corresponding T-locus is shown in Figure 7.2(b). The K-locus is also along the real axis but begins only at the

open loop pole location. To find the limiting value of T for stability, let z = -1 in (7.20). Then

$$e^{-aT_{lim}} = \frac{K - 1}{K + 1} \tag{7.21}$$

and consequently the limiting value of T is

$$T_{lim} = - \frac{1}{a} \ln \left[\frac{K - 1}{K + 1} \right] \tag{7.22}$$

Note that while the K-locus continues to $-\infty$, the T-locus terminates at -K. Thus, for K = 1, there is no value of T which can cause instability.

As an additional example of this type consider a system with no hold circuit, and

$$G(s) = \frac{Ka}{s(s+a)} \tag{7.23}$$

Then the pulse transfer function is

$$G(z) = \frac{K(1-e^{-aT})z}{(z-1)(z-e^{-aT})} \tag{7.24}$$

We have T-dependence in the gain factor and in the location of one pole. The two loci are indicated in Figure 7.2(c) for particular values of K, a, and T. It can be seen that both loci are circles in the z-plane.

(3) No resemblance between K and T loci

In the general case, of course, there will be no necessary resemblance between the two loci. Thus, if we have a zero-order hold and the plant of (7.23), the resulting pulse transfer function is

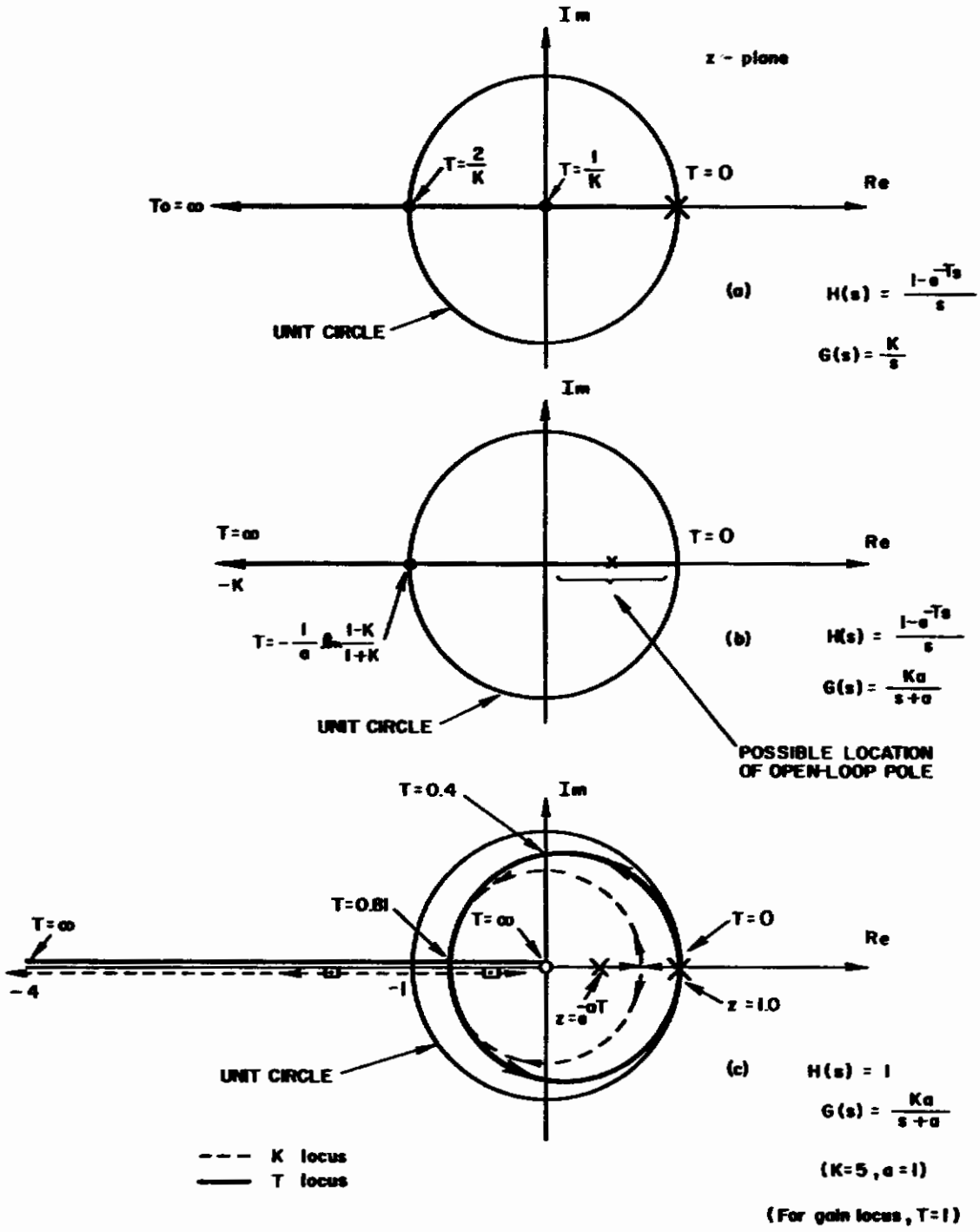


Figure 7.2 T-loci for Sample Systems

$$(7.25) \qquad GH(z) = K(1-z^{-1})\, Z\left\{\frac{a}{s^2(s+a)}\right\}$$

$$= \frac{K\left\{(aT-1+e^{-aT})z + \left[1-e^{-aT}(1+aT)\right]\right\}}{(z-1)(z-e^{-aT})}$$

The K-locus of this system is again a circle. However, the T-locus is by no means easy to determine, and must be obtained by numerical substitution.

It appears that a catalog of T-loci for various types of systems would be quite helpful in stability studies of sampled-data systems with adjustable sampling rates. It is possible that a systematic method for construction of the loci can be obtained, in the course of developing such a catalog of loci.

7.5 Asymptotic Stability Investigation Using the Direct Method of Lyapunov

The so-called "Direct" or "Second-Method" of Lyapunov for determining asymptotic stability of nonlinear differential equations has become quite important in recent years (see, for example, Letov (1960), Kazda (1960), Kalman and Bertran (1959). The method is based on finding a scalar function of the state variables of the system which satisfies certain conditions. If such a function, called a Lyapunov function, does indeed exist, then the null solution of the differential equation is asymptotically-stable in the large. The importance of the method is based on the fact that the stability information is obtained without a need to solve the differential equation.

Much less literature is available on the use of the Lyapunov method for determining asymptotic stability (either global or local) of difference equations. The basic references

are those of Hahn (1958) and (1959), Kalman and Bertram (1960) and Bertram (1960). From these references, the following stability theorem can be stated (without proof) and applied to determine the stability of the nonlinear system of Figure 7.1.

Stability Theorem for Nonlinear Difference Equations

If, for the vector difference equation

$$\bar{y}_{n+1} = \bar{f}(\bar{y}_n, t) \tag{7.26}$$

there exists a positive definite scalar function $V(y_n^1, y_n^2, \ldots y_n^k, t)$ of the discrete state variables $y_n^1, y_n^2, \ldots y_n^k$ (which are components of the k-dimensional state vector y_n) and time t_n such that $V(0,\ldots,0) = 0$ and the first difference

$$\Delta V(\bar{y}_n, t_n) \equiv V(y_{n+1}^1, y_{n+1}^2, \ldots y_{n+1}^k, t_{n+1}) - V(y_n^1, \ldots y_n^k, t_n) \tag{7.27}$$

of this function exists, then satisfaction of the conditions

$$\begin{aligned}
&\text{(i)} && V(\bar{y}_n, t_n) > 0 \quad \text{when } \bar{y}_n \neq 0 \\
&\text{(ii)} && V(\bar{y}_n, t_n) \text{ is continuous in } \bar{y}_n \\
&\text{(iii)} && V(\bar{y}_n, t_n) \to \infty \text{ as } \|\bar{y}_n\| \to \infty \\
&\text{(iv)} && \Delta V(\bar{y}_n, t_n) < 0 \quad \text{when } \bar{y}_n \neq 0
\end{aligned} \tag{7.28}$$

implies that the null solution $\bar{y}_n = 0$ of the aforementioned vector difference equation is asymptotically stable in the large and that $V(\bar{y}_n, t_n)$ is a Lyapunov function for this system.

Example. Let us now apply this method to the problem of Figure 7.3, where T_n, the n-th sampling

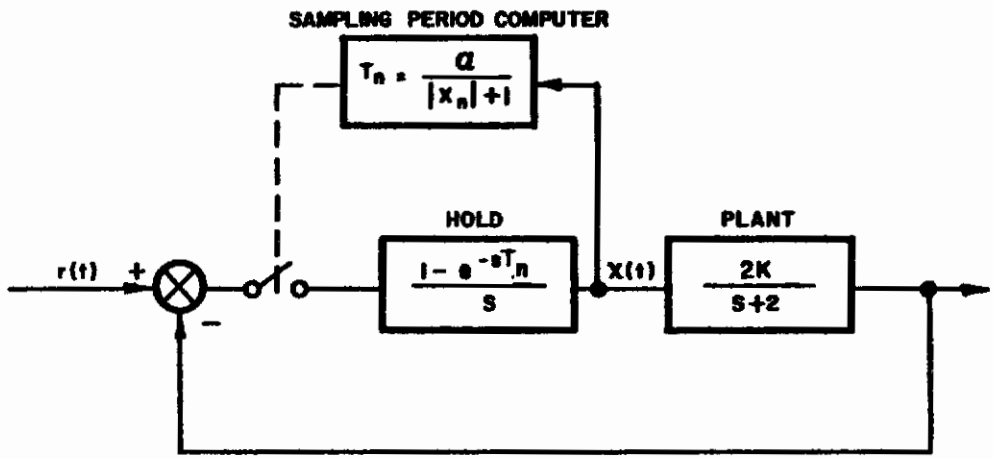


Figure 7.3 Nonlinear Sampled-Data System

interval, is defined as

$$T_n \triangleq t_{n+1} - t_n \tag{7.29}$$

as in Section 7.3 above.

We assume that the system is "adaptive" in the sense that a large error results in an increase in sampling frequency, and choose a control law of the form

$$T_n = \frac{\alpha}{|x_n| + 1} \tag{7.30}$$

where allowable values of α for stability are to be found. The system is clearly nonlinear. The equations at the sampling instants are obtained as in Section 7.3.

$$X_n = E_n \qquad t_n \leq t < t_{n+1} \tag{7.31}$$

$$C_{n+1} = e^{-2T_n} C_n + K(1-e^{-2T_n}) X_n \tag{7.32}$$

and

$$X_n = R_n - C_n \tag{7.33}$$

where the subscript n denotes the value at the n-th sampling instant. That is

$$C_n \triangleq C(t_n) \tag{7.34}$$

Since the sampling intervals are not constant, we cannot replace t_n by nT. Substituting (7.31) and (7.33) in (7.32) we obtain the state equations for the system:

$$C_{n+1} = \left[(1 + K)\, e^{-2T_n} - K\right] C_n + K(1-e^{-2T_n})R_n \tag{7.35}$$

$$T_n = \frac{\alpha}{|R_n - C_n| + 1}$$

We investigate the stability of the null solution and therefore let $R_n = 0$. Then the equations reduce to:

$$C_{n+1} = \left[(1+K)e^{-2T_n} - K\right] C_n \tag{7.36}$$

$$T_n = \frac{\alpha}{|C_n| + 1}$$

Based on suggestions in the References cited above, let us choose a Lyapunov function of the form

$$V(\bar{C}_n, t_n) = \|C_n\|^2 \tag{7.37}$$

where $\|\bar{C}_n\|$ represents the Euclidean norm of the vector $\bar{C}_n$.

$$V(C_n) = C_n^2 \tag{7.38}$$

The first difference of the Lyapunov function will be

$$\begin{aligned} V(C_n) &= C_{n+1}^2 - C_n^2 \\ &= C_{n+1}^2 - C_n^2 \end{aligned} \tag{7.39}$$

Now, since $V(C_n)$ is clearly positive definite, continuous in C_n and $\to \infty$ as $\|C_n\| \to \infty$, then for stability we must simply find a region where $\Delta V(C_n)$ is negative definite. Now, for $\Delta V(C_n)$ to be negative, we must have a region where

$$C_{n+1}^2 < C_n^2 \tag{7.40}$$

Substituting (7.36) in this relation

$$C_n^2 \left[(1 + K)^2 e^{-4T_n} - 2K(1 + K)e^{-2T_n} + K^2 \right] < C_n^2 \tag{7.41}$$

Or, since $C_n^2 > 0$

$$(1 + K)^2 e^{-4T_n} - 2K(1 + K)e^{-2T_n} + K^2 < 1 \tag{7.42}$$

If we let K = 2, this inequality becomes

$$e^{-2T_n} (9e^{-2T_n} - 12) < -3 \tag{7.43}$$

or, since e^{-2T_n} is greater than zero for all $T_n > 0$, we must have

$$3e^{-2T_n} - 4 < -\frac{1}{e^{-2T_n}} \tag{7.44}$$

If we interpret the two sides of this inequality as equations defining two functions $f_1(e^{-2T_n})$ and $f_2(e^{-2T_n})$ we can obtain a graphical interpretation of the stability requirement by plotting f_1 and f_2 vs e^{-2T_n}, as done in Figure 7.4.

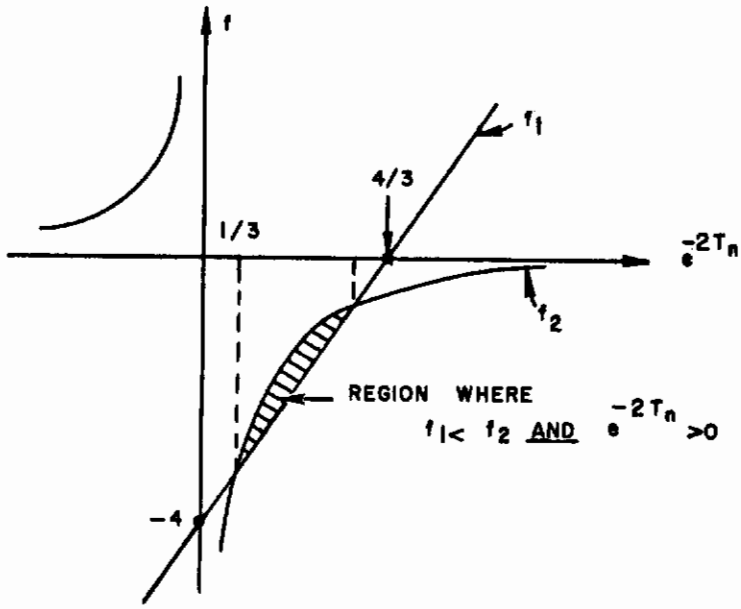


Figure 7.4 Stability Region for System of Figure 7.3

To find the values a and b which define the limits within which stability is assured, we solve the equation $f_1 = f_2$, to obtain

$$3(e^{-2T_n})^2 - 4e^{-2T_n} + 1 = 0 \tag{7.45}$$

The roots of this equation are

$$r_1, r_2 = \frac{4 \pm \sqrt{16-12}}{6} = \begin{cases} 1/3 \\ 1 \end{cases} \tag{7.46}$$

Therefore, $\Delta V(C_n) < 0$ for

$$\frac{1}{3} < e^{-2T_n} < 1 \tag{7.47}$$

The upper limit clearly cannot be exceeded, since for all

$$T_n > 0, \; e^{-2T_n} < 1$$

The lower limit means that

$$e^{-2T_n} > \frac{1}{3}$$

or

$$-2T_n > \ln \frac{1}{3} ; \qquad T_n < \frac{1}{2} \ln 3$$

Using the "control law" for T_n as given by (7.36), we have

$$T_n = \frac{\alpha}{|C_n| + 1} < \frac{\ln 3}{2} \tag{7.48}$$

This expression assumes its maximum value for $|C_n| = 0$, and consequently

$$\alpha = \frac{\ln 3}{2} \tag{7.49}$$

Therefore, provided that α stays below the limit of (7.49), the nonlinear system (7.36) will have an asymptotically stable solution if disturbed from the equilibrium position $C_n = 0$.

It is interesting to compare this stability region with that obtained by plotting the T-locus for the corresponding constant sampling rate system. The open-loop transfer function of the system of Figure 7.3 is given by

$$GH(s) = \left(\frac{1-e^{-sT}}{s}\right)\left(\frac{2K}{s+2}\right) \tag{7.50}$$

The corresponding pulse transfer function is

$$GH(z) = (1-z^{-1})\, Z\left[\frac{2K}{s(s+2)}\right] = \frac{2(1-e^{-2T})}{z - e^{-2T}} \quad (\text{for } K = 2) \tag{7.51}$$

We are interested in the T-locus,, i.e. the locus of the closed loop roots as a function of sampling period T.

The characteristic equation is given by

$$1 + GH(z) = 0 \tag{7.52}$$

or

$$(z-e^{-2T}) + (2-2e^{-2T}) = 0$$

and the one closed-loop root is given by

$$z = 3e^{-2T} - 2 \tag{7.53}$$

The T-locus is shown in Figure 7.5. The critical values for stability occur for

$$z = 3e^{-2T} - z = \pm 1$$

or

$$T = 0 \quad \text{and} \quad T = \frac{\ln 3}{2}$$

Thus, for stability in the constant sampling rate case, $T < \frac{\ln 3}{2}$ is required. Note that this agrees with Equation (7.48), but that (7.48) has much more generality since it includes the nonlinear case.

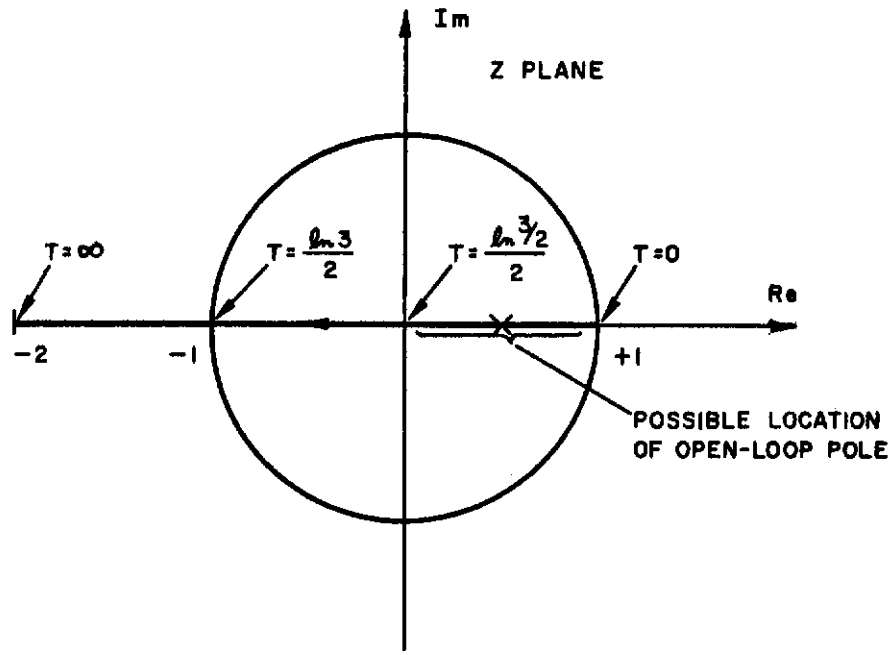


Figure 7.5 T-locus for Example System

7.6 Summary

In this chapter several techniques for the analysis of sampled-data systems with variable sampling rates have been presented. The methods indicate how transient response and system stability can be studied for such systems.

The mathematical models of human operators discussed in detail in preceding chapters have been based on constant sampling rates. The methods discussed in this chapter are offered as a possible contribution to the future study of new models based on variable sampling.

Chapter 8

CONCLUSIONS AND RECOMMENDATIONS

8.1 Conclusions

The major concern of this work has been an analytical and experimental investigation of a class of mathematical models for the human operator based on the theory of linear sampled data control systems. The results show that for the particular tracking system considered, the sampled-data models do indeed result in input-output behavior which more closely approximates experimental results than is true of linear continuous models. In particular, it has been shown that the models which include sampling and first-order hold circuits are consistent with a large body of evidence in the literature on tracking, that they do produce energy at frequencies beyond the range of a band-limited input, that they are consistent with the results of tracking an intermittent display, and finally, that the analysis of such models results in spectral characteristics which check closely with experiment, at least under certain conditions.

The combination of analysis and experiment used for the study of the sampled-data systems proposed in this report as human operator models probably represents the major contribution of this work. The fact that the spectral characteristics of human operator output and error signals when the input signal consists of simple sum of sine waves could be predicted from an analysis based on stationary random processes with continuous spectral density functions,

could not be known a priori. A combination of analysis and experiment was required to determine the validity of such an analytical procedure. The closed-form expressions for error and output power spectral densities proposed in this report, then, while lengthy and cumbersome, should make it possible to generalize the work to other input functions than those used here, if they are programmed for digital computation as has been done in this work.

The results of the study can be viewed as a logical extension of previous work with quasi-linear continuous models. The continuous models were considered adequate representations of tracking behavior when the input function bandwidth did not exceed approximately 3/4 cps. In the present study the band extended to 1.6 cps; spectral peaks were noted in the range of 1 to 1.6 cps (which are consistent with previous data); and these peaks were shown to be consistent with linear sampling models as well. In other words, the sampled models result in a decrease in "remnant" power, for difficult tasks, where the remnant is considered to be that component of the operator's output which the model does not explain. The limitations of the study are reviewed in the next section.

For the particular tracking system and a particular operator, it has been shown that the sampled-data model can be constructed by following a systematic sequence of spectral measurements, which yield the parameters of the continuous model and the sampling frequency. Unfortunately the experimental study was not sufficiently extensive to make definitive statements about a synthesis procedure. However, the rules suggested in the report for determination of model sampling frequencies were satisfactory for those

experimental runs where high frequency spectral peaks could be clearly observed.

The study has placed considerable emphasis on the measurement of the error spectral density in human tracking, as a sensitive indicator of tracking performance. In particular, it is concluded on the basis of the present study that a peak occurs in the error spectrum (here called the "sampling peak") which is related not only to the sampling frequency of the mathematical models, but also to the display frequency when the display is intermittent.

The major conclusion of this work, then, is that the application of sampled-data theory to the study of man-machine systems is well justified; that the results of the present study give an indication that sampled-data models can be used to predict several important aspects of tracking behavior; and that much more work remains to be done.

8.2 Limitations and Suggestions for Further Research

The extensions of this work become readily apparent when one examines the restrictions which have been placed upon it. The removal of the restrictions is the logical direction for new research. The restrictions upon the present study fall into two logical groups: those governing the experimental situation and those related to the analytical tools which have been employed. Let us consider these in turn.

8.2.1 Extensions in Tracking Research. Several directions for new work stem directly from the restrictions placed on the situation itself:

(1) The entire study has been restricted to tracking systems with no controlled element dynamics. From

past work, as summarized by McRuer and Krendel (1959), it is clear that the transfer characteristics of human operators are affected by the dynamics of the controlled element, but that in general this variation is predictable in terms of quasi-linear continuous models. Whether the straightforward extension of the techniques of this report to the case where significant dynamics are present can be made will require further analysis.

(2) The entire study has been restricted to one-dimensional compensatory tracking. Elkind (1956) has treated both pursuit and compensatory tracking, and the error spectra in the two cases are similar. Pursuit tracking spectra, however, do not show the peaks discussed here until the input frequencies extend past 1 cps. Higher sampling frequencies will be required in the sampled-data models if they can be shown to be applicable. The extension to two-dimensional tracking will be considerably more difficult, since the interaction between the two channels in the operator is not understood at present. However, it is possible that adequate values of sampling frequency and "time-sharing" between the two channels will result in better models than those presently available.

(3) The models proposed in this study are primarily applicable to trained operators. From the experimental results of Chapter 6 we have seen that the data would support a tentative hypothesis that "sampling peaks" and consequently model sampling frequencies increase with practice, along with increases in model gain. The change in model characteristics during learning should make an interesting experimental project. Except for Sheridan's work (1960) on time variation, very little has been done

with models in this area.

(4) The present study made use of only eight subjects, clearly too small a sample to justify extensive conclusions. While past work in the field has often been based on work with a few or even a single operator, this procedure does not appear to be completely justified for tasks of the difficulty of those encountered here. The tracking task used in this study was close to the performance limitations of some of the subjects, consequently the variation among subjects was quite large, and much more extensive experimental work will be required to ascertain the range and statistical parameters of this variation.

(5) Variations in the model for adaptation to different inputs is also a logical extension of this work. It has been shown that higher model sampling frequencies were required with more difficult inputs. It should be possible to design an experiment in which the task difficulty changes during a run, and to measure the resulting model parameters. The length of time required for adaptation may be in conflict with the time required for measurement, but this will require investigation.

(6) Perhaps most important, the conclusions of this report will require further study. It has been noted that the selection of a sampling frequency for the model has been hampered by both the limited frequency range of the input function and the limitations of the analog spectral analyzer. In future studies, even with compensatory tracking, it is recommended that input signals contain some energy to at least 2.5 or 3 cps, and that

spectral measurement techniques should be adequate for the detection of peaks in the range of 1 to 2 cps.

8.2.2 Extensions in Analytical Work. The analysis in this report has been based almost entirely on the study of linear sampled data systems with "impulse sampling". The following extensions readily suggest themselves:

(1) Impulse sampling does not seem to be a very good representation of any conceivable physiological process. The analysis reported here could be repeated using the finite-pulse-width "P" transforms developed by Farmanfarma (1958). The first stage in such a study could be based upon a fixed but finite pulse width, and the resulting behavior examined as a function of pulse width. Later work could lead into pulse-width modulation as a possible form of adaptation.

(2) The hold circuits considered here were clearly inadequate. The first extension of the work should give greater attention to partial velocity extrapolation circuits, possibly extending to systems with some capability for sensing the 2nd derivative of the project error.

Secondly, it may be possible to study the effects of "leaky hold circuits" upon model behavior. If the hold circuit represents some physiological process, it is reasonable that it should not have perfect time domain properties, and therefore, that it should not necessarily have zeros in the frequency domain exactly at intervals of $2\pi/T$ rad/sec. A tentative study has been made of "approximate hold circuits" based on the Padé approximation to e^{-Ts} and the results appear to be interesting.

(3) The modification of the sampled-data models required to take "adaptation" into account has been mentioned repeatedly throughout the report. In particular, some attention has been given to the adjustment of loop gain and sampling frequency as possible forms of adaptive behavior. However, only some preliminary steps have been taken in the analysis of sampled-data systems with variable sampling rates. In Chapter 7 a "first look" has been taken at transient and stability analyses. Both the "T locus" and the Lyapunov methods are promising, and their extension would probably be fruitful for sampled-data theory even if not for human operator models. The evaluation of nonlinear human operator models would be a very complex problem since frequency spectra would have no meaning in the ordinary sense.

(4) The method suggested in the report for construction of the sampled-data model requires fairly complex measurement techniques, followed by a sequence of curve fitting operations. Recently Ornstein (1960) has reported on the use of a technique developed by Margolis and Leondes (1959) for determining parameters of an assumed "transfer function" for the human operator by a modification of the method of "steepest descent". It may be possible to extend the "learning model" concept to the sampled data case, where the sampling frequency is one of the parameters to be tracked. Once again, such research would be extremely useful to control theory as well to the understanding of human tracking.

(5) The sampling periods suggested in this report have been deterministic in nature, whether they occur at fixed intervals of time or at times dependent on the system state variables. It is also possible to consider that the

sampling interval is a stochastic function of the accuracy of tracking, for example. Thus, it is conceivable that the sampling period is given with a certain probability. Then, a shorter sampling interval could become more probable if the tracking error is large and less probable if it is small. The formulation of such a probabilistic model would be an interesting extension of this work.

(6) The first-order-hold circuits suggested in this report suffer from the serious limitation of including predictors which are physically unrealizable and thus make analog simulation impossible. The "Modified Hold Circuits" were introduced to eliminate this problem, but resulted in excessive low frequency gain. More research is needed on the types of extrapolation circuits which are best suited for human operator models.

8.3 Speculations and Implications

The discussion of the preceding chapters has been almost exclusively concerned with measurements performed at the input and output "terminals" of the human operator, treated as a system element. Almost nothing has been said about possible physiological mechanisms which could account for sampling behavior.

While it is certainly clear that the tracking behavior described in the report can be produced by processes not involving sampling (such as continuous but non-linear operations, for example), it should be indicated that there is some physiological evidence which strongly supports the sampling hypothesis. Pitts and McCulloch (1947) suggested that something akin to sampling takes place in the central

nervous system. More recently Verzeano and Negishi (1960) have shown on the basis of studies with multiple micro-electrodes that there are waves of inhibition and excitation which travel in the brain under control of the thalamus. These waves, in effect, act as "gating" signals which turn groups of cortical networks "on" and "off". The rate of these excitation (followed by inhibition) signals is approximately that of the alpha rhythm, 7 to 10 cps. Based on such findings, it may be reasonable to speculate that the signals from a particular sensory inputs are admitted to the associative and interpretive areas of the cortex only intermittently. Furthermore, if the basic "clock" frequency of the intermittency is approximately constant, one or more cycles of the "clock" may elapse before the arrival of feedback information makes it possible to admit new data. While this account is clearly speculative, and reminiscent of the "moment" theory of Stroud (1955), it is certainly interesting since it does suggest that the intermittency hypothesis may indeed be founded on physiological phenomena in the brain.

The research reported here may have some significant implications in the design of future man-machine systems. Thus, for example, television displays in future lunar-landing vehicles may be intermittent in nature and the behavior of the human operator with intermittent displays will be of importance. The possibility of time-sharing a sampled display between two channels of information is also conceivable. Furthermore, as future space systems will be both extremely costly and of relatively high risk to the human operator, it will be desirable to use analytical techniques for design

and evaluation to as large an extent as possible. Simulation of various possible failure modes may make it imperative to use representations of the human pilot which are adequate under conditions which approximate the limits of performance. The sampled models presented here, may prove to be applicable for this purpose.

But, regardless of practical application, it is hoped that this work, as an example of a methodology for the development of mathematical models, will be a small contribution to the scientific study of human behavior.

BIBLIOGRAPHY

1. Adams, J. A. (1961) "Human Tracking Behavior", Psych. Bulletin, 58:55-79, January 1961

2. Ashby, W. R. (1960) Design for a Brain, 2nd Edition, New York: John Wiley, 1960

3. Balabanian, N. (1958) Passive Network Synthesis, New York: Prentice-Hall, 1958

4. Bates, J. A. V. (1947) "Some Characteristics of the Human Operator", Jour. Institution Elec. Engrs. (London) 94 (IIA): 298-304, 1947

5. Bekey, G. A. (1960) "Adaptive Control System Models of the Human Operator", Proc. 1960 Symposium on Adaptive Control Systems, to be published, Pergamon Press, 1962

6. Bekey, G. A. (1961) "A Survey of Techniques for the Analysis of Sampled-Data Systems with a Variable Sampling Rate", to be published as ASD Technical Report 62-35 in 1962

7. Bendat, J. S. (1958) Principles and Applications of Random Noise Theory, New York: John Wiley, 1958

8. Bennett, W. R. (1956) "Methods of Solving Noise Problems", Proc. IRE, 44:609-38, May 1956

9. Bertram, J. E. (1960) "The Direct Method of Lyapunov in the Analysis and Design of Discrete-Time Control Systems", Work Session in Lyapunov's Second Method, ed. by L. F. Kazda, Ann Arbor: University of Michigan Press, 1960

10. Blackman, R. B. and Tukey, J. W. (1958) The Measurement of Power Spectra, New York: Dover Publications, 1958

11. Booton, R. C. (1953) "The Analysis of Nonlinear Control Systems with Random Inputs", Proc. Symposium on Nonlinear Circuit Analysis, 2:369-391, April 1953

12. Broadbent, D. E. (1958) Perception and Communication New York: Pergamon Press, 1958

13. Brown, R. H. (1961) "Visual Sensitivity to Differences in Velocity" Psych. Bulletin, 58:89-103, March 1961

14. Churchill, R. V. (1941) Fourier Series and Boundary Value Problems, New York: McGraw-Hill, 1941

15. Davenport, W. B. and Root, W. L. (1958) Introduction to the Theory of Random Signals and Noise, New York: McGraw-Hill Book Co., 1958

16. Diamantides, N. D. (1957) "Informative Feedback in Jet-Pilot Control Stick Motion" AIEE Applications and Industry, 243-49, November 1957

17. Diamantides, N. D. (1958) "A Pilot Analog for Airplane Pitch Control" J. Aero. Sciences, 25:361-71, June 1958

18. Elkind, J. I. (1956) "Characteristics of Simple Manual Control Systems" Lincoln Laboratory Report No. 111, M. I. T. Lincoln Laboratory, Lexington, Mass., April 1956

19. Ellson, D. G., Gray, F., et al (1947) "Wavelength and Amplitude Characteristics of Tracking Error Curves", AAF-AMC Engr. Report TSEAA-694-2D, Wright Field, Ohio, April 1947

20. Farmanfarma, G. (1958) "General Analysis and Stability Study of Finite Pulsed Feedback Systems", AIEE Applications and Industry, 148-162, July 1958

21. Fogel, L. J. (1956) "An Analysis for Human Flight Control", IRE Convention Record, pt. 8, 69-88, 1956

22. Fogel, L. J. (1957) "The Human Computer in Flight Control" IRE Trans. on Elec. Computers, EC-6:195-201, September 1957

23. Friedland, B. (1960) "Sampled Data Control Systems Containing Periodically Varying Elements", Proc. International Federation for Automatic Control Congress, Moscow, 1960; to be published by Butterworths, 1961

24. Goldman, S. (1953) Information Theory, New York: Prentice-Hall

25. Goodyear Aircraft Corp. (1952) "Final Report: Human Dynamics Study", GAC Report GER-4750, April 1952

26. Goodyear Aircraft Corp. (1957) "Human Response Dynamics: Geda Computer Application", GAC Report GER-8033, January 1957

27. Gottsdanker, R. M. (1952) "The Accuracy of Prediction Motion", Jour. Exp. Psych., 43:26-36, 1952

28. Grenander, U. and Rosenblatt, M. (1957) Statistical Analysis of Stationary Time Series, New York: John Wiley, 1957

29. Guillemin, E. (1957) Synthesis of Passive Networks, New York: John Wiley, 1957

30. Hahn, W. (1958) "On the Application of the Method of Lyapunov to Difference Equations" (in German) Mathematische Annalen 136:402-441, 1958 (Translated into English by G. A. Bekey and available as Report No. 61-27, Department of Engineering, University of California, Los Angeles).

31. Hick, W. E. and Bates, J. A. V. (1950) The Human Operator of Control Mechanisms", Permanent Records of Research and Development, No. 17-204, Ministry of Supply, England, May 1950

32. Higgins, T. J. and Holland, D. B. (1959) "The Human Being as a Link in an Automatic Control System", IRE Trans. on Medical Electronics, ME-6:125-133, September 1959

33. Hufnagel, R. E. (1958) "Analysis of Cyclic-Rate Sampled-Data Feedback Control Systems" AIEE Applications and Industry, 421-425, November 1958

34. Hufnagel, R. E. (1959) Analysis of Aperiodically Sampled-Data Feedback Systems , Ph. D. Dissertation, Electrical Engineering Department, Cornell University, New York, June 1959

35. Hyndman, R. W. and Beach, R. K. (1958) "The Transient Response of the Human Operator" IRE Trans. on Medical Electronics, PGME-12:67-71, December 1958

36. Jackson, A. S. (1958) "Synthesis of a Linear Quasi-Transfer Function for the Operator in Man-Machine Systems" WESCON Convention Record, pt. 4, 2:263-271, 1958

37. Johnson, C. L. (1956) Analog Computer Techniques, New York: McGraw-Hill, 1956

38. Joseph, P., Lewis, J., and Tou, J. (1961) "Plant Identification in the Presence of Disturbances and Application to Digital Adaptive Systems" AIEE Applications and Industry, 18-24, March 1961

39. Jury, E. I. (1960) "Status of Sampled Data Systems" AIEE Communications and Electronics, 769-777, January 1960

40. Jury, E. I. and Mullin, F. J. (1959) "The Analysis of Sampled-Data Control Systems with Periodically Time-Varying Sampling Rate" IRE Trans. on Automatic Control, AC-4:15-20, May 1959

41. Kalman, R. E. and Bertram, J. E. (1959) "A Unified Approach to the Theory of Sampling Systems", Jour. Franklin Institute 267:405-436, May 1959

42. Kalman, R. E. and Bertram, J. E. (1960) "Control System Analysis and Design via the Second Method of Lyapunov- II. Discrete Time Systems", Jour. Basic Engineering, 82D:394-400, 1960

43. Kazda, L. F. (1960) Work Session in Lyapunov's Second Method, Ann Arbor: University of Michigan Press, September 1960

44. King, W. J., (1961) "Continuous Compensatory Tracking by a Cebus Monkey", Science, 134:947-948, 29 September 1961

45. Krendel, E. and Barnes, G. H., (1954) "Interim Report on Human Frequency Response Studies" WADC Technical Report TR 54-370, June 1954

46. Laning, J. H. and Battin, R. H. (1956) Random Processes in Automatic Control, New York: McGraw-Hill, 1956

47. Letov, A. M. (1961) Stability in Nonlinear Control Systems (Translated from Russian) Princeton University Press, 1961

48. Licklider, J. C. R. (1960) "Quasi-linear Operator Models in the Study of Manual Tracking", pp. 171-279 in Developments in Mathematical Psychology, ed. by R. D. Luce, Glencoe, Illinois: The Free Press, 1960

49. Lighthill, M. J. (1960) Fourier Analysis and Generalized Functions, Cambridge University Press

50. Margolis, M. and Leondes, C. T. (1959) "A Parameter Tracking Servo for Automatic Control Systems", IRE Trans. on Automatic Control, AC-4:100-111, 1959

51. Mayne, R. (1951) "Some Engineering Aspects of the Mechanism of Body Control" Electrical Engineering, 70:207-212, March 1951

52. McRuer, D. T. and Krendel, E. (1957) "Dynamic Response of Human Operators" WADC Technical Report TR 56-524, October 1957

53. McRuer, D. T. and Krendel, E. (1959) "The Human Operator as a Servo System Element" Jour. Franklin Institute 267:381-403, May 1959 and 267:511-536, June 1959

54. McRuer, D. T. (1961) "Some Statistical Properties of Time Functions Composed of Sinusoids" Technical Memo No. 80, Systems Technology, Inc., Inglewood, Calif., March 1961

55. Noble, M., Fitts, P. M. and Warren, C. E. (1955) "The Frequency Response of Skilled Subjects in a Pursuit Tracking Task", Jour. Exp. Psych., 49-249-56, 1955

56. North, J. D. (1952) "The Human Transfer Function in Servo Systems" in Automatic and Manual Control, ed. by A. Tustin, 473-502, London: Butterworths, 1952

57. Ornstein, G. N. (1961) "Applications of a Technique for the Automatic Analog Determination of Human Response Equation Parameters", Report NA61H-1, North American Aviation, Columbus, Ohio, January 1961 (Also available as Ph. D. Thesis, Ohio State University, Department of Psychology, 1960)

58. Phillips, R. S. (1947) "Manual Tracking", in Theory of Servomechanisms, ed. by H. M. James, N. B. Nichols and R. S. Phillips, pp. 360-368, New York: McGraw-Hill, 1947

59. Pitts, W. and McCulloch, W. S. (1947) "The Perception of Visual and Auditory Forms", Bull. Math. Biophysics, 9:124-147

60. Platzer, H. L. (1955) "A Non-Linear Approach to Human Tracking" Franklin Institute Interim Technical Report No. I-2490-1, December 1955

61. Ragazzini, J. R. (1948) "Engineering Aspects of the Human Being as a Servomechanism", Unpublished Paper, presented at the American Psychological Association Meeting, 1948

62. Ragazzini, J. R. and Franklin, G. (1958) Sampled-Data Control Systems, New York: McGraw-Hill, 1958

63. Russell, L. (1951) Characteristics of the Human as a Linear Servo Element, M. S. Thesis, Massachusetts Institute of Technology, May 1951

64. Sheridan, T. B. (1960) "Time Variable Dynamics of Human Operator Systems", MIT Dynamic Analysis and Control Lab Report AFCRC-TN-60-169, Cambridge, Massachusetts, March 1960

65. Siskind, R. K. (1961) "Probability Distributions of Sums of Independent Sinusoids" Technical Memo No. 83, Systems Technology, Inc., Inglewood, California, March 1961

66. Stroud, J. M. (1955) "The Fine Structure of Psychological Time", pp. 174-205 in Information Theory in Psychology, ed. by H. Quastler, Glencoe, Ill.: Free Press

67. Tartakovskii, G. P. (1957) "The Stability of Linear Pulse Systems with Variable Parameters" Radio Engineering and Electronics, 2:15-22, January 1957

68. Tou, J. T. (1959) Digital and Sampled-Data Control Systems New York: McGraw-Hill, 1959

69. Tustin, A. "The Nature of the Operator's Response in Manual Control and its Implications for Controller Design", Jour. Institution Elec. Engrs. (London) 94 (IIA):190-202, 1947

70. Verzeano, M. and Negishi, K. (1960) "Neuronal Activity in Cortical and Thalamic Networks", Journal Gen. Physiology, 43:177-195, July 1960

71. Walston, C. E. and Warren, C. E. (1953) "Analysis of the Human Operator in a Closed-Loop System", Research Bulletin 53-32, Human Resources Center, ARDC, Ohio State University, August 1953

72. Ward, J. R. (1958) The Dynamics of a Human Operator in a Control System: A Study Based on the Hypothesis of Intermittency Ph. D. Dissertation, Department of Aeronautical Engineering, University of Sydney, Australia, May 1958

73. Widrow, B. (1959) "Adaptive Sampled Data Systems: A Statistical Theory of Adaptation", IRE- Wescon Convention Record, 1959, pt. IV:74-85

Appendix 1

INTERMITTENCY IN HUMAN TRACKING: A REVIEW OF THE EVIDENCE

A1.1 Introduction

The purpose of this Appendix is to examine the evidence for the "intermittency hypothesis" in human tracking, i.e., the hypothesis that certain aspects of the input-output behavior of the human operator in a closed loop tracking situation can be represented by a model which includes sampled as well as continuous information. An attempt is made in the following pages to collect and integrate the evidence from a variety of sources (primarily psychological journals) by relating it to a background of engineering concepts on the implications of sampling in a physical system. The major evidence comes from examination of tracking records, from studies connected with the so-called "psychological refractory period," from studies concerned with the perception of discrete events and from reaction time data. Following an examination of the evidence, some deductions are made on the basis of the hypothesis and their implications examined. The concluding portion of the Appendix is a critique of the methods of experimental psychology.

A1.2 Examination of Tracking Records

One of the earliest arguments in favor of the existence of an intermittent process in the human tracker is the appearance of tracking records (Bates, Craik, Tustin, 1947). To many observers the human response curves (and consequently the error curves) have pronounced periodicity, with a strong component in the vicinity of 2 cps, even if this component is not present in the input. Craik (1947, 1948) contends that the spectrum of tracking records actually contains a predominant frequency around 2 cps. Ellson, Hill, and Gray (1947) performed some 15,000 measurements on about 3,000 sections of tracking record and found that over 80% of the

"wavelengths" of the correcting responses of trained trackers ranged between 0.2 and 0.6 seconds. The major observation the above authors and others make is that visual inspection of the tracking data suggests that the tracker waits until an error has increased beyond a given threshhold, and then acts. Increasing the magnification of the display, however, has little or no effect on the frequency of corrections (Craik 1947, Hick 1948). See also Searle and Taylor (1948).

The evidence is not quite as clear as the above paragraph might suggest. Careful spectral analyses of "remnant" data (i.e., the difference signal between human output and outputs of "best fit" linear continuous models) do not reveal strong peaks in the vicinity of 2 cps (Elkind, 1956), unless the input is very difficult (i.e., of high bandwidth). In these cases (such as rectangular input spectra with cutoff beyond 1 cps) there are well defined peaks between 2 and 4 cps. (See discussion in Chapters 4 and 6.) However, in all cases, spectral analysis of human operator outputs, when the input consists of band-limited signals, reveals energy at frequencies beyond the bandwidth of the input. Elkind, McRuer and Krendel (1957) and others suggest that these additional frequencies are due to random signals generated by the operator. The Goodyear studies (1952) were based on the assumption of harmonics of the input frequencies. Ward (1958) did some analog computer studies to support the hypothesis that sampled models could account for the harmonic content of the output signal. Clearly, all three of these hypotheses appear reasonable a priori: in order for frequencies beyond the range of input frequencies to appear in the output, the operator could behave in a nonlinear manner, in an intermittent manner, or he might simply generate and introduce additional frequencies. In any case, the generation of harmonics by the sampler is a possible explanation for the frequency content of human tracking outputs. Furthermore, the discrete appearance of many of the correcting responses would be consistent with the presence of sampling. Typical tracking records are shown in Figures in Chapter 6.

Arguments such as the above led Craik (1947, 1948) in his classical article, to postulate that human tracking consists of a series of "ballistic responses," which are triggered intermittently and run to completion regardless of changes in the input process which may occur in the following 1/2 second, approximately. This suggested explanation was unfortunately linked with the work of Telford (1931) thus was at least partly responsible for the controversy we shall examine in the next section.

A1.3 The "Psychological Refractory Period"

C. W. Telford published an article in 1931 in which he coined the phrase "psychological refractory period" to explain increased reaction times observed in certain kinds of step-tracking situations. We shall examine these briefly below. Basically, Telford suggested that a process analogous to the refractory period in a nerve fiber operates in the central nervous system during tracking tasks. In a nerve fiber refractoriness refers to the brief time interval following the occurrence of a nerve impulse during which the fiber is incapable of responding to any stimulus no matter how intense. Unfortunately, Telford's analogy to tracking processes, where he suggested the "refractory period" was of the order of 1/2 second, was interpreted very rigidly by a number of psychologists, and the resulting controversy has continued to the present day; a symposium on the "Psychological Refractory Period" was held in London in January 1959.

The "PRP" is most commonly used to interpret data from pursuit tracking of steps, such as those typical responses illustrated in Figure A1.1, which are taken from Vince (1949). The subject is asked to track step changes in the input which follow each other at random (or at least unknown and unpredictable) intervals of time. The response in one direction is essentially the same as in the opposite, provided the time interval is sufficiently long, and consists of a reaction time delay of 0.20-0.25 seconds followed by a corrective maneuver which generally resembles the step response

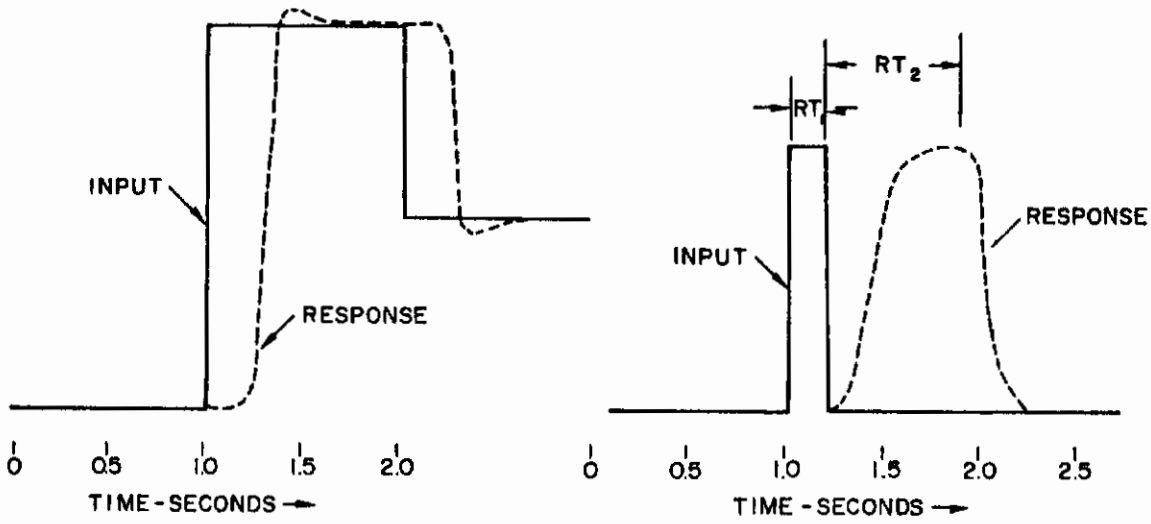


Figure A1.1 Typical Step-Tracking Responses

of a second order system with less than critical damping. However, if the return step follows the first step by less than about 0.5 seconds, the apparent reaction time to the second step is lengthened. A number of similar experiemnts were performed by Vince (1948, 1950). Telford's original work (1931) showed that key-pressing reactions to sounds were slowed significantly if the interval between successive stimuli was 1/2 second or less. Poulton (1950) showed that the stopping of a movement which is carried out zig-zag between two rows of contacts is difficult if less than about 1/2 second is allowed between the stop-signal and the time the contact is reached. We shall consider these and other similar experiments carried out by Davis, Welford, Elithorn and Lawrence, Hick and others, which concerned the "lengthened reaction time to a second stimulus" as part of the problem of the "psychological refractory period" (PRP).

The fact that the corrective movement apparently is unable to utilize visual feedback after it is initiated was shown by Vince (1949) who concluded that movements lasting less than about 0.4 seconds suffered no degradation in accuracy when the stimulus

was removed after the beginning of the movement. The results obtained by Vince were used by Welford (1952) to suggest that the human operator is a "single channel" information processing system and therefore that a new stimulus cannot be dealt with while the system is occupied either with receiving information, processing it, or monitoring the responding movement. Similar suggestions had been made previously in a series of papers by Hick (1948-50).

Much of the discussion which appears in the psychological journals concerning the "PRP" is concerned with how generally applicable some of the above statements are. Davis (1956) indicated that an increase in the second reaction time (RT_2) will occur only if the interval (I) between stimuli is less than RT_1. Vince's own experiments (1950) proved that the crude interpretation of Telford's "PRP" of about 1/2 second was simply not true since under proper conditions corrective movements were possible within approximately 0.25 sec after the arrival of the first stimulus. The delays were considerably reduced when the second movement was in the same direction as the first.

When the two stimuli follow each other by less than about 0.4 seconds, they tend to be "grouped" and reacted to as a single stimulus, according to Vince (1950) and Welford (1959). However, Halliday, Kerr, and Elithorn (1960) find some exceptions to the expected "grouping" which leads them to question the entire theory.

It appears clear from a reading of the actual experimental data that the basic result is not being questioned, i.e., that when two stimuli follow each other by less than about 0.4 seconds the RT to the second one is lengthened, as indicated by the approximate equation of Welford (1952):

$$RT_2 = RT_n + RT_1 - I \qquad \text{for} \qquad I < RT_1$$

where RT_n is the "normal" reaction time and I is the interval between stimuli. Unfortunately, the interpretations of these basic

data are not so clear. Hick (1948), Poulton (1950) and later Elithorn and Lawrence (1955) attempt to explain the effect by introducing a number of new hypothetical concepts, such as the "expectation" of the subject for stimulus 2 given stimulus 1, or the "inhibition" of a stimulus. Broadbent (1958) discusses the "expectancy" theory at some length. A number of writers including Hich (1948), Hick and Bates (1950), Welford (1952 and 1959) and Davis (1957) argue for the "single channel" theory of data processing.

The philosophical discussions concerning the "psychological refractory period" and its interpretation do not seem as important to the present author as the basic data being discussed. The basic results concerning the apparent lengthening of RT's are consistent with a hypothesis that a sampled model can produce similar results. As has been stated above, this model will make no pretense of being unique, but the evidence summarized in this section makes it appear reasonable.

A1.4 Perception of a Number of Discrete Events

Additional supporting evidence for the "intermittency hypothesis" comes from a series of experiments concerned with the perception of the number of a series of closely spaced stimuli, which are denoted by the title "Numerosity" in the psychological literature (Cheatham and White 1952, 1954; White, Cheatham and Armington, 1953). The subjects were presented with a series of light flashes at rates from 10 to 30 flashes/sec, the light-dark ratio being selected so as to avoid fusion. The rate reported by the subjects did not exceed 6 to 8/sec regardless of the actual rate. A similar result was produced with auditory stimuli. Apparently 1 stimulus per 100 millisec is the limit of subjective perception, even though these rates are well below the physiological limits for the sense organs.

Such a grouping of discrete stimuli into "time quanta" is an expected consequence of sampling processes which admit new

information to the central processors only intermittently. The apparent rate of 10/sec (rather than 2-3/sec as observed in tracking) is reasonable since no monitoring of a controlled muscular response is required in this case. Broadbent (1958, p. 280) has suggested that there may be a quantizing of perception into samples about 1/3 sec in length if monitoring of response is required, shorter otherwise. A similar theory was advanced by Stroud (1955) who suggested that subjective or psychological time is discrete rather than continuous, calling these units "moments" of experience. While these suggestions are largely speculative in nature, they grow from the same background of accumulated experimental evidence and thus they form part of the rationale for using a sampled-data model to represent the human operator.

A1.5 Prediction of Target Motion

As discussed in Chapter 1, one of the features which characterizes the human operator is his ability to "predict," i.e., to extrapolate from the present and past history of an event to the future over a limited range of time. Thus, in pursuit tracking for example, the subject may see the target "vanish behind a cloud" and be instructed to continue tracking in order to pick up the target later. Clearly, such behavior requires some ability to formulate and store the laws of target motion, as understood by the tracker, for use in extrapolation. Is this behavior consistent with the sampling hypothesis?

One of the characteristics of sampling systems discussed above in Section 3.2 is that if such a system has a continuous output (as the human operator does) then the system must contain a data reconstruction element, a digital-to-analog converter which uses values of a signal obtained at sampling instants and produces an output at all values of time. Such "hold" circuits may produce outputs having discontinuities at the sampling instants, but if they are followed by low-pass elements the system output may be continuous. Consider now the action of a first-order hold upon a sampled signal. This circuit

extrapolates between sampling instants with a constant slope, as determined by the two past samples, as indicated in equation A1.1 below:

$$(A1.1) \quad f(t) = f(t_n) + \frac{f(t_n) - f(t_{n-1})}{T} (t - t_n) ;$$

$$t_n \leq t < t_{n+1}$$

Thus, we would expect that a system containing sampling followed by a first-order hold would be capable of extrapolating perfectly a constant-rate input, but that it would always undershoot a constantly accelerating input and overshoot a constantly decelerating input.

Let us now examine the psychological evidence in this area. Much of the work on extrapolation and anticipation has been done by Gottsdanker (1952, 1955, 1956) and Poulton (1950, 1952, 1957). Gottsdanker (1952a) finds that extrapolation of constant rate movements can be continued by many subjects with average deviations of less than 1% from the guided rate; that accelerating movements tend to be under-estimated and decelerating movements over-estimated. In this early work, Gottsdanker found no evidence that his subjects were able to use the acceleration information in the signal for extrapolation. Later studies (1956), Poulton's work (1957c) as well as some early studies of tracking behavior such as those of Taylor and Birmingham (1948) tend to indicate that with sufficient practice, operators are capable of using some acceleration information. With relatively little training, however, operators will tend to extrapolate at the average rate of target movement during the last 1/2 sec or so of target visibility. It is interesting to note that Hill, Gray and Ellson (1947) in their analysis of tracking records found that trackers will tend to accept above-threshold position errors for several seconds if they have achieved a near perfect match in rate.

Prediction and anticipation are very complex problems. They have been introduced here only to illustrate that results in this field are not in disagreement with the intermittency hypothesis.

A1.6 Precognitive Tracking

This name is given by McRuer and Krendel (1957) to the tracking of targets which follow simple analytical movements, such as sinusoidal motion, and thus are perfectly predictable. While this problem is part of the general area of prediction, it has been separated from it since we refer here to the matching of the movement of a visible target, rather than prediction of the motion of an invisible one. Typically, trackers here will begin to track sine waves as they do any unknown signal, with movements which appear intermittent, attempting to match rate, and gradually, with training, learning to match acceleration. Following the early phase of tracking, the operator will tend to lock into synchronism with the input and neglect amplitude errors for a short time while he adjusts his frequency, finally attempting to adjust amplitude errors to zero as well. Noble, Fitts, and Warren (1955) had subjects tracking sine waves from very low frequencies up to 4 cps and found no average phase shift when the subject could stay in synchronism, no obvious attenuation and no obvious resonances in their response. The time-on-target, however, decreased monotonically with increased frequency. Poulton (1957a) in similar experiments with a simple harmonic input of 1 cps found that constant errors in amplitude or phase would arise slowly, might remain undetected for a number of cycles, and finally be corrected. Apparently, the tracking is considered satisfactory if rate and acceleration are matched.

The relevance of these experiments to this Appendix arises from possible interpretations of the results of Noble, Fitts, and Warren mentioned above. The increase in average tracking error as a function of frequency in this experiment was due to the fact that larger phase errors would appear before being corrected at high than at low frequencies. The authors point out that the subject

apparently requires a relatively constant time to observe and correct phase errors, irrespective of input frequency. This fact would be in agreement with a sampled-data hypothesis. Since in this kind of experiment subjects are capable of continuing their tracking for a short time with their eyes closed, it is suggested by the authors that the operator samples and predicts, comparing prediction with actual performance, so that his motion is not based on the stimulus directly but on a continuously predicted response which is corrected by observed errors.

A1.7 Other Implications of the Intermittency Hypothesis

The major psychological background and justification for the intermittency hypothesis has been given in the preceding portions of this Appendix. In this section we shall survey briefly other relevant material which pertains to various deductions made on the basis of the hypothesis and their verification.

(a) Periodicities in task performance. Stroud's "Moment" theory of psychological time (1955) has been mentioned previously. An experiment reported by Augenstine (1955) concerned with scanning of rows of letters to find a number and the deciphering of anagrams shows evidence of the solution times clustering at certain times, as if the solution times were quantized.

(b) Aided-tracking time constants. Aided tracking refers to providing the operator with a control device whose output contains not only a position term proportional to the operator's handle displacement, but also a rate and sometimes an acceleration term. Clearly, the choice of the position, rate, and acceleration constants is a typical control system design problem; in tracking systems it has been solved primarily by experiment since purely analytical approaches have sometimes given misleading results, such as that of Phillips (1947). Searle (1951) used the assumption of intermittent corrections at 1/2 sec intervals, along with several

other assumptions, to design some aiding constants and obtained excellent agreement with experimental results for his particular experiment.

(c) Explanation of remnant data. As outlined in Chapter 1, linear continuous models of human operators are usually obtained by measurements of cross-correlation functions (or cross-spectral densities) so designed that the resulting model best approximates the operator in the mean square sense. The difference between actual human output and model output was termed the "remnant" by Tustin (1947) and the name is still used. Elkind (1956) and McRuer and Krendel (1957) have analyzed several possible explanations for the remnant term. One of these explanations assumes that the human's output motion $x(t)$ can be approximated by a series of discrete steps, each of which includes a part linearly related to the forcing function and a "noise component" or random error which is not linearly related to the input in any deterministic way. While this model cannot be considered completely satisfactory even on an a priori basis, its analysis produced results within an order of magnitude of experimental results, and having the right general trends.

(d) Eye movements in tracking. This is another extremely complex field, partly due to the difficulties of measurement and instrumentation and partly due to the complex mechanism of the retina. There is a wealth of published material on this and related fields. For example, Tinker (1958) in his review of recent work on eye movements in reading lists 72 references. Only two aspects of the work relevant to this study (because visual inputs to human operators are considered) will be mentioned. First, eye movements are always jerky, consisting of rapid position changes followed by fixation periods of various durations. During free search (Ford, White and Lichtenstein, 1959) there tend to be 2 to 4 fixations per second, of about 1/4 sec duration each. Rather than making a smooth circular movement in radar tracking for example, the eyes tend to approximate the circle by jerky movements (Grotewohl, 1952). The ability of the

eyes to move rapidly from position to position is due to the fact that they have probably the highest force to inertia ratio (from driving muscles to the organ itself) of all body parts. Secondly, the eye movements (known as "saccadic" movements) are remarkably constant in velocity. Tinker (1958) reports that during 90° sweeps at rates from 2 to 3 per second there was no observable change in velocity between reversals of direction. Apparently such constant velocity tracking is possible through retinal control alone.

(e) Delayed perceptual feedback. If one assumes that central processes are intermittent in their operation because of the Welford "single channel of information processing" theory, then the length of the sampling interval is determined by the time required from perception through output monitoring. One could deduce from such a hypothesis that if the feedback monitoring is delayed by one or more "sampling periods" that performance would be extremely difficult. It can be anticipated that the subject would tend to slow or continue to repeat his performance in order to allow for a lengthening of the effective "sampling interval" to include the additional delay. As a matter of fact, experimental data confirm these deductions. If a delay of 1/4 sec is introduced between the external world and a subject's ears very characteristic changes occur. With simple tasks such as key tapping the subject tends to hold the key down longer, to press harder, and to tap more times than requested. With attempts to speak, the subject typically increases his vocal intensity, his speech slows down, and he tends to repeat sounds (Chase, 1959). These facts are now well documented and there is an extensive bibliography available on delayed auditory feedback. Recent studies of delayed visual feedback using video-tape recording and special TV presentations so that the subject sees only a delayed picture of his performance have similar results, (Smith, McCrary, and Smith, 1960). Each task takes longer, the increase in time being related to the difficulty of the task. Tasks, such as writing a letter of the alphabet, normally requiring about 0.5 seconds, required about 1.2 seconds. More difficult tasks required as much as 10 times

normal times for execution. Furthermore, there was a definite tendency to repeat movements and duplicate errors twice in succession, which is analogous to the sound repetition occurring in delayed auditory feedback.

A1.8 Summary and Critique of the Psychological Literature

In the preceding pages a brief survey of the psychological literature related to intermittency in human behavior has been presented. The emphasis has been on material growing directly from or closely related to the tracking situation to illustrate that ample experimental justification is available for the use of a sampled-data model for the human operator in a tracking loop. The evidence summarized has included examination of typical tracking records; the "psychological refractory period" literature, with a particular emphasis on the results of step-tracking experiments; a review of the "numerosity" experiments concerned with the perception of a number of discrete events in succession; a brief study of the relevance of prediction experiments and pre-cognitive tracking data to the intermittency problem; and some sketches of other related phenomena such as eye movement studies, delayed feedback experiments, and some preliminary studies of remnant data. With the exception of the Elkind-McRuer-Krendel remnant analysis and the preliminary and unquantitative work reported in the dissertation of Ward (1958), the intermittency hypothesis has not been used as the foundation for construction of mathematical models of the human operator in the past.

While the list of references and number of discussions of the general topics of this chapter in the psychological literature is quite impressive, one is left with a singularly unsettled feeling after reading it. In much of the literature there is more heat than light and considerable careful reading is required to separate the significant from the trivial in the refutations and counter-refutations of previous experiments. To this writer, it appears that the literature of experimental psychology suffers from the following problems:

(1) Premature publication. Many papers are published without a careful analysis of the limitations and range of applicability of the results. Consequently, future papers in the same field often devote considerable space to analyzing the limitations of previous papers. While publication of preliminary results is justifiable in any scientific field, in psychology the situation goes to extremes. For example, Battig et al (1954) examined the effect of intermittent displays on tracking proficiencies, and concluded that there is an optimum display presentation frequency and that higher frequencies than the optimum may actually decrease "time-on-target" scores. Within a year (Battig, 1955) a new paper indicated that the so-called optimum was an artifact of the experimental procedure and was due to brightness variations rather than intermittent presentation. Garvey, Knowles and Newlin (1956) use most of a paper on prediction of radar target track positions to refute another paper which also used "prediction" in the title, but actually was a test of the subjects ability to remember a previously shown trajectory.

(2) The experiments do not justify the claims. The controversy over the "psychological refractory period" is an illustration of this problem. The early papers by Hick, Craik, and Vince presented results of step-tracking experiments and used them to substantiate the PRP. Poulton (1950) showed that a significant part of the PRP was due to causing the second movement to be in the opposite direction to the first; if the movement is in the same direction as the first, the second reaction time shows very little PRP. It would appear that steps in both directions should have been used before publication. Vince (1948) indicated that ungraded or unskilled responses, such as key tapping, appear to be exceptions to the PRP. Hick (1949) repeated the experiment and indicated that Vince's findings were not an exception. In recent years the controversy has concerned the problem of "grouping," i.e., the condition under which two stimuli following each other very closely are reacted to as a single stimulus. Vince (1950) indicated such an effect; Elithorn and Lawrence (1955) question its existence. Arguments by Hick and Davis have also entered the scene. Now this is an

experiment involving human subjects and the procedure must be carefully analyzed before claiming conclusions. Welford (1959) analyzes some of the past experiments in this area and shows that in some cases the subjects knew that stimuli came in pairs and thus could "prepare" for the second one, while in others they did not. As another example, Gottsdanker (1952) published material in which he claimed that subjects are not capable of extrapolating target acceleration but only velocity. As Poulton (1957) points out, many other experiments have shown that with sufficient practice and knowledge of results, acceleration can be used by trackers.

(3) Excessive categorizing. There appears to be a prevalent tendency for experimental psychologists to look for "laws" of human behavior; even on the basis of fragmentary evidence these "laws" are given names which are seldom placed in quotation marks. Consequently, as pointed out by many writers in the field of semantics (such as Hayakawa, Lee, Rapoport and others), these constructs tend to become stratified and much discussion ensues concerning their validity. Telford's "psychological refractory period" is a case in point; Elithorn and Lawrence's "central inhibition" and "internal anticipatory set" are another. In the discussions concerning the validity of the "laws" some authors seem to forget that hypotheses are not substantiated by calling them "laws" or "mechanisms" but by the results of experiments. In reading some of the defenses of the "laws" of human behavior, one is reminded of the words of Lewis Carroll:

> "When I use a word," Humpty Dumpty said, rather in a scornful tone, "it means just what I choose it to mean -- neither more nor less."

BIBLIOGRAPHY FOR APPENDIX I

1. Adams, Jack A. (1961) "Human Tracking Behavior," Psych. Bull. 58:55-79, January 1961.

2. Augenstine, L. G. (1955) "Evidences of Periodicities in Human Task Performance," in Information Theory in Psychology, ed. by H. Quastler, Glencoe, Ill.: Free Press, 1955.

3. Bates, J. A. V. (1954) "Some Characteristics of the Human Operator," Jour. I.E.E. (London), 94:298-304, pt. IIA, 1947.

4. Battig, F.W., et al (1954) "Tracking and Frequency of Target Intermittence," J. Exp. Psych., 47:309-314, 1954.

5. Battig, F. W., Voss, J. F., and Brogden, W. J. (1955) "Effect of Frequency of Target Intermittence upon Tracking," J. Exp. Psych., 49:244-248, 1955.

6. Bennett, C. A. (1956) "Sampled-Data Tracking: Sampling of the Operator's Output," J. Exp. Psych., 51:429-438, 1956.

7. Broadbent, D.E. (1958) Perception and Communication, 268-296, 268-296, New York: Pergamon Press, 1958.

8. Broadbent, D. E. (1952) "Speaking and Listening Simultaneously" J. Exp. Psych., 43:267-273, 1952.

9. Callaway, E. and Yeager, C. L. (1960) "Relationship between Reaction Time and Electroencephalographic Alpha Phase," Science, 132:1765-1766, December 1960.

10. Chase, R. A., Sutton, S., First, D. (1958) "Bibliography on Delayed Auditory Feedback," Research Report, Communications Laboratory, Columbia University College of Physicians and Surgeons, New York, December 1958.

11. Chase, R. A. et al (1959) "Comparison of the Effects of Delayed Auditory Feedback on Speech and Key Tapping," Science, 129:903-905, April 1959.

12. Cheatham, P.G., and White, C. T. (1952) "Temporal Numerosity: I. Perceived Number as a Function of Flash Number and Rate," J. Exp. Psych., 44:447-451, 1952.

13. Cheatham, P. G. and White, C. T. (1954) "Temporal Numerosity: II. Auditory Perception of Number," J. Exp. Psych., 47:425-428, 1954.

14. Chernikoff, R., Birmingham, H. P., and Taylor, F. V. (1955) "A Comparison of Pursuit and Compensatory Tracking under Conditions of Aiding and No Aiding," J. Exp. Psych., 49:55-59, 1955.

15. Chernikoff, R. and Taylor, F. V. (1952) "Reaction Time to Kinaesthetic Stimulation Resulting from Sudden Arm Displacement," J. Exp. Psych., 43:1-8, 1952.

16. Conklin, J. E. (1957) "Effect of Control Lag on Performance in a Tracking Task," J. Exp. Psych., 53:261-268, 1957.

17. Conrad, R. (1954) "Missed Signals in a Sensori-Motor Skill" J. Exp. Psych., 48:1-9, 1954, Also published as APU report 187/53, Medical Research Council (England), 1953.

18. Conrad, R. (1956) "The Timing of Signals in Skill," J. Exp. Psych. 51:365-370, 1956.

19. Craik, K. J. W. (1947) "Theory of the Human Operator in Control Systems: I. The Operator as an Engineering System," Brit. J. Psych., 38:56-61, 1947.

20. Craik, K. J. W. (1948) "Theory of the Human Operator in Control Systems: II. Man as an Element in a Control System" Brit. J. Psych. 38:142-148, 1948.

21. Davis, R. (1956b) "Comments on 'Central Inhibition: Some Refractory Observations' by Elithorn and Lawrence," Quart. J. Exp. Psych., 8:39, 1956.

22. Davis, R. (1957) "The Human Operator as a Single Channel Information System," Quart. J. Exp. Psych., 9:119-129, 1957.

23. Davis, R. (1956a) "The Limits of the Psychological Refractory Period," Quart. J. Exp. Psych., 8:24-38, 1956.

24. Elkind, J. I. (1956) "Characteristics of Simple Manual Control Systems," Tech. Report No. 111, M.I.T. Lincoln Laboratory, April 1956.

25. Elkind, J. I. (1953) "Tracking Response Characteristics of the Human Operator," Human Factors Operation Research Laboratories, U. S. Air Force ARDC, Report HFORL 40, September 1953.

26. Elithorn, A. and Lawrence, C. (1955) "Central Inhibition: Some Refractory Observations," Quart. J. Exp. Psych., 7:116-127, 1955.

27. Ford, A., White, C. T. and Lichtenstein, M. "Analysis of Eye Movements During Free Search," J. Opt. Soc. of Am., 49:287-292, March 1959.

28. Garvey, W. D. and L. L. Mitnick (1957) "An Analysis of Tracking Behavior in Terms of Lead-Lag Networks," J. Exp. Psych. 53:372-378, 1957.

29. Garvey, W. D., Knowles, W. B. and Newlin, E. P. (1956) "Prediction of Future Position of a Target Track on Four Types of Displays," Naval Research Laboratory: NRL Report 4721, Washington, D. C., April 1956.

30. Gerathewohl, S. F. (1952) "Eye Movements During Radar Operations," J. Aviation Med., 23: 597-607, December 1952.

31. Goodyear Aircraft Corp. (1952) "Final Report: Human Dynamics Study," Report GER-4750, Akron, Ohio, April 1952.

32. Gottsdanker, R. M. (1956) "The Ability of Human Operators to Detect Acceleration of Target Motion," Psych. Bull., 53:477-487, 1956.

33. Gottsdanker, R. M. (1952a) "The Accuracy of Prediction Motion" J. Exp. Psych., 43:26-36, 1952.

34. Gottsdanker, R. M. (1955) "A Further Study of Prediction Motion," Amer. J. Psych., 68:432-437, 1955.

35. Gottsdanker, R. M. (1952b) "Prediction of Motion with and without Vision," Amer. J. Psych., 65: 533-543, 1952.

36. Halliday, A. M., Kerr, M. and Elithorn, A. (1960) "Grouping of Stimuli and Apparent Exceptions to the Psychological Refractory Period," Quart. J. Exp. Psych., 12:72-89, May 1960.

37. Hick, W. E. and Welford, A. T. (1956) "Comments on 'Central Inhibition: Some Refractory Observations' by Elithorn and Lawrence," Quart. J. Exp. Psych., 8:39-41, 1956.

38. Hick, W. E. (1947) "Discontinuity and the Human Operator of Machine Controls" Medical Research Council (England), APU Report 67, June 1947.

39. Hick, W. E. (1948) "The discontinuous Functioning of the Human Operator in Pursuit Tasks," Quart. J. Exp. Psych., 1:36-51, 1948.

40. Hick, W. E. and Bates, J. A. V. (1950), The Human Operator of Control Mechanisms, London: Ministry of Supply, 1950.

41. Hick, W. E. (1949) "Reaction Time for the Amendment of a Response," Quart. J. Exp. Psych., 1:180-192, 1949.

42. Hick, W. E. (1950) "The Threshold for Sudden Changes in the Velocity of a Seen Object," Quart. J. Exp. Psych., 2:33-41, 1950.

43. Hill, H., Gray, F., and Ellson, D. G. (1947) "Wavelength and Amplitude Characteristics of Tracking Error Curves," AAF-AMC Engr. Report TSEAA-694-2D, Wright Field, Ohio, 22 April 1947.

44. Holland, J. G. (1958) "Human Vigilance," Science 128:61-67, July 1958.

45. Klemmer, E. T. (1957) "Simple Reaction Time as a Function of Time Uncertainty," J. Exp. Psych., 54:195-200, 1957.

46. Licklider, J. C. R. (1960) "Quasi-Linear Operator Models in the Study of Manual Tracking," in Developments in Mathematical Psychology, ed. by R. D. Luce, 171-280, Glencoe, Ill.: Free Press, 1960.

47. McRuer, D. T. and Krendel, E. (1957), "Dynamic Response of Human Operators," WADC Technical Report 56-524, October 1957, (Particularly pp. 91-106 discuss a "discrete step" model).

48. Noble, M., Fitts, P. M. and Warren, C. E. (1955) "The Frequency Response of Skilled Subjects in a Pursuit Tracking Task," J. Exp. Psych., 49:249-56, 1955.

49. North, J. D. (1952) "The Human Transfer Function in a Servo-Systems," 473-502, in Automatic and Manual Control, ed. by A. Tustin, London: Butterworths, 1952.

50. Phillips, R. S., "Application of the Rms-Error Criterion in Determining the Best Aided-Tracking Time Constant," 363-368, in Theory of Servomechanisms, ed. by H. M. James, N. B. Nichols, and R. S. Phillips, New York: McGraw-Hill (1947).

51. Poulton, E. C. (1952b) "The Basis of Perceptual Anticipation in Tracking," Brit. J. Psych., 43:295-302, 1952.

52. Poulton, E. C. and Gregory, R. L. (1952a) "Blinking During Visual Tracking," Quart. J. Exp. Psych., 4:57-65, 1952.

53. Poulton, E. C. (1957b) "Learning the Statistical Properties of the Input in Pursuit Tracking," J. Exp. Psych., 54:28-32, 1957.

54. Poulton, E. C. (1950) "Perceptual Anticipation and Reaction Time," Quart. J. Exp. Psych., 2:99-112, 1950.

55. Poulton, E. C. (1957c) "On Prediction in Skilled Movements," Psych. Bull., 54:467-478, 1957.

56. Poulton, E. C. (1957a) "On the Stimulus and Response in Pursuit Tracking," J. Exp. Psych., 53:189-194, 1957.

57. Searle, L. V. (1951) "Psychological Studies of Tracking Behavior: Part IV. The Intermittency Hypothesis as a Basis for Predicting Optimum Aided-Tracking Time Constants," NRL Report 3872, Naval Research Laboratory, Washington, D.C. October 1951.

58. Searle, L. V. and Taylor, F. V. (1948) "Studies of Tracking Behavior: I. Rate and Time Characteristics of a Simple Corrective Movements," J. Exp. Psych., 38:615-631, 1948.

59. Smith, W. M., McCrary, J. W. and Smith K. U. (1960) "Delayed Visual Feedback and Behavior," Science, 132:1013-1014, October 1960.

60. Stroud, J. M. (1955) "The Fine Structure of Psychological Time," 174-205, in Information Theory in Psychology, ed. by H. Quastler, Glencoe, Ill.: Free Press, 1955.

61. Succi, G. J., Davidoff, M. D. and Surwillo, W. W. (1960) "Reaction Time as a Function of Stimulus Information and Age," J. Exp. Psych., 60:242-244, October 1960.

62. Taylor, V. F. and Birmingham. H. P. (1948) "Studies in Tracking Behavior: II. The Acceleration Patterns of Quick Manual Corrective Responses," J. Exp. Psych., 38:783-795,1948.

63. Telford, C. N. (1931) "Refractory Phase of Voluntary and Associative Responses," J. Exp. Psych., 14:1-35, 1931.

64. Tinker, M. A. (1958) "Recent Studies of Eye Movements in Reading," Psych. Bull., 55:215-231, 1958.

65. Tustin, A. (1947) "The Nature of the Operator's Response in Manual Control and Its Implications for Controller Design," Jour. Instn. of Electr. Engrs., (London), 94(II-A):190-202, 1947.

66. Vince, M. A. (1949b) "Corrective Movements in a Pursuit Task," Quart. J. Exp. Psych, 1:85-103, 1949.

67. Vince, M. A. (1948) "The Intermittency of Control Movements and the Psychological Refractory Period," Brit. J. Psych., 38:149-157, 1948.

68. Vince, M. A. (1949b) "Rapid Response Sequences and the Psychological Refractory Period," Brit. J. Psych. 40:23-40, 1949.

69. Vince, M. A. (1950) "Some Exceptions to the Psychological Refractory Period in Unskilled Manual Responses," Medical Research Council (England), APU Report 124, 1953.

70. Ward, John R. (1958) The Dynamics of a Human Operator in a Control System: A Study Based on the Hypothesis of Intermittency, Ph.D. Dissertation, University of Sydney, Australia, May 1958.

71. Welford, A. T. (1952) "The 'Psychological Refractory Period' and the Timing of High-Speed Performance: A Review and a Theory," Brit. J. Psych., 43:2-19, 1952.

72. White, C. T., Cheatham, P. G. and Armington, J. C. (1953) "Temporal Numerosity: II. Evidence for Central Factors Influencing Perceived Number," J. Exp. Psych., 46:283-287, 1953.

Appendix 2

STATISTICAL ANALYSIS OF SAMPLED-DATA SYSTEMS

A2.1 Introduction

The purpose of this appendix is to present the analytical background for the study of linear sampled-data systems with stationary random inputs. The emphasis will be on the development of input-output relations as well as relations for computation of mean squared error and power spectral density at various points in a feedback sampled-data system. Problems concerning optimum filtering of sampled-data will not be discussed in this section; they are covered in References 1-5, 10, 11. The present work is based primarily on the papers by Franklin (1955), Trembath (1957), Johnson (1957) and Mori (1958).

The presence of the sampler in sampled-data systems renders signals in such systems non-stationary. Consequently, the statistical characteristics of even a linear, constant coefficient sampled-data system differ in their ensemble and time-averages. In general, the statistical characteristics vary periodically at the sampling frequency since the sampler can be considered to be a periodically time-varying amplifier.

A2.2 Probability Distribution and Density Functions

Consider a stationary random process (with the ergodic property) of which the time function $r(t)$ is a member. If we sample this function periodically (every T seconds) for γ seconds, we obtain a sequence $r^*(t)$, as indicated in Figure A2.1. Let the height of each pulse be given by $r(nT)/\gamma$ so that the area of the n-th pulse, $A_n = \frac{r(nT)}{\gamma}(\gamma)$, i.e. the areas will not depend on the pulse width.

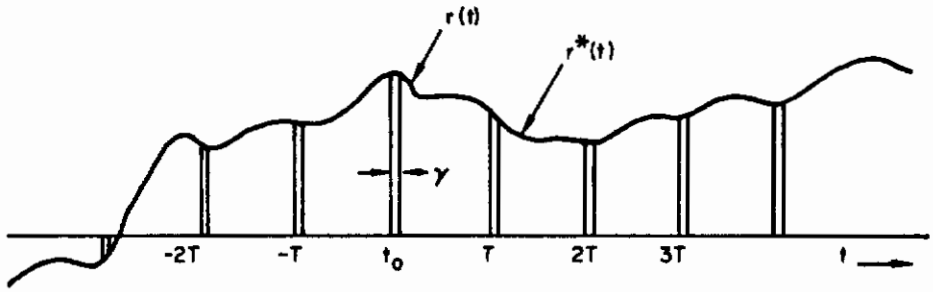


Figure A2.1
Sampled Random Function

Now if the time origin t_o of the modulated pulse train is random (i.e. the initial phase δ is uniformly distributed in the interval (0, T)) then the sampled process (of which the sequence $r^*(t)$ is a member) will be stationary and ergodic if the continuous process is stationary and ergodic, provided that $r^*(t)$ is assumed to exist only at the sampling instants. Instead of making this assumption and proceeding directly to time averages (as Franklin does) we shall begin with ensemble averages.

Since $r(t)$ is assumed to be stationary (and at this point not necessarily ergodic) it can be seen that

$$\text{(A2.1)} \qquad \text{Prob}\left[x < r(t) \leq x + dx\right] = \text{Prob}\left[x < r^*(t) \leq x + dx\right]$$

for $nT \leq t \leq nT + \gamma$; that is the probability density function of the sampled random function evaluated at the sampling instants is the same as that of the original continuous function $r(t)$.

Symbolically

(A2.2) $$f_{r^*(nT)}(x) = f_r(x)$$

where $f_r(x) \triangleq$ first probability density function of the continuous process $r(t)$

$f_{r^*(nT)}(x) \triangleq$ first probability density function of sampled process evaluated at sampling instants.

However, since $r^*(t) = 0$ for $nT + \gamma < t < (n+1)T$, we have

(A2.3) $$f_{r^*}(x_1 t) = \delta(x), \qquad nT + \gamma < t < (n+1)T$$

i.e. the probability that $r^*(t)$ has any value other than zero between sampling instants is zero.

The corresponding probability distribution functions are

(A2.4) $$F_{r^*}(x,t) = \text{Prob}\left[r(t) \leq x\right] = F_r(x) \text{ for } nT \leq t \leq nT + \gamma$$

(A2.5) $$F_{r^*}(x,t) = u(x); \qquad nT + \gamma < t < (n+1)T$$

where F denotes the probability distribution function and $u(x)$ is the unit step function, indicating that the probability is always 1 that $r^*(t)$ will be less than or equal to any value x in the sample space. Thus, since the probability distribution functions are dependent on the time origin, the sampled random process is non-stationary.

Let us now fix our attention at the sampling instants, nT, and assume that the function $r(t)$ changes negligibly during the pulse-on-time γ. Then, from equation A2.2, $f_{r^*}(x, nT) = f_r(x,t)$. Now, the mean value of the continuous process can be written as

$$E\left\{r(t)\right\} = \int_{-\infty}^{+\infty} x f_r(x,t)\, dx$$

$$= \int_{-\infty}^{+\infty} x f_{r^*}(x,nT)\, dx = E\left\{r^*(nT)\right\}$$

That is, the ensemble average of the continuous function equals that of the sample function at the sampling instants. Now, if the continuous process is both stationary and ergodic, then the ensemble and time averages are equivalent, i.e.

$$(A2.7) \qquad E\left\{r(t)\right\} = \int_{-\infty}^{+\infty} x f_r(x,t)\, dx = \lim_{T \to \infty} \int_{-T}^{T} r(t)\, dt$$

Now, since the time average of a sampled process can be written as

$$(A2.8) \qquad \overline{r^*(nT)} = \lim_{N \to \infty} \frac{1}{2N+1} \sum_{n=-N}^{+N} r^*(nT)$$

and therefore

$$(A2.9) \qquad \overline{r^*(nT)} = r(nT)$$

where the bar denotes time averaging. Therefore, averages of the sampled sequences are equivalent to continuous time or ensemble averages for ergodic stationary processes. It should be noted, however, that attention is being focussed only on the sampling instants nT. The complete time average of the sampled signal $r^*(t)$ will clearly differ from that of $r(t)$ since the sampled signal contains less average power. Thus, the time average of the

sampled signal is given by $\overline{r^*(t)}$ (rather than $\overline{r^*(nT)}$ for the sequence):

$$(A2.10)\quad \overline{r^*(t)} = \lim_{N \to \infty} \frac{1}{(2N+1)T} \sum_{n=-N}^{N} \gamma r^*(nT) = \frac{\gamma}{T}\ r^*(nT) = \frac{\gamma}{T}\,\overline{r(t)}$$

The factor γ/T represents the fraction of the time when the sampled signal exists, and thus represents a "duty factor". Similarly, the mean-square value of the pulse-amplitude modulated signal $r^*(t)$ differs from the mean square value of the continuous signal from which it was obtained by the same "duty factor", i.e.:

$$(A2.11)\qquad \overline{\left[r^*(t)\right]^2} = \frac{\gamma}{T}\quad \overline{r^2(t)}$$

A2.3 Correlation Functions

As in the previous section, a distinction must be made between signals in sampled-data systems on the one hand and number sequences on the other hand. As mentioned previously, if the continuous signal $x(t)$ is a member of a stationary (and ergodic) random process, the sampled sequences $x(nT)$ will likewise be stationary (and ergodic). In general, however, the pulse modulated signal $x^*(t)$ and signals obtained from linear operations on it will be non-stationary. In other words, signals in sampled-data systems with stationary inputs will be stationary only at the sampling instants. Therefore, ensemble-averaged and time-averaged correlation functions will generally not be equal.

Consider the open-loop sampled-data system shown in Figure A2.2 where $r(t)$ can be considered as stationary and ergodic, the sampler is periodic and closes for γ seconds every T seconds,

and $h(\tau)$ is the impulse response of the linear continuous system.

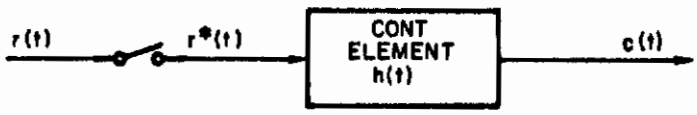


Figure A2.2
Open-Loop System

Let the autocorrelation function of the input be $R_{rr}(\tau)$. Then the autocorrelation of the sequence $\{r(nT)\}$ is defined as

$$R_{rr}(nT) = \lim_{N \to \infty} \frac{1}{2N+1} \sum_{k=-N}^{+N} r(kT)\, r(kT - nT) \tag{A2.12}$$

Let us now examine the relation of this function to the time-averaged autocorrelation of $r^*(t)$, which is defined as

$$R_{r^*r^*}(\tau) = \lim_{T_o \to \infty} \frac{1}{2T_o} \int_{-T_o}^{T_o} r^*(t)\, r^*(t - \tau)\, dt \tag{A2.13}$$

Since pulse-amplitude modulated signal $r^*(t)$ is given by

$$r^*(t) = \sum_{n=-\infty}^{+\infty} r(nT)\, p(t - nT) \tag{A2.14}$$

where $p(t)$ is a pulse of width γ occurring at $t = 0$, we can substitute this relation in (A2.13) to obtain

$$R_{r^*r^*}(\tau) = \lim_{T_o \to \infty} \frac{1}{2T_o} \int_{-T_o}^{T_o} \sum_n r(nT)\, p(t-nT) \sum_m r(mT) p(t+\tau-mT)\, dt$$

(A2.15)

If we let $m = n + k$ and interchange summation and integration operations where permissible we obtain

$$(A2.16) \quad R_{r^*r^*}(\tau) = \sum_{k=-\infty}^{+\infty} \lim_{N \to \infty} \frac{1}{2N+1} \sum_{n=-N}^{+N} r(nT)\, r(nT+kT) \left\{ \lim_{T_o \to \infty} \frac{1}{2T_o} \int_{-T_o}^{T_o} p(t-nT)\, p(t-nT+\tau-kT)\, dt \right\}$$

Now, since $p(t-nT)$ is a periodic function with period T, it can be averaged simply over one period T. Therefore, if we let $t_1 = t - nT$, and average from $-T/2$ to $+T/2$; equation (A2.16) becomes

$$(A2.17) \quad R_{r^*r^*}(\tau) = \sum_{k=-\infty}^{+\infty} \lim_{N \to \infty} \frac{1}{2N+1} \sum_{n=-N}^{+N} r(nT) r(nT+kT) \left\{ \frac{1}{T} \int_{-T/2}^{T/2} p(t_1)\, p(t_1+\tau-kT)\, dt_1 \right\}$$

The second summation is equal to the input autocorrelation evaluated at the sampling instants, by equation (A2.12), and thus (A2.17) becomes:

$$(A2.18) \qquad R_{r^* r^*}(\tau) = \sum_{k=-\infty}^{+\infty} R_{rr}(kT) \; \frac{1}{T} \int_{-T/2}^{T/2} p(t_1)\, p(t_1+\tau-kT)\, dt_1$$

If we now let $\gamma \to 0$, the convolution of pulse trains reduces to an impulse train, and the time averaged autocorrelation function is

$$(A2.19) \qquad \overline{R_{r^* r^*}(\tau)} = \frac{1}{T} \sum_{k=-\infty}^{+\infty} R_{rr}(kT)\, \delta(\tau - kT)$$

where the bar is used to emphasize the time averaging.

Consider now the ensemble-averaged autocorrelation function, which we shall denote by $\widetilde{R_{r^* r^*}(\tau,t)}$. It is defined by the average

$$(A2.20) \qquad \widetilde{R_{r^* r^*}(\tau,t)} = \widetilde{r^*(t)\, r^*(t+\tau)}$$

again substituting the equivalent summations, but going directly to the impulse approximation, we obtain

$$(A2.21) \qquad \widetilde{R_{r^* r^*}(\tau,t)} = \widetilde{\sum_{n=-\infty}^{+\infty} \sum_{m=-\infty}^{+\infty} r(nT)\, r(mT)\, \delta(t-nT)\, \delta(t+\tau-mT)}$$

Since the only random variables are the functions $r(nT)$ and $r(mT)$, the operations of summing and averaging can be interchanged to yield

$$(A2.22) \qquad \widetilde{R_{r^* r^*}(\tau,t)} = \sum_n \sum_m \widetilde{r(nT)\, r(mT)}\, \delta(t-nT)\, \delta(t+\tau-mT)$$

The ensemble average $\overline{r(nT)\, r(mT)}$ is recognized as the correlation function of the input (or of the number sequence $r(nT)$), i.e.

$$\overline{r(nT)\, r(mT)} = R_{rr}(mT - nT) = R_{rr}(mT - nT) \tag{A2.23}$$

since the input process was assumed to be ergodic. If we let $(m-n) = k$ equation (A2.22) becomes

$$\overline{R_{r^* r^*}(\tau,t)} = \sum_n \sum_k R_{rr}(kT)\, \delta(t-nT)\, \delta(\tau+t-mT-kT) \tag{A2.24}$$

and since k takes on all values from $-\infty$ to $+\infty$ and the time shift τ is arbitrary, this can be written as

$$\overline{R_{r^* r^*}(\tau,t)} = \sum_{k=-\infty}^{+\infty} R_{rr}(kT)\, \delta(\tau-kT) \sum_{n=-\infty}^{+\infty} \delta(t-nT) \tag{A2.25}$$

Clearly, this function is <u>not</u> identical to the time-averaged auto-correlation function, $\overline{R_{r^* r^*}(\tau)}$ obtained in (A2.19). Rather, and this is typical of ensemble averages in sampled-data systems, $\overline{R_{r^* r^*}(\tau,t)}$ is seen to be a periodic function of t, with period T. One way of eliminating this periodicity is to average $R_{r^* r^*}(\tau,t)$ over one sampling period: this process leads us again to the time-averaged correlation function. Thus:

$$\begin{aligned} \frac{1}{T} \int_{-T/2}^{T/2} \overline{R_{r^* r^*}(\tau,t)}\, dt &= \frac{1}{T} \int_{-T/2}^{+T/2} \left\{ \sum_k R_{rr}(dT)\delta(\tau-kT) \sum_n \delta(t-nT) \right\} dt \\ &= \frac{1}{T} \sum_{k=-\infty}^{+\infty} R_{rr}(kT)\, \delta(\tau-kT) \end{aligned} \tag{A2.26}$$

This is the expected result, since due to the ergodic nature of $r(t)$, the statistical characteristics of the ensemble are considered the same, including their time variations; therefore a time average of single record equals the time average of the ensemble average.

It should be noted that the ensemble-averaged autocorrelation function of the output $c(t)$ of the system of Figure A2.2 is also a periodic function of t. Consequently, for simplicity, we shall concentrate on time-averaged properties in what follows. It should also be noted that the factor of $1/T$ which arises in equation (A2.26) as a result of time averaging, represents the "duty factor" of Section A2.1 for the limiting case as the pulse width $\gamma \to 0$.

In summary, the autocorrelation of the sequence $r(nT)$ is given by the sequence

$$R_{rr}(kT) = \lim_{N \to \infty} \frac{1}{2N+1} \sum_{n=-N}^{n=+N} r(nT)\, r(nT + kT) \tag{A2.27}$$

and the "sampled autocorrelation function" of the signal $r^*(t)$ will be defined in terms of its time average as

$$R_{r^*r^*}(\tau) = \frac{1}{T} \sum_{k=-\infty}^{+\infty} R_{rr}(kT)\, \delta(\tau - kT) \tag{A2.28}$$

The cross-correlations are defined in a corresponding manner. Thus, the sequence cross-correlation is given by:

$$(A2.29 \qquad R_{rc}(kT) = \lim_{N \to \infty} \frac{1}{2N+1} \sum_{n=-N}^{+N} r(nT)\, c(nT + kT)$$

A2.4 Pulse-Spectral Densities

The discrete equivalent of the power spectral density is variously denoted as "pulse-spectral density", "sampled spectral density", or "sampled power spectrum". Once again, as in the case of correlation functions, the definitions must be carefully interpreted since the literature is by no means consistent in this regard.

Consider first the number sequence $r(nT)$, i.e. the values of the sampled function at the sampling instants. The autocorrelation function of the sequence is given by (A2.27) above, and we can call this $R^*_{rr}(kT)$. Then the sampled power spectrum is defined by the z-transform of $R^*_{rr}(kT)$, i.e.

$$(A2.30) \qquad S_{rr}(z) = \sum_{k=-\infty}^{+\infty} R^*_{rr}(kT)\, z^{-k}$$

Now consider the autocorrelation $R_{r^*r^*}(\tau)$ as defined by (A2.28), for the impulse modulated signal $r^*(t)$. We take the two-sided Laplace transform to obtain the power spectrum:

$$(A2.31) \qquad S_{r^*r^*}(s) = \int_{-\infty}^{+\infty} R_{r^*r^*}(\tau)\, e^{-\tau s}\, d\tau$$

Substituting (A2.28) into (A2.31) we obtain

$$(A2.32) \qquad S_{r^*r^*}(s) = \frac{1}{T} \int_{-\infty}^{+\infty} \sum_{k=-\infty}^{-\infty} R_{rr}(kT)\, \delta(\tau - kT)\, e^{-s\tau}\, d\tau$$

Since $R_{rr}(kT)$ does not depend on τ, the summation and integration can be interchanged and we obtain

$$(A2.33) \qquad S_{r^*r^*}(s) = \frac{1}{T} \sum_{k=-\infty}^{+\infty} R_{rr}(kT)\, e^{-skT}$$

Letting $z = e^{sT}$, equation (A2.33) can be written as

$$(A2.34) \qquad S_{r^*r^*}(z) = \frac{1}{T} \sum_{k=-\infty}^{+\infty} R^*_{rr}(kT)\, z^{-k}$$

which, as expected, differs from the sequence spectrum, $S_{rr}(z)$, by $\frac{1}{T}$. Thus, except for a proportionality constant, the sampled spectrum can be obtained as the z-transform of an appropriately defined sampled autocorrelation function.

The sequence spectral density can also be defined in terms of truncated series which insure finite energy and consequently convergence. Again, this is analogous to the continuous case. One introduces the auxiliary series $r_N(nT)$, as

$$(A2.35) \qquad r_N(nT) = r(nT) \qquad \text{when} \quad -N \leq n \leq N$$

$$r_N(nT) = 0 \qquad \text{elsewhere.}$$

Then a "truncated -z- transform" can be obtained as follows:

$$(A2.36) \qquad A_N(e^{sT}) = T \sum_{n=-N}^{N} r_N(nT)\, e^{-nsT}$$

and finally

$$(A2.37) \qquad S^*_{rr}(s) = \lim_{N \to \infty} \frac{1}{(2N+1)T} \left| A_N(e^{sT}) \right|^2$$

or, in "z-transforms"

$$(A2.38) \qquad S_{rr}(z) = \lim_{N \to \infty} \frac{1}{(2N+1)T} \left| A_N(z) \right|^2$$

where $A_N(z)$ is the finite sum z-transform of $r(nT)$.

Relationships (A2.37) and (A2.38) relate the sampled power spectrum directly to the time function $r(nT)$, and must be interpreted with caution, as in the continuous case. (cf. Laning & Battin)

Finally, the sampled power spectrum can be simply related to the power spectral density of the continuous functions $r(t)$. If we recall the alternate representations of the sampled function $r^*(t)$ in the frequency domain, it can be shown easily that (for zero initial conditions):

$$(A2.39) \qquad \mathcal{L}\left[r^*(t)\right] = \sum_{n=-\infty}^{+\infty} r(nT)\, e^{-nTs} = \frac{1}{T} \sum_{m=-\infty}^{+\infty} R\left(s + j\,\frac{2\pi m}{T}\right)$$

where the "shifting theorem" is used and $R(s) = \mathcal{L}\left[r(t)\right]$. Using the same rule on equation (A2.33) for the sampled spectrum we obtain

$$S_{r^*r^*}(s) = \frac{1}{T} \sum_k R_{rr}(kT)\, e^{-skT}$$

(A2.40)

$$= \frac{1}{T^2} \sum_n S_{rr}\left(s+j\,\frac{2\pi n}{T}\right) = \frac{1}{T}\, Z\left[S_{rr}(s)\right]$$

where equation (A2.40) relates the sampled spectrum to the continuous spectrum and harmonics generated by the sampling process.

It should be noted that all the definitions introduced in this section are based on time-averaging, i.e. that the correlation functions on which the power spectra are based are time-averaged over a sampling period.

By analogy with continuous processes, sampled cross-spectral densities can be defined as z-transforms of sampled cross-correlation functions.

For our future work we shall base the definitions of sampled spectra on equations (A2.33) and (A2.40); i.e. whenever possible we shall evaluate sampled spectra directly as z-transforms of continuous spectra times 1/T.

A2.5 Mean-Square Values

The mean-square value of a sampled function $r^*(t)$ can be evaluated either directly by time averaging or from correlation functions or from power spectra as in the continuous case.

Considering impulse modulated signals again, it was pointed out in Section A2.1 that the mean-squared value of the function $r^*(t)$ is given by

$$\overline{\left[r^*(t)\right]^2} = \frac{1}{T} \overline{\left[r(nT)\right]^2} \tag{A2.41}$$

From the sequence autocorrelation, equation (A2.27), we can write (A2.41) as

$$\overline{\left[r(nT)\right]^2} = R_{rr}(0) = \lim_{N \to \infty} \frac{1}{(2N+1)} \sum_{n=-N}^{N} \left[r(nT)\right]^2 \tag{A2.42}$$

An alternate expression for mean-squared value can be obtained from the sampled spectral density, again by analogy with continuous systems, by obtaining the average power from integration of the spectrum over all frequencies. In continuous systems we write

$$\overline{r(t)^2} = \frac{1}{2\pi} \int_{-\infty}^{+\infty} S_{rr}(\omega)\, d\omega \tag{A2.42'}$$

However, the sampled power spectrum, from equation (A2.33), is a periodic function of frequency, with period $\frac{2\pi}{T}$. Consequently, the mean-square is obtained by integration over a single period:

$$\overline{\left[r(nT)\right]^2} = \frac{T}{2\pi} \int_{-\pi/T}^{+\pi/T} S^*_{rr}(\omega)\, d\omega \tag{A2.43}$$

Relations (A2.41), (A2.42), and (A2.43) are very important in design of sampled data systems for minimum mean square error.

A2.6 Relation Between Sampled Correlation and Spectrum

In the continuous case, the autocorrelation function and the power spectral density are Fourier transforms of each other, and

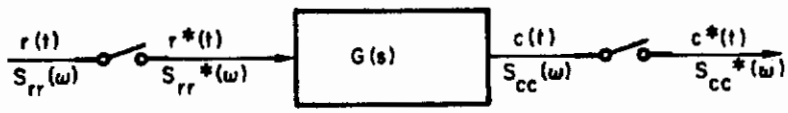
thus constitute the so-called Wiener-Khinchine relations. Similar relations exist for sampled-data systems, where Fourier series appear since the data are discrete. For sequence correlations $R^*_{rr}(kT)$ and sequence sampled spectral density $S^*_{rr}(\omega)$ we can write, on the basis of the above derivations:

$$(A2.44) \qquad S^*_{rr}(\omega) = \sum_{k=-\infty}^{+\infty} R^*_{rr}(kT)\, e^{-j\omega kT}$$

$$(A2.45) \qquad R^*_{rr}(kT) = \frac{T}{2\pi} \int_{-\pi/T}^{+\pi/T} S^*_{rr}(\omega) e^{j\omega kT}\, d\omega$$

A2.7 Input-Output Relations - (Open Loop)

Consider again the simple open loop system of Figure A2.3, where the power spectral density of the input is given by $S_{rr}(\omega)$. It is desired to compute both the continuous output spectrum, $S_{cc}(\omega)$ and the sampled output spectrum, $S_{c^*c^*}(\omega)$.



Open Loop System
Figure A2.3

In Figure A2.3 the sampled spectral density is denoted by $S_{rr}{}^*(\omega)$ and $S_{cc}{}^*(\omega)$ respectively, instead of $S_{r^*r^*}$ or $S_{c^*c^*}$ in order to simplify the notation.

If we assume impulse modulation, then the sampled spectrum $S_{rr}{}^*(\omega)$ is given by equation (A2.34) or (A2.40), i.e.

$$(A2.46) \qquad S_{rr}^{*}(\omega) = \frac{1}{T} Z\left[S_{rr}(s)\right]_{z = e^{j\omega T}}$$

Since $S_{rr}^{*}(\omega)$ is the power spectral density at the input of a linear continuous filter with transfer function $G(s)$, the output spectrum can be written immediately as

$$(A2.47) \qquad S_{cc}(\omega) = S_{rr}^{*}(\omega) \left|G(j\omega)\right|^{2}$$

If $G(s)$ represents a zero-order hold, the output spectrum is given by

$$(A2.48) \qquad S_{cc}(\omega) = S_{rr}^{*}(\omega) \left(\frac{1 - e^{-j\omega T}}{j\omega}\right)\left(\frac{1 - e^{j\omega T}}{-j\omega}\right)$$

Since the sampled output $c^{*}(t)$ can be related to the sampled input $r^{*}(t)$ directly by z-transforms, i.e.

$$(A2.49) \qquad C(z) = R(z)\, G(z)$$

where $G(z)$ is the z-transform of the filter described by the continuous transfer function $g(s)$, it can be expected that the sampled output spectrum $S_{cc}^{*}(\omega)$ can be obtained directly from $S_{rr}^{*}(z)$ by z-transform operations.

The output sequence, $c(nT)$, can be obtained from the convolution summation as

$$(A2.50) \qquad c(nT) = \sum_{k=0}^{\infty} h(kT)\, r(nT - kT)$$

where $h(t)$ is the impulse response of the filter in Figure A2.3. In accordance with equation (A2.28), the sampled autocorrelation function of $c^*(t)$ is given by the time average:

$$R^*_{cc}(nT) = \lim_{N \to \infty} \frac{1}{(2N+1)T} \sum_{k=-N}^{+N} c(kT)\, c(kT+nT)$$

$$= \lim_{N \to \infty} \frac{1}{(2N+1)} \sum_{k=-N}^{+N} \sum_{m=0}^{\infty} h(mT) r(kT-mT) \sum_{i=0}^{\infty} h(iT) r(kT+nT-iT) \qquad \text{(A2.51)}$$

Interchanging the order of summation

$$R^*_{cc}(nT) = \sum_{m=0}^{\infty} h(mT) \sum_{i=0}^{\infty} h(iT) \lim_{N \to \infty} \frac{1}{(2N+1)T} \sum_{k=-N}^{+N} r(kT-mT) r(kT+nT-iT)$$

(A2.52)

The last term on the right is easily recognized as the sampled autocorrelation of the input, and thus (A2.52) becomes

$$R^*_{cc}(nT) = \sum_{m=0}^{\infty} h(mT) \sum_{i=0}^{\infty} h(iT)\, R^*_{rr}(mT + nT - iT) \qquad \text{(A2.53)}$$

If we now multiply both sides of (A2.53) by z^{-n} and sum over all n we obtain

$$\sum_{n=-\infty}^{+\infty} R^*_{cc}(nT)\, z^{-n} = \sum_{n=-\infty}^{+\infty} z^{-n} \sum_{m=0}^{\infty} h(mT) \sum_{i=0}^{\infty} h(iT)\, R^*_{rr}(mT+nT-iT) \qquad \text{(A2.54)}$$

The left hand side of (A2.54) is the sampled power spectrum, and the right hand side can be rewritten as follows:

$$(A2.55)\quad S^*_{cc}(z) = \sum_{m=0}^{\infty} h(mT)z^{m} \sum_{i=0}^{\infty} h(iT)z^{-i} \sum_{n=-\infty}^{-\infty} R^*_{rr}(mT+nT-iT)z^{-(m+n-i)}$$

and consequently we have as the desired relationship

$$(A2.56)\qquad S_{cc}(z) = G(z)\, G\left(\frac{1}{z}\right) S_{rr}(z)$$

This relationship among the sampled spectra is the discrete equivalent to the relation

$$(A2.57)\qquad S_{cc}(s) = G(s)\, G(-s)\, S_{rr}(s)$$

which applies to the input and output of continuous systems.

A2.8 Relations in Closed-Loop Systems

Based on the definitions of the above paragraphs, it is now possible to present relationships for the statistical analysis of closed-loop sampled-data systems. Equation (A2.56), which relates the input and output sampled spectra, will be one of the key equations in this development. It should be noted again, that equation (A2.56) presents statistical properties which have been time-averaged over a sampling period. If the relation of (A2.56) had been derived without such averaging, it would exhibit time-dependence, and it could have been written in terms of modified z-transforms as follows:

$$(A2.58)\qquad S_{cc}(z,m) = |G(z,m)|^2 S_{rr}(z)$$

where $m = \frac{\Delta T}{T}$, $(0 = m = 1)$ represents a fractional sampling period, and $S_{cc}{}^*(z,m)$ represents the sampled spectral density

at the output at instants delayed by (ΔT) from the sampling instants (nT). The time dependence m disappears when (A2.58) is averaged from 0 to T (see Mori, 1958). In terms of real frequency, (A2.57) can be written as

$$S^*_{cc}(\omega) = \left|G^*(e^{j\omega T})\right|^2 S_{rr}(\omega) \tag{A2.59}$$

Consider now the closed-loop system shown in Figure A2.4.

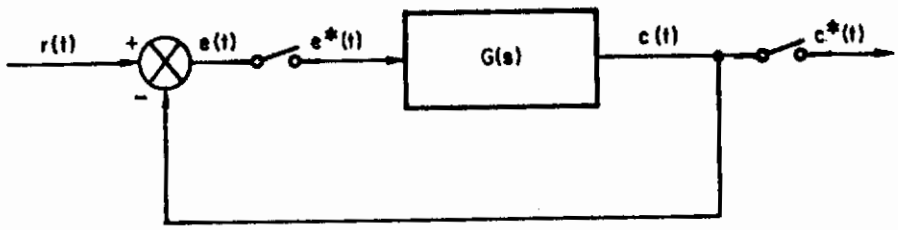


Figure A2.4
Closed Loop System

We are interested in the statistical properties of the error and output, both the continuous error and the sampled error. We again assume $r(t)$ to be a member of a stationary random process. We further assume the process to be Gaussian and concentrate on power spectra.

The output spectrum is obtained easily by replacing Figure A2.4 by the equivalent series open loop arrangement of Figure A2.5. The output power spectral density, $S_{cc}(\omega)$, is then obtained as a result of the following series of operations:

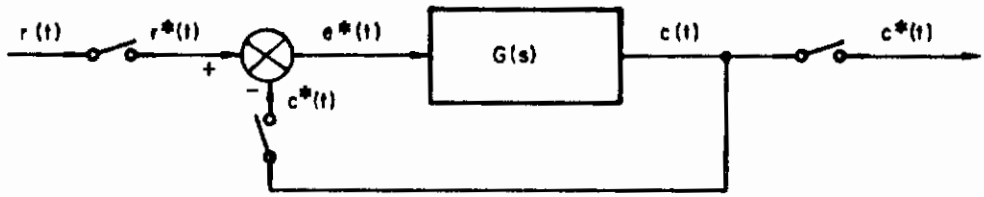


Figure A2.5 (a)

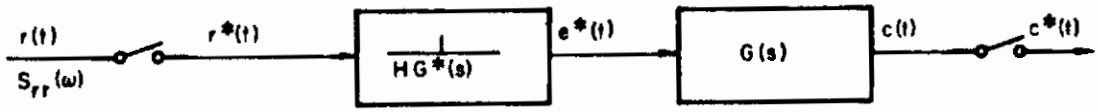


Figure A2.5 (b)
Equivalent Forms to Figure A2.4

Assume that $S_{rr}(\omega)$ is given. Then the sampled input spectrum is

$$S^*_{rr}(\omega) = \frac{1}{T} \; Z \left[S_{rr}(s) \right]_{z=e^{j\omega T}} \tag{A2.60}$$

The sampled error spectrum is given by Equation (A2.59), and becomes

$$S^*_{ee}(\omega) = \left| \frac{1}{1 + G^*(j\omega)} \right|^2 S^*_{rr}(\omega) \tag{A2.61}$$

and finally the output spectrum is given by

$$(A2.62) \quad S_{cc}(\omega) = |G(j\omega)|^2 S^*_{ee}(\omega) = \left| \frac{G(j\omega)}{1 + G^*(j\omega)} \right|^2 S^*_{rr}(\omega)$$

to which (A2.60) can be applied again to obtain the sampled output spectrum, or, the sampled output spectrum can be obtained directly from the sampled input spectrum as

$$(A2.63) \quad S^*_{cc}(\omega) = \left| \frac{G^*(\omega)}{1 + G^*(\omega)} \right| S^*_{rr}(\omega)$$

The mean-square error sequence $\overline{e(nT)^2}$ can be obtained directly from equation (A2.61) by using the relationships of Section A2.4 above:

$$(A2.64) \quad \overline{e(nT)^2} = \frac{T}{2\pi} \int_{-\pi/T}^{+\pi/T} S^*_{ee}(\omega)\, d\omega$$

It is still necessary to find the continuous error spectral density, $S_{cc}(\omega)$ and mean-squared continuous error, $\overline{e(t)^2}$. It should be noted that the continuous error does not appear explicitly in Figure A2.5, and therefore it must be obtained from Figure A2.4. To obtain the spectral density of $e(t)$ we define the truncated function $e_{t_1}(t)$:

$$(A2.65) \quad e_{t_1}(t) = e(t) \qquad -t_1 \leq t \leq t_1$$

$$e_{t_1}(t) = 0 \qquad \text{elsewhere.}$$

and then define the power spectral density in terms of the Fourier transform of the truncated signal:

$$(A2.66)\quad S_{ee}(\omega) = \lim_{t_1 \to \infty} \frac{1}{t_1}\left|\int_{-\infty}^{+\infty} e_{t_1}(t)\epsilon^{-j\omega t}dt\right|^2 = \lim_{t_1 \to \infty} \frac{1}{t_1}\left|E_{t_1}(j\omega)\right|^2$$

Since $e(t) = r(t) - c(t)$, equation (A2.66) becomes

$$(A2.67)\quad S_{ee}(\omega) = \lim_{t_1 \to \infty} \frac{1}{t_1}\left[\left|R_{t_1}(j\omega)\right|^2 + \left|C_{t_1}(j\omega)\right|^2 - 2\left|R_{t_1}(j\omega)C_{t_1}(j\omega)\right|\right]$$

and therefore we obtain, if the limit indicated above exists,

$$(A2.68)\quad S_{ee}(\omega) = S_{rr}(\omega) + S_{cc}(\omega) - 2\mathrm{Re}\left[S_{rc}(\omega)\right]$$

The output spectrum is obtained from (A2.61) and (A2.62) as

$$(A2.69)\quad S_{cc}(\omega) = \frac{1}{T} Z\left[S_{rr}(s)\right]_{z=\epsilon^{j\omega T}}\left|\frac{G(j\omega)}{1 + G^*(j\omega)}\right|^2$$

$$= S^*_{rr}(\omega)\left|\frac{G(j\omega)}{1 + G^*(j\omega)}\right|^2$$

Since the cross-spectral density between input and output is defined as

$$(A2.70)\quad S_{rc}(\omega) = \frac{1}{T}\left[\frac{G(j\omega)}{1 + G^*(j\omega)}\right] S_{rr}(\omega)$$

we have as the expression for the power spectral density of the continuous error:

$$(A2.71)\quad S_{ee}(\omega) = S_{rr}(\omega) - \frac{2}{T} S_{rr}(\omega)\mathrm{Re}\left[\frac{G(j\omega)}{1+G^{*}(j\omega)}\right] + S^{*}_{rr}(\omega)\left|\frac{G(j\omega)}{1+G^{*}(j\omega)}\right|^{2}$$

and consequently the continuous mean-squared error is

$$(A2.72)\qquad \overline{e(t)^2} = \frac{1}{2\pi}\int_{-\infty}^{+\infty} S_{ee}(\omega)d\omega$$

Since $S_{ee}(\omega)$ depends on $e^{j\omega T}$, it will in general be a function of T. Then (A2.72) could be used to set the sampling period for minimum mean square error.

REFERENCES

1. Blum, M. "An Extension of the Minimum Mean Square Prediction Theory for Sampled Input Signals", IRE Trans. on Information Theory, IT-2: 176-184, Sept. 1956.

2. Chang, S. S. L. "Optimum Transmission of Continuous Signal over a Sampled Data Link", AIEE Trans. (Applic. and Industry) 79, (pt. II): 538-542, Jan. 1961.

3. DeRusso, P. M. "Optimum Linear Filtering of Signals Prior to Sampling", AIEE Trans. (Applic. and Industry) 79, (pt. II): 549-555, Jan. 1961.

4. Franklin, G. F. (1955) "Linear Filtering of Sampled Data", IRE Convention Record, 1955, (pt. 4): 119-28.

5. Johnson, G. W. "Statistical Analysis of Sampled-Data Systems", WESCON Convention Record, 1957, (pt. 4): 187-195.

6. Lees, A. B. "Interpolation and Extrapolation of Sampled Data" IRE Trans. on Information Theory IT-2: 12-17, 1956.

7. Lloyd, S. P. and B. McMillan "Linear Least Squares Filtering and Prediction of Sampled Data", Proc. Symposium on Network Theory (Brooklyn Polytechnic Inst.) 221-47, 1955.

8. Mori, M. "Statistical Treatment of Sampled-Data Control Systems for Actual Random Inputs", Trans. ASME 80: 444-456, 1958.

9. Pyatnitskii, G. I. "The Action of Random Processes on Discontinuous Control Systems", Automation and Remote Control, 21: 585-594, May 1960.

10. Stewart, R. M. "Statistical Design and Evaluation of Filters for the Restoration of Sampled Data", Proc. IRE, 44: 253-257, Feb. 1956.

11. Tartakovskii, G. P. "Stationary Random Processes in Linear Pulse Circuits with Variable Parameters", Radio Eng. and Electronics 2: 380-388, 1957.

12. Trembath, N. W. "Random Signal Analysis of Linear and Nonlinear Sampled Data Systems", DACL Report No. 108, Mass. Inst. of Tech., June 1957.

Appendix 3

ANALOG MEASUREMENT OF POWER SPECTRA

A3.1 Introduction

The purpose of this Appendix is to present a brief review of the theory of analog spectral analysis and to discuss the construction and calibration of the device constructed. The discussion will concern only signals of the type used in the present study, i.e., signals composed of sums of sinusoids with an additive "noise" component. The input functions will be deterministic in all cases and frequencies of the sinusoidal components known; the purpose of the analyzer, then, is to provide a reliable measure of the amplitude of the sinusoidal components.

The material which follows is based on many sources, The theoretical background is taken from Grenander and Rosenblatt (1957), Davenport and Root (1958) and Bendat (1958). The measurement (or, more properly, "estimation") of power spectra is carefully treated by Blackman and Tukey (1958) and Parzen (1960). A number of practical problems are mentioned by Chang (1961) in a chapter on spectral estimates. The discussion on measuring spectra of signals composed of sums of sinusoids of known frequency by analog methods is based heavily on Seltzer and McRuer (1959) and McRuer (1960).

Power spectral densities (and therefore correlation functions) are defined theoretically on the basis of integrals with doubly infinite limits. In the time domain, this implies that the function f (t) is available for all time, i.e., for $-\infty < t < +\infty$. In the practical measurement problem, we only have a function $f_T(t)$ which is defined for a finite interval of time, say T_R seconds. In the case of human operator studies the length of this interval is governed by factors such as the desire to maintain essentially stationary conditions as well as economic factors. Furthermore, the extraction of frequency information in finite frequency bands requires filters, and physically realizable filters cannot have perfect cutoff

characteristics. Thus, the characteristics of finite data and physically realizable filters make the practical situation vastly different from the theoretical. To these two major problems, the imperfections of other components (such as amplifiers and multipliers) must be added.

A3.2 Power Spectra of Periodic Functions

Before examining the measurement problem, the properties of periodic functions will be reviewed briefly. We are interested in the function

$$f(t) = \sum_{n=1}^{N} c_n \cos(n\omega t - \theta_n) \tag{A3.1}$$

for two reasons:

(1) The input function utilized in the experiment consisted of a finite time sample of such a function, and

(2) Since experimental functions will always necessarily be of finite duration $2T_R$, they can be expanded in a Fourier series in the interval $(-T_R, T_R)$ i.e., a series of the form (A3.1) with $N = \infty$.

We assume that $f(t)$ satisfies the Dirichlet conditions (see Churchill, 1941) and therefore, that an expansion in a Fourier series will be valid almost everywhere in the interval.

Consider the function $f(t)$ expressed as a Fourier series in the interval $(-T, T)$:

$$f(t) = \frac{a_o}{2} + \sum_{n=1}^{\infty} \left[a_n \cos\left(\frac{n\pi}{T}\right) t + b_n \sin\left(\frac{n\pi}{T}\right) t\right] \tag{A3.2}$$

where the coefficients a_n and b_n are given by

$$(A3.3) \qquad a_n = \frac{1}{T} \int_{-T}^{T} f(\tau) \cos\left(\frac{n\pi}{T}\right) \tau \, d\tau$$

$$b_n = \frac{1}{T} \int_{-T}^{T} f(\tau) \sin\left(\frac{n\pi}{T}\right) \tau \, d\tau$$

Alternatively, we can write

$$(A3.4) \qquad f(t) = \sum_{n=0}^{\infty} c_n \cos\left(\frac{n\pi}{T} t - \theta_n\right)$$

where

$$c_n = (a_n^2 + b_n^2)^{1/2} \; ; \; c_o = \frac{a_o}{2} \; ; \; \theta_n = \tan^{-1} \frac{b_n}{a_n}$$

or, for the complex form

$$(A3.5) \qquad f(t) = \sum_{n=-\infty}^{+\infty} \alpha_n \, e^{j\left(\frac{n\pi}{T}\right) t}$$

where

$$\alpha_n = \frac{a_n - jb_n}{2} \; ; \qquad \alpha_{-n} = \frac{a_n + jb_n}{2}$$

There are several ways of approaching the definition of the power spectral density of $f(t)$.

A3.2.1 Definition from autocorrelation function. From the Wiener-Khinchine relations, the power spectral density is defined in terms of the Fourier transform of the autocorrelation function as

$$(A3.6) \qquad S_{xx}(\omega) = \frac{1}{\pi} \int_{-\infty}^{+\infty} R_{xx}(\tau)\, e^{-j\omega\tau}\, d\tau$$

or using the Fourier cosine transform (since $R_{xx}(\tau)$ is an even function of τ)

$$(A3.7) \qquad S_{xx}(\omega) = \frac{2}{\pi} \int_{0}^{\infty} R_{xx}(\tau) \cos \omega\tau \, d\tau$$

Now, since the process f (t) is assumed stationary, the autocorrelation function can be defined in terms of the time average as

$$(A3.8) \qquad R_{ff}(\tau) = \overline{f(t)\, f(t-\tau)}$$

$$= \lim_{T_o \to \infty} \frac{1}{2T_o} \int_{-T_o}^{+T_o} f(t)\, f(t-\tau)\, dt$$

Substituting from (A3.4) we have

$$(A3.9) \quad R_{ff}(\tau) = \lim_{T_o \to \infty} \frac{1}{2T_o} \int_{-T_o}^{+T_o} \left\{ \sum_{n=1}^{\infty} c_n \cos(\omega_n t - \theta_n) \sum_{i=1}^{\infty} c_i \cos[\omega_i (t-\tau) - \theta_i] \right\} dt$$

where $\omega_n = n\pi/T$. Therefore, evaluating this expression

$$R_{ff}(\tau) = c_o^2 + \frac{1}{2} \sum_{n=1}^{\infty} c_n^2 \cos \omega_n \tau$$

Thus, as is well known, the autocorrelation of a periodic function is seen to be periodic as well (i.e., it contains all the periodicities present in f (t)) and it is independent of the phase relations θ_n in f (t).

To obtain the power spectral density we substitute (A3.10) into the defining equation (A3.7)

$$(A3.11) \qquad S_{ff}(\omega) = \frac{2}{\pi} c_o^2 \int_0^\infty \cos \omega\tau d\tau + \frac{1}{\pi} \sum_{n=1}^{\infty} c_n^2 \int_0^\infty \cos \omega_n \tau \cos \omega\tau d\tau$$

To evaluate this expression, one can express $\cos \omega\tau$ as

$$(A3.12) \qquad \cos \omega\tau = \frac{e^{j\omega\tau} + e^{-j\omega\tau}}{2}$$

and proceed formally to evaluate the integrals in (A3.11) by using the relation

$$(A3.13) \qquad \frac{1}{2} \int_{-\infty}^{+\infty} e^{j\omega t} d\omega = \delta(t)$$

where $\delta(t)$ is the Dirac impulse function. Actually, $\delta(t)$ is not a "function" in the ordinary sense and must be used carefully. However, its use is common in the study of random processes and we shall use it formally in these notes. The treatment can be rigorized by following Lighthill (1958). Using the relation (A3.13) to define the inverse Fourier transform of unity as being $\delta(t)$, we can substitute this result into (A3.11) to obtain

$$(A3.14) \quad S_{ff}(\omega) = 2 c_o^2 \delta(\omega) + \frac{1}{2} \sum_{n=1}^{\infty} c_n^2 [\delta(\omega - \omega_n) + \delta(\omega + \omega_n)]$$

Thus, the power spectral density of a periodic function of infinite extent in the time domain consists of a row or "comb" of δ-"functions"

in the frequency domain, i.e., all the power is concentrated at the frequencies $\pm \omega_n$ $(n = 0, 1, \cdots)$.

The <u>average power</u> contained in the spectrum of (A3.14) is given by integration of the spectral density over all frequencies

$$(A3.15) \qquad P_{av}\left[f(t)\right] = \frac{1}{2\pi}\int_{-\infty}^{+\infty} S_{ff}(\omega)\, d\omega = c_o^2 + \sum_{n=1}^{\infty} \frac{c_n^2}{2}$$

which is the expected result, since the average power contributed by a sinusoid of amplitude c_i is $P_i = c^2/2$.

A3.2.2 <u>Definition from truncated Fourier transform</u>. A second form of the definition of the spectral density focuses attention on the finite length of a record. If $f(t)$ is an infinite length record, its total energy

$$(A3.16) \qquad \int_{-\infty}^{+\infty} f^2(t)\, dt$$

will not be bounded unless $f(t) \longrightarrow 0$ as $t \to \infty$. Periodic functions do not behave in this way, and therefore we define a truncated function

$$(A3.17) \qquad f_T(t) = f(t) \qquad |t| \leq T$$

$$= 0 \qquad |t| > T$$

Then the Fourier transform of $f_T(t)$ exists and can be written as

$$(A3.18) \quad F_T [f(t)] = \int_{-\infty}^{+\infty} f_T(t)\, e^{-j\omega t}\, dt = \int_{-T}^{T} f(t)\, e^{-j\omega t}\, dt$$

Now, Parseval's theorem for Fourier transforms (see, for example, Bendat (1958, p. 41) states that the total energy can be evaluated in the time or frequency domain, i.e.,

$$(A3.19) \quad \int_{-\infty}^{+\infty} x^2(t)\, dt = \frac{1}{2\pi} \int_{-\infty}^{+\infty} |F(\omega, x)|^2\, d\omega$$

where $F(\omega, x)$ is the Fourier transform of $x(t)$. Applying this form to (A3.16) with $f_T(t)$ instead of $f(t)$ we obtain for the total energy in $f_T(t)$

$$(A3.20) \quad E_{total} = \int_{-\infty}^{+\infty} f_T^2(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} |F_T(\omega, f)|^2\, d\omega$$

Therefore, the <u>average power</u> in $f(t)$ can be defined as

$$(A3.21) \quad P_{av} = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} f^2(t)\, dt$$

$$= \lim_{T \to \infty} \int_{0}^{\infty} \frac{F_T(\omega, f)^2}{2\pi T}\, d\omega$$

since $F_T(\omega)^2$ can be shown to be an even function.

Now, relation (A3.21) gives the average power for <u>all</u> frequencies. The power spectral density is <u>defined</u> as the average power in a narrow frequency band per unit frequency. Let us consider an ideal filter $Y_i(j\omega)$ which is perfectly selective such that

$$(A3.22)\quad |Y_i(j\omega)| = 1 \quad \text{for} \quad 0 \le \omega_r \le \omega \le \omega_s$$

$$= 0 \quad \text{elsewhere.}$$

Then, if the signal $f_T(t)$ is applied to the filter $Y_i(j\omega)$, its output Fourier transform, $B(j\omega)$, will be given by

$$(A3.23)\qquad B_T(\omega, f) = Y_i(j\omega)\, F_T(\omega, f)$$

Using the ideal frequency selective filter defined by (A3.22), the average output is from (A3.21)

$$(A3.24)\qquad P_{av\text{-}out} = \lim_{T\to\infty} \int_{\omega_r}^{\omega_s} \frac{|F_T(\omega, f)|^2}{2\pi T}\, d\omega$$

Many authors at this point reverse the limiting and integral operation in (A3.24) and define the power spectral density as

$$(A3.25)\qquad S_{ff}(\omega) = \frac{1}{2\pi} \lim_{T\to\infty} \frac{|F_T(\omega, f)|^2}{T}$$

Unfortunately, this procedure is not mathematically justified and fails occasionally, as outlined by Davenport and Root (1958). However, if we define:

$$(A3.26)\qquad S_{ff_T}(\omega) = \frac{|F_T(\omega, f)|^2}{T}$$

with no limiting operations and then obtain the power P_T:

$$(A3.27)\qquad P_T(\omega_c, f, \omega) = \int_{\omega_c - \frac{\Delta\omega}{2}}^{\omega_c + \frac{\Delta\omega}{2}} S_{ff_T}(\omega)\, d\omega$$

This quantity represents the average power, in a record of finite length $2T$, in a range of frequencies of width $\Delta\omega$ centered about ω_c. The power spectral density is then obtained (for this single record $f(t)$) by a sequence of two limiting operations, i.e.,

$$S_{ff}(\omega) \triangleq \lim_{\Delta\omega \to 0} \frac{P(\omega_c, f, \Delta\omega)}{\Delta\omega} \tag{A3.28}$$

where

$$P(\omega_c, f, \Delta\omega) = \lim_{T \to \infty} P_T(\omega_c, f, \Delta\omega) \tag{A3.29}$$

Equation (A3.25) can be used to lead to the same result obtained previously for the spectral density of infinite-length sine waves, if convergence is assumed.

Working with the complex form of the Fourier series, we have

$$F_T[f(t)] = \int_{-T}^{T} \left\{ \sum_{-\infty}^{+\infty} \alpha_n e^{j\omega_n t} \right\} e^{-j\omega t}\, dt \tag{A3.30}$$

$$= \sum_{-\infty}^{+\infty} \alpha_n \int_{-T}^{T} e^{j(\omega_n - \omega)t}\, dt$$

Consequently:

$$S_{ff_T}(\omega) = \frac{1}{2\pi T} \sum_n |\alpha_n|^2 \left\{ \left[\int_{-T}^{T} e^{j(\omega_n - \omega)t}\, dt \right] \right. \tag{A3.31}$$

$$\left. \left[\int_{-T}^{T} e^{-j(\omega_n - \omega)t}\, dt \right] \right\}$$

Now

$$(A3.32)\quad \int_{-T}^{T} e^{j(\omega_n - \omega)t}\, dt = \frac{e^{j(\omega_n - \omega)T} - e^{-j(\omega_n - \omega)T}}{j(\omega_n - \omega)} \quad \text{for } \omega \neq \omega_n$$

$$= 2T \quad \text{for } \omega = \omega_n$$

Going to the limit, as in (A3.25) results in

$$(A3.33)\quad \lim_{T\to\infty} S_{ff_T}(\omega) = 0 \quad \text{for } \omega \neq \omega_n$$

$$= \lim_{T\to\infty} \frac{4T}{2\pi} \sum_n |\alpha_n|^2 \quad \text{for } \omega = \omega_n$$

This last expression increases without bound, since it corresponds to the impulses encountered in (A3.14). They can be introduced here as well by defining the power spectral density as follows:

Definition:

$$(A3.34)\quad \text{Let} \quad S_{ff}(\omega) = \pi \lim_{T\to\infty} \frac{1}{T^2} \left| F_T [f(t)] \right|^2 \delta(\omega - \omega_n)$$

at all points where this expression is not zero; and

$$S_{ff}(\omega) = \lim_{T\to\infty} \frac{1}{T} \left| F_T [f(t)] \right|^2 \quad \text{elsewhere,}$$

provided these limits exist. Using this definition, we obtain

$$(A3.35)\qquad S_{ff}(\omega) = 2 \sum_{n=-\infty}^{+\infty} |\alpha_n|^2 \delta(\omega - \omega_n)$$

which agrees with (A3.14) if the Fourier coefficients are expressed in the corresponding form.

A3.3 The Measurement Problem

The problem of evaluating spectral density functions can be outlined with reference to the preceding development. Clearly, we can only obtain a measured approximation to $S_{xx_T}(\omega)$, as defined in (A3.26), which may also be a valid estimate of $S_{xx}(\omega)$. The measurement is dependent on 3 major factors:

(1) The manner of obtaining the truncated function $x_T(t)$ from $x(t)$.

(2) The characteristics of the actual bandpass filter $Y_a(j\omega)$ in relation to the ideal filter $Y_i(j\omega)$.

(3) The relation of run length and filter resolution.

These factors apply regardless of the nature of the input process. If the input function $x(t)$ is deterministic it is possible to obtain theoretically exact estimates of the power spectrum (i.e., exact with perfect components), provided that certain relations of frequency spacing, run length and filter characteristics are observed. If however, the input function is a sample function of a random process, the situation is considerably more complex. The expected value $E\left\{S_{xx_T}(\omega)\right\}$, obtained as an ensemble average of many runs, may still not converge to the actual spectral density $S_{xx}(\omega)$ of the process. The variance of the estimate will, however, still depend heavily upon run length and filter characteristics, which we shall consider below.

A3.3.1 The time "window." The function $x_T(t)$ can be viewed as the result of an operation on an infinite-length function $x(t)$ by a "window" function $w(t)$. The simplest window is that used in defining $f_T(t)$ in the previous section, i.e.,

$$w_o(t) = \begin{cases} 1 & |t| \leq T \\ 0 & |t| > T \end{cases} \tag{A3.36}$$

and then writing

$$x_T(t) = w_o(t)\ x(t) \tag{A3.37}$$

This product of time functions results in a convolution in the frequency domain, i.e., the Fourier transform of $x_T(t)$ is given by

$$F[x_T(t)] = F_T[x(t)] = \int_{-\infty}^{+\infty} W_o(j\omega - j\omega_1)\, X(j\omega_1)\, d\omega_1 \tag{A3.38}$$

Thus, the "time window" $w_o(t)$ has a corresponding "frequency window" $W_o(j\omega)$ which affects <u>all</u> frequencies in $X(j\omega)$. It is therefore clear that the choice of $w(t)$ will affect the performance of a spectral analyzer very significantly.

Consider the effect of the particular window $w_o(t)$ on the power spectrum of particular sine wave given by

$$p(t) = c_p \cos(\omega_p t - \theta_p) \tag{A3.39}$$

which can be assumed to be one of many in a complex input function. The Fourier transform $F_T(x, \omega)$ was obtained above in equation (A3.32) as

$$F_T(X_p, \omega) = c_p \left[\frac{e^{j(\omega_p - \omega)T} - e^{-j(\omega_p - \omega)T}}{j(\omega_p - \omega)} \right] \tag{A3.40}$$

$$= \frac{2 \sin(\omega_p - \omega)\, T}{(\omega_p - \omega)}$$

and hence the finite length spectral estimate $S_{xx_T}(\omega)$ is, from equation (A3.26):

$$S_{xx_T}(\omega) = \frac{T c_p^2}{2\pi} \frac{\sin^2[(\omega_p - \omega)\, T]}{[(\omega_p - \omega)\, T]^2} \tag{A3.41}$$

At $\omega = \omega_p$, this reduces to

$$(A3.42) \quad S_{xx_T}(\omega_p) = \frac{T c_p^2}{2\pi} \lim_{\omega \to \omega_p} \frac{\sin^2 (\omega_p - \omega) T}{(\omega_p - \omega)^2 T^2} = \frac{T c_p^2}{2\pi}$$

This latter value can be seen to correspond to the average power $x(t)$ per "unit bandwidth," where the unit bandwidth is defined in terms of the frequency separation of the Fourier series in the interval $(-T, T)$. Since the fundamental period is $2T$, we have the elementary bandwidth

$$(A3.43) \quad \Delta\omega = \omega_n - \omega_{n-1} = 2\pi \left(\frac{n}{2T} - \frac{n-1}{2T} \right) = \frac{\pi}{T}$$

The average power in the frequency bandwidth ΔP is the power of the sine wave, i.e., $1/2\ c_p^2$. Thus, the power spectral density can also be defined as

$$(A3.44) \quad S_{xx_T}(\omega) = \frac{\Delta P_n}{\Delta \omega_n} = \frac{1/2\ c_p^2}{\pi/T} = \frac{T c_p^2}{2\pi}$$

which agrees with (A3.42) above.

Now, it should be noted that equation A3.41 indicates that even with a single sine wave there is power present at all frequencies, having the characteristic spectrum of Figure A3.1. As $T \to \infty$, $S_{xx_T}(\omega) \to \infty$ for $\omega = \omega_p$ and $S_{xx_T}(\omega) \to 0$ for $\omega \neq \omega_p$, as expected.

The result of Figure A3.1 was obtained with one particular "time window," namely $w_o(t)$ as given by (A3.36). A considerable reduction in the side-lobes of the spectrum can be obtained by using more sophisticated windows, as outlined by Blackman and Tukey (1958) and Parzen (1960). Since $w_o(t)$ was used in the present study, we shall not pursue this matter further.

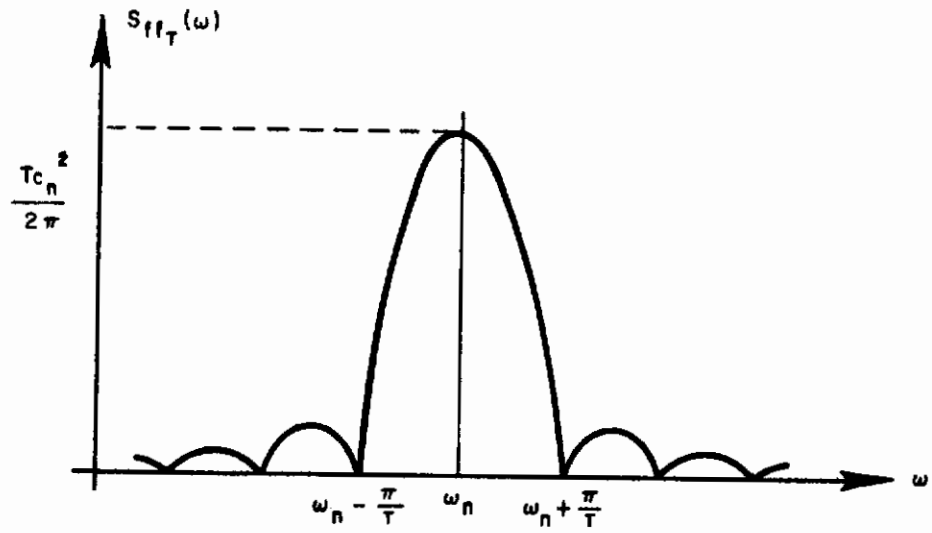


Figure A3.1 Spectrum of Finite Duration Sine Wave

A3.3.2 The filter problem. The result on frequency spreading of the previous section was derived assuming a perfect rectangular filter $Y_i(j\omega)$ for the recovery of the average power of a signal in an elementary bandwidth $(\Delta\omega)$. In practice, of course, any attempt to extract the fundamental peak $Tc_p^2/2$ will be subject to error since physically realizable filters do not have perfect cutoff. Note that the area of the fundamental peak in Figure A3.1 is given by

$$(A3.45) \quad P_{av}(2\Delta\omega) = \frac{Tc_n^2}{2\pi} \int_{\omega_n - \Delta\omega}^{\omega_n + \Delta\omega} \frac{\sin^2[(\omega_n - \omega)T]\, d\omega}{[(\omega_n - \omega)T]^2} =$$

$$= \frac{c_n^2}{2\pi} \int_{-\pi/T}^{+\pi/T} \frac{\sin^2 X}{X^2}\, dX$$

But, since

$$\int_{-\infty}^{+\infty} \frac{\sin^2 X}{X^2}\, dX = \pi; \quad \text{then} \quad P_{av}(2\Delta\omega) < \frac{c_n^2}{2}$$

In this relation c_n^2 is used to denote an arbitrary sinusoid. Since, clearly it is impossible to obtain a perfect filter, the power cannot be recovered perfectly. It is shown by Blackman and Tukey (1958) and McRuer (1961a) that the signal $x(t)$ is acted upon by an "effective filter" which is the result of the combined effects of the physical filter and the spectral window. Furthermore, the "pass band" of this effective filter is at least as wide as the wider of the filter and window spectra. (See Section 3.6 below.)

The effective bandwidth of the filter is sometimes taken as the distance between half-power points on the filter frequency diagram. Parzen (1960) suggests that it be considered as the length of the base of a rectangle which has the same area and same maximum height as the graph of the filter $Y(j\omega)$; i.e.,

$$(BW)_{eff} = \frac{\int_{-\infty}^{+\infty} Y(j\omega)\, d\omega}{\max_{\omega} |Y(\omega)|}$$

A3.3.3 On frequency resolution. We have noted above that the effect of finite run length is to cause a spread of the spectrum of a single sinusoid over the frequency domain. In addition, from the sampling theorem in the frequency domain (see Bendat (1958) or Goldman (1953)), if a function $x(t)$ exists only in the time interval $(-T, T)$ and is zero otherwise, its Fourier transform is completely determined by the values of the transform $F_T(x, \omega)$ evaluated at a series of points $2\pi/2T$ rad/sec apart. Therefore, the run length determines the frequency separation of sinusoids which can be considered independent (McRuer, 1961b). If the truncated signal $x(t)$ had, in fact, been sampled from a process with a continuous spectrum (thus containing an infinite number of frequencies,) the finite duration $(2T)$ would make such a signal indistinguishable from a sum of sinusoids separated by frequency intervals of π/T rad/sec.

For the present experiment, these comments indicate the minimum separation of sinusoidal components in the input for a particular run length.

A3.3.4 Ensemble averaging. The analysis of the spectral measurement method outlined in the previous paragraphs was based on a single run. If the functions $f(t)$ to be analyzed are random or contain random components, the operations discussed above must be repeated many times and an average of the results taken across the ensemble. Such an analysis is given below for the spectral analysis method actually used in the study.

A3.4 Selection of Experimental Values

Based on the previous analysis, the following values were selected for the experiment:

Minimum frequency in input function: $f_{min} \cong 1.1$ rad/sec.

Run length $T_R = 180$ sec $\cong$ 32 cycles of f_{min}.

Frequency resolution $= 2\pi/T_R \cong 0.035$ rad/sec.

Distance between input sine waves $\cong$ 1 rad/sec.

Filter effective bandwidth: (See below for description)

2nd order filter with half power frequency at .07 rad/sec,

$BW_{eff} \cong 0.14$ rad/sec.

Based on these considerations, the experimental situation was characterized by the sketch of Figure A3.2.

This particular filter bandwidth was based on Blackman and Tukey's suggestion that the filter be "several to many times $1/T_R$ wide" (p. 114).

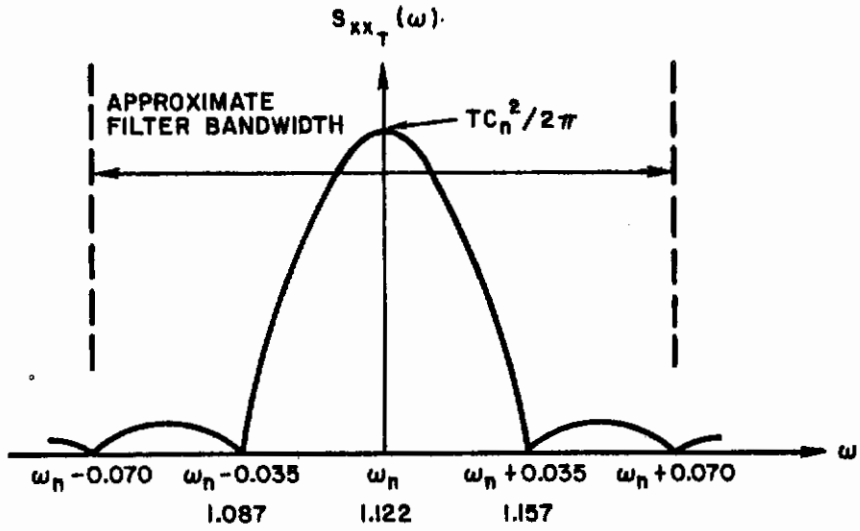


Figure A3.2

A3.5 Description of Measurement Method

The signals we are concerned with are of the form

$$(A3.46) \qquad x(t) = \sum_{n=1}^{10} c_i \cos(\omega_i t - \theta_i) + N(t)$$

where $N(t)$ is a noise component assumed to have zero mean and we desire to obtain estimates of the spectral density of $x(t)$ at the frequencies ω_i; i.e., to measure a value proportional to $Tc_i^2/2\pi$. The method selected for the measurement is based on Seltzer and McRuer (1959) and is discussed in detail in that reference, as well as in McRuer (1961a). What follows is a heuristic explanation of its operation, with a more rigorous analysis given in Section A3.6. The method is based on low-pass rather than band-pass filtering.

We consider a function $f(t)$ given by a finite sum of sine waves:

$$(A3.47) \qquad f(t) = \sum_{i=1}^{N} (a_i \sin \omega_i t + b_i \cos \omega_i t)$$

where the ω_i are not necessarily multiples of some basic frequency. Multiplication by $\sin \omega_j t$ and $\cos \omega_j t$ respectively produces:

$$(A3.48) \quad f(t) \sin \omega_j t = \sin \omega_j t \sum_i (a_i \sin \omega_i t + b_i \cos \omega_i t)$$

$$f(t) \cos \omega_j t = \cos \omega_j t \sum_i (a_i \sin \omega_i t + b_i \cos \omega_i t)$$

Now, if there is a frequency component ω_j present in $f(t)$, then the product results in the following:

$$(A3.49) \quad f(t) \sin \omega_j t = \frac{a_j}{2} - \frac{a_j \cos 2\omega_j t}{2} + \sin \omega_j t \sum_i (N - 1) \text{ terms}$$

and similarly

$$(A3.50) \quad f(t) \cos \omega_j t = \frac{b_i}{2} + \frac{b_j \cos 2\omega_j t}{2} + \cos \omega_j t \sum_i (N - 1) \text{ terms}$$

We now note that these two expressions consist of a d.c. term and an oscillatory component and we use low-pass filters to extract the d.c. term. If the filters were perfect, we would extract the d.c. components $a_j/2$ and $b_j/2$ exactly. Since filters of zero-width and rectangular cutoff are not physically realizable, we obtain instead a ripple component, the magnitude of which depends on the separation of the components ω_i in the original signal and the run length T_R. Averaging the resulting value decreases the effect of the ripple. If we denote the filter outputs as

$$(A3.51) \quad q_s(t) = \frac{a_j}{2} + \epsilon_a(t)$$

$$q_c(t) = \frac{b_j}{2} + \epsilon_b(t)$$

Then we can square these terms to obtain, after adding:

$$(A3.52) \qquad q_s^2 + q_c^2 = \frac{1}{4}(a_j^2 + b_j^2) +$$

$$+ [(a_j \epsilon_a + b_j \epsilon_b) + (\epsilon_a^2 + \epsilon_b^2)]$$

Filtering the result tends to reduce the effect of the oscillatory terms due to the errors ϵ. As noted in the previous section, the effectiveness depends on the filter bandwidth and averaging time as well as on the separation of frequencies ω_i. Thus we obtain an estimate of the spectral density as

$$(A3.53) \qquad S_{ff_T}(\omega_j) = \frac{1}{\pi} \int_0^{T_R} (q_s^2 + q_c^2)\, dt \cong \frac{T_R\, c_j^2}{2\pi}$$

A more rigorous analysis of the operation of this analyzer is given in Section 3.6 below.

In block diagram form this method of spectral analysis is given in Figure A3.3. Since the driving function used in the experiments consisted of 10 sine waves, 10 identical circuits to Figure A3.3 were required.

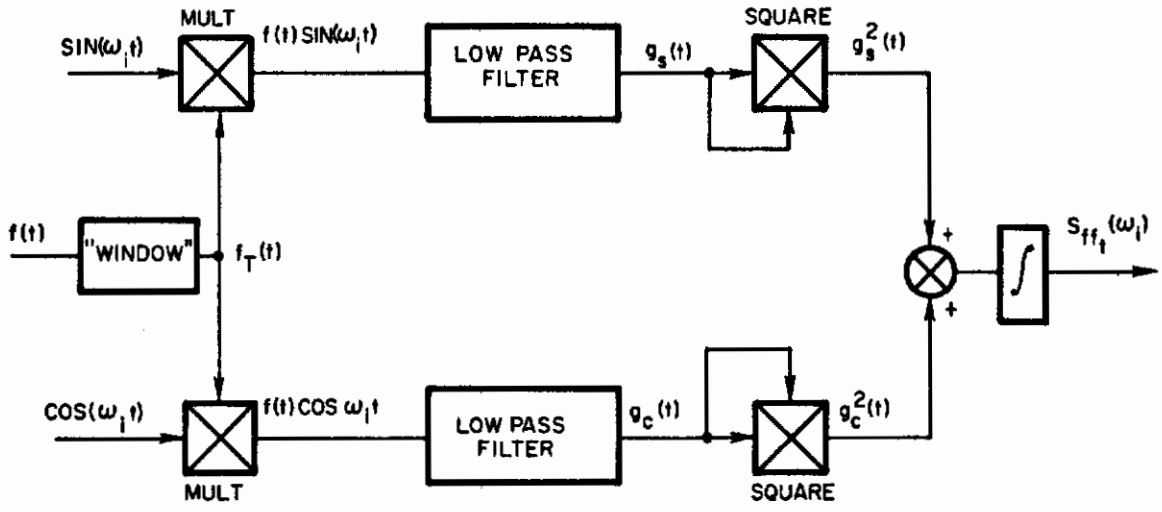


Figure A3.3 Block Diagram of Analog Spectral Analyzer

A3.6 Analysis of the Spectral Analyzer

Consider now a somewhat more rigorous presentation of the operation of the spectral analysis method of Figure A3.3 based on McRuer (1961a). Let the input function be $f(t)$ as before. If the time window is denoted by $w(t)$ then the inputs to the filters will be

$$\text{(A3.54)} \qquad \begin{cases} f_T \cos \omega_o t = f(t)\, w(t) \cos \omega_o t \\ f_T \sin \omega_o t = f(t)\, w(t) \sin \omega_o t \end{cases}$$

If the filter weighting functions are given by $h(t)$, the filter outputs will be

$$\text{(A3.55)} \qquad q_c(t) = \int_{-\infty}^{t} h(\tau)\, w(t-\tau)\, f(t-\tau) \cos \omega_o (t-\tau)\, d\tau$$

and

$$\text{(A3.56)} \qquad q_s(t) = \int_{-\infty}^{t} h(\tau)\, w(t-\tau)\, f(t-\tau) \sin \omega_o (t-\tau)\, d\tau$$

The upper limit in the integrations of (A3.55) and (A3.56) can be replaced by ∞ with the understanding that $f(t-\tau) = 0$ for negative values of its argument. The measured power spectral density $S_{xx_T}(\omega)$, as shown in Figure A3.3, is given by

$$\text{(A3.57)} \qquad S_{ff_T}(\omega) = \frac{1}{2T} \int_{-\infty}^{+\infty} [q_c^2(t) + q_s^2(t)]\, d\tau$$

Now since

$$\text{(A3.58)} \qquad \left[\int_{-\infty}^{+\infty} f(t,\tau_1)\, d\tau_1\right]^2 = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(t,\tau_1)\, f(t,\tau_2)\, d\tau_1\, d\tau_2$$

we can write (A3.57) in the form

$$(A3.59)\quad S_{ff_T}(\omega) = \frac{1}{2T}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\Big\{h(\tau_1)\,h(\tau_2)\,w(t-\tau_1)\,w(t-\tau_2)\Big[\cos\omega_o(t-\tau_1)\cos\omega_o(t-\tau_2) + \sin\omega_o(t-\tau_1)\sin\omega_o(t-\tau_2)\Big]\,f(t-\tau_1)\,f(t-\tau_2)\Big\}\,d\tau_1\,d\tau_2\,dt$$

Now, if $f(t)$ is a sample function of a stationary random process, we perform an ensemble average of $S_{ff_T}(\omega)$. The only statistical terms appearing in (A3.59) are $f(t-\tau_1)$ and $f(t-\tau_2)$. The expected value of their product is precisely the autocorrelation function of $F(t)$, i.e.,

$$(A3.60)\qquad R_{ff}(\tau_1-\tau_2) = E\Big\{f(t-\tau_1)\,f(t-\tau_2)\Big\}$$

and consequently we have

$$(A3.61)\quad \overbrace{\qquad}\,S_{ff_T}(\omega) = \frac{1}{2T}\iiint_{-\infty}^{+\infty}\Big\{h(\tau_1)\,h(\tau_2)\cos\omega_o(\tau_1-\tau_2)\,R_{ff}(\tau_1-\tau_2)\,w(t-\tau_1)\,w(t-\tau_2)\Big\}\,d\tau_1\,d\tau_2\,dt$$

Then the "window" function $w(t)$ can be integrated by noting that, from Parseval's relation:

$$\int_{-\infty}^{+\infty} w(t-\tau_1)\,w(t-\tau_2)\,dt$$

$$= \frac{1}{2}\int_{-\infty}^{+\infty}\left|W(\omega_1)\right|^2 e^{-j\omega_1(\tau_1-\tau_2)}\,d\omega_1$$

where $W(j\omega_1)$ is the Fourier transform of $w(t)$. The autocorrelation function appearing in (A3.61) can be replaced by the Fourier transform of the "true" power spectral density

$$(A3.63) \qquad R_{ff}(\tau_1 - \tau_2) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} S_{ff}(\omega)\, e^{j\omega(\tau_1 - \tau_2)}\, d\omega$$

The integration with respect to τ_o and τ_2 of the filter weighting function $h(t)$ results in an expression involving its Fourier transform $H(j\omega)$:

$$(A3.64) \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} h(\tau_1)\, h(\tau_2) \cos \omega_o (\tau_1 - \tau_2)\, e^{j(\omega - \omega_1)(\tau_1 - \tau_2)}$$

$$d\tau_1\, d\tau_2 = \frac{1}{2} \left\{ \left| H[(\omega - \omega_1) + \omega_o] \right|^2 + \left| H[(\omega - \omega_1) - \omega_o] \right|^2 \right\}$$

Using the above results, equation (A3.61) can be written in the form

$$(A3.65) \qquad \widetilde{S_{ff_T}(\omega_o)} = \int_0^{\infty} B_T(\omega, \omega_o)\, S_{ff}(\omega)\, d\omega$$

where $B_T(\omega, \omega_o)$ can be considered as an effective filter, by which the "true" spectral density is weighted. $B_T(\omega, \omega_o)$ is actually the convolution of two filters: the "window" or filter due to run length, and the physical filter used to extract $S_{ff_T}(\omega_o)$; i.e.,

$$(A3.66) \quad B_T(\omega, \omega_o) = \frac{1}{2T} \left\{ \int_{-\infty}^{+\infty} \left| W(\omega_1) \right|^2 \left| H(\omega + \omega_o - \omega_1) \right|^2 + \left| W(\omega_1) \right|^2 \left| H(\omega - \omega_o - \omega_1) \right|^2 d\omega \right\} +$$

$$+ \frac{1}{2T} \left\{ \int_{-\infty}^{+\infty} \left| H(\omega_1) \right|^2 \left| W(\omega + \omega_o - \omega_1) \right|^2 + \left| H(\omega_1) \right|^2 \left| W(\omega - \omega_o - \omega_1) \right|^2 d\omega \right\}$$

As we have seen in Section A3.4, if simplifying assumptions are made concerning $H(t)$ or $w(t)$, it is possible to obtain a simple analytic expression for the estimated spectral density. It is also interesting to note that because of the form of $B_T(\omega, \omega_o)$, the effective "passband" of B_T will be at least as wide as the wider of the filter passband and the spectral window $W(j\omega)$. Consequently, the resolution of a particular analyzer cannot be improved by choice of a narrower filter, but only by a greater run length.

A3.7 Design of Low Pass Filters

As discussed above in Section A3.4, the run length has been selected as 180 sec, and the lowest frequency is approximately 1/6 cps, therefore 30 cycles of f_1 will occur in one run; from this information we can compute the width of the spectral peak at f_1. The filter bandwidth (the frequency spacing between half power points) should be wider than the main lobe of the spectrum, and cutoff rapidly enough to avoid adjoining peaks. Filters of approximately 0.02 cps width and attenuation of 12 db octave were selected, resulting in a situation such as shown in the sketch of Figure A3.4 below. A filter meeting these requirements is a second-order system with a damping ratio of 0.7 and a resonant frequency of 0.01 cps. The transfer function of such a filter is given by:

$$G_F(s) = \frac{\omega_o^2}{s^2 + 2\xi\,\omega_o s + \xi\,\omega_o^2}$$

$$\omega_o \cong 0.07 \text{ rad sec}, \quad \xi \cong 0.7$$

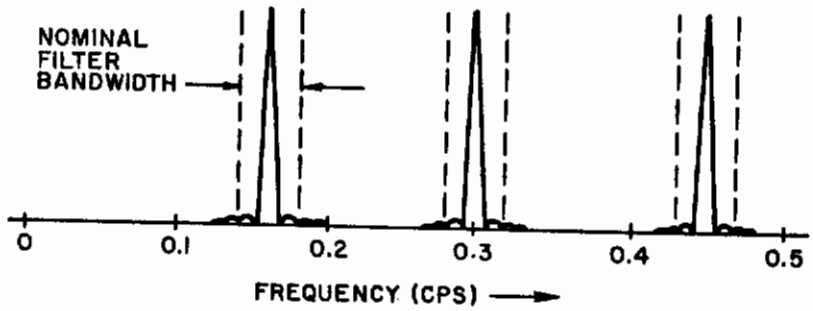


Figure A3.4 First Three Spectral Peaks of f(t)

There are a number of ways to realize such a filter, both passive and active. An active realization using operational amplifiers was selected since (1) the entire experiment was done on an analog computer and operational amplifiers are available, and (2) no consideration of loading problems is required, since the operational amplifiers have extremely low output impedance (less than 0.01 ohm for high-quality chopper stabilized amplifiers). The following methods were considered:

(1) Direct solution of the differential equation.

For this method of simulation we cross-multiply equation (A3.49) to yield

$$(s^2 + 2\xi\,\omega_o\,s + \omega_o^2)\;E_o(s) = \omega_o^2\,E_i(s)$$

Dividing by s^2 we obtain, after rearrangement of terms:

$$E_o(s) = [E_i(s) - E_o(s)]\,\frac{\omega_o^2}{s^2} - \frac{2\xi\,\omega_o\,E_o(s)}{s} \qquad \text{(A3.68)}$$

which can be represented by the analog circuit of Figure A3.5.

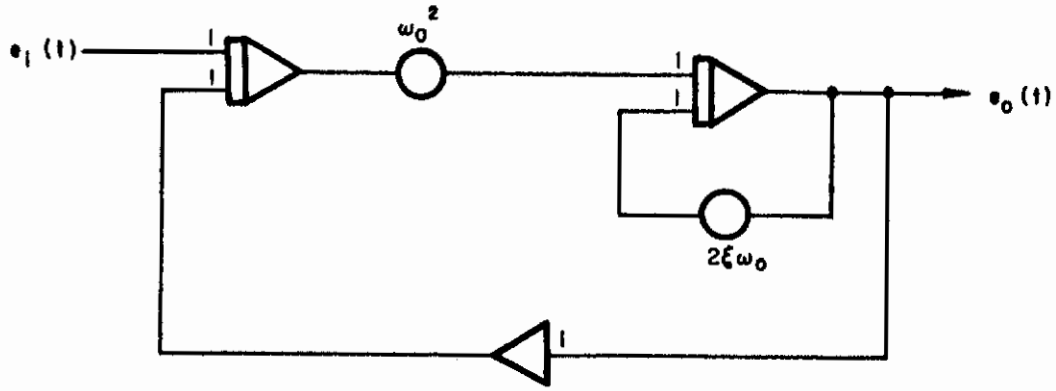


Figure A3.5 Second-Order Filter

The obvious disadvantage of this circuit is that it requires 3 amplifiers, and since 20 filters are required for 10 channels of spectral analysis, an alternate circuit with less amplifiers is desirable.

(2) <u>One-amplifier circuit using T-networks</u>

A one-amplifier circuit is shown in Figure A3.6. The feedback impedance is, in complex frequency notation:

$$Z_f(s) = \frac{A_f (1 + T_{2f} s)}{1 + sT_{1f} + s^2 T_{1f} T_{2f}} \qquad \text{(A3.69)}$$

and the input network impedance is given by

$$Z_i(s) = A_i (1 + T_i s) \qquad \text{(A3.70)}$$

where

$A_f = 2R_f$, $A_i = 2R_i$

$T_i = (R_i C_i)/2$

$T_{1f} = 2R_v C_{2f}$

$T_{2f} = (R_f C_{1f})/2$

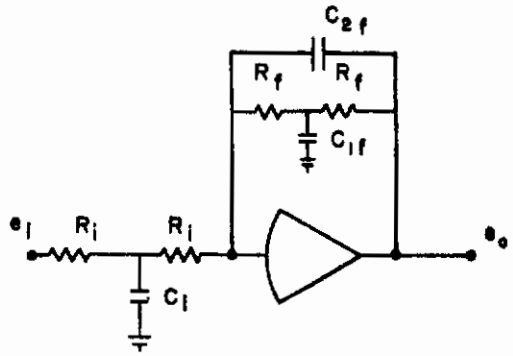


Figure A3.6 One-Amplifier Filter

The transfer function of the circuit of Figure A3.6 is then

$$(A3.71)\quad \frac{E_o(s)}{E_i(s)} = -\frac{A_f}{A_i}\,\frac{(1 + s\,T_{2f})}{(1 + s\,T_{1f} + s^2\,T_{1f}\,T_{2f})\,(1+T_i\,s)}$$

If we select $A_f = A_i$ and $T_i = T_{2f}$ then (A3.71) reduces to

$$(A3.72)\quad \frac{E_o(s)}{E_i(s)} = -\frac{1}{1 + s\,T_{1f} + s^2\,T_{1f}\,T_{2f}}$$

which is of the desired form.

Circuit A3.6 (and other similar circuits) have a major drawback, namely, that they rely on cancellation of a pole by a zero, and such cancellation is always imperfect due to component tolerances. Consequently, by raising the order of the system new problems may be introduced. Furthermore, many elements are required for this circuit (4 resistors and 4 capacitors) and this number can be reduced somewhat for other circuits.

(3) <u>One-amplifier realization using interconnected input and feedback networks</u>.

The circuit selected for experimental purposes is given by Johnson (1956) and shown in Figure A3.7a. The input-output relationship for this circuit is established by conventional circuit analysis techniques, such as application of Kirchhoff's laws. The resulting transfer function is given by:

$$(A3.73)\quad \frac{E_o(s)}{E_i(s)} = -\frac{R_3/R_1}{(R_2\,R_3\,C_1C_2)\,s^2 + R_2\,C_1\,(1 + R_3/R_1 + R_3/R_2)\,s + 1}$$

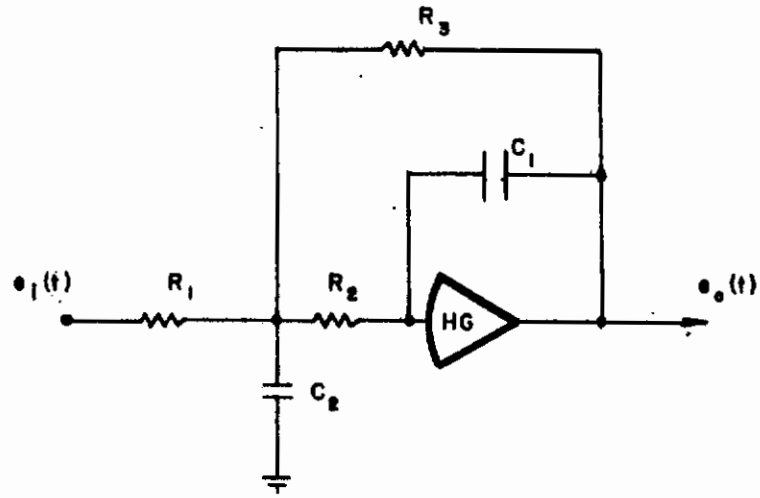


Figure A3.7a One-Amplifier Realization of 2nd Order Filter

Expression (A3.73) corresponds to the standard form

$$G(s) = -\frac{K}{s^2/\omega_o^2 + 2\xi\, s/\omega_o + 1}$$

and consequently we have the relations:

$$K = \frac{R_3}{R_1}\,; \quad \omega_o^2 = \frac{1}{R_2 R_3 C_1 C_2}; \quad \frac{2\xi}{\omega_o} = R_2\, C_1\, (1 + \frac{R_3}{R_1} + \frac{R_3}{R_2})$$

The choices of particular component values were governed by the following two factors:

a) Availability

b) Simplicity (i.e., integral or easy fractional values)

in addition to the obvious need to yield the correct values of ω_o and ξ. As outlined above, the desired values are $\omega_o = .07$, $\xi = .7$.

On the basis of availability of 1% resistors and 5% capacitors, the following values were selected:

$$R_3 = 20\ M, \qquad R_2 = R_1 = 10\ M$$

$$C_1 = 0.33 \mu fd, \qquad C_2 = 3\ \mu fd.$$

Consequently, substituting into (A3.74) we find:

$$\omega_o = \frac{1}{200} = 0.0707 \text{ rad/sec} \tag{A3.75}$$

$$\xi = (3.3)\ (5)\ \frac{\omega_o}{2} = 0.58$$

A3.8 Calibration and Adjustment of Analog Spectral Analyzer

Basically, the analog spectral analyzer constructed had 10 identical circuits and thus was capable of computing spectral estimates of 10 frequencies, as indicated in Figure A3.7b.

The signal to be analyzed, f (t), was constructed by summing the 10 sine waves required for the analyzer circuits and passing the resulting signal through a low pass filter, as indicated below in Figure A3.8: where ω_B was selected to be one of three values (∞ , 1.5 rad/sec and 3.0 rad/sec). Relay 1 in Figure A3.7 b remained open for 120 seconds, sufficiently long for the low-pass filter transients to attenuate to about 1% of their initial value then closed and remained closed for 180 seconds. The final values $k_e S_{xx_T}(\omega_i)$ were read on a digital voltmeter and recorded. The gains at the summer input (nominally 0.8) were adjusted to correct for d.c. gain errors in the low-pass filters.

We can compute the theoretically expected output by using the relationships developed above. Consider first the situation where no filter is used on the input (i.e., $\omega_B = \infty$). Then

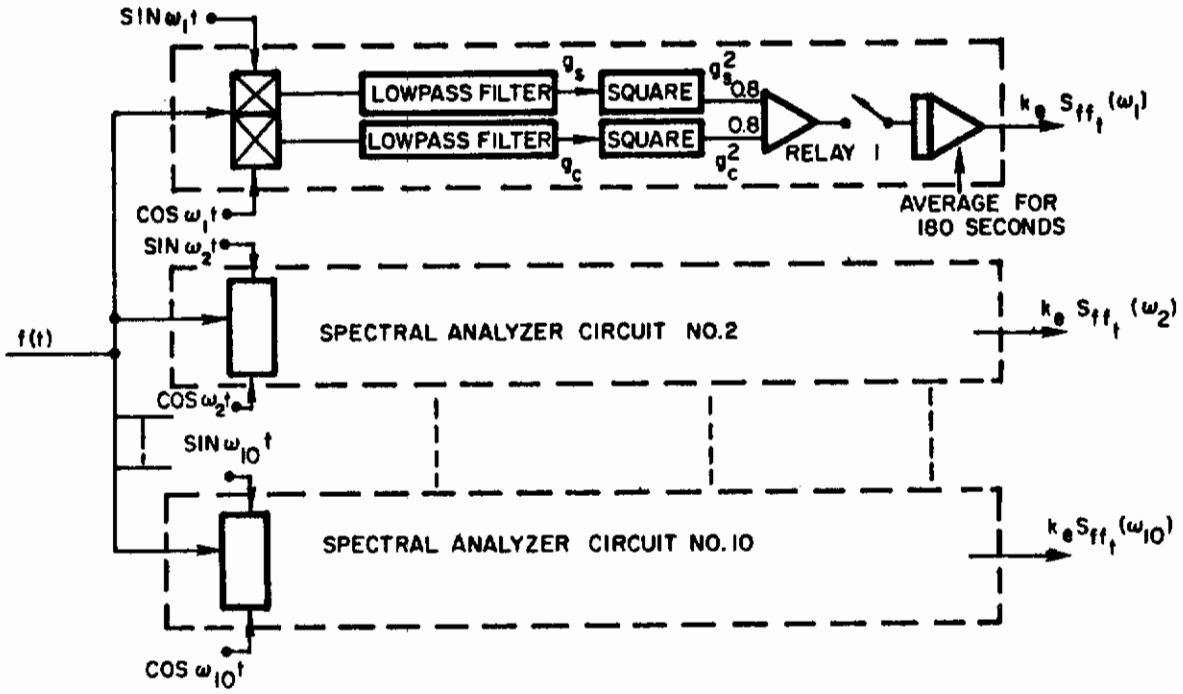


Figure A3.7b Block Diagram of Analog Spectral Analyzer

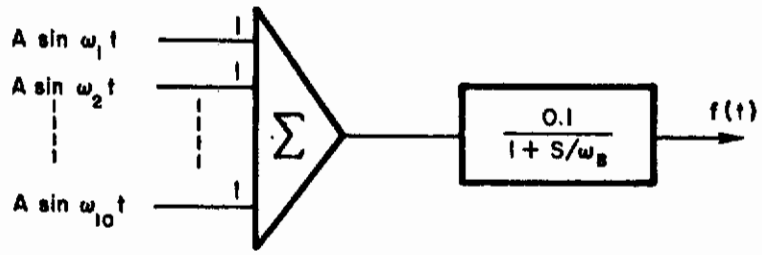


Figure A3.8 Circuit Used to Obtain Input Function

$$(A3.76)\quad f(t) = \frac{1}{10}\sum_{i=1}^{10}(A_i \sin \omega_i t + B_i \cos \omega_i t); \qquad B_i = 0$$

(The factor of 1/10 was needed to prevent overloads when summing 10 sinusoids, each with a 65V peak value, on the computer.) The multiplier outputs are:

$$(A3.77)\quad \frac{1}{100} f(t)\,[C_1 \cos \omega_o t] = \frac{C_1}{100}\left[\sum_{i=1}^{10} A_i \sin \omega_i t\right] \cos \omega_o t$$

$$(A3.78)\quad \frac{1}{100} f(t)\,[C_1 \sin \omega_o t] = \frac{C_1}{100}\left[\sum_{i=1}^{10} A_i \sin \omega_i t\right] \sin \omega_o t$$

where C_1 is the amplitude of the sinusoids supplied to the individual analyzer circuits. Now equation (A3.77) has no d.c. component, but from (A3.78) we get

$$(A3.79)\quad \frac{1}{100} f(t)\,[C_1 \sin \omega_o t] = \frac{C_1 A_o}{200} - \frac{C_1 A_o}{200} \sin 2\omega_o t$$

$$\left[\frac{C_1}{100} \sin \omega_o t \sum_i \sin \omega_i t\right]_{i \neq o}$$

From (A3.75) the low-pass filters have a gain of 2, consequently we have for the filter outputs q_s and q_c:

$$(A3.80)\quad q_s(t)\Big|_{\omega_o} = \frac{CA_o}{100} + R_s; \qquad q_c(t)\Big|_{\omega_o} = R_c$$

where R represents a ripple component due to the physical characteristics of the low-pass filters which do not reject perfectly all frequencies outside of their pass-band. The squarer output is then

$$(A3.81) \qquad [q_s(t)]^2\Big|_{\omega=\omega_o} \cong \frac{(C_1 A_o)^2}{10^4} \times 10^{-2}; \quad [q_c(t)]^2 \cong 0$$

Consequently the summer output is, including the gain factor of 0.8:

$$(A3.82) \qquad \Sigma_{out} = \frac{0.8}{10^6} C^2 (A_o^2) \text{ (+ Ripple)}$$

The integrator output, then, assuming the average value of the ripple to be zero, is

$$(A3.83) \quad I_o = k_e S_{ff_T}(\omega_o) = \frac{0.8}{10^6} C_1^2 \int_o^{180} A_o^2 \, dt = \frac{(0.8)(180) C_1^2}{10^6} A_o^2$$

Since the desired estimate is: $S_{ff}(\omega_o) = \frac{1}{2} c_o^2$ where c_o is the Fourier component amplitude, we have from the above $k_e = 2(0.8)(180)C_1^2/10^6$. Since for the simulation $C_1 = 65.0$ volts, $A_o = 6.5V$, we get from (A3.83) the expected output for all 10 channels:

$$(A3.84) \qquad I = K_e S_{ff_T}(\omega_i) = 25.8 \text{ volts}$$

When the sum of sine waves is processed through a low-pass filter,

$$(A3.85) \qquad G_1(j\omega) = \frac{1}{1 + j\omega/\omega_B}$$

then we can compute the output power spectrum of the filter knowing the input, as done previously. The magnitude of the spectral peaks of the input function is given in equation (A3.86)

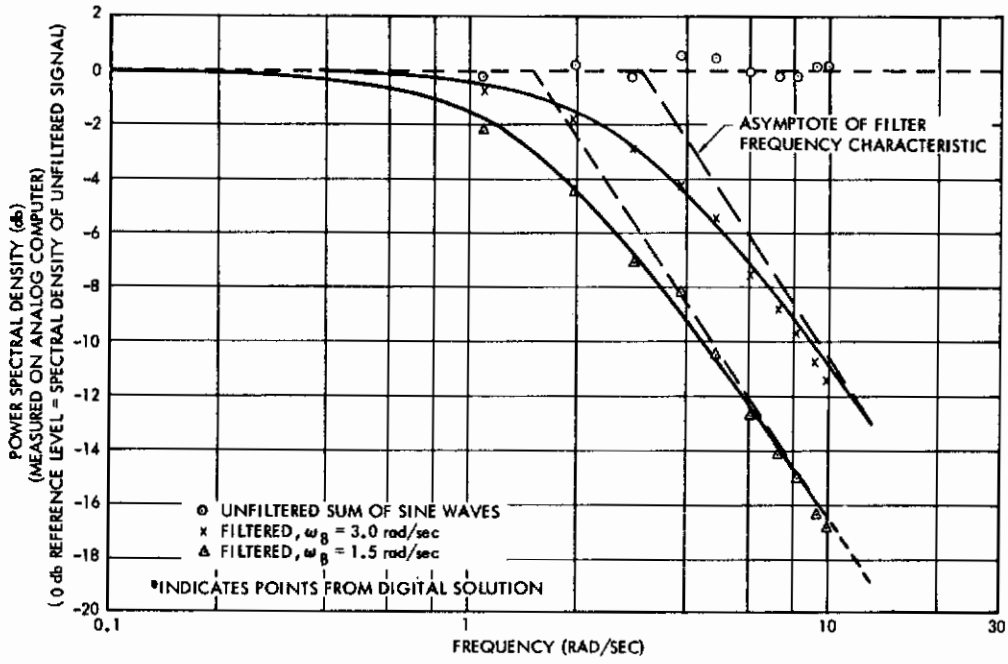
$$(A3.86) \qquad S_{ff_T}(\omega_i) = \frac{25.8V}{k_e} \text{ volts/rad/sec}$$

Then the filter output magnitude is given by

$$S_{ff_T}(\omega_i)\Big|_{\text{filtered}} = \frac{25.8/k_e}{1 + \omega_i^2/\omega_B^2}$$

Letting $\omega_B = 1.5$ rad/sec and $\omega_B = 3.0$ rad/sec results in the 3 calibration curves shown in Figure A3.9. It can be seen that the measured points agree with the theoretical curves within approximately 1/2 db which is considered adequate accuracy for the present experiment.

Points obtained from the digital computation of power spectra (discussed in Appendis 5) are also indicated in Figure A3.9.



Calibration Curves for Analog Spectral Analyzer
Figure A3.9

REFERENCES ON
ANALOG MEASUREMENT OF POWER SPECTRA

1. Bendat, J. S. (1958) Principles and Applications of Random Noise Theory, New York: John Wiley, 1958.

2. Benepe, O. J., Narasinham, R. and Ellson, D. (1954) "An Experimental Evaluation of the Application of Harmonic Analysis," WADC Technical Report 53-384, May 1954.

3. Bennett, W. R. (1956) "Methods of Solving Noise Problems," Proc IRE, 44: 609-38, May 1956.

4. Blackman, R. B. and Tukey, J. W. (1958) The Measurement of Power Spectra, New York: Dover Publications, 1958.

5. Burke, C. J., Narasinham, R. and Benepe, O. T. (1953) "Some Problems in the Spectral Analysis of Human Behavior Records," WADC Technical Report 53-27, Wright Air Development Center, July, 1953.

6. Chang, S. S. L. (1954) "On the Filter Problem of the Power Spectrum Analyzer," Proc. IRE, 42: 1278-82, 1954.

7. Chang, S. S. L. (1961) Synthesis of Optimum Control Systems, New York: McGraw-Hill, 1961. (In particular Chapter 7: "Spectral Density Estimates.")

8. Davenport, W. B. and Root, W. L. (1958) An Introduction to the Theory of Random Signals and Noise, New York: McGraw-Hill, 1958.

9. Gilbert, E. G. (1960) "Computation of Correlation and Spectral Functions by Orthogonal Filtering," AIEE Communication and Electronics, pp. 954-954, January 1960.

10. Goldman, S. (1953) Information Theory, New York: Prentice-Hall, 1953.

11. Grenander, U. and Rosenblatt, M. (1957) Statistical Analysis of Stationary Time Series, New York: John Wiley and Sons.

12. Johnson, C. L. (1956) Analog Computer Techniques, New York: McGraw-Hill, 1956.

13. Krendel, E. (1951) "A Preliminary Study of the Power Spectrum Approach to the Analysis of Perceptual Motor Performance," AF Technical Report No. 6723, Wright Air Development Center, October, 1951.

14. Krendel, E. (1952), "The Spectral Density Study of Tracking Performance," WADC Technical Report 52-11, Wright Air Development Center, January, 1952.

15. Lighthill, M. J. (1960), Fourier Analysis and Generalized Functions, Cambridge University Press, 1960.

16. McRuer, D. T. (1961a) "Autocorrelation and Spectral Density of Periodic Functions," Systems Technology, Inc. Technical Memo No. 78, March 1961.

17. McRuer, D. T. (1961b) "Computation of Spectra with Finite Run Length," Systems Technology, Inc. Technical Memo No. 81, Los Angeles, March, 1961.

18. Narasinham, R. and O.J. Benepe (1951) "The Use of Autocorrelation Functions in the Harmonic Analysis of Human Behavior," AF Technical Report No. 6529, Wright Air Development Center, October 1951.

19. Parzen, E. (1960) "Mathematical Considerations in the Estimation of Spectra," Technical Report No. 3, Applied Math. and Statistics Laboratories, Stanford University, August 9, 1961.

20. Seltzer, J.L. and McRuer, D.T. (1959) "Survey of Analog Cross-Spectral Analyzers," WADC Technical Report 59-241, Wright Air Development Division, December, 1959.

Appendix 4

ANALOG SIMULATION OF SAMPLED DATA SYSTEMS

The use of an analog computer for representation of sampled-data systems is well established, with the first references to such work dating to about 1954 (See Bibliography at end of Appendix). However, the simulation of human operator models has required the development of some novel techniques which are summarized in this Appendix.

A4.1 Introduction and Background

The basic circuit presented in the majority of the references is the approximate simulation of a sampler followed by a zero-order hold, as shown in Figure A4.1.

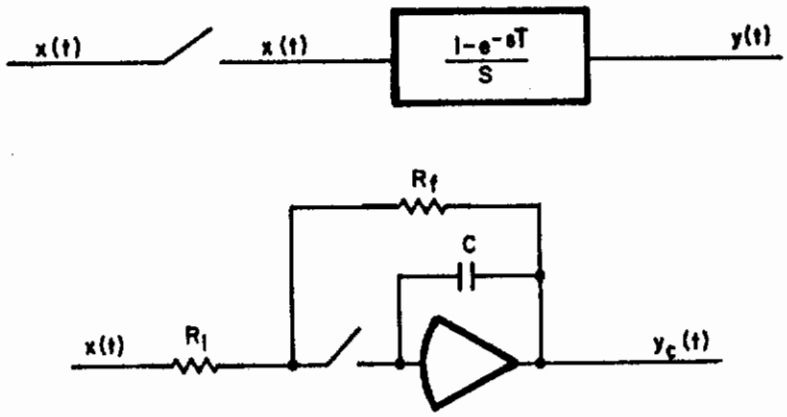


Figure A4.1 Simulation of Sampler and Zero-Order Hold

When the switch S in Figure A4.1(b) is closed, the circuit has the transfer function

$$\frac{Y_c(s)}{X(s)} = -\frac{R_f}{R_i}\frac{1}{1 + R_f C s} \tag{A4.1}$$

and consequently the transient solution is

$$y_c(t) = -\frac{R_f}{R_i}\left(1 - e^{-t/R_f C}\right) + y_c(o) \tag{A4.2}$$

If the time constant $(R_f C)$ is sufficiently small compared to the sampling period T, the rise time may be negligible. In the sampled-data simulator used for the work reported here (see Reich and Perez, 1961) the choice of parameters was the following:

$$R_f = R_i = 50 \text{ K}\Omega$$

and $$C = 0.001 \text{ mfd (or .01 mfd by selection)}$$

Thus the time constant becomes:

$$R_f C = (.05M)(.001 \text{ } \mu\text{fd}) = 50 \text{ } \mu \text{ sec.}$$

The gain of the circuit is unity, since $R_f = R_i$ and the maximum allowable sampling rate is determined by the time constant of 50 μ sec. If we assume that the input changes insignificantly during the period of switch closure, and arbitrarily let the switch closure time τ be restricted to

$$\tau \leq .05T \tag{A4.3}$$

in order to approximate impulse sampling, then we have the following relations:

Time constant $R_f C = 50\ \mu$ sec

Switch closure time $\tau \geq 9.2$ RC in order to charge to within $\frac{1}{10,000}$ of input

$\therefore\ \tau \geq 9.2\ (50)\mu$ sec. If we choose $\tau = 500\mu\text{sec} = 0.5$ msec

then $T_{min} = 20\ (.5\ \text{msec})$ by relation (A4.3) or

$T_{min} = 0.01$ seconds and $\underline{f_{max} = 100\ \text{cps}}$

Since in the human tracking problem the maximum sampling frequencies expected are of the order of 3/second, we can use the larger value of capacitor (0.01 mfd) to obtain better stability and still obtain pulse durations considerably less than 5 % of the sampling period.

When the switch is open, the sampled voltage will be held by the capacitor. In practical present-day computers the leakage resistance of the capacitor is sufficiently large to make the exponential decay negligible during one sampling period.

It should be noted that the circuit of Figure 2.1(b) is <u>not</u> a simulation of a zero-order-hold, but rather "sampler-followed-by zero-order-hold", since the modulated pulse train $x^*(t)$ does not appear in the circuit.

A4.2 <u>Solution of Difference Equations</u>

Consider a difference equation of the form

$$(A4.4)\quad y(t_n) + a_1 y(t_{n-1}) + \ldots + a_{n-1}\, y(t_1) + a_n\, y(t_o) =$$

$$b_o x(t_n) + b_1 x(t_{n-1}) + \ldots + b_{n-1} x(t_1) + b_n x(t_o)$$

where t_n is the n-th sampling instant after time t_o. To solve such an equation on the analog computer, (2n + 1) sample-and-hold circuits are required to obtain and store the present and past samples of the input $x(t_n)$ and $y(t_n)$. In block diagram form the circuit is indicated in Figure A4.2 for a second order equation.

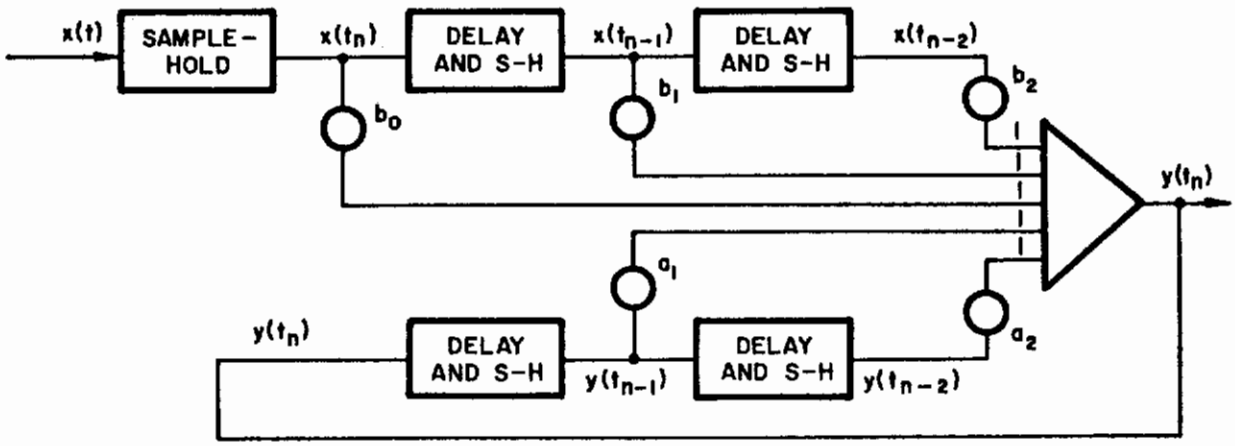


Figure A4.2 Block Diagram of Difference Equation Solution

Thus, the n-th sample of the voltage x or y is obtained by means of a sample-and-hold unit and passed on, at successive sampling instants, to additional storage units. To obtain the transmission of a sampled value, a cascade of sample-and-hold circuits of the type of Figure A4.1 can be used, in either of the two ways indicated in Figure A4.3.

In Figure A4.3(a) the chain of S-H amplifiers is activated by energizing the sampling relays in turn, beginning with the last and working backwards. This arrangement has the following features:

(a) Only one amplifier is required for obtaining and holding each past sample.

(b) The sampling times must be adjusted carefully in order to insure that a value is transmitted to the next channel fully before new information is admitted. If many channels are cascaded, the sum of the individual delays in each channel can amount to a significant proportion of the sampling period.

Example

In the simulator utilized in this work, as outlined above in Section A4.1 we have the amplifier time constant

$$RC = 50\ \mu \text{ sec}$$

and the switch closure time τ,

$$\tau = 10\ RC = 0.5 \text{ m sec}$$

Thus, if the n-th sampling switch closes at the sampling instants (kT), the (n+1)st (or previous switch in the chain) will close at the times (kT + 0.5) m sec, the (n+2)st switch at (kT + 1.0) m sec and so forth.

An alternate method of obtaining past samples from a cascaded chain of sample-and-hold circuits is shown in Figure A4.3(b). In this system a group of "sampling" relays close simultaneously and at least $t_D \geq 10\ RC$ seconds later, when the "sampling" relays have opened, the "present" relays close and transfer the sampled values to the next S-H circuit in each pair. This arrangement has the following features:

(a) All sampling operations take place simultaneously, rather than in sequence as in the method of Figure A4.3(a).

(b) Two amplifiers are required for each channel of sample-hold.

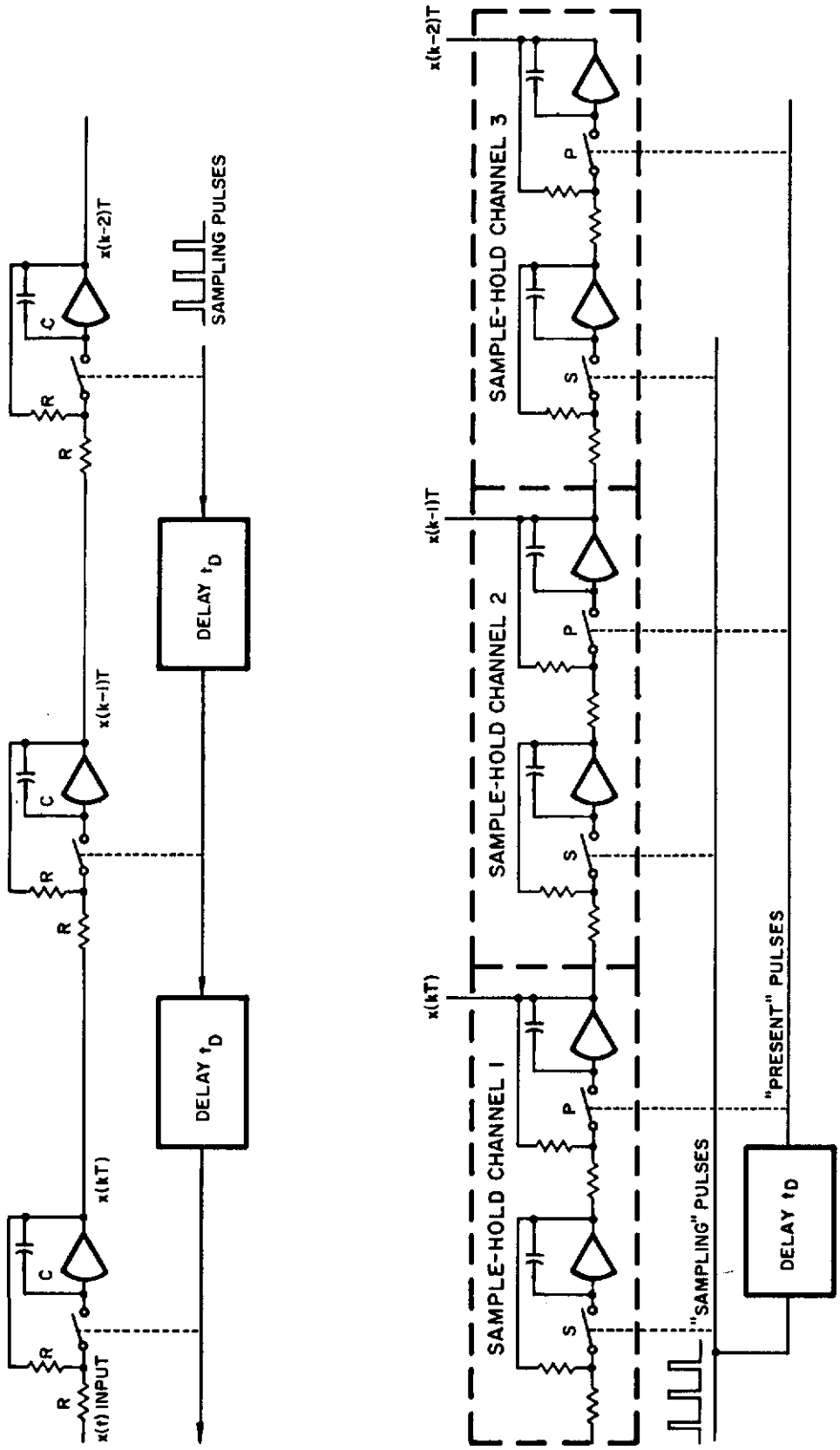


Figure A4.3 Alternate Methods of Computing Present and Past Samples

The simulator used in these experiments uses the two-amplifier method of simulating sample-and-hold circuits, and future references will be to this system exclusively, unless clearly stated otherwise.

A4.3 Simulation of a First-Order Hold

The circuits discussed above are used for simulation of a sampler and zero-order hold. The human operator models discussed in this report make use of a first-order-hold to obtain exact extrapolation of constant-velocity inputs. The first order hold can be described by its transfer function

$$H_1(s) = T(1 + sT)\left(\frac{1 - e^{-sT}}{sT}\right)^2 \tag{A4.5}$$

or by its output time response between sampling instants:

$$y(t) = \frac{y(t_n) - y(t_{n-1})}{T}(t - t_n) + y(t_n) \quad (t_n \leq t < t_{n+1}) \tag{A4.6}$$

Note that the output is continuous between sampling instants and depends only on the present and past sample values of the input. An analog computer circuit for simulation of a sampler and first-order-hold is given in Figure A4.4, both in block diagram form and as a detailed computer diagram, based on the availability of the sampling pulse train used in the S-H circuits.

Two major points can be observed in this circuit:

(a) The integrator must be reset to zero at the beginning of each sampling period. In the system of Figure A4.4(b) the sampling pulse used for the S relay in the S-H circuits is also used for the reset operation by means of a relay amplifier. (An electronic switch

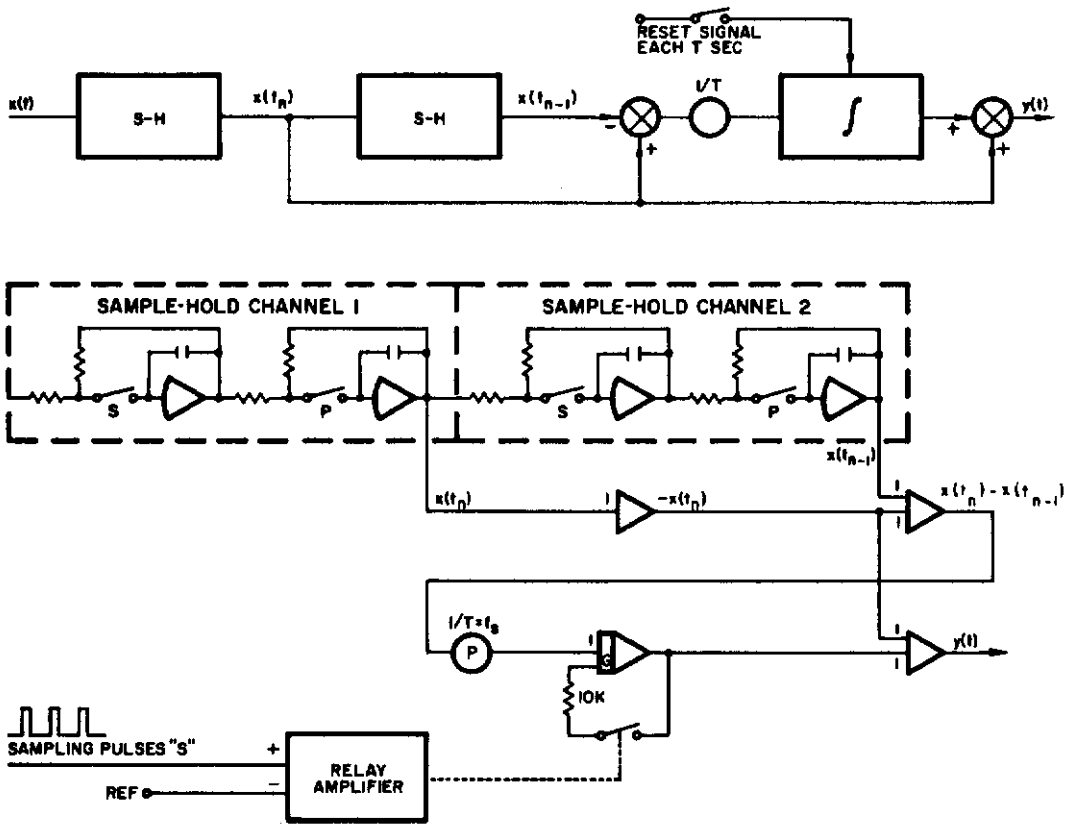


Figure A4.4. Simulation of Sampler and First-Order-Hold

could be used here to advantage, but the speed requirements in the problem do not make it mandatory).

(b) Potentiometer "P" can be used to transform this circuit into a simulation of a "partial velocity correction hold", if it is set to k/T instead of $1/T$ as indicated.

Then for $k = 0$ we have a zero-order hold,
for $k = 1$ we have a full first-order-hold, and
for $0 < k < 1$ we have partial velocity correction.

A4.4 Simulation of Modified Hold Circuits

In Chapter 3 (Section 3.4) the concept of a "modified" first order hold circuit was introduced. This circuit is based on continuous rather than discrete slope estimation. The conventional and modified hold circuits with partial velocity correction are shown in block diagram form in Figure A4.5.

The simulation of the modified PVH circuit can be accomplished directly by using the 2-amplifier sample-and-hold channels discussed previously. The simulation diagram is shown in Figure A4.6. The prominent features of this circuit are the following:

(a) The use of S-H channels makes it possible to introduce time delays in series with the hold circuit of any magnitude (for $D \neq T$) as required for simulation of the "reaction time". The sampling pulse delay is set to the minimum relay closure time (approximately 50 μ sec) if no additional time delay is desired. If an additional delay is desired, the "present" pulses S_p are correspondingly delayed.

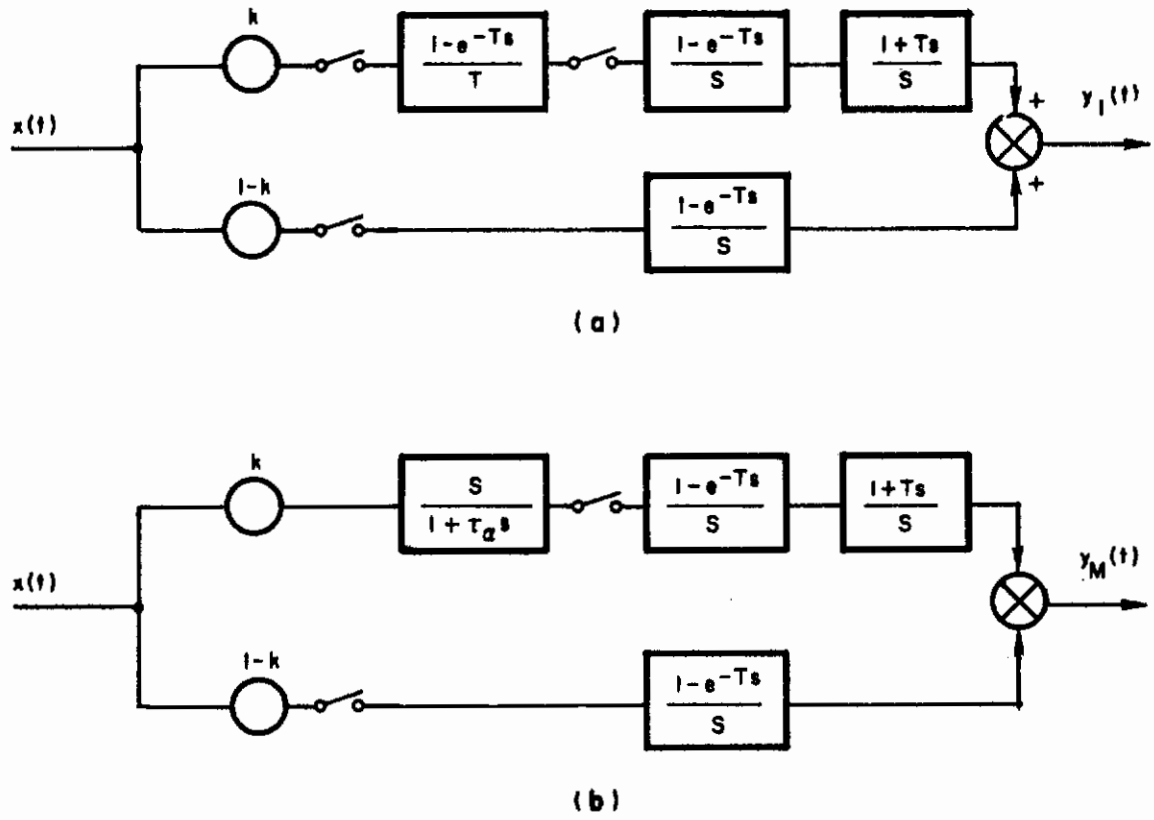


Figure A4.5 (a) Partial Velocity Hold Circuit
(b) Modified Partial Velocity Hold Circuit

(b) Relay S_1 closes when the computer is "Reset" in order to discharge the capacitor.

(c) The approximate differentiation circuit used can be described by the transfer function $G(s) = s/(1 + \tau_d s)$ where the time constant τ_d is given by $\tau_d = R_d C_d$ in Figure A4.6.

Time responses of the modified hold circuits for various values of partial correction k are shown in Figure A4.7.

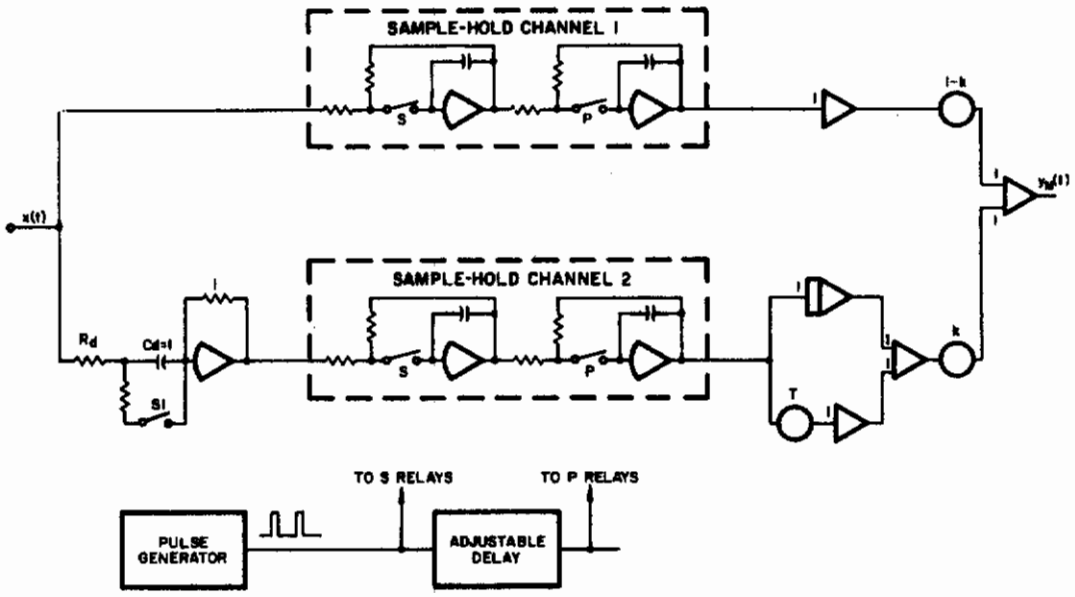


Figure A4.6 Analog Computer Simulation of Modified Partial Velocity Hold

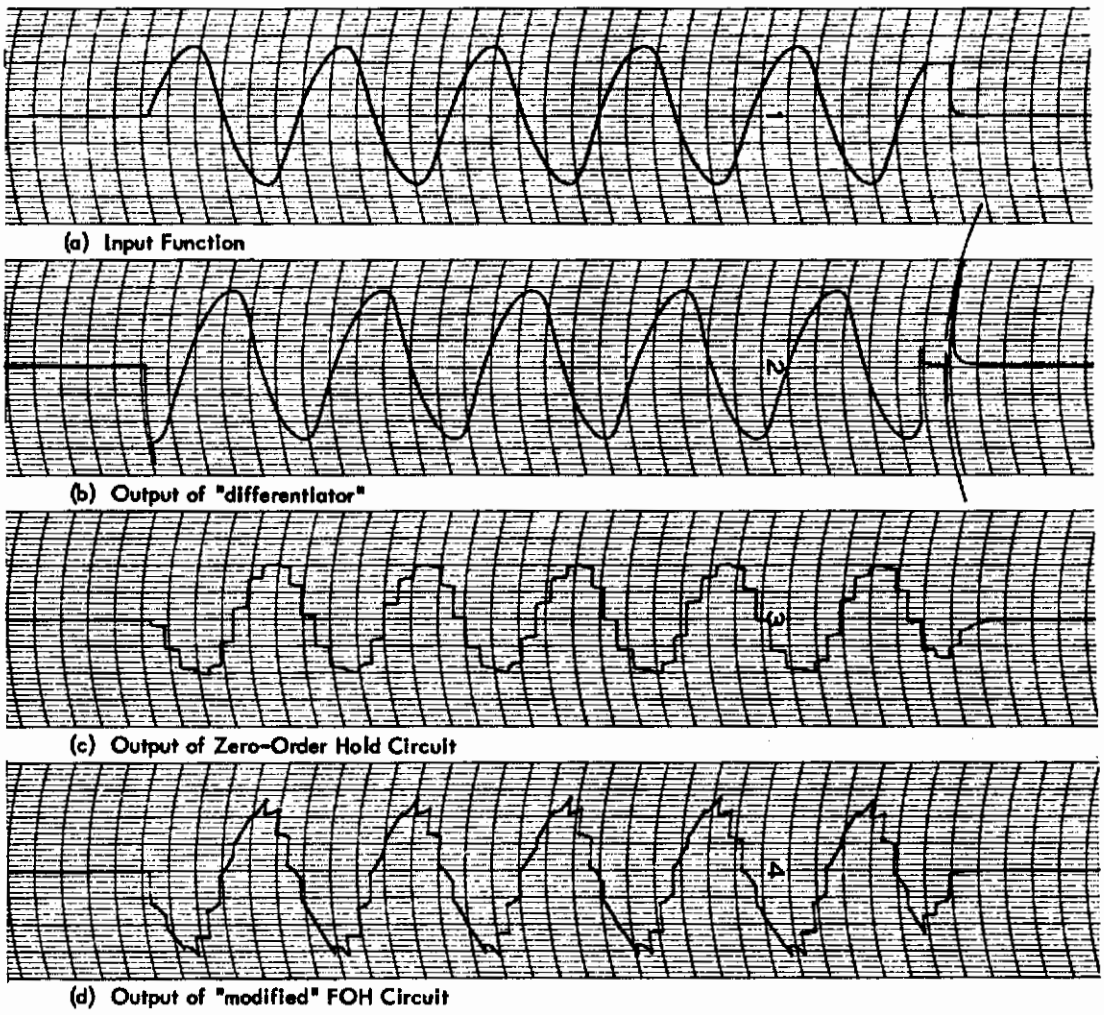


Figure A4.7 Time Response of Modified FOH Circuit

REFERENCES ON

ANALOG COMPUTER STUDY OF SAMPLED-DATA SYSTEMS

1. Andeen, R. E. (1959) "A Wide Range Sampler with Zero Order Hold", Control Engr., 6:151, May 1959.

2. Andrews, J. A. (1960) "The Dynamic Storage Analog Computer - DYSTAC", Proc. Western Joint Computer Conf., May 1960.

3. Bekey, G. A. (1959) "Generalized Integration on the Analog Computer", IRE Trans. on Electr. Computers, EC-8:210-217, June 1959.

4. Blanyer, C. G. (1959) "Sampled-Data Techniques in Complex Simulation", Instruments & Control Systems, (Simulation Council Newsletter), June 1959, 907-908.

5. Bourbeau, F. J. "Theory, Applications, and Bandwidth Limitations of an Analog Sampled-Data Simulator", presented Third International Conference on Analog Computation, September 1961.

6. Chestnut, H., Dabul, A. and Leiby, D. (1957) "Analog Computer Study of Sampled-Data Systems", Proc. Conf. on Computers in Control Systems, (AIEE), Atlantic City, October 1957, 71-77.

7. Diamantides, N. D. (1956) "A Multi-purpose Electronic Switch for Analog Computer Simulation and Autocorrelation Applications", IRE Trans. on Elect. Comp., EC-5:197-202, December 1956.

8. Elgerd, O. I. (1958) "Analog Computer Study of Transient Behavior and Stability Characteristics of Serial-Type Digital Data Systems", Communication and Electronics, (Trans. AIEE), 358-366, July 1958.

9. Findeisen, W. and Mardal, W. (1960) "Investigation of Digital Speed Servomechanism with a Conventional Analog Computer", Proc. IFAC, Moscow, June 1960.

10. Freilich, A. (1960) "Analog Simulation of Digital Control Loops", Proc. Fifth Symp. on Ballistic Missiles and Space Technology, 1:457-469, Academic Press, 1960.

11. Giloth, P. K. (1956) "A Simulator for the Analysis of Sampled-Data Control Systems", Proc. Nat'l. Simulation Conf., Dallas, January 1956.

12. Jury, E. I. and Semelka, F. W. (1958) "Time Domain Synthesis of Sampled-Data Systems", Trans. ASME, November 1958, 1827-1838.

13. Klein, R. C. (1955) "Analog Simulation of Sampled-Data Systems", IRE Trans. on Telemetry and Remote Control, May 1955, 2-7.

14. Korn, G. A. (1952) "The Difference Analyzer: A Simple Differential Equation Solver", MTAC, January 1952.

15. Rawdin, E. (1959) "Time Multiplexing Applied to Analog Computation", IRE Trans. on Electr. Computers, EC-8: 42-47.

16. Reich, J. E. and Perez, J. J. (1961) "Design and Development of a Sampled-Data Simulator", Proc. Western Joint Computer Conf., May 1961.

17. Schneider, G. (1960) "On the Development of Linear Difference Equation Units", Proc. IFAC, Moscow, June 1960.

18. Shumate, M. S. (1959) "Simulation of Sampled-Data Systems Using Analog to Digital Converters", Proc. Western Joint Computer Conf.,1959.

19. Tou, J. T. (1959) Digital and Sampled-Data Control Systems, New York: McGraw-Hill, 1959, 130-132.

20. Wadel, L. B. (1955) "Analysis of Combined Sampled and Continuous Data Systems on an Electronic Analog Computer", IRE Conv. Record, pt. 4, 3-7, 1955.

21. Wadel, L. B. (1954) "An Electronic Differential Analyzer as a Difference Analyzer", Jour. ACM., 1:128-36, July 1954.

22. Wadel, L. B. (1956) "Simulation of Digital Filters on the Electronic Analog Computer", Jour. ACM., 3:16-21, January 1956.

Appendix 5

DIGITAL COMPUTATION OF POWER SPECTRAL DENSITY

A5.1 Introduction

The purpose of this Appendix is to outline briefly the digital computer program used for computation of spectral density in this study.

The original input to the digital computer was a magnetic tape recording of error and output signals in the tracking loop. This data was digitized as indicated below.

The outputs of the program were used for three major purposes:

(1) To check on the accuracy of analog measurements of error spectra;

(2) To provide information on the presence of energy between and beyond the 10 spectral peaks which characterize the input function;

(3) To provide cross-spectral densities $S_{re}(jw)$ and $S_{rc}(jw)$ (between input and error, and input and output respectively), required for synthesis of the continuous operator model. The program was written by Negron (1961) at Space Technology Laboratories, Inc. and all the runs were made on an IBM 7090 computer at STL.

A5.2 Basic Equations of the Program

The equations are taken from Blackman and Tukey (1958). The autocorrelation function is computed as

$$(A5.1) \qquad \phi_i(\tau) = \frac{1}{T - |\tau|} \int_o^{T_R - \tau} g(t)\, g(t + \tau)dt \qquad (T_R = 2T)$$

where g(t) is the function being analyzed and 2T is the total length of the record. The time-window used is

$$w(\tau) = 0.54 + 0.46 \cos \frac{\pi\tau}{T}, \qquad |\tau| < T$$

$$(A5.2) \qquad = 0.04 \qquad |\tau| = T$$

$$= 0 \qquad |\tau| > T$$

which is called "hamming" by Tukey. The modified autocovariance function is then computed as

$$(A5.3) \qquad R_i(\tau) = \phi_i(\tau)\, w_i(\tau)$$

for the single i-th function.

The spectral density is then estimated by the Fourier cosine transform

$$(A5.4) \qquad S_{gg}(f) = 4 \int_o^{\infty} R_i(\tau) \cos \omega\tau d\tau$$

which is related to the spectral density $S_{gg}(\omega)$ discussed in Appendix 3 by the factor 2π.

For computation of cross-spectra between two records x(t) and y(t) the program first computes the modified cross-correlation function using the same time window:

$$(A5.5) \qquad R_{xy}(\tau) = w(\tau) \left[\frac{1}{2(T-\tau)} \int_{-T+|\tau|}^{T-|\tau|} x(t)\, y(t + \tau)dt \right]$$

and hence the cross spectral density is estimated as

$$(A5.6) \quad S_{xy}(f) = \int_{-\infty}^{+\infty} R_{xy}(\tau) \cos \omega\tau d\tau + j \int_{-\infty}^{+\infty} R_{xy}(\tau) \sin \omega\tau d\tau$$

Thus, the cross spectrum is complex. The program computes the magnitude and phase of $S_{xy}(f)$. By letting x(t) be the input r(t) and letting y(t) be the error e(t) and operator's output c(t) respectively, it was possible to compute the cross spectral densities $S_{re}(f)$ and $S_{rc}(f)$ which are required for synthesis of the human operator models.

In addition, the program computes the coherence function C(f) which is a measure of the degree of linearity in the relationship, i.e.

$$(A5.7) \quad C(f) = \frac{|S_{xy}(f)|^2}{S_x(f)\, S_y(f)}$$

A5.3 Preparation of Input Functions

The original input functions to the program were recorded on a General Precision Type FM magnetic tape recorder. Both man and model parameters were recorded as shown in Figure A5.1.

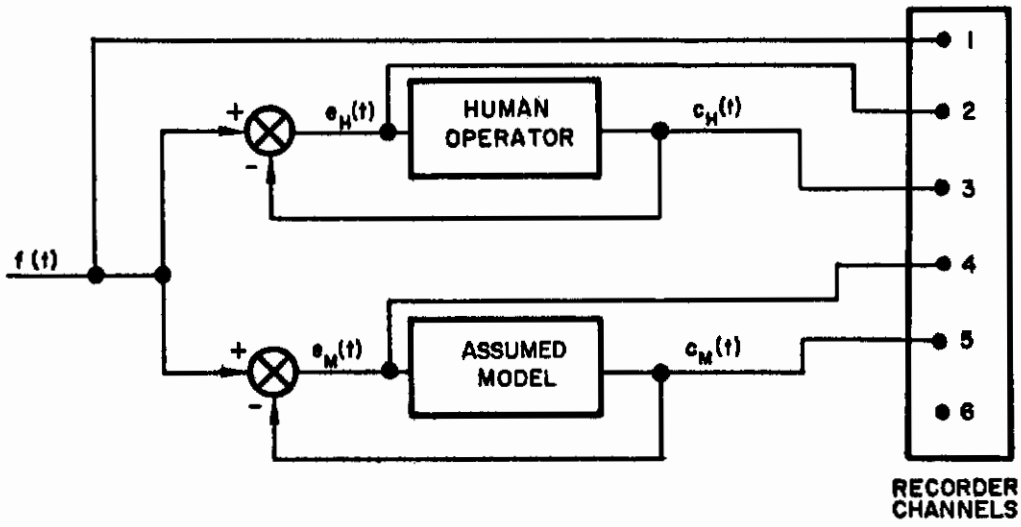


Figure A5.1 Tape Recording Arrangement

The analog data were digitized in preparation for the digital program. The analog data originally recorded at 3 3/4"/sec and played back at 60"/sec (a speed-up factor of 16) for processing through an analog to digital converter were digitized at 1666 operations per sec for 4 commutated channels. Thus, each function was sampled at

$$f_{samp} = \frac{1666}{16x4} \simeq 26 \text{ samples/sec} \tag{A5.8}$$

Since the relevant frequencies are below 3 cps, this rate is sufficiently above high to avoid difficulties due to Nyquist folding frequencies.

A5.4 Selection of Parameters for Programs

The operation of the program requires selection of Δt, the interval at which the correlation function is computed and the number of data points (n) to be used. These two numbers determine the run length and the frequency resolution respectively.

Five minutes of data per subject per run were available. Thus, there were (300 sec x 26 samples/sec) or 7800 data points available. To obtain a comparison with analog data (even though filter transients do not enter into the digital method) the first 2000 points were skipped and 5000 points selected.

The minimum Δt possible was

$$(\Delta t)_{min} = \frac{1}{26} \text{ sec} = .0384 \text{ sec} \tag{A5.9}$$

since the data were sampled at 1/26 sec intervals. The interval used in the computation was 1/13 sec or $\Delta t = .0768$ seconds; the maximum frequency computed was (from the sampling theorem)

(A5.10) $$f_{max} = \frac{1}{2(\Delta t)} = 6.5 \text{ cps}$$

The frequency resolution follows from the selection of the number of lag values (m) and the size of the interval Δt. With m = 260 and $(\Delta t) = .0768$, we have

(A5.11) $$\Delta f = \frac{1}{2(\Delta t)m} = \frac{1}{2(.0768)(260)} = \frac{1}{2(20)} = .025 \text{ cps}$$

Since the 10 spectral peaks in f(t) were approximately 1 rad = .16 cps apart, this value of (Δf) is adequate to obtain good separation in the computed spectrum.

From the fact that spectral estimates are distributed as chi-square, Blackman and Tukey give confidence limits for various measurement techniques. For the method used here, the width of the 90% confidence band of the computed spectral density, in db, is given by

(A5.12) $$20\sqrt{\frac{m}{2n - m}}$$

where m = no. of lag values
n = no. of data points used

If m = 260 and n = 5000, we obtain

(A5.13) 90% conf. band $\cong$ 3.3db

i.e., there is only a 10% chance that the true value of spectral density would be outside of the band $(S_{ff_T} - 1.65\text{db}) < S_{ff_T} < S_{ff_T} + 1.65\text{db}$. Of course, these confidence limits apply to random

processes. For deterministic inputs, such as utilized in our study, where the output can be assumed to consist of a deterministic plus a random component, these limits properly apply only to the measurement of the random component.

A5.5 Results

The digitally computed power spectral densities were reduced to IBM cards and plotted on a Benson-Lehner plotter. The resulting plots of error and output spectral density for the 7 runs discussed in Chapter 6 are given in Figures A5.2 to A5.8. Two major conclusions can be drawn from these curves:

(1) The amount of energy present between the spectral peaks is 10 to 20 db less than the peak power. Consequently, the operator does not introduce large amounts of power at frequencies where there is no input power.

(2) The error spectral density is significantly lower, again by 10-20 db, beyond the range of the input function. Therefore, any evidence of periodicity due to sampling is extremely difficult to detect. This problem is discussed in Chapter 6 where computed values are compared with measured ones. However, it is apparent that there are dominant peaks in the "noise" spectrum and they are at approximate harmonics of the input frequency.

Cross-spectral densities computed from the program were used to synthesize linear operator models, as outlined in Appendix 7.

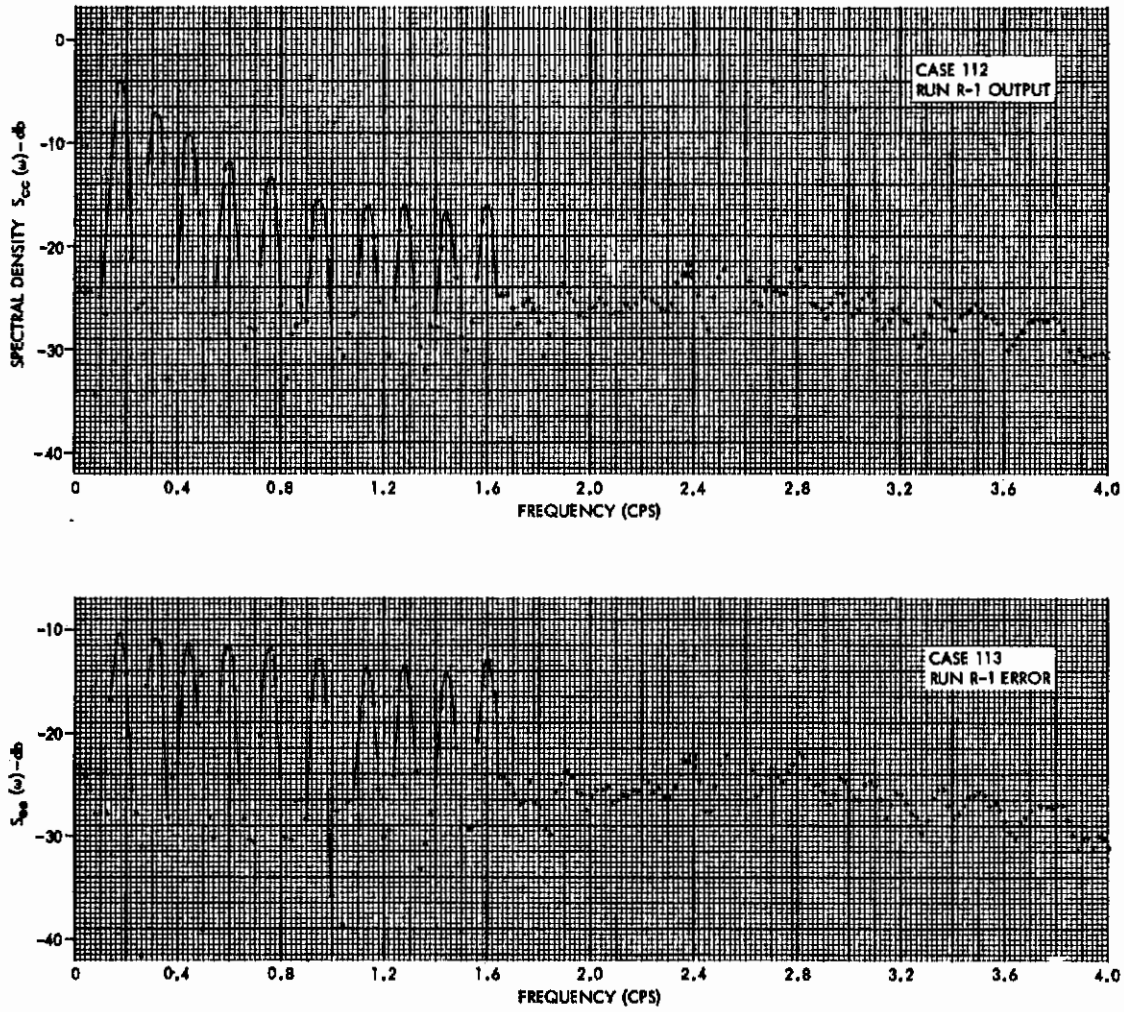


Figure A5.2. Output Spectral Density from Run R-1

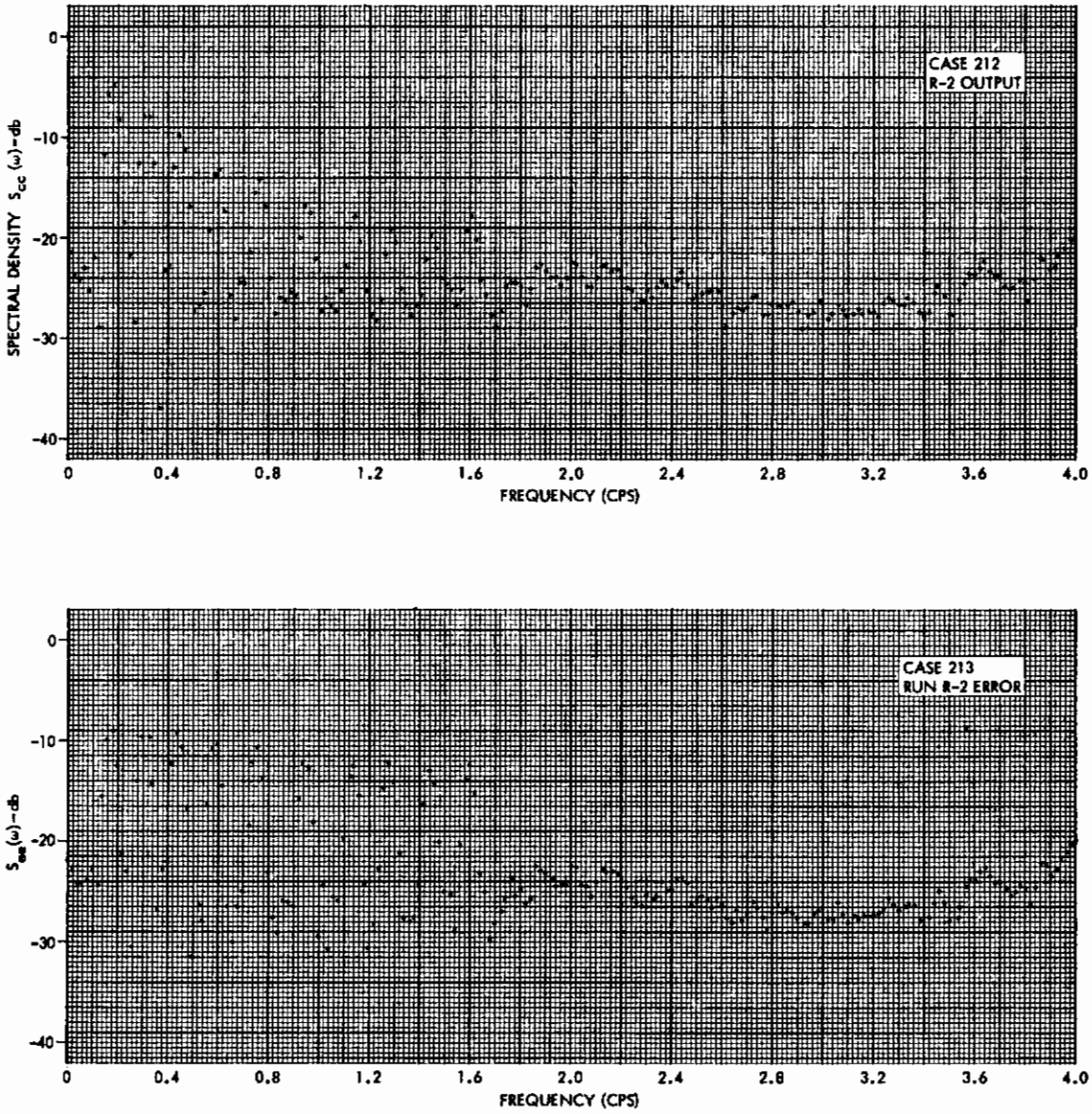


Figure A5.3. Output Spectral Density from Run R-2

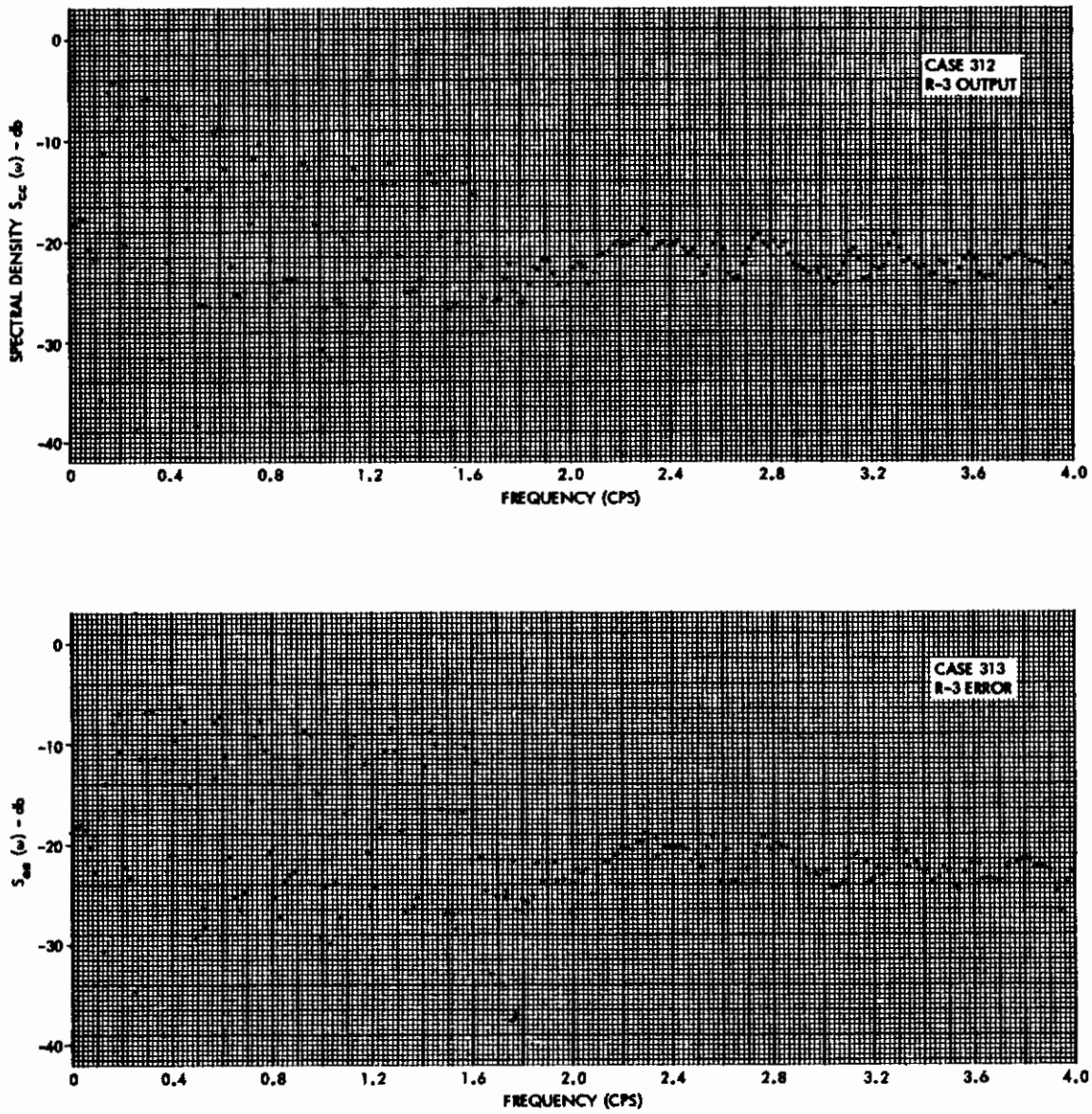


Figure A5.4. Output Spectral Density from Run R-3

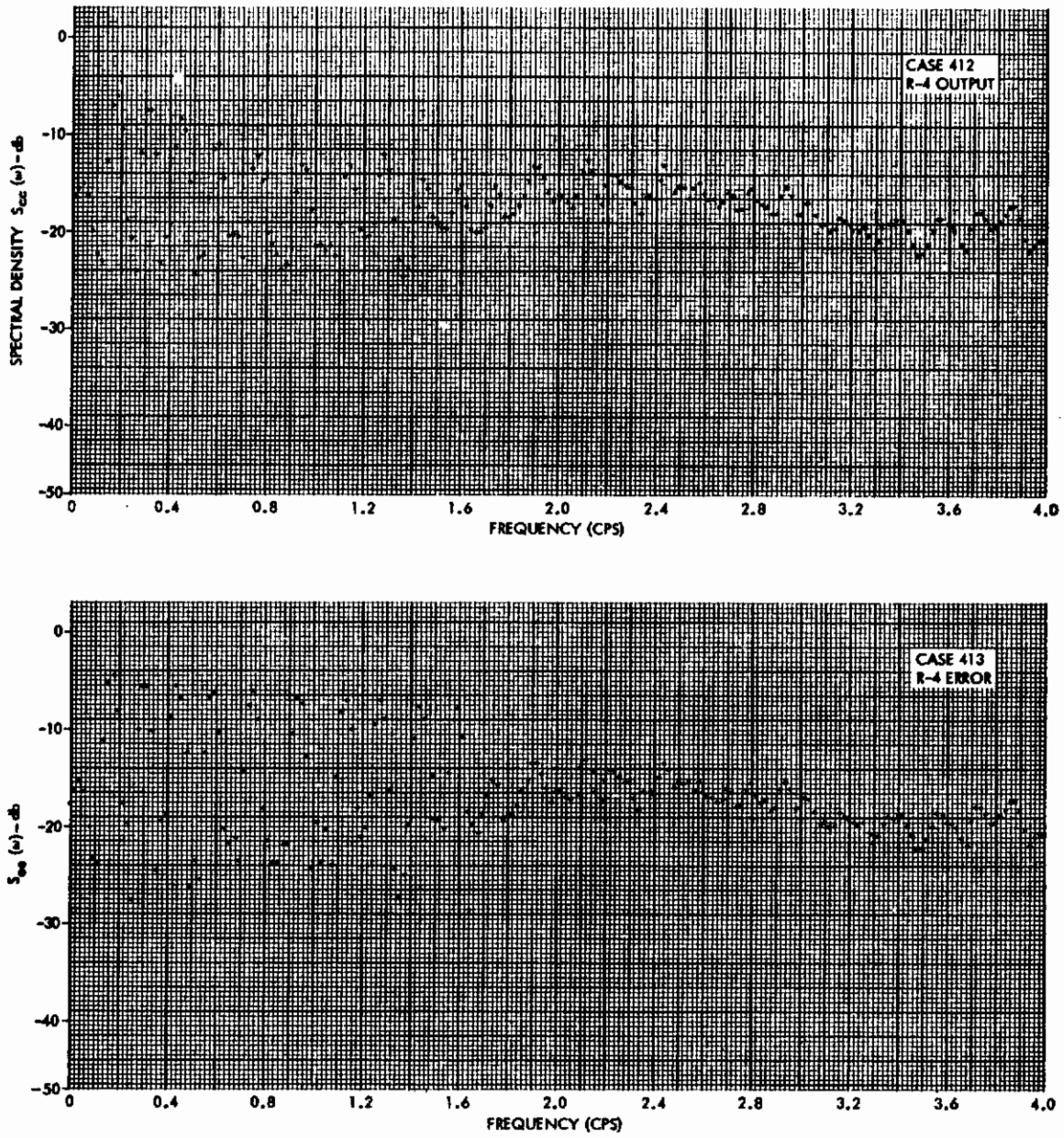


Figure A5.5. Output Spectral Density from Run R-4

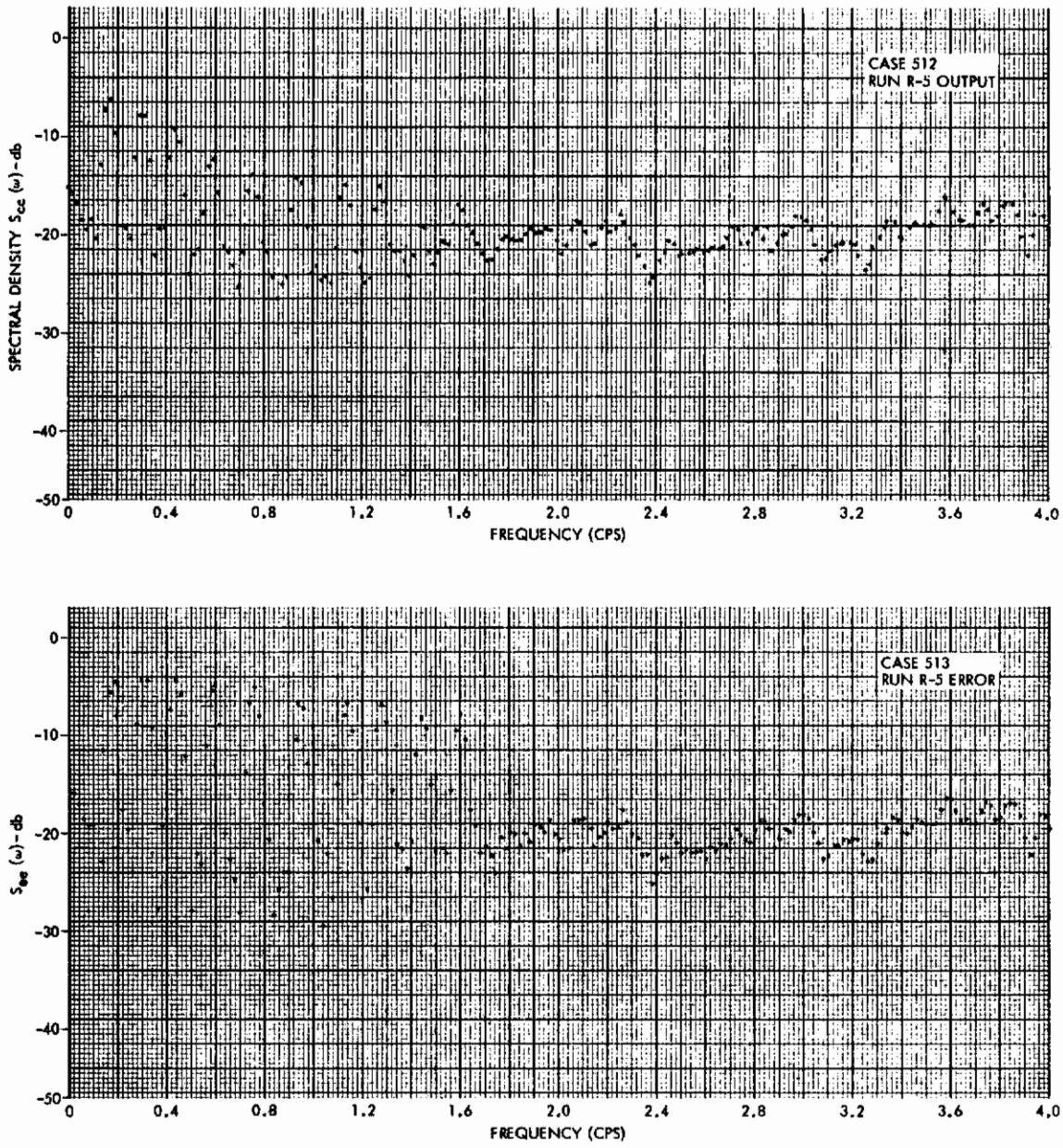


Figure A5.6. Output Spectral Density from Run R-5

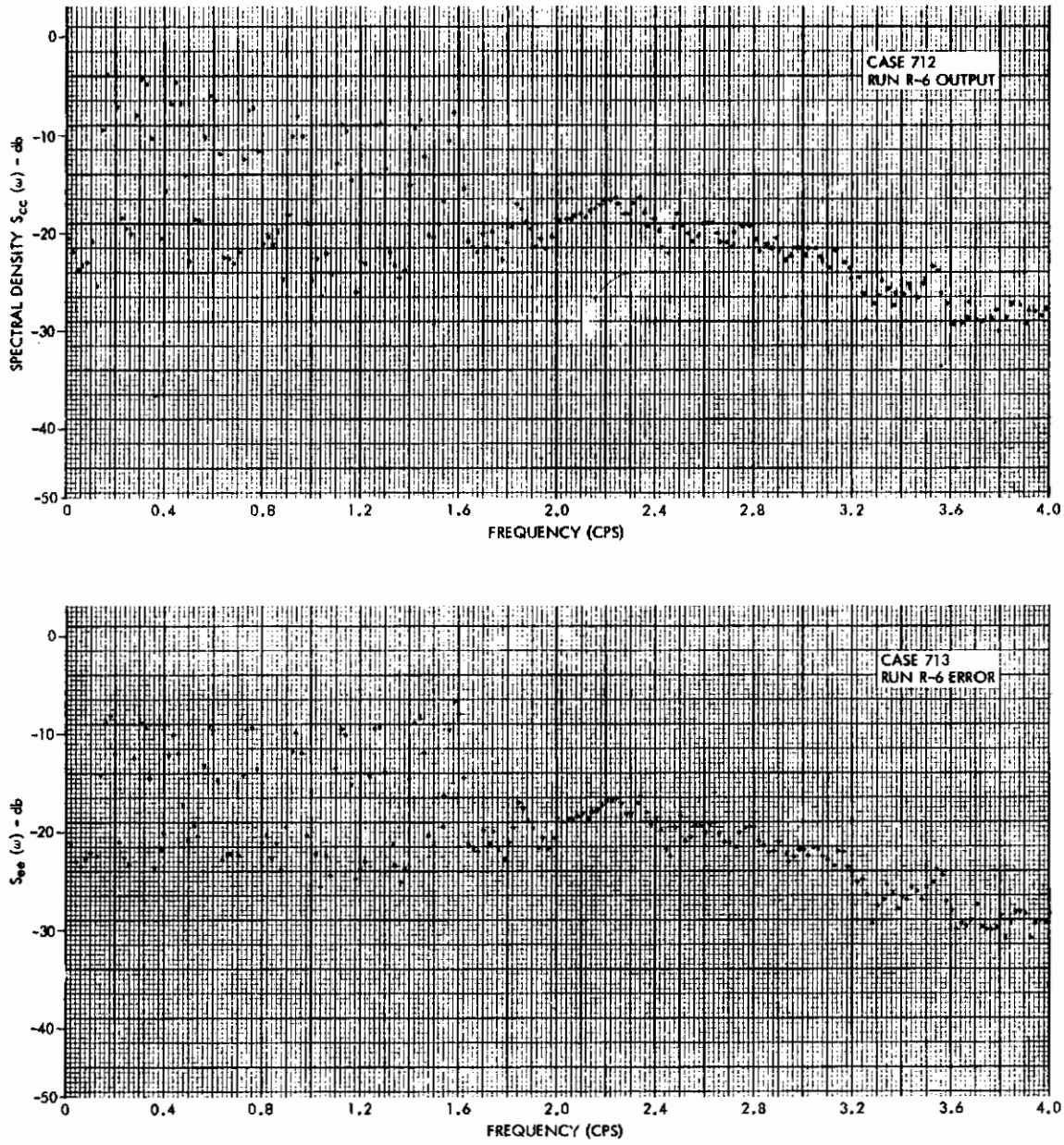


Figure A5.7. Output Spectral Density from Run R-6

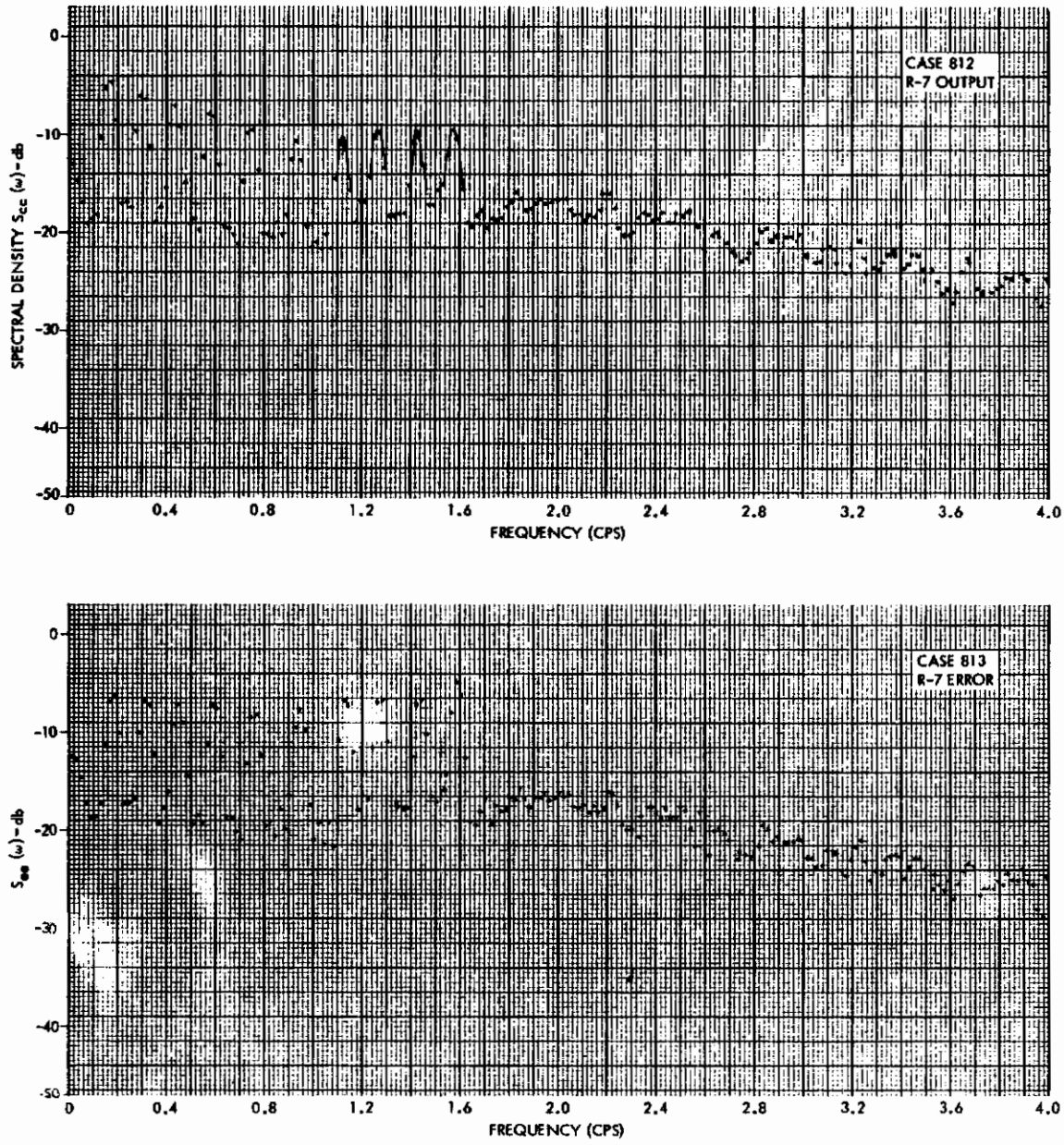


Figure A5.8. Output Spectral Density from Run R-7

References

1. Negron, C. D. (1961), "Final Problem Statement: Power Spectral Density and Cross-Correlation Analysis", Internal Memorandum Number 9840.3-15, Space Technology Laboratories, Inc., 20 June 1961.

2. Blackman, R. B. and Tukey, J. W. (1958), The Measurement of Power Spectra, New York: Dover Publications 1958.

Appendix 6
EXPERIMENTAL RESULTS

The results of the experimental portion of the study are presented in the following pages in the form of graphs of error spectral density for the various cases considered and for all subjects. The graphs are organized in five groups:

(1) Results for Case 3 (ω_B = 1.5, continuous display)

(2) Results for Case 4 (ω_B = 3.0, continuous display)

(3) Results for Case 5 (ω_B = 1.5, sampled display)

(4) Results for Case 6 (ω_B = 3.0, sampled display)

In each case all the runs are presented for each subject, followed by curves which represent the averaged spectrum for each subject. Two points should be noted concerning interpretation:

(1) The data are given in the form of experimental points connected by dotted lines. The lines are used to connect a particular set of points from one run for easier visualization, but they do <u>not</u> imply that there is energy present between the experimental points.

(2) Normalization of vertical scale. The power spectral density scale is referred to the measured values of the unfiltered input function. Thus, on the same scale, the spectral density $S_{ff_T}(\omega_i)$ will yield 10 points at the frequencies ω_i, all lying on the 0 db line.

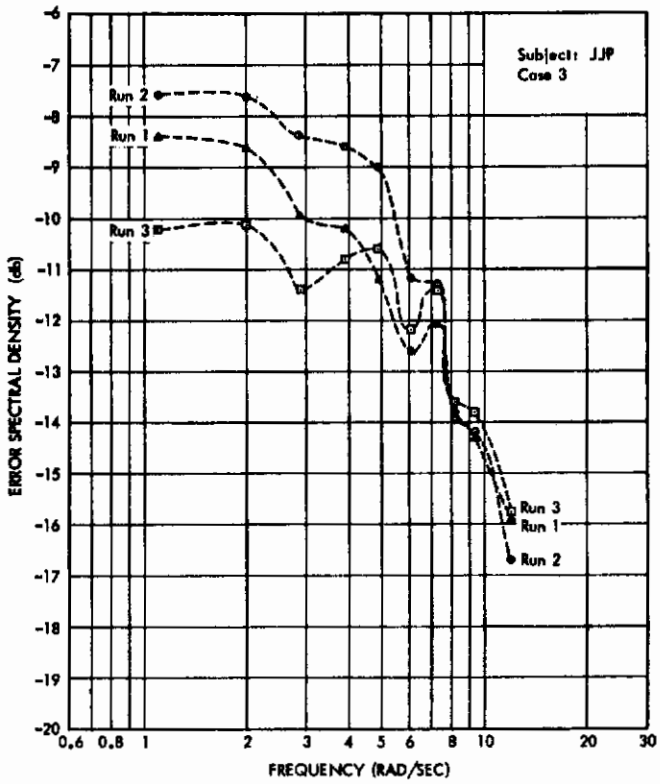


Figure A6.1

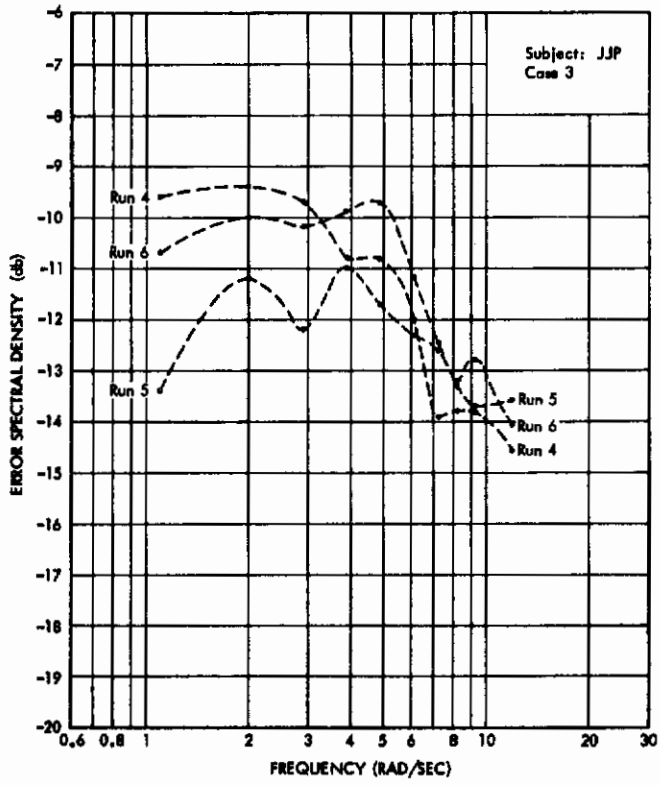


Figure A6.2

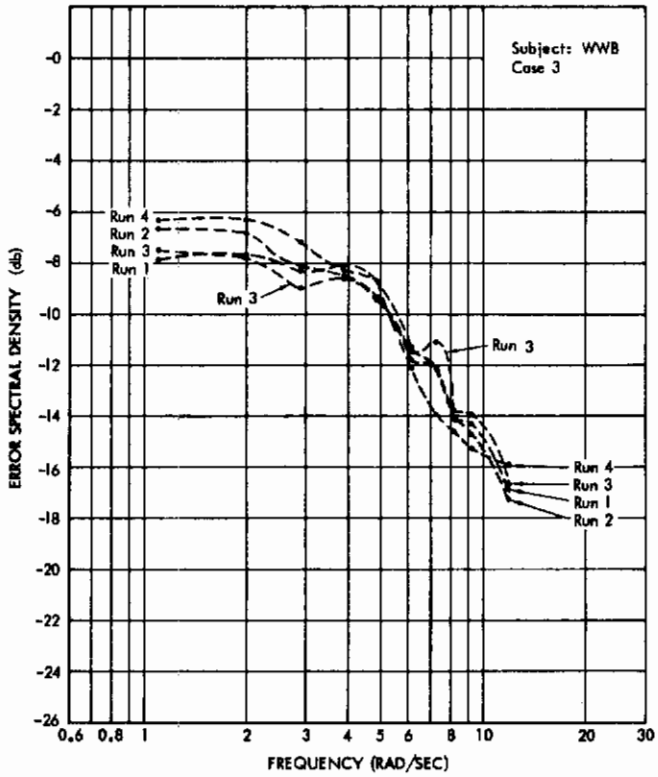


Figure A6.3

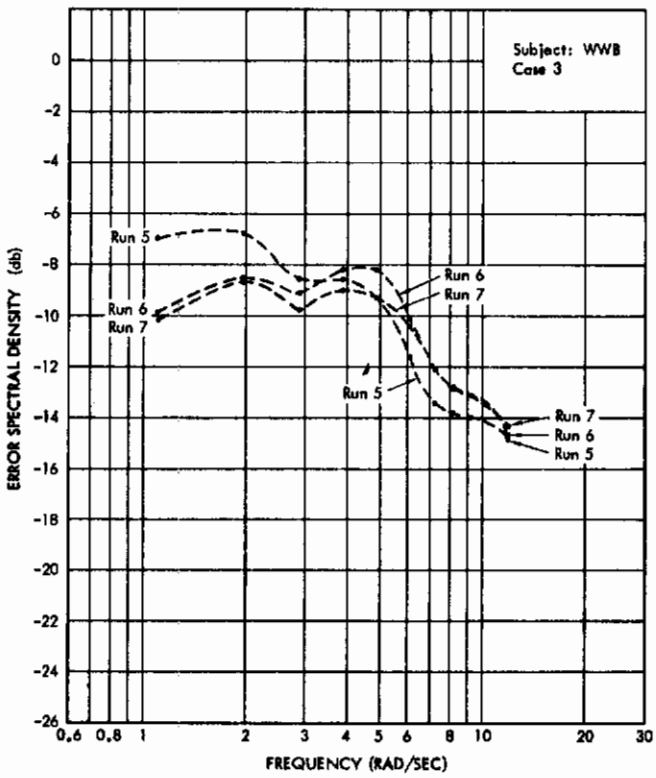


Figure A6.4

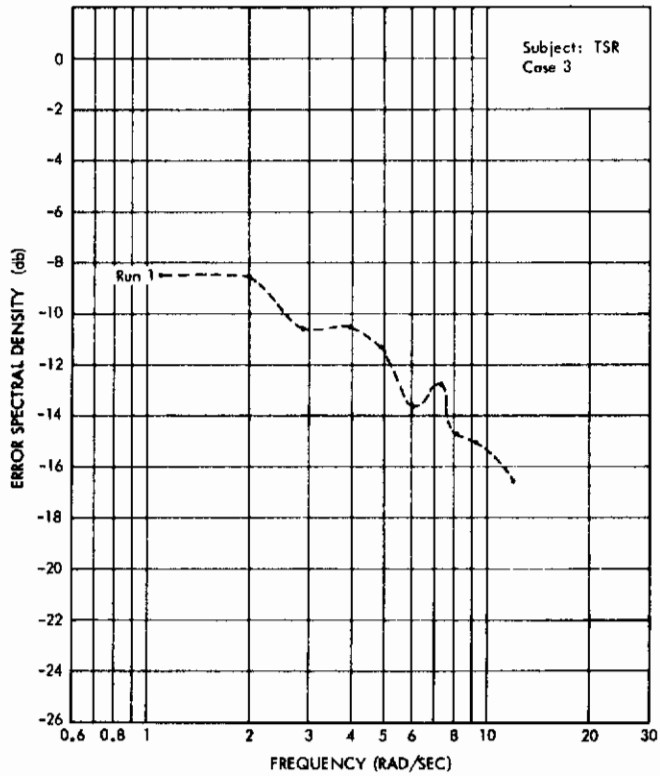


Figure A6.5

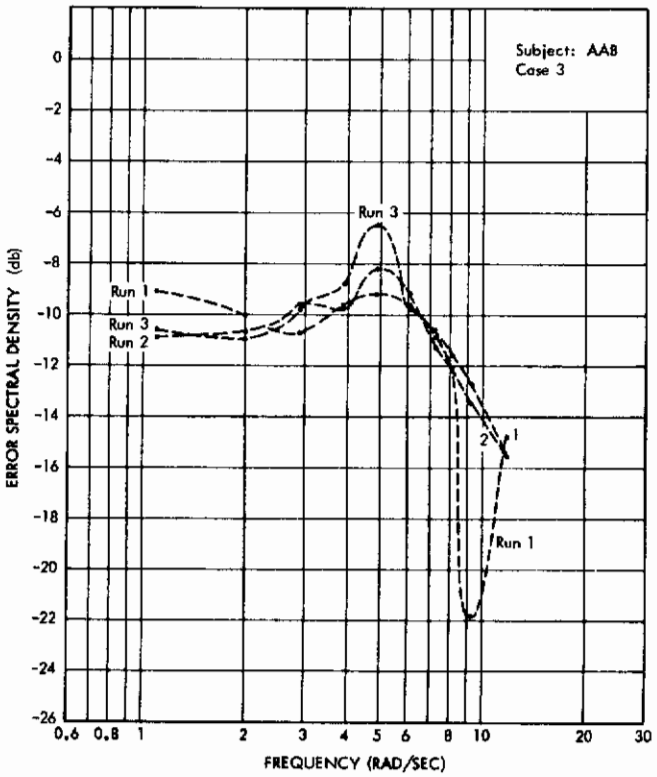


Figure A6.6

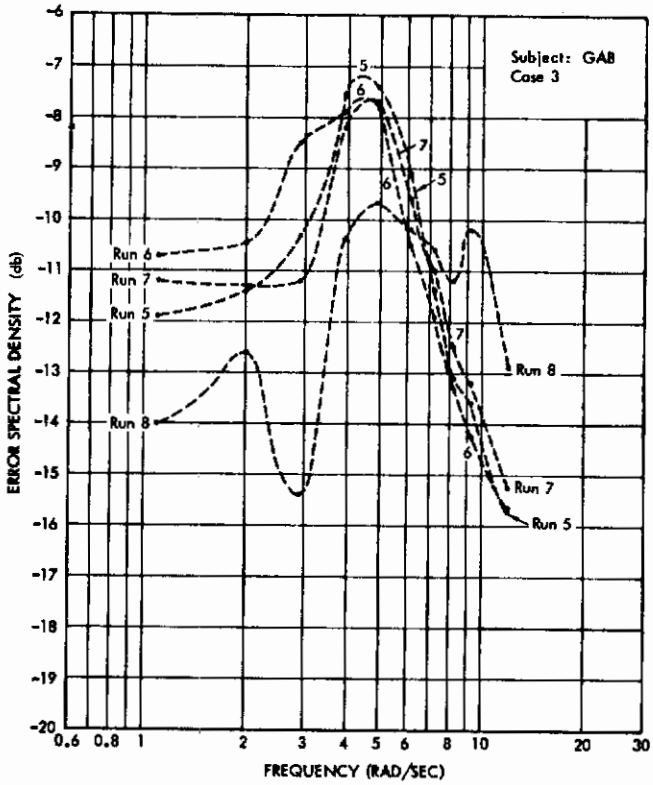


Figure A6.7

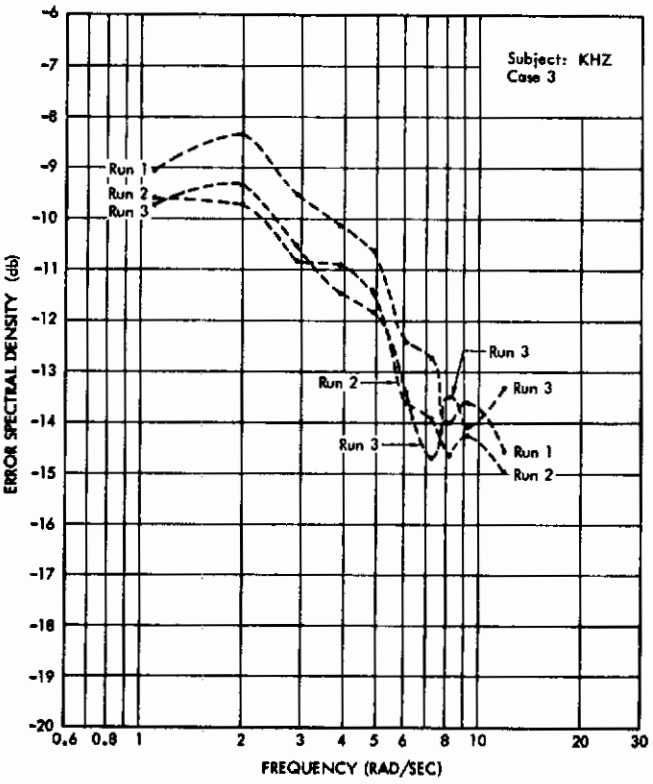


Figure A6.8

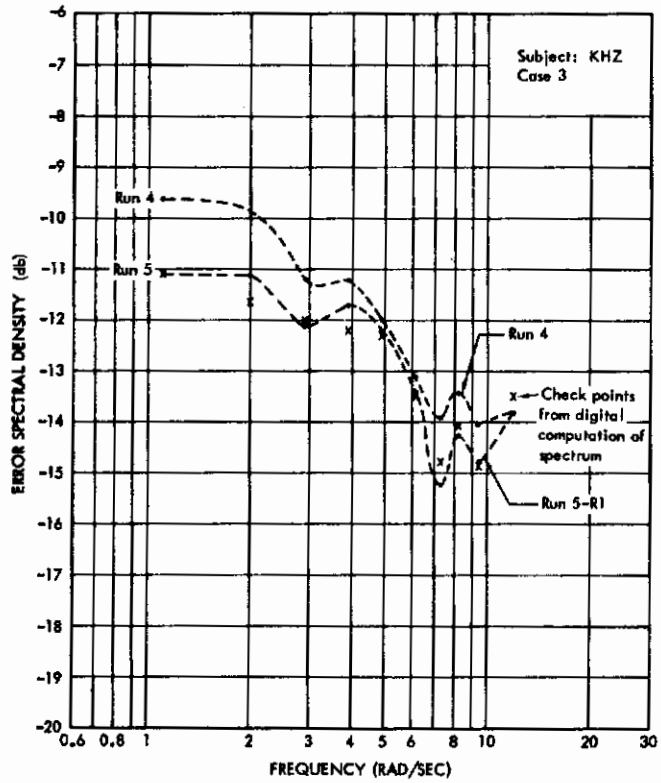


Figure A6.9

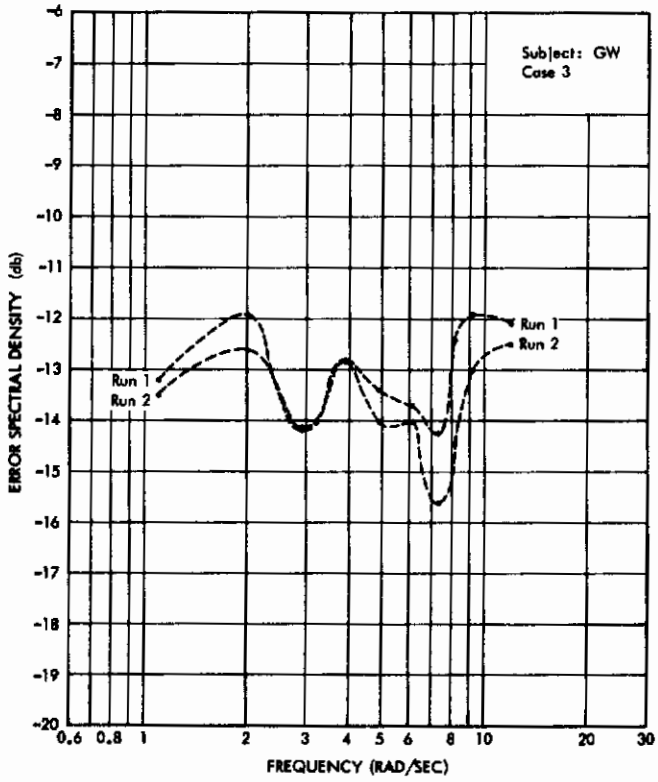


Figure A6.10

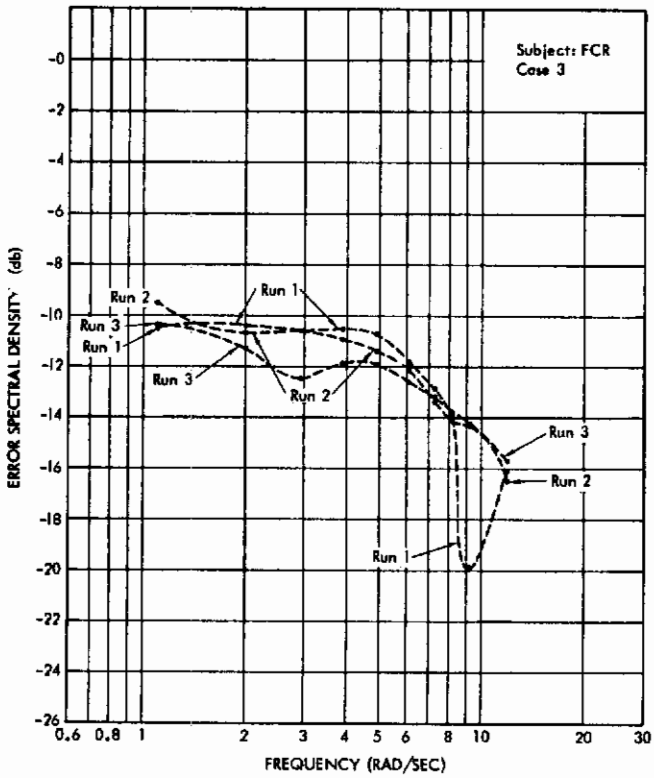


Figure A6.11

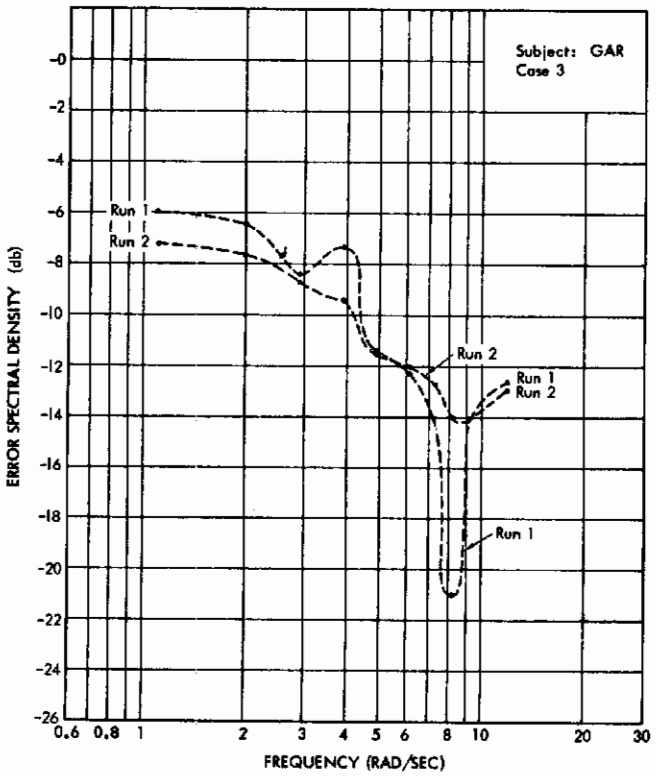


Figure A6.12

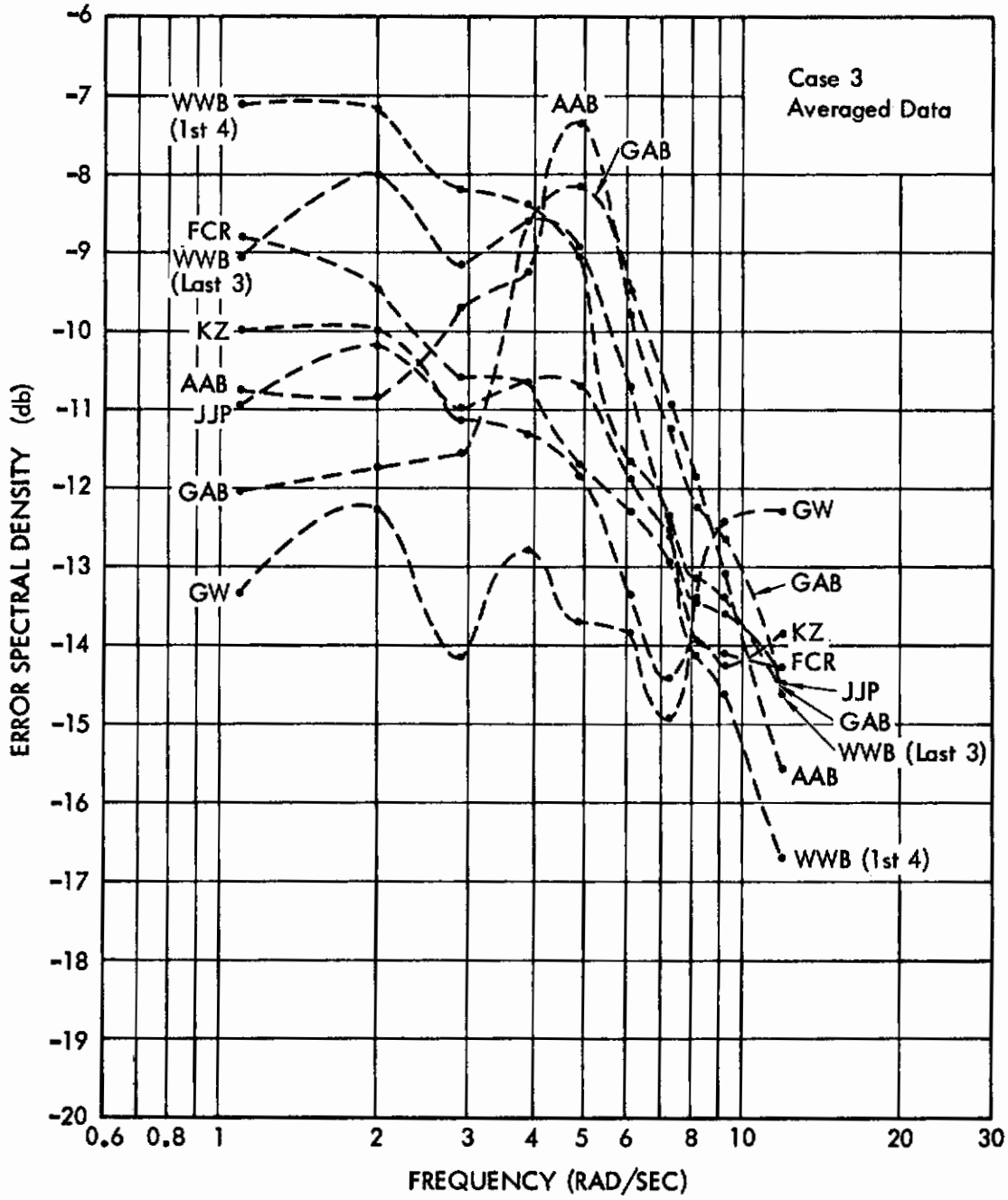


Figure A6.13

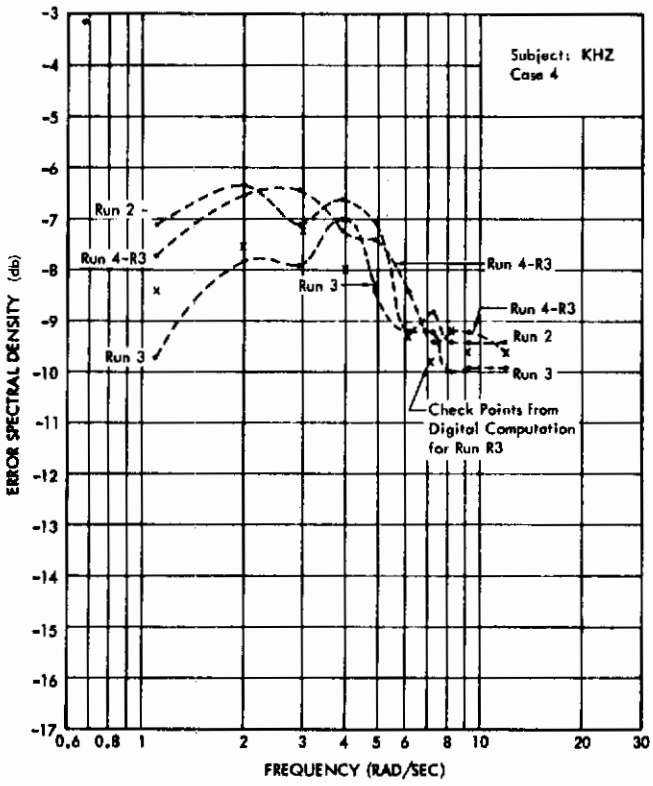


Figure A6.14

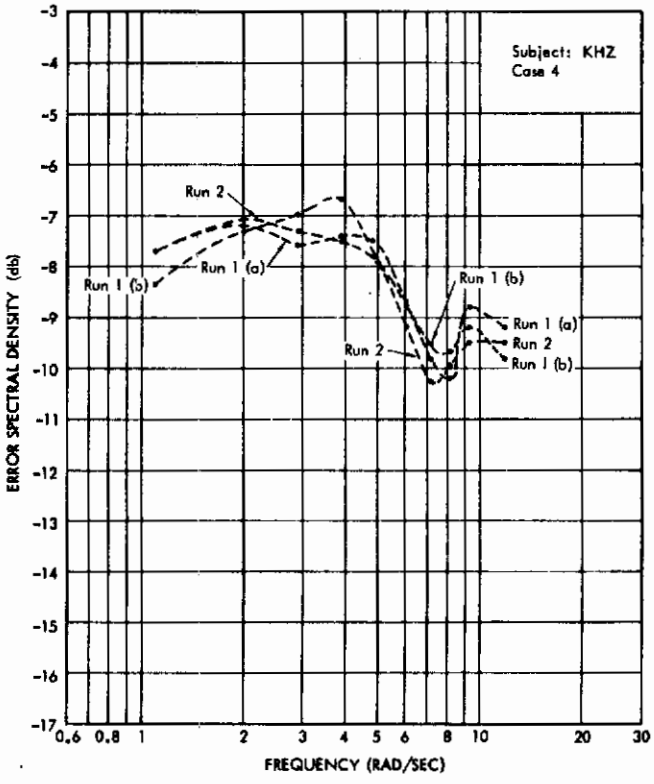


Figure A6.15

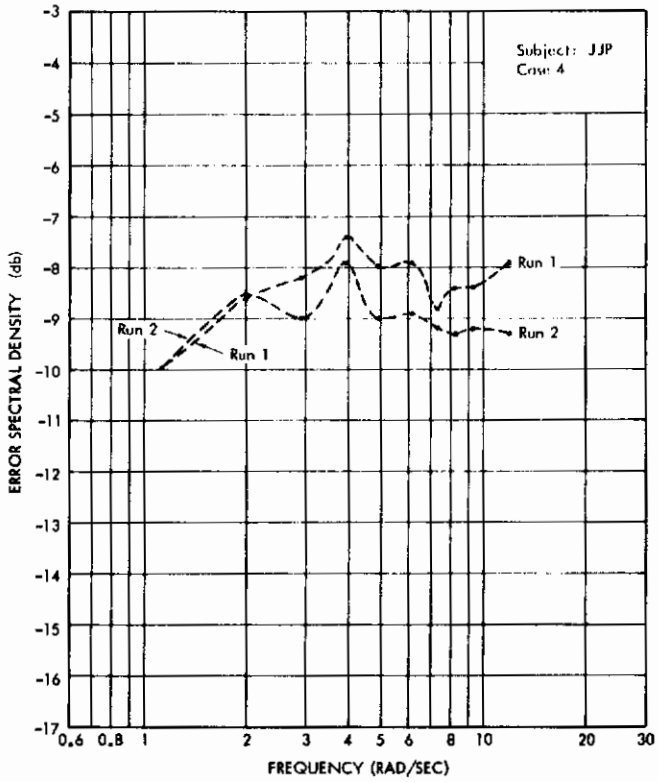


Figure A6.16

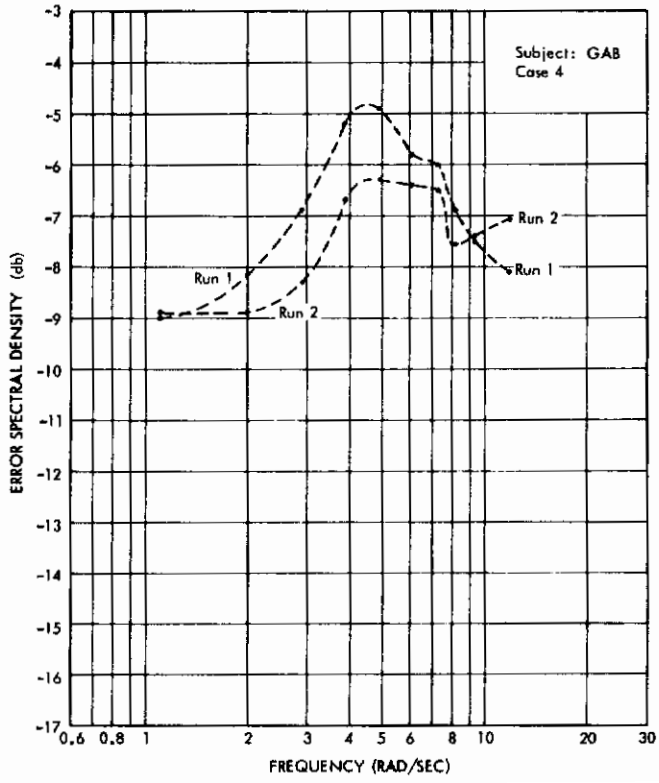


Figure A6.17

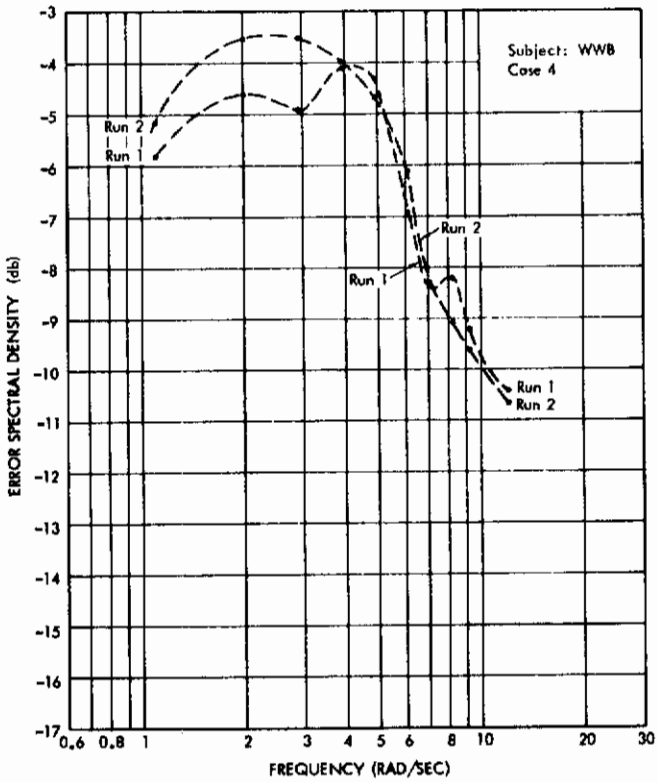


Figure A6.18

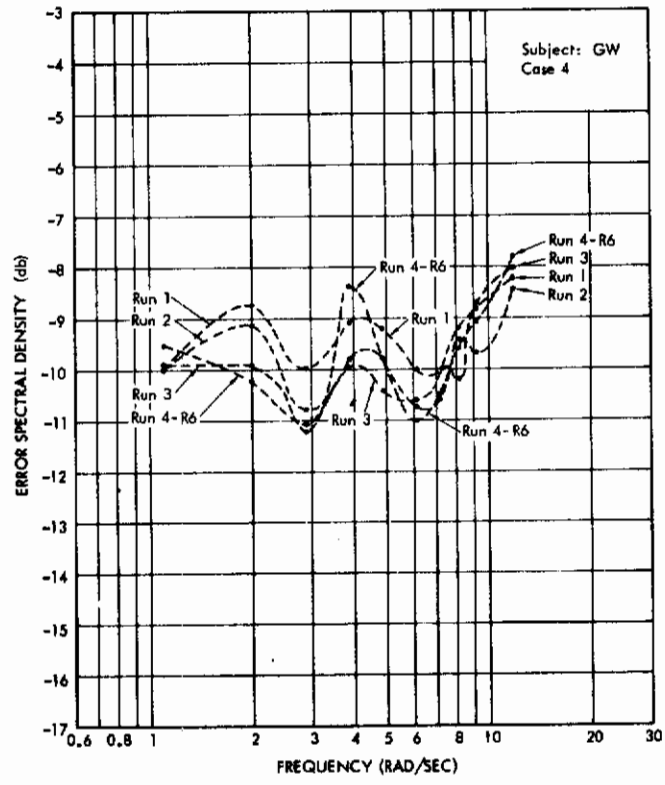


Figure A6.19

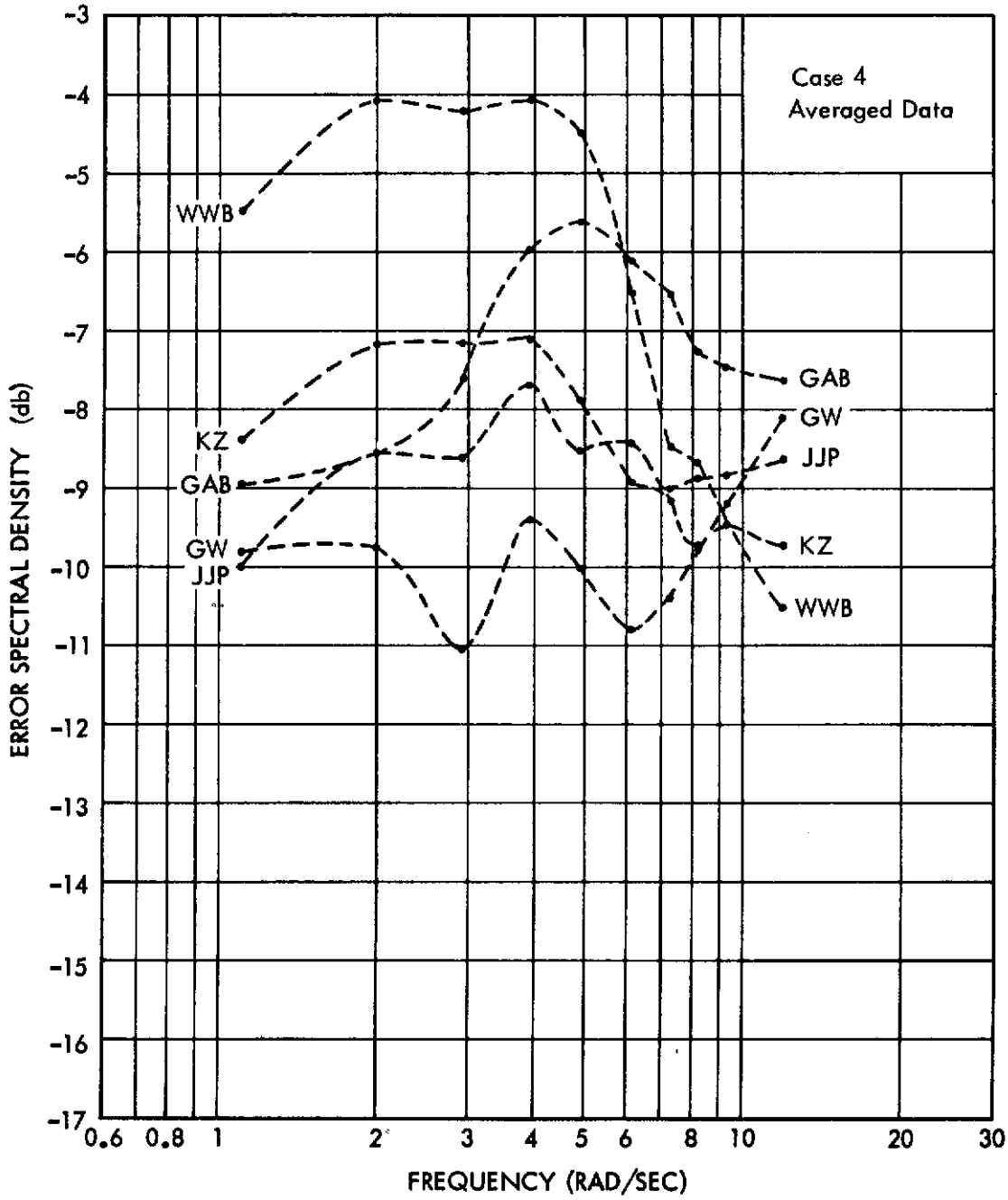


Figure A6.20

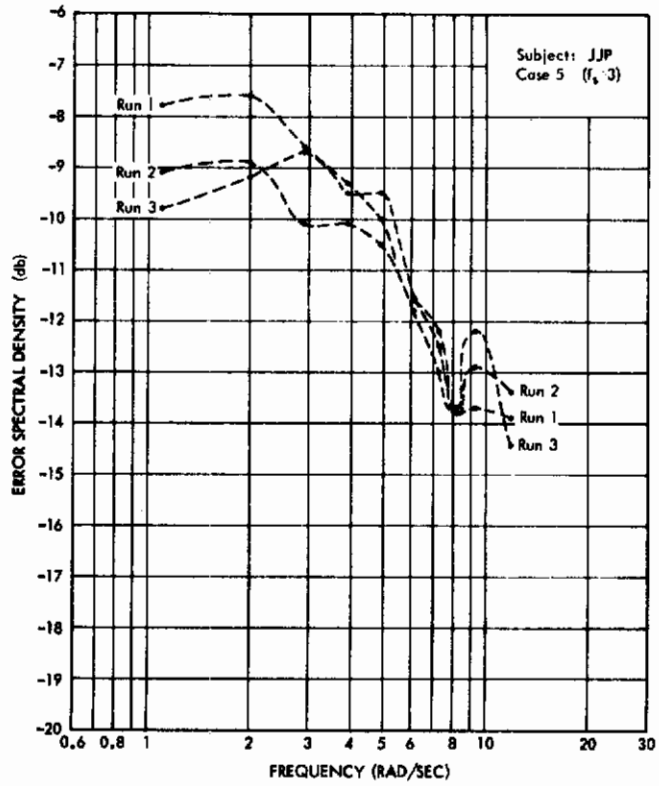


Figure A6.21

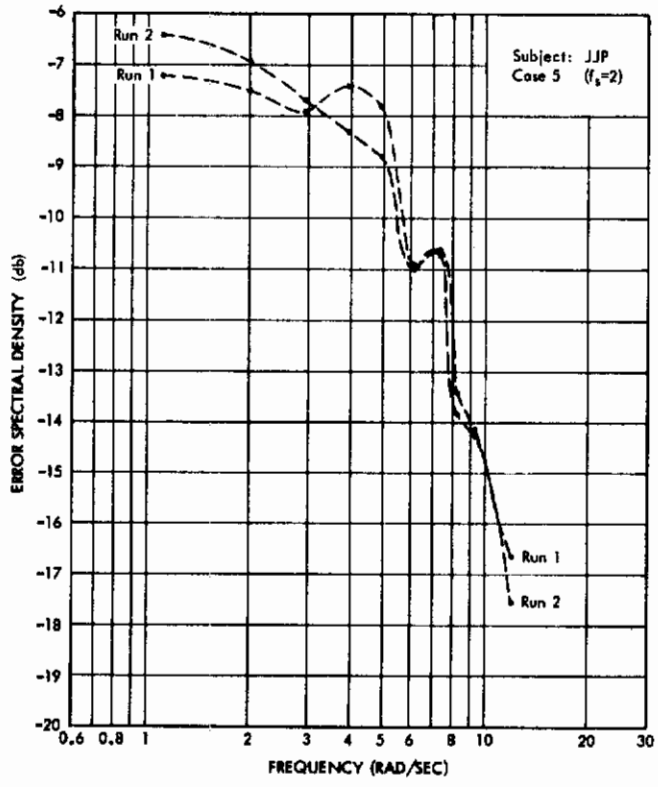


Figure A6.22

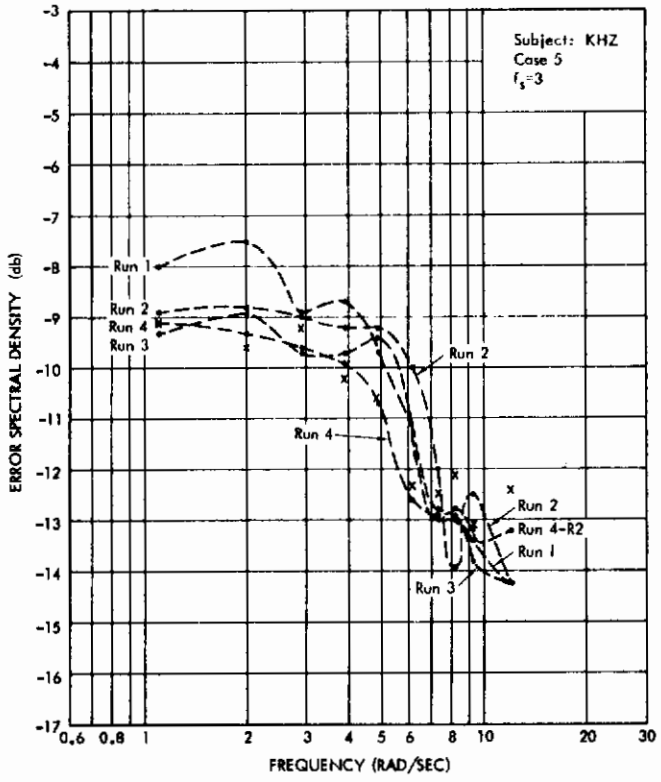


Figure A6.23

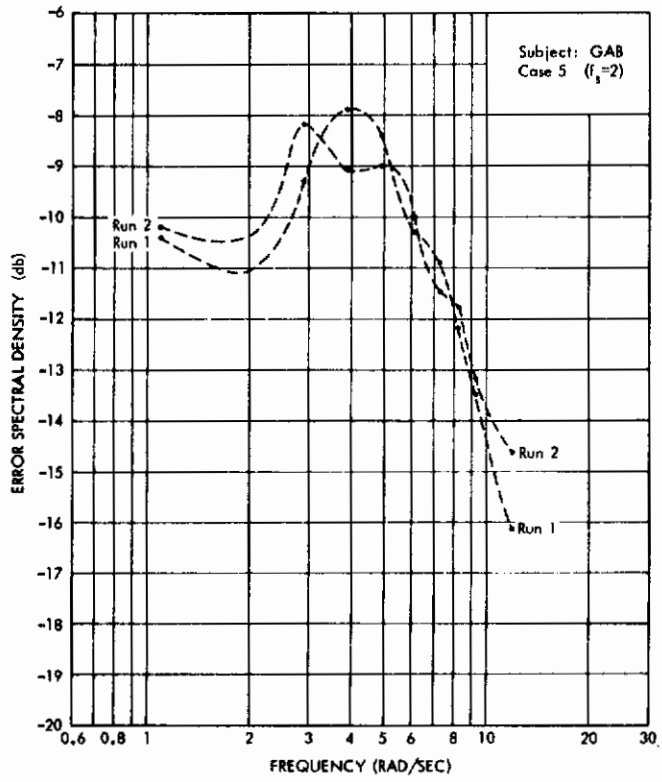


Figure A6.24

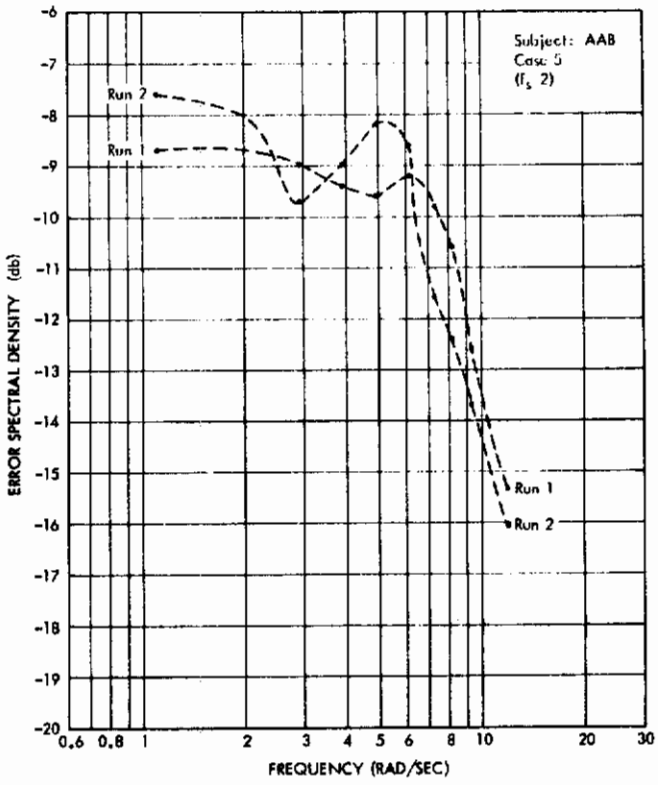


Figure A6.25

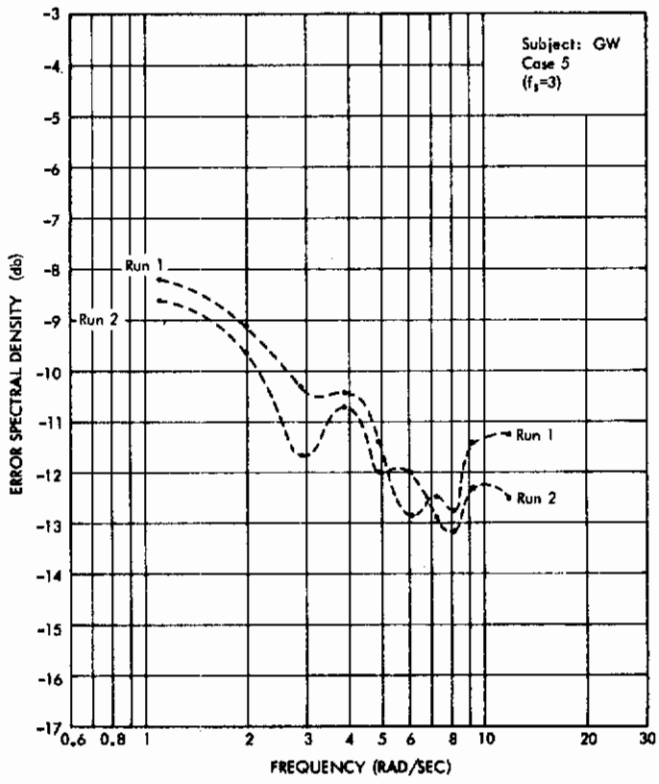


Figure A6.26

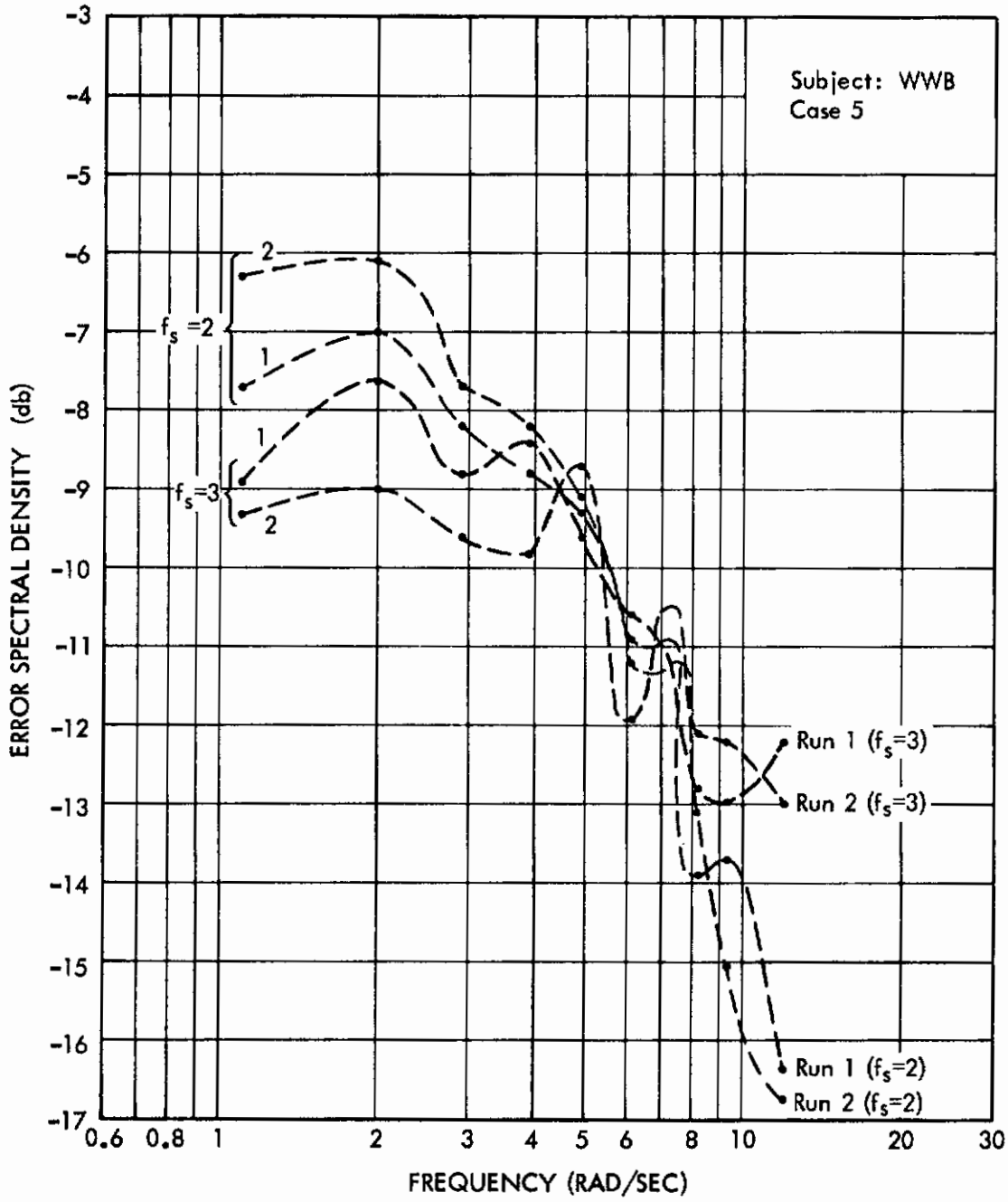


Figure A6.27

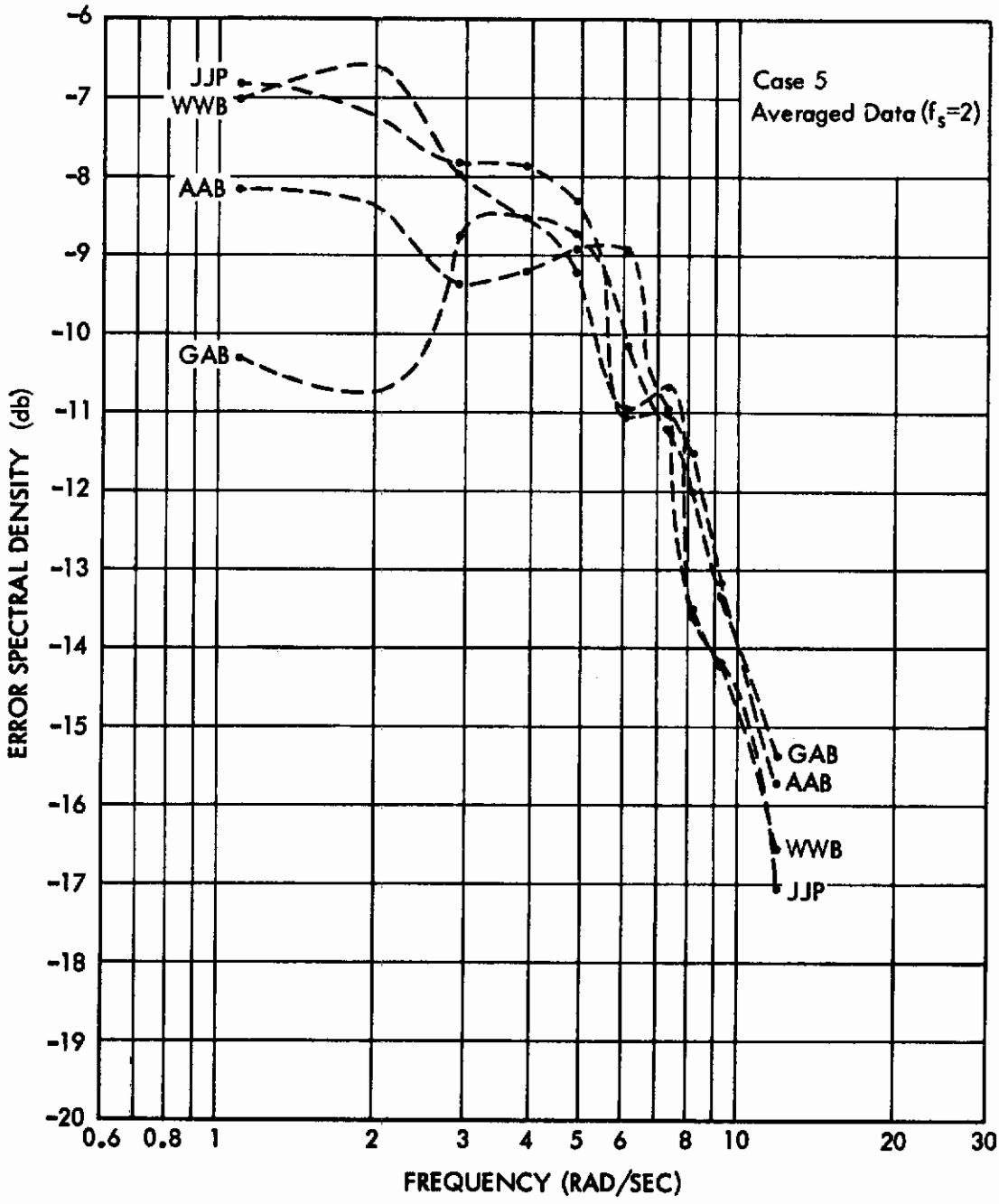


Figure A6.28

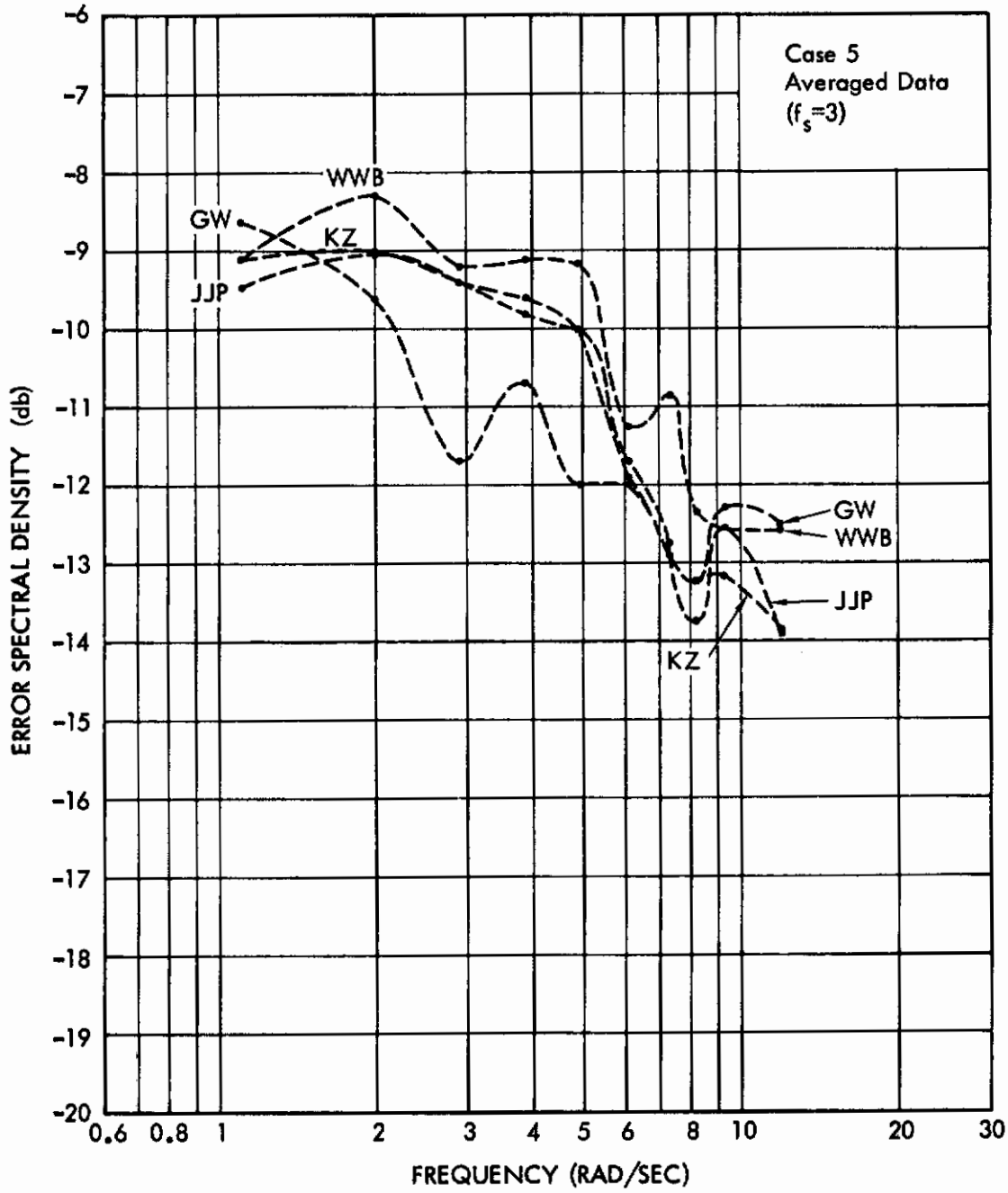


Figure A6.29

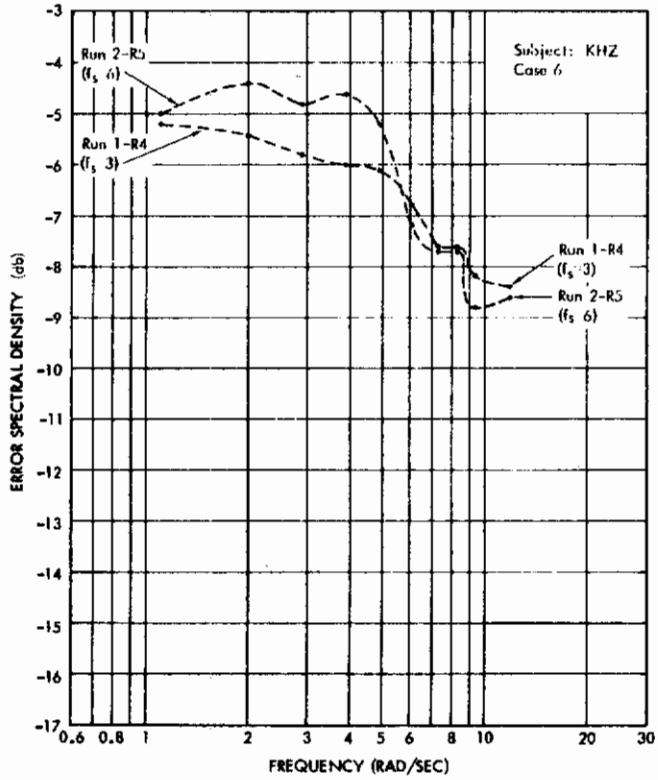


Figure A6.30

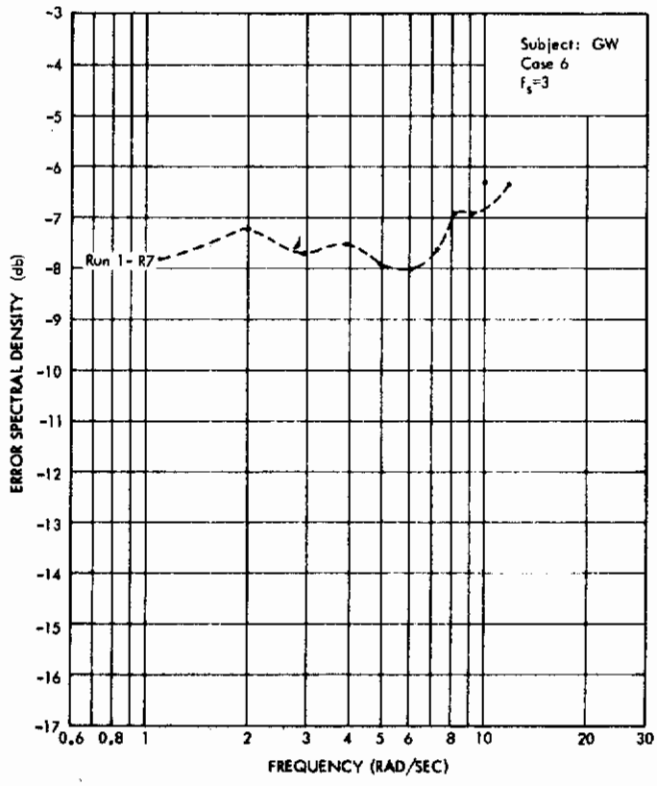


Figure A6.31

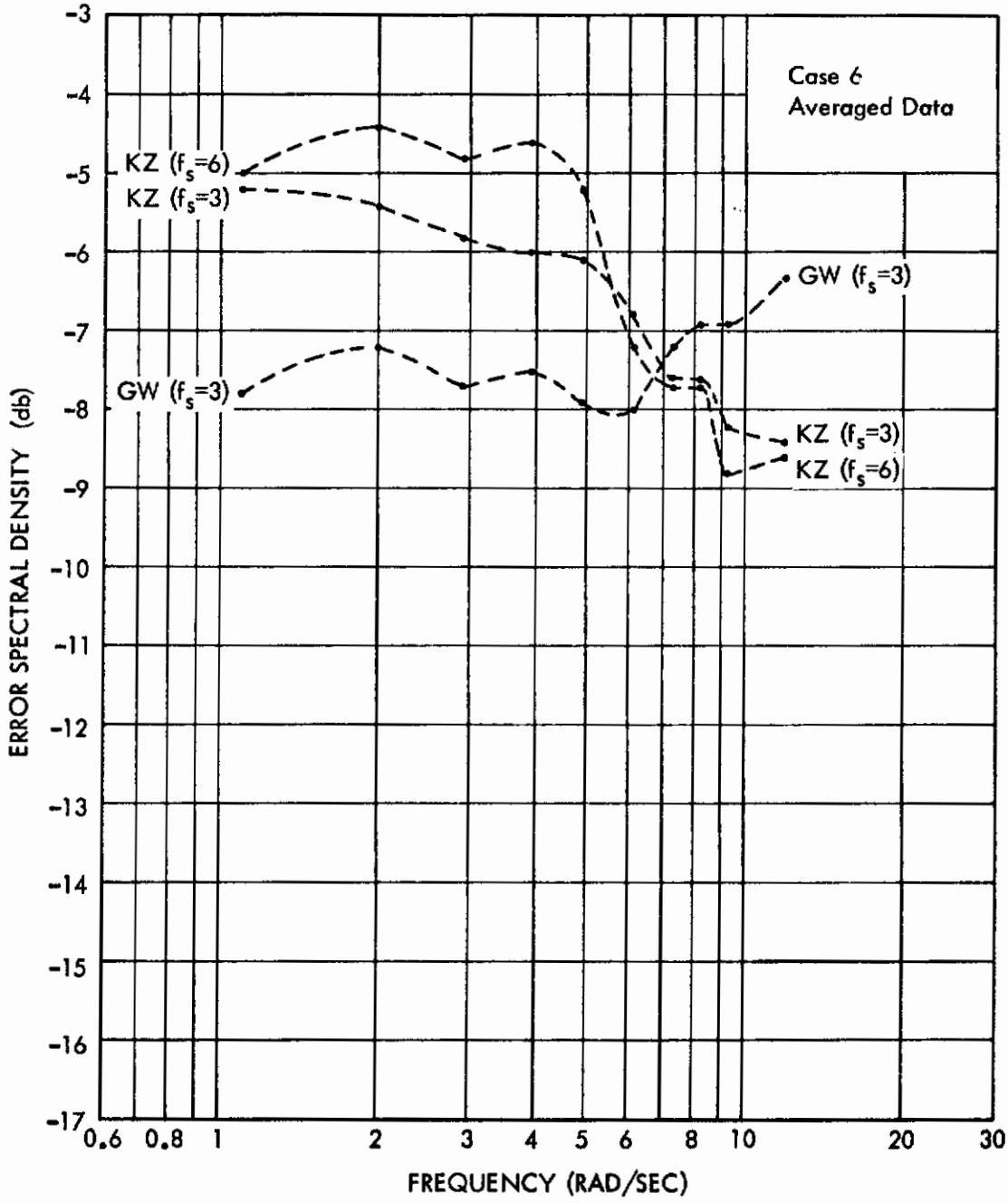


Figure A6.32

Appendix 7
COMPLETE ANALOG COMPUTER DIAGRAM

The complete schematic diagram of the analog computer setup utilized during the experimental portion of the study is shown in Figure A7.1. The diagram consists of the following basic portions:

(1) 10 oscillator circuits used for construction of the input function $f(t)$ and for the spectral density measurement circuits. Each circuit is provided with a damping adjustment to maintain constant amplitude for the 5 min. runs. At some frequencies, the damping required was positive, at others, negative, depending on the amplifier characteristics.

(2) 10 spectral density measurement circuits which are fully described in Appendix 3.

(3) The tracking system circuits, which include the summation and filtering of the sine waves to produce $r(t)$, the error measurement, scope input, and following "stick" output.

(4) The sampled-data model of the human operator, using the "modified" partial velocity hold circuit. The simulation of this circuit is fully described in Appendix 4.

(5) A clock consisting of an integrator-relay system which started the spectrum averaging circuits 120 seconds after the start of a run, and terminated the run 180 seconds later.

(6) An intermittent display circuit which made use of a sampling pulse from the Sample-Hold Computer to actuate a power relay which added a 1M resistor to the "intensity" circuit of the tracking oscilloscope, thus blanking the spot completely.

The equipment consisted of three basic computer consoles: a 60-amplifier Electronic Associates Model 16-131R, a 48-amplifier modernized 16-24D, and a Sample-Hold computer with 24 Beckman/Berkeley amplifiers fabricated at Space Technology Laboratories. In addition, 15 extra dual-channel electronic multipliers were "borrowed" from other consoles for use in the spectral density measurement circuits.

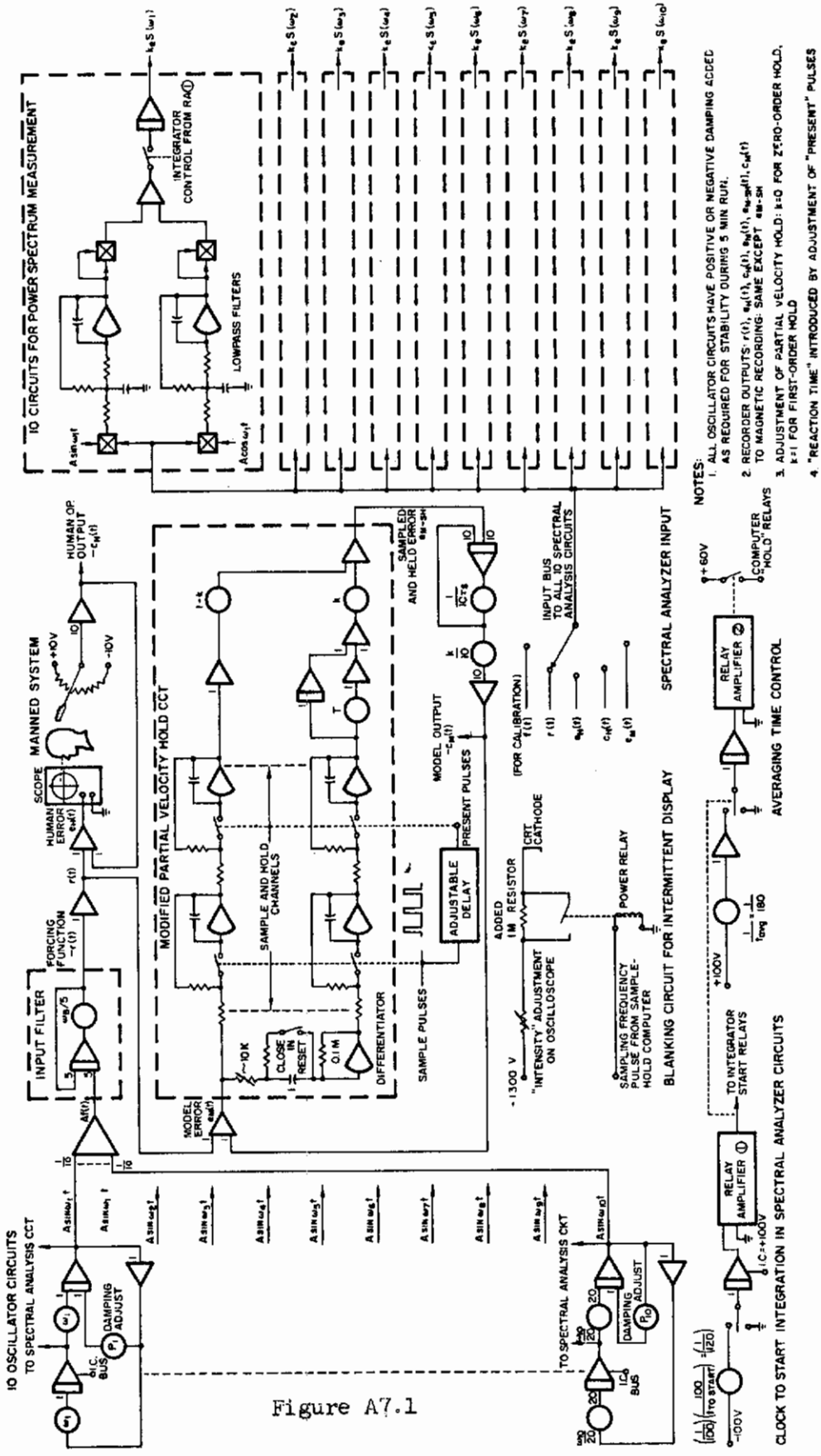


Figure A7.1

Appendix 8

"BEST FIT" DESCRIBING FUNCTIONS

The selection of analytical forms for continuous operator models, $G_c(j\omega)$, is illustrated by the curves in this Appendix. The method is described in Chapters 2 and 6. Basically, the function $G(j\omega)$ is obtained as the ratio of two cross-spectral densities obtained from the digital program described in Appendix 5, i.e.,

$$G(j\omega) = \frac{S_{rc}(j\omega)}{S_{re}(j\omega)} \tag{A8.1}$$

The magnitude of G is plotted on semi-log paper and the phase of G is plotted separately on linear paper to facilitate the fitting process. Since the time delay $e^{-Dj\omega}$ does not affect the amplitude characteristics, the fitting can be done separately for amplitude and phase.

Based on the work of McRuer and Krendel described previously, it is assumed that G can be fitted with the form

$$G_c(j\omega) \cong \frac{K\, e^{-jD\omega}}{1 + j\omega/a} = G_c'(j\omega)e^{-jD\omega} \tag{A8.2}$$

Using a template for the amplitude of G_c, a value of "a" is selected. The phase characteristic for the resulting lag term, $1/(1 + s/a)$ is subtracted from the phase plot of G. If the residual phase plot approximates a straight line on the linear phase-frequency plot, it represents a pure time delay, and the "fit" is considered satisfactory. If not, a new value of "a" is selected, until a satisfactory fit is obtained.

The value of D in (A8.2) is determined by using the phase relation of the pure time delay:

(A8.3) $$\phi_R = \arg e^{-j\omega D} = -\omega D$$

At $\omega = 2\pi$ we have

(A8.4) $$\phi_R = -2\pi D$$

and consequently the value of D in seconds is obtained from:

(A8.5) $$D = \frac{(\phi_R)_{f=1}}{2\pi(57.3)}$$

Figures A8.1 through A8.3 which follow show the experimental points plotted separately for amplitude and phase, the fitted amplitude characteristic of $G_c'(j\omega)$ and the residual phase ϕ_R for the recorded runs R-1 through R-3.

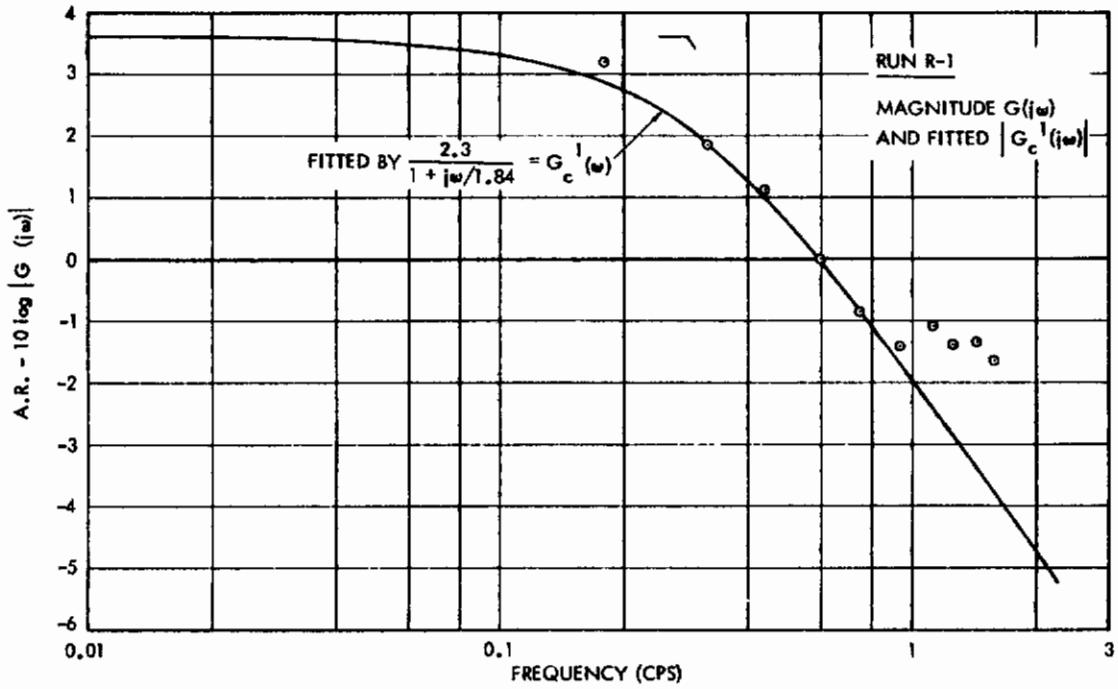


Figure A8.1 (a) Fitted Magnitude for Run R-1

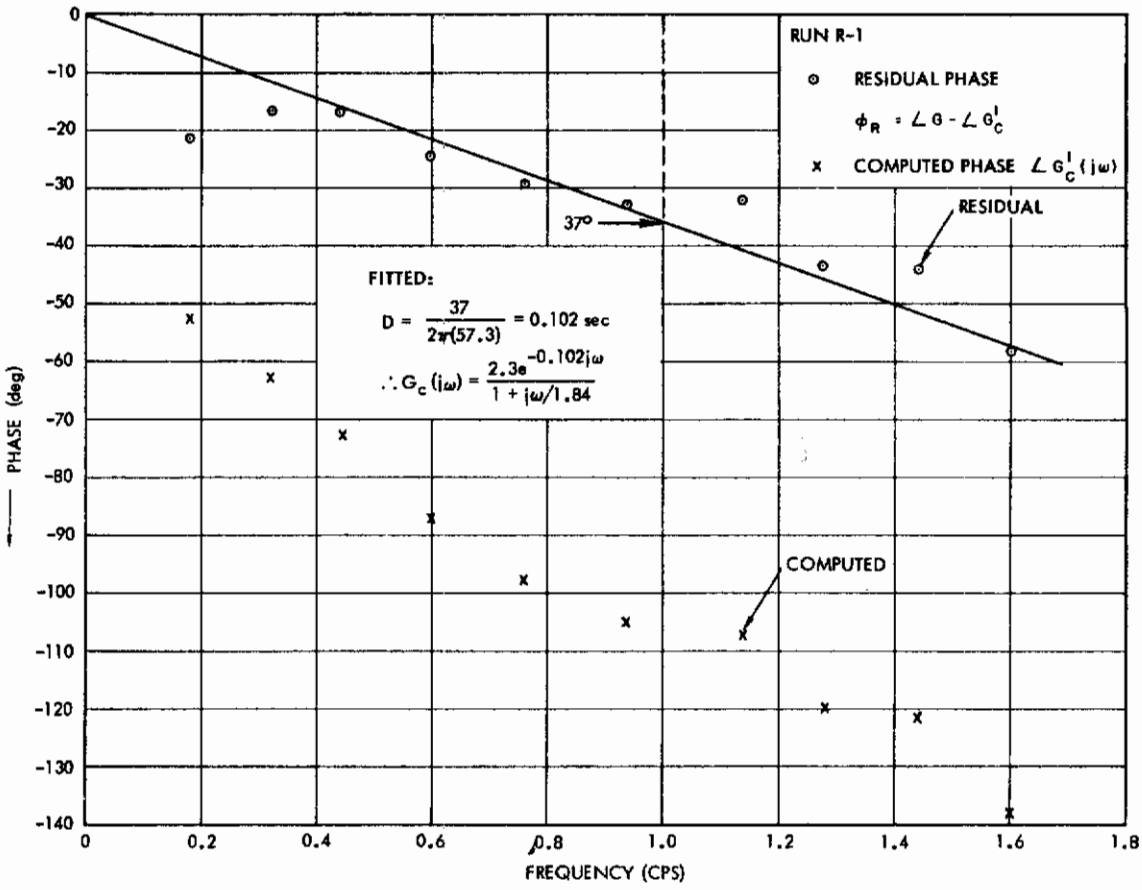


Figure A8.1 (b) Fitted Phase for Run R-1

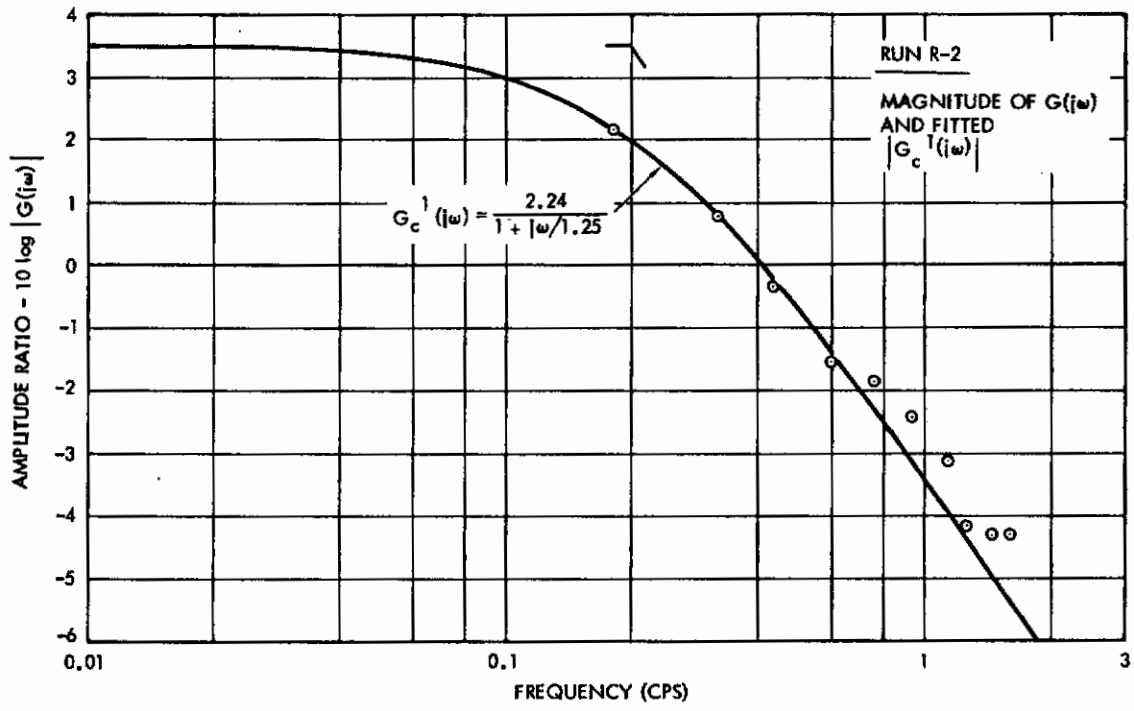


Figure A8.2 (a) Fitted Magnitude for Run R-2

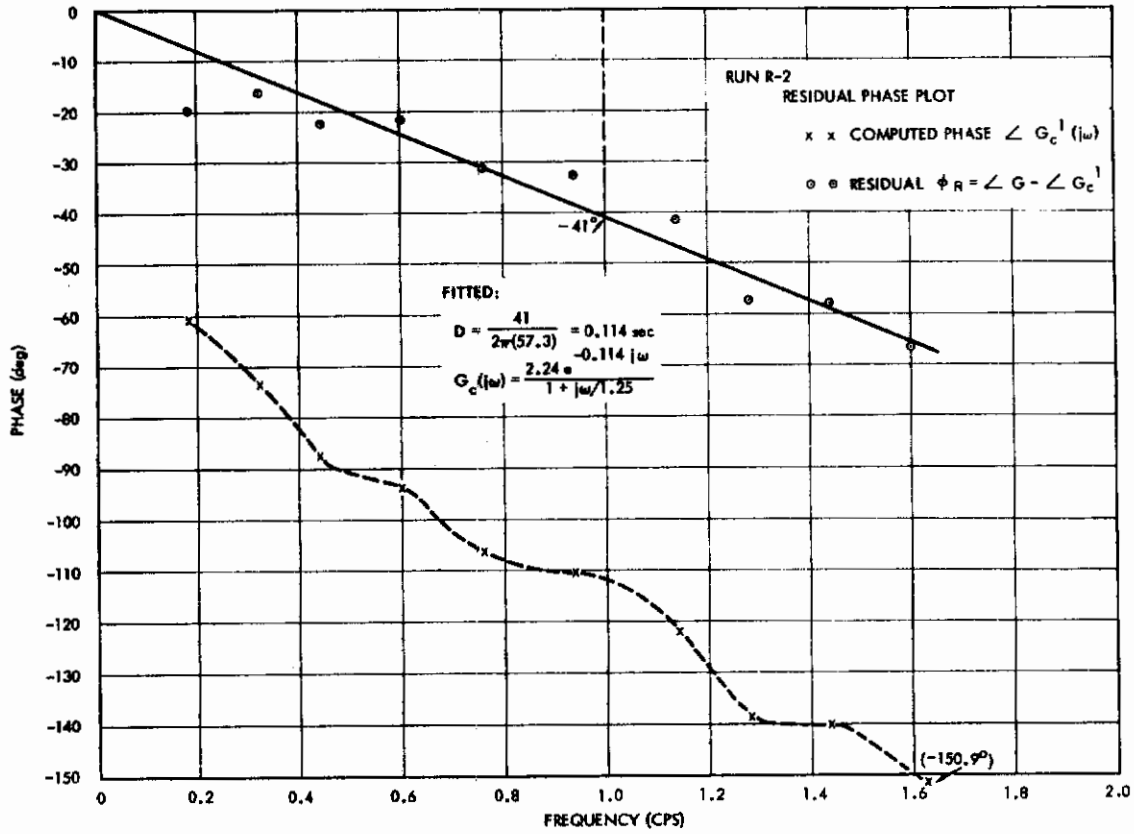


Figure A8.2 (b) Fitted Phase for Run R-2

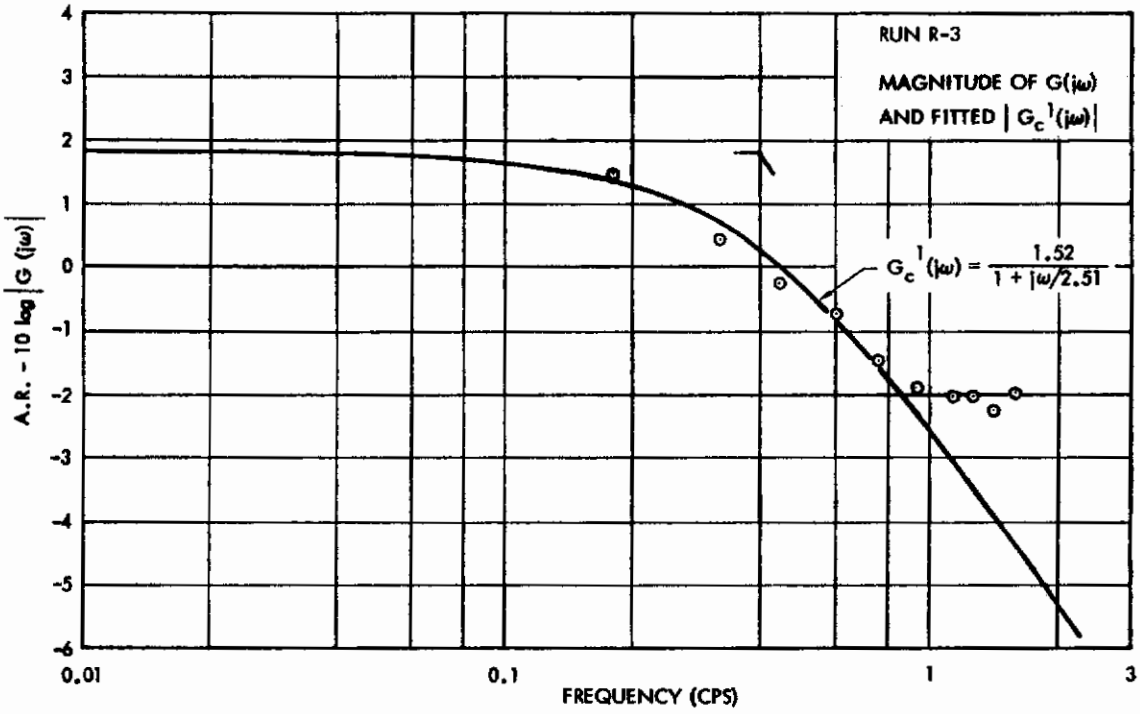


Figure A8.3 (a) Fitted Magnitude for Run R-3

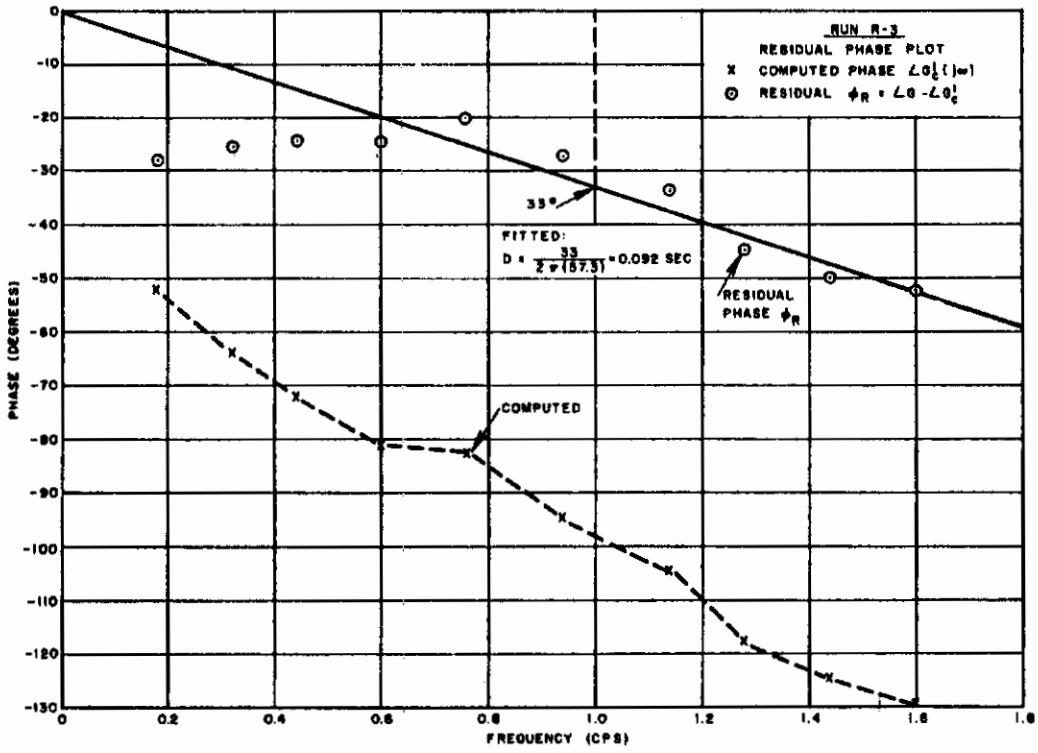


Figure A8.3 (b) Fitted Phase for Run R-3