UNIVERSITY OF LONDON

FACULTY OF ENGINEERING

Syntheses of a Model of the Human Operator Engaged in a Tracking Task

A thesis presented for the Degree of

Doctor of Philosophy

by,

Graham Warren Lange

October 1965

SUMMARY

This thesis presents the results of investigations aimed at obtaining an accurate representation of the characteristics of human operators relating to compensatory tracking with simple unity feedback, visual error display, manual output, and Gaussian random input.

The study of relevant literature, and of operators' step and ramp responses led to the initial formulation of a sampled data model, incorporating the hypotheses that the operator sampled instantaneous position and velocity error at intervals of about .15 second, and formulated programmes of hand movement over the same intervals. Study of continuous tracking, using an analogue representation of this model, led to the incorporation of internal feedbacks representing more sophisticated prediction. The extended model gave an excellent match to operators' tracking, and was in agreement with other workers' results.

This model was then utilised to perform a series of tests aimed at obtaining operators' open loop responses. Results emphasised the idealised nature of the hypotheses embodied in the model, while confirming its basic accuracy. They also showed that simple models of forward transfer, as proposed in the literature, could be fitted quite well to error-output covariance function data.

Further investigation led to the postulation of a random sampling hypothesis, with sampling operations lasting 20 - 45 ms., and occurring at intervals from 80 to 220 ms.

It was found that operators' transfers approached quite closely to the optimum, according to the 'least square error' criterion. Further, operators clearly exhibited features associated with self-organising systems, even in the simple task studied. It was concluded that a model of sampling and 'optimum' digital processing, combined with an adaptive form of output response, was capable of accurately representing operators' one-dimensional tracking characteristics.

CONTENTS

	Page
Summary	2
Acknowledgements	7
List of Symbols	8
Chapter 1 : INTRODUCTION	
1.1. Motivation and Aims of Investigations	15
1.2. General Consideration of Human Tracking Tasks	15
1.3. Representation of Human Tracking in Terms of Mathematical and Analogue Models	16
1.4. Factors Limiting the Scope of the Present Study	20
1.5. Aims of the Present Study	20
1.6. Arrangement of Thesis	22
Chapter 2 : REVIEW OF LITERATURE AND PHYSIOLOGY PERTINENT TO HUMAN TRACKING	
2.1. Review of General Literature Concerning Human Tracking Tasks	24
2.2. Review of Work at Imperial College	31
2.3. Aspects of Physiology Related to Human Tracking	32
2.4. Summary	35
Chapter 3 : PRELIMINARY INVESTIGATIONS	
3.1. Formulation of the Tracking Task	37
3.2. Selection of the Statistical Structure of a Continuous Input	38
3.3. Choice of Continuous Input Statistics and Duration of Tracking Runs	41
3.4. Selection of Discontinuous Random Inputs	43
3.5. Probability Density Functions of Loop Variates	44
3.6. Selection of Performance Criterion	44
3.7. Learning Phenomena and Training	47
3.8. Investigation of the Effects of varying Hand-to-Display Gain	48
3.9. Investigation of the Effects of Display Nonlinearities	52
3.10. Summary of Results and Conclusions from Preliminary Investigations	55
Chapter 4 : EXPERIMENTS LEADING TO THE INITIAL FORMULATION OF A SAMPLED DATA MODEL	
4.1. Specific Investigations Leading to the Initial Formulation of the Model	56
4.2. Experimental Investigation of the Effects of Sampled Error Display	56
4.3. Implications of the Results of Sampled Display Tests for the Hypothesis of Continuous Action	62

4.4. Implications of the Results of the Sampled Display Tests for the Sampling Hypothesis 66
4.5. Formulation of a Qualitative Model to Accord with Results of Sampled Display Experiments 69
4.6. Experimental Study of Operators' Responses to Random Step and Ramp Inputs 70
4.7. Quantitative Results of Step and Ramp Tests 71
4.8. Conclusions from Quantitative Results of Step and Ramp Experiments 75
4.9. Implications of the Quantitative Results of Step and Ramp Tests for the Proposed Model 79
4.10. Initial Formulation of a Model Structure from Qualitative Features of Step Responses 79
4.11. Further Formulation of Model Structure from Qualitative Features of Ramp Responses 84
4.12. Evaluation of Model as Initially Formulated 93

Chapter 5 : REFINEMENT OF THE SAMPLED DATA MODEL

5.1. Experimental Evaluation of the Ramp Response Model . 94
5.2. Evaluation of Ramp Response Model in Continuous Tracking Simulation 96
5.3. Extension of the Predictive Capabilities of the Model . 98
5.4. Linear Prediction of Position 99
5.5. Linear Prediction of Velocity from Velocity Information 99
5.6. Linear Prediction of Velocity from Position Information 103
5.7. Experimental Evaluation of Model Modified to Allow for Prediction 107
5.8. Summary 111

Chapter 6 : Z-TRANSFORM ANALYSIS OF THE SAMPLED DATA MODEL

6.1. Theoretical Analysis of the Sampled Data Model by Use of the Modified Matrix Z-Transform 112
6.2. Derivation of the Model Z-Transfers 117
6.3. Calculation of the Closed Loop Frequency Domain Characteristics of the Model 119
6.4. Calculation of the Forward Sequence Frequency Domain Characteristics of the Model 122

Chapter 7 : COMPARISONS OF THE CLOSED LOOP CHARACTERISTICS OF OPERATORS WITH THOSE OF FITTED MODELS

7.1. Introduction 126
7.2. Real Time Comparison of Operators' and Models' Responses to Common Random Ramp Inputs 126
7.3. Consideration of Continuous Tracking Response of Operators and Model 130
7.4. Comparison of Operator and Model Covariance Functions Calculated from Records of Continuous Tracking 138
7.5. Fitting Model and Operator Covariance Functions 148

7.6. Consideration of the Calculated Spectral Characteristics of the Model 150
7.7. Summary 155

Chapter 8 : QUALITATIVE AND QUANTITATIVE RESULTS OF OPEN LOOP TESTS

8.1. Motivations for Conducting Open Loop Tests 157
8.2. Method of Conducting Open Loop Experiments 157
8.3. Consideration of Qualitative Results of Open Loop Tests. 158
8.4. Consideration of Quantitative Results of Open Loop Tests 165
8.5. The Fitting of Models to the Open Loop Test Data ... 173
8.6. Models Fitted to the Open Loop Test Covariance Data . 177
8.7. Implications of Simple Models Fitted to Open Loop Covariance Data 182
8.8. Summary 187

Chapter 9 : FURTHER ASPECTS OF HUMAN TRACKING BEHAVIOUR

9.1. Randomness of Sampling 189
9.2. Consideration of Model Structure in Relation to Compensatory and Pursuit Tracking 193
9.3. Relation of the form of the Model's Response to Remnant Data 194
9.4. Relation of Operators' Responses to Impulse Response Optimisation 195
9.5. Relation of Experimental Results to the Data of Other Workers in the Light of Impulse Response Optimisation . 198
9.6. Consideration of Operators' Optimum Seeking Behaviour . 199
9.7. Summary 203

Chapter 10 : CONCLUSIONS AND RECOMMENDATIONS

10.1. Summary of Results and Conclusions Drawn from Investigations 205
10.2. General Conclusions 209
10.3. Recommendations for Further Investigations 212

References 213

APPENDIX I : Apparatus for Investigation of Compensatory Tracking

I.1. The Basic Operator Feedback Loop 216
I.2. Arrangements for Conducting a Tracking Run 218
I.3. Circuit for Sampled Error Display 219
I.4. Error Switching Circuit for Open Loop Tests 219

APPENDIX II : Low Frequency Gaussian Random Signal Generator

II.1. Requirements for the Statistics of the Random Signal . 223
II.2. Methods of Generation of a Low Frequency Random Signal. 223
II.3. Probability Density Function of Intervals of a Pulse Sequence Generated by Level-Sampling 224

II.4. Spectral Density of a Random Square Wave with Approximately Independent Transitions 226
II.5. Practical Scheme for Generation of Random Square Wave . 230
II.6. Filter for Deriving Very Low Frequency Random Signal from the Random Square Wave 232

APPENDIX III : Generation of Random Step and Ramp Functions

III.1. Generation of Random Step Function 239
III.2. Statistics of Step Generator Output 241
III.3. Random Ramp Function 241

APPENDIX IV : Estimation of Variance and Cross-Correlation from Integrated Moduli

IV.1. Estimation of Variance 244
IV.2. Estimation of Cross-Correlation from Integrated Moduli. 248

APPENDIX V : Practical Realisation of the Elements of the Model

V.1. Generation of Velocity Signal 250
V.2. Velocity and Position Computation Loops 250
V.3. Generation of Velocity-Triangle Response 252
V.4. Overall Analogue Model Circuit 254

APPENDIX VI : Forms and Stability of Model Z-Transfers

VI.1. Transfer of the Velocity-Triangle Generating Circuit . 256
VI.2. Explicit Forms of the Model Transfers 258
VI.3. Qualitative Examination of Closed Loop Stability of Model 261
VI.4. Evaluation of the Roots of h(z) and q(z) 262

APPENDIX VII : Optimum Linear Filters Including Pure Delay . 264

APPENDIX VIII : Distortion of Error-Output Statistical Functions Due to Recirculation of Remnant . 267

APPENDIX IX : Spectral Analysis and Model Fitting to Time Domain Data

IX.1. Causation of Difficulties of Model Fitting in the Spectral Domain 270
IX.2. Model Fitting by Direct Use of Time Domain Data ... 270

APPENDIX X : Calculation of Probability Density Functions Associated with Random Sampling 273

APPENDIX XI : Frequency Response of Sampled Data Systems . 276

ACKNOWLEDGEMENTS

The investigations described in this thesis were carried out in the Electrical Engineering Department of the Imperial College of Science and Technology . The author wishes to express his gratitude to Professor J.H.Westcott,for initiating and supervising the study, and to the Head of the Department,Sir Willis Jackson, for provision of facilities for research.

Thanks are due to Dr.W.F.Fincham for reading the original draft of this thesis , and for his many helpful suggestions.

The author would like to thank Messrs. A.G.Forton, J.M.C.Clarke, W.R.Atkins,and D.M.Fellows for acting as volunteer operators,and Mr.E.Sigurdson for assistance with digital computations.

Grateful acknowledgement is made to the University of London and to the DSIR for their awards of postgraduate studentships.

Finally,the author would like to thank other members of the control group,and in particular Dr.D.G.Watts , for helpful and informative discussions.

LIST OF SYMBOLS

a ; Filter break point (rads/sec)

a(n.Δt)
b(n+k.Δt) ; General representation of zero mean time functions, at discrete times separated by k.Δt .

a,b ; Parameters of models fitted to open loop covariance data

$[A(z,m)]$; Matrix of modified Z-transfers (see equations 6.144.,5.,6.)

$[B(z,m)]$; Ditto

$b_{1.23}$ etc. ; Partial regression coefficients

C(s) ; Overall output of continuous control system with noise added at output, as transform

C(z) ; Output (Z-transformed) of simple sampled data system.

$C_p(n), C_p(nT)$; Output position at time nT

$C_v(n), C_v(nT)$; Output velocity at time nT

$\underline{C}(nT)$; Vector of output position and velocity at time nT

$\underline{C}(z), \underline{C}(z,m)$; Z-transform and modified Z-transform of output vector

${}_pC_v(n)$; Prior prediction of output velocity at time nT

$\underline{C}_d(z)$; Vector (Z-transformed) of disturbance added at output.

$C_p(jw)$; Overall closed loop frequency response of model, in terms of sinusoid fitted to $C_p(nT)$; $-\infty<n<\infty$

$C_p^f(jw)$; Actual open loop frequency response of model, in terms of sinusoid fitted to open loop output at sampling instants

D ; Part of reaction time delay to a step input due to its occurring between sampling instants.

D(s) ; Output of continuous control system prior to the addition of noise at the output, (as transform).

D(z) ; Transfer function denominator polynomial in z .

$D_p(m)$; Difference between position output at time (n+m)T and position output at time nT

D_m ; Average magnitude of deviation of R_t from $\bar{R}_t$

$d(t)$; Instantaneous difference between operator and model errors

κ_{YX} ;Apparent coefficient of non-determination between X and Y

E ; Expectation of term in brackets

E_1, E_3 ; Errors of estimation of position and velocity error, respectively.

E_2, E_4 ; Errors of execution of position and velocity programmes, respectively

$E_p(n), E_p(nT)$; Position error at time nT

$E_v(n), E_v(nT)$; Velocity error at time nT

$\underline{E}(nT)$; Vector of position and velocity errors at time nT

$\underline{E}(z), \underline{E}(z,m)$; Z-transform and modified Z-transform of error vector

e ; Continuous loop error (as suffix)

$E(s)$; Error in continuous control system, with noise added at output, (as transform)

$_pE_v(n)$; Prior prediction of velocity error at time nT

$\underline{E}_d(z)$;Vector (Z-transformed) of disturbance added in error channel

$E_p(jw)$; Closed loop frequency response of model error, in terms of sinusoid fitted at sampling instants

$Erf(\ )$; Gaussian error function

$f_k(x)$; k th. member of Sturm sequence

f_1, f_2, f_3, etc,; Weighting factors for linear prediction of velocity from past samples

f_b ; Break point (cps.) of general noise generator filter.

$F_p(jw)$; Equivalent forward frequency response from position error to position output (both in terms of sinusoids fitted at sampling instants), under closed loop conditions.

$g_p, g_v, g_s, g_d, g_1, g_3, g_{mp}, g_{mv}$; model gain parameters

$G(s)$; General Laplace transfer function

$G(z)$; General Z-transfer function (ratio of polynomials)

$G_h(z)$; Z-transfer of continuous model with fictitious sampler and data reconstruction

$G_1(s), G_2(s)$; Continuous model transfers fitted to open loop test covariance data

$G_n(s)$; Optimum Wiener transfer for prediction of an input consisting of n th. order filtered (low-pass) white noise

$G_a(s), G_b(s)$; Constituent transfers of optimum Wiener predictor

$_{npr}G_b(s)$; Optimum form of $G_b(s)$ without regard to physical realisability

$_{npr}g_b(t)$; Impulse response corresponding to $_{npr}G_b(s)$

$h(z)$; Polynomial in z ,the roots of which correspond to poles of the actual open loop model Z-transfer

$[H(z,m)]$; Matrix of modified Z-transfers ,elements $h_{pp}, h_{pv}, h_{vp}, h_{vv}(z,m)$ giving zeros of actual open loop model Z-transfer

i ; continuous loop input (as suffix)

I ; Length of operator's sampling interval (random variate)

I_s ; Length of operator's sampling interval succeeding that in which a step input has occurred

I_{ss} ; Interval between successive stimuli

j ; complex operator,defined by $j^2 = -1$

k ; integer (as dummy variable)

k ; gain constant

K_T ; Gain describing function due to threshold effect.

L ; Pure delay of the operator's response to a step input

m ; Parameter between 0 and 1,used to define time between sampling instants

$[M]$; Particular form of companion matrix to a given polynomial

n ; An integer, including 0 .

n! ; Factorial n

N(s) ; Noise added at output of continuous control system (as transform)

N(z) ; Transfer function numerator polynomial in z .

o ; Continuous output (as suffix)

p ; Sampled position input (as suffix)

P ; Performance index (ratio of variances)

P_a ; Probability of a step in any sampling interval

$p(I)$; Probability density function of operator's sampling intervals

$p(|X|)$; Probability density function of $|X|$

$p(n)$; Probability of n transitions of a random square wave in time

$P_1, P_2, \ldots P_9$; Products of combinations of model parameters

$p(w)$; Polynomial in the complex variable w ,of Bilinear transform

$P_v(n)$; Programme of velocity correction released at time nT

$P(z,m)$; Overall modified Z-transfer of velocity triangle generator

$_fP_e$; Computed apparent fresh position error

$_mP_e$; Computed modified fresh position error

$q(D)$; Probability density function of delay D

$q_k(x)$; Quotient derived in forming Sturm sequence

$q(z)$; Polynomial in z ,the roots of which correspond to the poles of the closed loop model Z-transfer

$[Q(z,m)]$; Matrix of modified Z-transfers, elements $q_{pp}, q_{pv}, q_{vp}, q_{vv}(z,m)$ giving zeros of closed loop model Z-transfer

$r(L)$; Probability density function of delay L

$R(s)$; Input to continuous control system ,(as transform)

$R(z)$; Z-transform of input to simple sampled data system

$r(x)$; Polynomial in complex variable x ,derived from p(w)

$R(o)$; Correlation function of filtered thyratron noise

r_{YX} ; Biassed coefficient of correlation between X and Y

r ; Unbiassed coefficient of correlation between X and Y

R_t ; Reaction time delay measured in step and ramp responses

$\overline{R}_t$; Mean of R_t

$R_p(n), R_p(nT)$; Input position at time nT

$R_v(n), R_v(nT)$; Input velocity at time nT

$\underline{R}(nT)$; Vector of input position and velocity at time nT

$\underline{R}(z), \underline{R}(z,m)$; Z-transform and modified Z-transform of input vector

${}_pR_v(n)$; Prior prediction of input velocity at time nT

${}_rR_v(n)$; Reconstruction of input velocity at time nT

s ; Laplace operator, complex frequency

S_a ; Sinusoid fitted to position output over sampling interval

S_f ; Sinusoid fitted to position output at sampling instants

s_n ; Normalised standard deviation

S_k ; k th. Newton sum of a polynomial

sgn.( ) ; Denotes the operation of taking the sign of ( )

t ; time

T ; Model sampling interval

nT ; Times nT coincide with model sampling instants

Ψ ; Covariance data sampling interval

T_N ; Length of integration time for estimating error variance

T_d ; Time delay

T_s ; Error display sampling interval

Δt ; Record sampling interval

T_1, T_L, T_N ; Time constants associated with linear continuous models of the human operator

T_{r_1} ; First reaction time

T_{r_2} ; Second reaction time

T_o ; Organisation time

$[\ ]^T$; Transpose of matrix

v ; Velocity of input at sampling instants (as suffix)

V_i ; Variance of input

V_e ; Variance of error

$_iV_e$; Variance of error linearly correlated with input

$_nV_e$; Variance of error not linearly correlated with input

v_r ; Relative variation

var(|X|) ; Variance of |X|

$V(x_o)$; Sign change function associated with Sturm sequence

$V_i(n)$; Input to velocity triangle generator at time nT

$V_o(n+m)$; Output of velocity triangle generator at time (n+m)T

$_fV_e$; Computed fresh velocity error

W ; Output power of optimum Wiener predictor

$_{npr}W$; Output power of non-physically realisable Wiener predictor

w ; Complex variable used in Bilinear Transformation

x ; Complex variable derived from w above (x = jw);Appendix VI

x ; Zero mean Gaussian variate (Appendix IV)

X(t) ; Value of random square wave at time t (Appendix II)

X(t),Y(t) ; Model and operator errors for computing variance and correlation (Appendix IV)

X_1,X_2,X_3,X_4; Regression variables,as defined in Section 4.7.

z ; The operator e^{Ts}

z_v ; The v th.root of a polynomial in z .

.....

α,β ; Weighting factors

δt ; Infinitesimal interval of time

$\lambda(max)$; Eigenvalue of [M] with largest modulus

$\eta(n)$; Error of prediction of $E_v(n)$

ζ(zeta) ; Damping factor

$\Lambda(z,m)$; Discontinuous function of m in transfer of velocity-triangle generator

λ ; No. of pulses/sec. of a Poisson pulse train

μ ; No. of pulses/sec. of a random pulse train with second Poisson interval probability density function

$\mu(n)$; Output at time nT of velocity computation loop with loop gain of (1-k)

σ_a ; Standard deviation of variate a .

τ ; Dummy time parameter , time lag (covariance)

$\rho_{xx}(\tau)$; Correlation function of random square wave

$\Phi_{xx}(w)$; Power spectral density of random square wave

${}^n\Phi_{rr}(s)$; Two sided Laplace transform of input to Wiener predictor : white noise through an n th. order low-pass filter

w(omega) ; Angular frequency ,rads/sec.

Notation for Covariance and Spectral Density Functions

$\sigma_{ab}(\tau)$; Covariance of time functions a(t) and b(t+τ)

$\Phi_{ab}(w)$; Cross power spectral density of time functions a(t) and b(t)

$\Phi_{ab}(z)$; Cross pulse spectral density of sampled time functions a(nT) and b(nT) ; $-\infty < n < \infty$

$[\Phi_{ab}(z)]$; Matrix of pulse spectral densities formed from position and velocity of a(t) and b(t) at sampling instants,as specified in Section 7.5.

Chapter 1 ; INTRODUCTION

1.1. Motivation and Aims of Investigations

The investigations described in this thesis were mainly concerned with the description of the characteristics of the human operator performing a tracking task. The chief desire was to obtain analogue and mathematical models capable of a good representation of human control actions. The viewpoint throughout has been that of the control engineer. The results should,however,be of interest to those working on similar problems in other fields of study.

The motivation for carrying out such research springs naturally from a consideration of the large number of practical situations in which the human being acts upon information which is partially dependent on his previous actions - i.e. the operator acts as an element in a feedback loop. The increasing variety of human control tasks,and their demands upon the operator,emphasise the need to understand the basic structure of the operator's tracking characteristics.

In a study of this nature it was not feasible to cover more than a small part of the field open to investigation. Experiments were therefore designed with a view to obtaining results possessing the greatest possible degree of general applicability. It was hoped that these results might lead to further advances in the status of human operator tracking research.

1.2. General Consideration of Human Tracking Tasks

A general definition of a human tracking task may be given as follows:-

"A human tracking task is a task requiring an operator to modify the state of his external environment in response to sensory inputs derived therefrom,so as to achieve as nearly as possible,a desired state,or succession of such states,in that environment. In particular the operator is required to counteract the effects of chronic variations and fluctuations in the state of the external environment."

Many tracking tasks involve the simultaneous utilisation of several sensory modalities. It is usually the case,however,that the predominant source of information about the external environment is provided by the visual channel. It is hardly surprising that this should be so,when consideration is given to the relative proportions of the human brain allocated to the processing of information from the various senses. The number of neurons in the visual cortex is a considerable fraction of the brain's entire complement of cells,and is,for example,about 100 times larger than the number concerned with the auditory channel.

In the majority of human tracking tasks the operator's output is in the form of muscular activity, usually causing movement of the limbs. This implies that the operator has to contend with the unavoidable dynamics of limb movements, in addition to those of the controlled element.

The class of all tracking tasks may be subdivided into two sub-classes, namely pursuit and compensatory tasks. In the pursuit task the operator has direct information concerning the state of his external environment, whereas in the compensatory task the operator's information is restricted, so that he only has direct knowledge of the difference between the actual and desired state of his external environment.

The two task configurations are represented diagrammatically in Fig.1.2.1, where the single, directed lines will normally represent a flow of several signals. The external environment consists of the controlled element, the input, and the display ; 'display' is here to be interpreted in the widest sense - e.g. in an aircraft it may include motion cues not displayed on any instrument panel. The input signal may be zero or constant, implying that the operator is only required to counteract disturbances entering via the controlled element. On the other hand it may consist of a deterministic or random function of time, or a mixture of both ; e.g. landing an aircraft along a predetermined flight path, or driving along a mountain road.

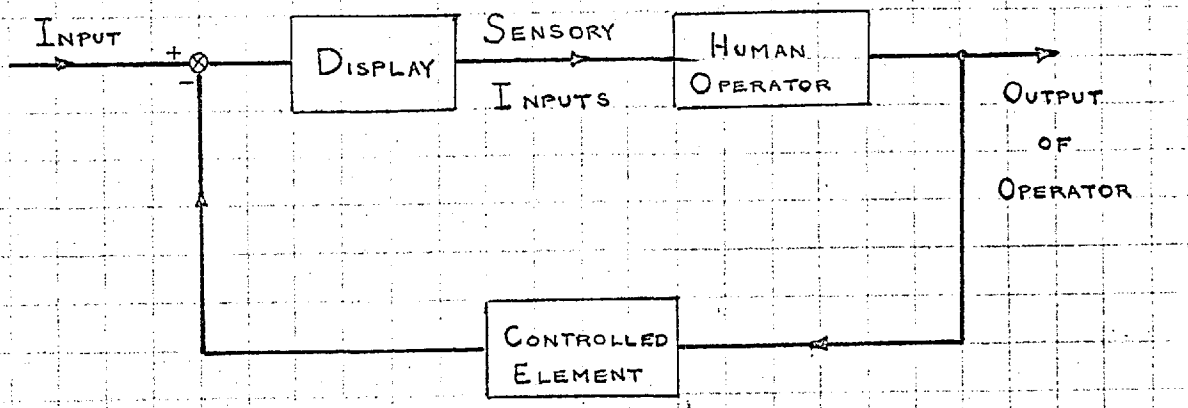
Whatever the task configuration and input may be, the operator is required to match the output of the controlled element to the input signal, by applying some specified criterion of error. The diagrams of Fig.1.2.1. illustrate that the operator can, in principle, only distinguish between fluctuations at the output of the controlled element and changes in the input signal under conditions of pursuit tracking. The difference between the pursuit and compensatory configurations is minimal when the input signal is a relatively simple deterministic function of time, because the operator is then able to predict the input quite accurately, even though tracking is compensatory.

1.3. Representation of Human Tracking in Terms of Mathematical and Analogue Models

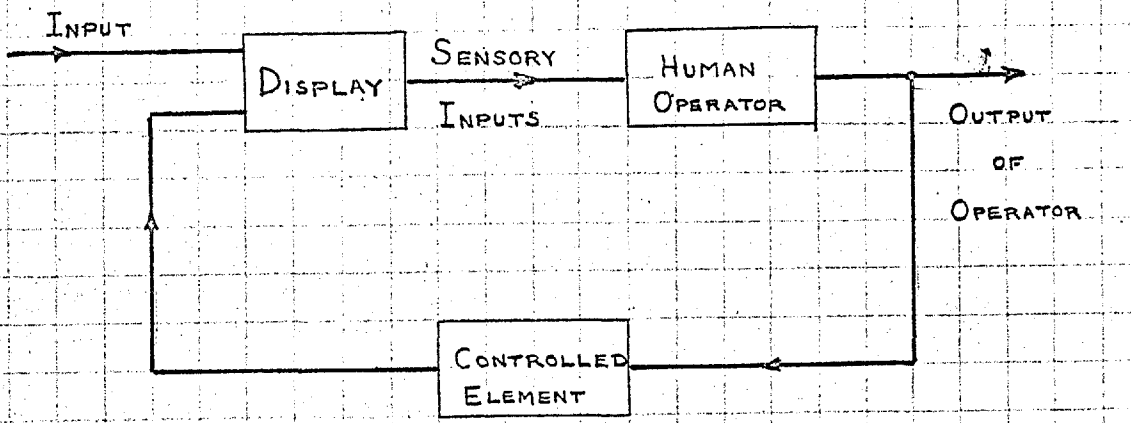
The desire to represent the human operator's performance, in a tracking task, in terms of a model is evidenced by the large number of such models described in the literature. The principal motivations behind such descriptions are:-

(1) The representation of the characteristics of an operator in terms of a model, even though such a model may be physically unrealisable, generally gives a clearer insight into the nature and essential features of the operator's response than it would be possible to convey by means of a set of disembodied data. Such an insight is an essential requirement for any attempt at generalisation of, or application of, the results.

Fig. 1.2.1. Diagrammatic Representation of Compensatory and Pursuit Tracking Tasks



Compensatory Tracking Configuration



Pursuit Tracking Configuration

(2) Very close approximation of the time and/or frequency domain characteristics of the human operator is often most conveniently effected through the experimental refinement of an analogue model ; this is especially true when the desired accuracy of representation demands the employment of a large number of parameters. Such analogue models need not necessarily operate in real time.

As the above considerations suggest, two main methods have been employed to arrive at models of human tracking. These are:-

(a) Input, error, and operator output data are gathered and analysed in terms of time and/or frequency domain functions. Examination of the experimental data then leads to the proposition of what is usually a relatively simple model. The parameters of this model are then chosen to yield a best fit (according to the investigator's criterion) to these data. Should the fit prove unacceptably poor, the model structure is refined, usually being elaborated by the addition of further elements, and parameters are re-computed for best fit. The process is continued until a model is obtained which shows a satisfactory fit to the experimental data. The criteria governing the choice of the best fit are sometimes a little arbitrary - often visual fits confirmed by a second, usually independent person.

(b) An 'a priori' model is proposed, taking into account those characteristics of the operator already known or established experimentally. The model is then set up on an analogue simulator and its parameters are adjusted to yield the best fit to the operator's response - again according to the experimenter's criteria of comparison. The procedure may be carried out in real time, with identical inputs to operator and model, or it may be carried out in analogue time by the use of recording apparatus. A detailed examination of the quality of fit of time and/or frequency domain characteristics is then carried out, and either suggests further refinements of the model, or aids in the formulation of further experiments - e.g. changing the form of the input signal or the controlled element. The complexity of the model finally derived is dependent on the aims of the investigation, and on the equipment and time available for its completion. There is a pay-off between model complexity, cost of implementation, and possible increase in accuracy of representation.

It is interesting to compare methods (a) and (b) in terms of the results they might be expected to yield. Method (a) possesses the advantages that a relatively simple model is usually obtained, and that, though the model must be mathematically tractable, it need not be physically realisable. The main disadvantages are that such models are generally not capable of giving a really good insight into the physical processes involved in human tracking, and that their components exhibit but a tentative relationship to corresponding components which might exist in the human operator.

Method (b) exhibits advantages and disadvantages which are practically the inverse of those possessed by method (a). Thus the simulator method allows of only physically realisable models, but they need not be mathematically tractable (in the restricted sense of representation by reasonably simple mathematical functions). For example, it is possible to include quite complex nonlinearities, if required. It may be noted that the restriction as to physical realisability is not very important, since any human operator must satisfy a similar restriction. If a model contains physically unrealisable elements, then this is essentially a reflection of the limitations of the model itself. A further point in favour of method (b) is that the structure of the model may be chosen to correspond as far as possible with the physical structure of the human operator, within the limits of prior knowledge and experimental facilities available to the investigator. This aspect of the method generally leads to a greater physical insight and intuitive appreciation of the operator's tracking characteristics.

The great majority of work concerning human operator models has been carried out under conditions of visual input and manual response, but some work has been done with more general experimental conditions - e.g. a study conducted at the Goodyear Aircraft Corporation (1) employed a moving mock-up of an aircraft cockpit. Models derived via method (a) have, to date, been quasi-linear in form ; i.e. they represent the operator's response in terms of a linear transfer function, the parameters of which are dependent on the statistics of the input, and on the form of the controlled element. The portion of the operator's output not linearly correlated with the input signal is ascribed to remnant noise. This noise is generally represented as entering the operator's output, and usually contains power over a frequency range considerably wider than that of the input signal spectrum. It is commonly attributed to the following causes:-

(1) Nonlinear operation on the input signal.

(2) Short term random inaccuracy of response.

(3) Longer term variation of the operator's transfer characteristics.

(4) A 'dither' signal which is sometimes injected by the operator. This occurs only with some operators in particular experimental configurations - often where the control lever exhibits stiction effects.

(5) On the basis of the sampling hypothesis (i.e. the hypothesis that the operator acts in such a fashion as to respond to his error only at discrete points in time) some of the remnant power would be due to ripple occuring as a result of the sampling operation. Bekey (2) has been the protagonist responsible for investigating this effect.

An important instance of the use of method (b) is provided by the studies conducted under the auspices of the Goodyear Aircraft Corporation (1). These involved the formulation of a nonlinear model to represent the human pilot controlling a simulated aircraft via compensatory feedback, and serve to exemplify the advantages of the method. Both Ward (3) and Lemay (4) used an analogue simulator to represent the human operator on the basis of the intermittency hypothesis, utilising direct comparison of model and operator responses.

1.4. Factors Limiting the Scope of the Present Study

The extent of the present study was limited by the following factors:-

(1) Only five experimental subjects were available. They were all young, male postgraduate students, and all were volunteers.

(2) The number of tracking runs which it was possible to perform in a given time was limited by considerations of subject availability and operator fatigue. Together with factor (1), this implied a restriction on the number of investigations which might be carried out.

(3) Limitations of apparatus restricted consideration to tasks involving purely visual display.

(4) It was decided to restrict the study to one dimensional tracking, partly because of limitations of apparatus, and partly because such tracking permitted a reasonably simple characterisation of the tracking task, and was therefore likely to provide the best opportunity of developing an accurate model.

(5) The chief concern of the study was with tracking tasks involving random inputs. Comparatively few practical tasks involve deterministic inputs, and when performing such tasks operators tend to exhibit special forms of behaviour. It might be possible to relate such behaviour to the operator's performance when following random inputs, but it seemed more appropriate to proceed in the reverse direction.

1.5. Aims of the Present Study

The primary aim of the experimental investigations was to develop an accurate analogue model of the human operator's tracking characteristics. The chief requirements governing the formulation of this model were as follows:-

(a) The model should be capable of giving a very good, detailed representation of human tracking, especially in the time domain.

(b) The components of the model should correspond as closely as possible with similar components (or representations of sets of components) as proposed or hypothesised from the results of physiological experimentation on the human subject. It was recognised that such correspondence would have to be fairly gross in nature.

(c) The model's representation of the predictive abilities of the operator was of particular interest.

(d) Subject to (a),(b),and (c) above,the number of model parameters should be kept to a minimum.

(e) It was desired to compare the model with others described in the literature,which are quasi-linear in character. It was therefore desirable to derive a linear model,subject to satisfying (a). This would also greatly facilitate mathematical analysis.

A further aim of the investigation was to derive simple, linear models of the operator's open-loop transfer characteristics by direct consideration of error and output functions related to compensatory tracking. The main snag is that operator 'noise' (i.e. components in the operator's output not linearly correlated with the loop input) recirculates around the control loop,and causes distortion of the statistical functions which are required to estimate the operator's forward transfer by this method (see Appendix VIII). The customary method of analysis circumvents this difficulty by considering only input-error and input-output cross-correlation or cross-spectral density functions. Information from remnant terms in the operator's error and output is thereby effectively discarded.

Inevitable noise recirculation under closed-loop conditions, could be overcome by the artifice of opening the operator's feedback loop,and injecting a suitable synthetic signal obtained from the model ; this was the aim of the present study. This procedure placed additional emphasis on the requirement that the model should be capable of a very good time domain simulation of the operator's response - otherwise the operator would soon detect that his feedback loop had been broken. Additional information concerning the quality of the analogue model would also be obtained from this part of the study,by comparing recordings of model and operator responses, generated whilst the latter was 'open-looped'.

The achievement of the above aims would represent a very useful contribution to knowledge concerning human tracking. To obtain a model sufficiently good to permit a period of open-loop operation during a task involving 1:1 compensatory feedback,it would be necessary to obtain a very high cross-correlation between operator and model outputs,and also that the form of the model's output should be very 'lifelike' ; i.e. the model should be capable of generating some of the higher frequency ripple effects observed in operator's outputs. The model would therefore need to be more advanced than the large number of linear continuous models so far proposed in the literature. Also,the study of open-loop tracking,as proposed above,

might be expected to throw fresh light on human tracking behaviour. A study of open-loop tracking is reported in (1),but this related to rather complicated controlled element dynamics.

1.6. Arrangement of Thesis

The thesis has been arranged in the form of ten chapters plus appendices. The contents of Chapters 2-10 are summarised below.

Chapter 2 contains a brief review of the salient features of previous research concerned with the study of human operator tracking,as reported in the literature. Emphasis is placed on the models which have been proposed or derived. It concludes with an outline of physiological considerations related to human tracking.

Chapter 3 deals with the precise formulation of the tracking tasks to be studied,and with preliminary investigations into the probability density functions of signals in the tracking loop,the effect of mild display nonlinearities,selection of a performance criterion,and training and learning phenomena.

Chapter 4 describes an investigation into the effects of sampled display presentation. It then describes how detailed features of of the operator's step and ramp responses led to the initial formulation of a sampled data model.

Chapter 5 shows how the initially formulated model was refined by considering prediction. The results of a series of trials, comparing model and operator under conditions of common input,are then presented.

Chapter 6 describes the analysis of the refined form of the sampled data model in terms of the matrix Z-transformation. The calculation of the model's closed-loop frequency domain characteristics is then considered. Finally,corresponding open-loop characteristics are derived, and the difference between equivalent and actual open-loop characteristics is illustrated.

Chapter 7 is concerned with the comparison of operator's characteristics with those of the corresponding 'best-fit' models. The data presented include actual time traces,covariance and correlation functions,and theoretically calculated frequency domain functions.

Chapter 8 discusses experiments involving opening the operator's feedback loop.Results of such experiments are presented and evaluated from the viewpoints of,(a) the implications relative to the analogue model,and (b) the utilisation of covariance function data to derive simple mathematical models,and the implications of these models.

Chapter 9 considers the operator's performance in relation to the performance of Wiener predicting filters. The implications of random sampling are discussed,and suggestions are made concerning the operator's methods and criteria for achieving the best possible tracking performance.

Chapter 10 presents the conclusions drawn from the investigation, and suggestions as to the lines along which further research might be fruitful.

Chapter 2 : REVIEW OF LITERATURE AND PHYSIOLOGY PERTINENT TO HUMAN TRACKING

2.1. Review of General Literature Concerning Human Tracking Tasks

The pioneering study of the characteristics of the human operator performing a tracking task was reported by Tustin (5), in 1947. The work was concerned with gun-turret control, where displacement of a spade-grip handwheel gave rise to a velocity of rotation - i.e. the effective controlled element dynamics included a pure integration. The input was a 'random appearing', but deterministic, function of time, and consisted of a sum of three sinusoids. The operator's display was effectively compensatory.

The basis of Tustin's method of analysis was to consider the operator's output as being composed of two parts, one linearly correlated with the loop input, the other having no linear correlation. The operator's equivalent linear transfer was then obtained by considering the frequency domain relation of the linearly correlated part of the operator's output to the loop input function. From knowledge of the overall loop transfer, the equivalent forward transfer of the operator could be calculated. This transfer was of the form:-

$$G(s) = k.\frac{(1 + sT_1)}{s}.e^{-.3s} \qquad 2.1.1.$$

Tustin used the term 'remnant' to describe that part of the operator's output which was not linearly correlated with the loop input; others have termed it operator 'noise'. Its causation is considered in Section 1.3.

The above method of analysis is, in effect, a crude form of cross-spectral analysis. It has formed the basis of many subsequent studies. Most workers would not agree with the perfect integration in expression 2.1.1., but it should be remembered that the form of the input would have led to difficulty in distinguishing between a pure and a quasi-integration.

An hypothesis of intermittent action to explain operators' compensatory tracking characteristics was proposed by Craik (6) in 1947-8. The hypothesis was based on the observation that the operator's error function tended to exhibit a cyclical pattern, even under conditions where the input was random. There were usually about 2 'cycles' per second. Craik therefore proposed that the operator sampled his input (i.e. the loop error) about twice per second. He processed each of the observed samples, and formulated an appropriate ballistic response. This response was then released, and ran to completion before any further response, formulated from processing of the next observed sample, could be effected.

Craik linked the above hypothesis of intermittency with the

phenomenon of the "psychological refractory period" (PRP), which has been the subject of some controversy in the psychological literature. The phenomenon is observed when the time interval between successive stimuli to which a response is required, is less than about ½ sec. Thus Vince (7) found that the presentation of successive steps within ½ sec. led to longer delay in response to the second stimulus. Vince (8) later concluded that movements lasting less than .4 secs. were not affected by removal of the stimulus, once response movement had been initiated. However, it was subsequently found that the apparent PRP was affected by the direction of the second step, and that normal response was possible after intervals as short as .25 sec.

Welford (9) used results concerning the PRP as the basis of the hypothesis that a common central mechanism was responsible for both formulating a response and monitoring its execution. The operator could not, therefore, appreciate a further input while engaged in monitoring a response to a previous input. He proposed a relation of the form :-

$$T_{r2} = T_{r1} + T_o - I_{ss} \quad ; \quad I < T_{r1} \qquad 2.1.2.$$

where T_{r1} = reaction time to first stimulus

T_{r2} = reaction time to second stimulus

T_o = organisation time

I_{ss} = interval between stimuli

Welford's hypothesis may be stated as follows:-

(a) The central mechanism requires a discrete interval of time to organise information and release an appropriate response.

(b) The central mechanism may become occupied in processing feedback stimuli due to inaccurate first response.

(c) The central mechanism cannot deal with input stimuli simultaneously - information is stored until it can be processed.

There is a plethora of papers concerning the PRP, which seems to evidence some discord. Tests were therefore carried out by the present investigator, as described in Chapter 4. These tests were conducted with trained operators under instructions to respond as quickly as possible to each stimulus. Results indicated a maximum PRP between .1 and .15 sec., which is a time considerably smaller than those mentioned above. It seems that the differences can probably be ascribed to the particular experimental conditions. It is particularly important to give operators clear instructions and adequate training when conducting this type of test.

Elkind (10) gave a very thorough account of the effect of the form of the loop input on human tracking characteristics. He studied both pursuit and compensatory tracking, with purely visual error sensing. Input and output (pursuit) or error (compensatory) were displayed on an oscilloscope. Manual control was effected by means of a light, virtually frictionless stylus, held in the manner of a pencil. The form of the controlled element was effectively that of a unity transfer. The class of inputs studied was that of random appearing , but essentially deterministic, functions composed of a large number of sinusoids (40-144), the phases of which were not related. Studies were conducted regarding operator variability, and the effects of input amplitude, bandwidth, and shape. Three operators participated, and each was given a period of training before data gathering runs, each of 4 minutes duration, were made.

The main conclusions of the investigation were as follows:-

(a) Not much variation was observed between operators, or between results for the same operator and different runs.

(b) Pursuit tracking was always better than compensatory ; the difference was most marked for inputs containing high frequency terms.

(c) Operators were fairly insensitive to input amplitude, over a range of standard deviations from .1 to 1 inch. There was slightly more operator noise associated with the lowest amplitude.

(d) Operators were sensitive to input bandwidth. The effect of increasing bandwidth was to reduce loop gain, especially at low frequencies. The equivalent open loop phase shift at frequencies below about 2 cps. was also reduced. The highest bandwidths - above 1½ cps. - were associated with rather erratic tracking, and showed poor linear input - output cross-correlation.

(e) Operators were sensitive to the shape of the input spectrum. Sharper cut-off from a given break frequency led to higher values of low frequency gain and equivalent open loop phase shift - i.e. the equivalent open loop transfer approached more closely to an integration, at least at low frequencies.

(f) The results obtained from bandpass inputs indicated that, as centre frequency increased, the operator displayed a reduction in absolute level of gain. However, the shape of his gain-frequency curve remained remarkably unaffected, being effectively translated up the frequency scale. Phase characteristics were also translated up the frequency scale with little modification, especially in the case of pursuit tracking.

(g) Pursuit tracking characteristics could be related quite well to a parameter optimised filter, involving a straight transfer plus

differentiation in cascade with a pure delay. Compensatory characteristics showed a general similarity at low frequencies, but exhibited divergence at higher frequencies, probably because of the poorer prediction associated with compensatory tasks.

(h) Operator remnant variance was greater for compensatory than for pursuit tracking of a given form of input. For both configurations the remnant variance increased with increasing input bandwidth.

The form of the observed noise spectra could be approximated quite well by a highly damped second order transfer. These noise spectra could be related to a representation of output error in the form of a random step function.

It was not possible to derive analytic models of the operator's forward transfer for the pursuit configuration. For the compensatory task simple models of the operator's equivalent open loop transfer were derived by a visual fitting process to spectral domain data. These models were of the general form:-

$$G(s) = \frac{k.e^{-sT_d}}{(s + a)(s + b)} \qquad 2.1.3.$$

One lag was usually sufficient, but it was necessary to specify the second lag to yield an apparently stable closed loop response for models fitted to data for low bandwidth input signals ; the parameter associated with this lag could not be measured directly.

For square and filtered input spectra, the fitted values of the gain parameter, k , ranged from -3 to +34.5 dB. Corresponding values of pure delay ranged from .116 to .64 secs., while lag break frequencies ranged from .6 to .035 cps. The parameters varied in a systematic way; as input bandwidth increased the lag break frequency increased, while both gain and delay decreased. The behaviour of model parameters reflected the operator's adaptation to changes in the input spectrum, and could be explained in terms of the optimum filter mentioned above.

Krendel & McRuer have written two important papers (11,12) reviewing work on human tracking. The investigations described cover a wide field of input characteristics and controlled element dynamics. Models fitted to operators' transfers for continuous random appearing inputs are presented in the form of describing functions representing the transfer from that part of the loop error function linearly correlated with the input to the corresponding part of the output function. The form of the models may be summarised as :-

$$G(s) = \frac{k.(1 + sT_L).e^{-sT_d}}{(1 + sT_1).(1 + sT_N)}.K_T \qquad 2.1.4.$$

where T_d represents reaction time delay

T_N represents neuromuscular lag

$(1 + sT_L)/(1 + sT_1)$ represents equalisation introduced by the operator as a result of training.

K_T represents a small threshold effect (not directly measurable)

k represents the inherent gain of the operator. It is a parameter particularly sensitive to variations in input structure and controlled element dynamics.

The authors state that there is little justification in considering more complex models of the same general form, because improvement in fit is likely to be marginal, and parameters of such a model would be unduly sensitive to fluctuations in the data used for their calculation.

With simple controlled elements (unity transfers in effect) the outputs of linear continuous models of the form described above show cross-correlations with operators' outputs of the order of .7 to .9 . More complex controlled element dynamics generally lead to lower values of this cross-correlation. The best fit which might be obtained from a linear model is indicated by Elkind's data concerning the coefficients of linear determination between input and operator output, for compensatory tasks. These ranged from .995 for the lowest bandwidth task to about .4 for the highest ; for a bandwidth of 1 cps. the value was around .9 . Unfortunately, no data are available concerning the fit of Elkind's analytic models in the time domain.

Data on remnant spectra are not plentiful. There is little to add to the results given by Elkind, and already described, save that where dither was observed, the frequency was around 1 to 1½ cps.

In addition to the models already summarised an interesting nonlinear model is described. This was derived as a result of studies conducted at the Goodyear Aircraft Corporation, and included a linear transfer, plus threshold and anticipation effects. The latter was represented as a perfect relay in parallel with the other model transfers. Thus its effect was to bias the output according to the sign of the error. Dither was allowed for by adding a 1.4 cps. sinewave at the output of the other elements of the model. The overall model was matched to the operator by direct analogue simulation, while both operator and model tracked the same input. After matching, the model was sufficiently good to be substituted in the control loop for somw time, without the operator's knowledge. The length of time for which this could be done was greatest in the case of purely visual display ; with a moving cockpit the substitution was much less easy. The study was concerned with the operator's control of a simulated airframe, so that the controlled element dynamics consisted of a lead plus third order lag. Unfortunately, no data are available in regard to the numerical cross-correlation of model and operator outputs, although the model gave a good visual match.

Krendel & McRuer also describe an interesting model of the process of learning and adaptation which operators exhibit during the course of training (13). They describe the operator's adaptation to a compensatory task in terms of a progression from a purely compensatory configuration towards an effectively pursuit configuration and finally,with very predictable inputs,to a precognitive mode of tracking. This latter term describes the process whereby the operator is able to predict the input quite accurately,and match his hand motion to it. The error signal then serves to provide a relatively long term measure of the accuracy of prediction. The authors term this representation of the learning process the "successive organisation of perception" model.

North (14) suggested,in effect, a similar organisation of the learning process. His results sprang from the use of logic to derive 'a priori' characteristics of human tracking,which he then compared to experimental results.

An example of the use of time domain functions in order to derive a model of the human operator,for compensatory tracking, is provided by an experimental study conducted by Henderson (15). He employed error autocorrelation and error-output cross-correlation functions. Instead of attempting a direct solution of the convolution equation relating these functions to the operator's transfer,he set up an assumed transfer of the form :-

$$G(s) = \frac{k.(1 + sT_1).e^{-sT_d}}{s(s^2 + 2\zeta w_o s + w_o^2)} \qquad 2.1.5.$$

on an analogue computer. He then fed this analogue model with a signal corresponding to the error autocorrelation,and varied its parameters so as to obtain a good fit between model output and observed error-output cross-correlation. Henderson argued that the distortion of the cross-correlation caused by the spurious cross-correlation between error and remnant,would be small at reasonable values of positive lag,because the remnant caused the error,but the reverse was not true (see Appendix VIII). Accordingly,values of cross-correlation for negative and small positive lags were ignored.

It is difficult to assign a likely error to the parameters determined by the above analysis,because no coefficient of determination was computed. However,the fitted model parameters showed the operator's adaptability. It was found that,when the display gain was doubled,the fitted model gain parameter, k ,was halved. Pure delay in continuous tracking was found to be .16 sec. in both cases...

The analysis employed by Henderson was a special case of the general method of time domain analysis proposed by Elkind & Green (16). This method depends on weighting the outputs of a set of orthonormal filters so as to give a best fit,in the least mean squares sense,to the output of the operator,when both filters and operator are fed by a common input function. The basis of the method lies in

the form of the multiple regression equations linking filter output weightings with the cross-correlations between individual filter outputs and the operator's output. The method is discussed further in Appendix IX. It may be extended to the fitting of piecewise linear nonlinear relations by the partitioning of input space.

So far, the models described have all been in the form of continuous transfers ; however, some work has been done in regard to the fitting of sampled data models to operator tracking characteristics. Thus Ward (3) made a study of time responses of a sampled data model to represent compensatory tracking, and attempted to match these to those generated by operators fed with the same input. The input consisted of a sum of three sinusoids, the highest frequency of which was .2 cps.

A very thorough study of a class of sampled data models to represent operators' tracking with unity feedback, was made by Bekey (2). This class of models was obtained by the insertion of a sampler plus data reconstruction circuit in cascade with a linear transfer, the form of which was similar to that of the linear continuous models already described. Bekey investigated three types of data reconstruction circuits, viz. the zero order hold, first order hold, and modified first order hold (both velocity and position directly sampled). Corresponding partial velocity holds were also considered.

The model's sampling frequencies were chosen on the basis that small peaks are often observed in operators' error spectra, at frequencies between 1 and 1½ cps. These spectral peaks were interpreted in terms of the effect of spectral folding due to the operators' sampling action. Sampling intervals were therefore chosen in the range from .3 to .5 sec. The input function was chosen so as to contain a fair amount of power at high frequencies, in order to show up the effects of spectral folding more clearly. The form of the input was that of a sum of 10 sinusoids, ranging up to a frequency of 12 radians/sec., filtered by an exponential lag with a break point set at .75, 1.5 or 3 rads./sec. as required.

Bekey studied both continuous and sampled displays ; the latter were sampled at the same frequency as proposed for the model. Spectral functions of model and operator did not show any convincing correspondence as regards peaks, in the case of continuous display - in fact peaks were not consistently observed in operators' averaged characteristics. The correspondence was better in the case of sampled display.

Parameters fitted to the first order hold model by a programme of digital computation, showed that a small pure prediction was needed, even though the 'plant' possessed a minimum phase linear transfer. Unfortunately, no figures were quoted for the coefficient of determination relating to the fitted parameters, or for time domain cross-correlation between model and operator errors.

Bekey investigated the implications of a variable model sampling frequency, in terms of T locus and Liapounoff function analyses, but gave no direct evidence to support the hypothesis that this actually occurred in the case of operators.

2.2. Review of Work at Imperial College

Human operator studies at Imperial College were initially conducted by Rogers (17), under the supervision of Prof.J.H.Westcott. Rogers studied operators' step tracking responses. These were obtained by passing a chart, on which a number of step functions had previously been drawn, beneath a slit, through which the operator both observed the input line and attempted to draw a line to coincide with it. Rogers formulated an initialytheory of sampling to explain the observed features of these step responses. He obtained a histogram for the pure delay associated with step response, which he fitted by means of a square probability density function. On the strength of this, he hypothesised that the operator sampled regularly in time, with an interval of about .15 sec.

The basic scheme suggested by Rogers was modified and elaborated by Wilde (18), using more elaborate apparatus, consisting of an oscilloscope for error display, and free-moving control lever, arranged to give unity feedback. Wilde suggested that the operator's response to a step input was in the form of a velocity-triangle, giving a double-parabola position output. He proposed a mechanism which forms the basis for that described in Section 4.10., and also suggested that the operator sampled his input every .15 secs., which corresponded to the average length of observed velocity-triangles. From a study of continuous tracking records Wilde suggested that the operator attempted to make his output velocity proportional to observed position error, with an inevitable lag. He thus acted as an integrator, or quasi-integrator, plus a delay corresponding to reaction time. This hypothesis was consistent with results obtained by other workers.

Johnson (19) conducted an enquiry into the effects of stick 'feel' on continuous compensatory tracking performance. He used an input effectively derived by passing white noise through a second order filter with a double break at .5 cps. Apart from the forces applied to the control lever, the experimental set-up was as used by Wilde. Johnson studied the effects of inertia, spring restoring force and viscosity in the control lever 'feel', and derived continuous models to represent their effects on operators' characteristics. These models were of the form described in Section 2.1., but were derived by considering correlation functions (error autocorrelation and error-output cross-correlation). No allowance was made for corruption, except that only positive values of lag were considered. No objective measure of fit of the derived models was available. Unfortunately, no study was made concerning the relation of the fine structure of the operator's output to the particular type of 'feel' on the stick. However, it was found that operators could adapt quite well to spring and inertial forces, but found it difficult to adapt to forces simulating viscosity, though they did eventually succeed in tracking reasonably well with this type of output restraint.

Lemay (4) developed a sampled data model to represent continuous compensatory tracking with unity controlled element

transfer. The input was a random function formed by passing white noise through a third order filter,with a triple break at .6 cps. The model structure was based on the hypothesis of sampling at .15 sec. intervals,and velocity-triangle responses,as proposed by Wide. The chief addition to Wilde's original conceptual model of step response was the addition of a prefilter,with transfer (a + sb),preceeding the sampler. This prefilter results in a model structure similar to that of Bekey's modified partial velocity hold model.

Fitting of model parameters was accomplished by a parameter variation technique. The model was constructed to operate in real time and both model and operator were fed with the same input. Correlation functions were computed between corresponding operator and model functions. These showed that ,at best fit settings, the cross-correlation between model and operator errors was of the order of .6,and between outputs it was of the order of .9 . Lemay also tested continuous models,as proposed in the literature,and found that they exhibited poorer cross-correlations between errors,of the order of .5,but similar cross-correlation between outputs. However, he concluded that the fine structure of the sampled model's output and error functions showed distinctly better correspondence to those of operators than did the analogous functions of the continuous models.

Haynes (20) performed a series of step response tests under compensatory tracking conditions similar to those employed by Wilde. He was mainly interested in the probability density functions of the pure delay observed in operators' responses to step inputs. For input step functions approximating a random telegraph wave - i.e. steps occurring between fixed levels with an approximate independence between successive times of occurrence - he found that the histograms of delay could be closely approximated by a slightly skewed normal distribution. A satisfactory degree of significance,according to the χ^2 test,was obtained by matching a normal distribution to the experimentally determined histograms. Haynes also studied the response of an operator to a regular square wave,and found that he was capable of predicting quite accurately ; he exhibited a similar shape of delay histogram,though with a mean delay of approximately zero. Haynes concluded that the operator's sampling action was not regular, as proposed by Rogers,but was random in nature.

2.3. Aspects of Physiology Related to Human Tracking

The consideration of physiological data is a natural step in the formulation of hypotheses to account for the observed characteristics of human tracking. As explained in Section 1.4.,the present study was restricted to visual display and manual response,but is reasonably representative of a large class of tracking tasks. Accordingly,aspects of visual perception,brain processing,and muscular

control are considered in the following paragraphs.

Visual and Cerebral Processing

The portion of the nervous system concerned with visual perception is composed of the retina, optic nerve and visual cortex, together with structures associated with the control of position, focus, and aperture of the eye. There is considerable information processing at the retinal level. Evidence in favour of this view is provided by the fact that there are many more receptors in the average retina than could be catered for by the half million or so fibres in each optic nerve.

The maximum static visual acuity is largely determined by the granular structure of the retina in the foveal region, which is the area concerned with the perception of fine detail. The generally accepted limit corresponds to an angular resolution of one minute of arc - approximately the angle subtended by an object 1"dia. at 300 ft. - but the resolution of lines can be up to ten times better than this. Areas of the retina outside the foveal region are much less capable of observing fine detail, but are well adapted to observing a moving object - a property which seems to be due to the retinal processing mechanism. Moreover, moving objects can only be observed in satisfactory detail if their images are maintained reasonably stationary on the retina, and within the foveal region. Thus, an eye tracking system is a necessary concomitant to retinal structure.

The mechanism of eye tracking has recently been elucidated by Young (21). The chief characteristics of the tracking action are that the eye velocity tends to be matched to that of the object being tracked by a series of quasi-ramp functions, while position corrections are made at intervals by means of saccadic jumps. On the basis of experimental data, Young evolved a nonlinear sampled data model with a sampling frequency of 5/sec. Nonlinearities were incorporated to account for threshold effects (the system was relatively insensitive to the precise position of the image, so long as it lay within the foveal region), and saturation effects (the system was not capable of following at velocities exceeding 30°/sec.). As visual tracking is effected directly by the eye muscles, whose source of activation lies in the visual cortex, it seems reasonable to hypothesise that information regarding both position and velocity of an observed object is readily available to the remainder of the cerebral cortex. The mechanism of eye tracking partly accounts for the erratic responses of operators attempting to track a very broad bandwidth input, because the following of such an input would represent quite a difficult task for the eye, and would be especially difficult under compensatory conditions.

Study of e.e.g. records of brain activity reveals the presence of a number of rhythms, occurring over a frequency range from ½ to 30 cps. (see Walsh (22)). When resting, but during normal wakefulness, the most prominent feature is that of the α-rhythm, which is characterised by a frequency in the range 8 to 13 cps. These waves

have a temporal pattern as well as a spatial distribution over the surface of the cerebral cortex.

It has been suggested by Wiener (23) that the operator exhibits a sampling action corresponding to the α-rhythm frequency, i.e. samples are taken every .1 sec. He accounts for the virtual disappearance of the α-rhythm during conscious tasks by proposing that it is a carrier, modulated by information being transmitted. In support of this hypothesis he quotes some results on reaction delay obtained by Lindsley, at UCLA., who considered that the observed delay was composed of a constant delay of .1 sec., due to cerebral processes, plus a delay due to sampling at 10/sec., plus a delay due to muscles only accepting messages every .1 sec.

Muscle Control

The mechanism of voluntary muscle movement is not as simple as might appear. Demands are passed from the cortex to the cerebellum, which is the organ responsible for muscular co-ordination. Messages then pass via the brain stem to the spinal cord. It appears that there are two routes available for muscle control, since each muscle is innervated by both α- and γ-fibres. The α-fibre route involves a subsidiary muscle-spinal cord feedback loop. Signals pass via α-fibres from muscle spindles to synapses at the spinal roots (which also receive the efferent α-fibres from the brain stem), whence further α-fibre efferents pass directly to the muscle. The γ-fibre route involves direct innervation of the muscle spindles, and appears to be in the nature of a 'reference' setting, determining the length of the muscle for a stable posture. α- and γ-fibres appear to work together in the manner of a feedback plus feedforward system, with sudden, coarse movement being effected via the α-fibre route. However, the mechanisms of muscle control are not, as yet, completely understood (32).

Experimental results show that the human subject is able to abolish the stretch reflex in response to instructions (24). (The stretch reflex is the build up of muscular force when a limb is stretched at a constant rate, after initially being at rest.) This reflex involves centres higher than the spinal roots, and, indeed, the cerebellum appears to be involved in its mediation (22). In addition to muscle spindles, there are tendon and joint receptors which are capable of influencing muscular contraction via this pathway. In view of the sophistication and complexity of the systems involved, it seems reasonable to postulate that the operator can effect a fine control of muscular movement, and can control its time structure quite closely. The consequences of this postulate are considered further in Chapter 9 .

2.4. Summary

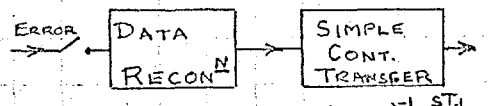
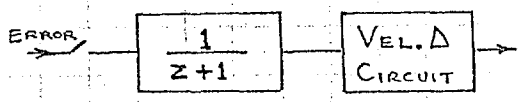
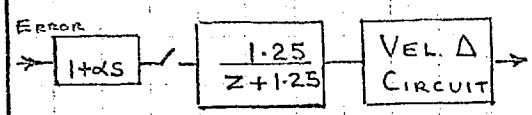
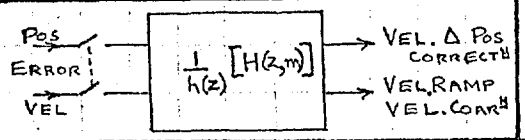
Much work has been devoted to describing human operator tracking characteristics in terms of linear continuous models. The general form of such models consists of a combination of pure delay, gain, 'neuromuscular' lag, and an equalisation consisting of lead and lag terms. These models are capable of generating outputs showing cross-correlations with operator outputs as high as .9 ,when both are fed with the same input. However,cross-correlation between errors tends to be much smaller.

The observation of the PRP led to the proposition and investigation of sampled data models with sampling intervals of .5 sec. However,the PRP observations appear to be very dependent on experimental conditions. Simple models involving sampling intervals of .3 to .5 sec. have also been fitted to operator spectral data,but such models required a small pure prediction to achieve the best fit.

A sampled data model proposed as a result of step response tests,with a sampling interval of .15 sec.,proved capable of achieving correlations with operator data fully comparable to those achieved by linear continuous models. Its simulation of the fine structure of operator output proved better than that of a continuous model,investigated under similar conditions of tracking. This model employed a regular sampling action,but careful analysis of step response data showed that a random sampling action was more likely. The above models are summarised in Table 2.4.I.

Considerations of physiology led to the postulation of a sampling action with an average interval of .1 sec.,corresponding to the α-rhythm,but experimental results indicated a somewhat lower frequency. Eye tracking results imply that both position and velocity information are directly available to the cerebral cortex. The sophistication of the muscular system appears to support the hypothesis that fine control of the time structure of muscle movement is possible.

TABLE 2.4.I. SUMMARY OF LINEAR MODELS OF THE OPERATOR

INVESTIGATOR	INPUTS USED	FORM OF MODEL PROPOSED	TYPE	METHOD OF FITTING	ACTUAL CORR^{N}. BETWEEN OUTPUTS
TUSTIN	SUM OF 3 SINUSOIDS	$\frac{k(1+sT_1)}{s} e^{-.3s}$	CONTINUOUS	SPECTRAL DOMAIN	N.A.
ELKIND	VARIOUSLY FILTERED SUMS OF SINUSOIDS	$\frac{k \cdot e^{-sT_d}}{(s+a)(s+b)}$	CONTINUOUS	SPECTRAL DOMAIN	N.A.
MC RUER & KRENDEL (REVIEWERS)	VARIOUS CONTINUOUS	$\frac{k(1+sT_L)}{(sT_1+1)(sT_N+1)}$	CONTINUOUS	SPECTRAL DOMAIN	UP TO .9 FOR SIMPLEST CONT^{L} ELEMENTS
HENDERSON	CONTINUOUS (COV^{CE} DATA TO MODEL)	$\frac{k(1+sT_1) \cdot e^{-sT_d}}{s(s^2+2\zeta\omega_0+\omega_0^2)}$	CONTINUOUS	COV. FUNCT^{NS} ϕ_{ee} AND ϕ_{eo}	N.A.
BEKEY	FILTERED SUMS OF 10 SINUSOIDS		SAMPLED SAMPLING INT^{L} .3 TO .5 SECS	SPECTRAL FUNCTIONS AND DIG. COMP.	N.A.
WILDE	RANDOM STEPS & CONTINUOUS RANDOM (FILT^{R})		SAMPLED SI ≐ 150 MS.	STUDY OF RECORDER TRACES	N.A.
LEMAY	RANDOM STEPS & CONTINUOUS RANDOM (FILT^{R})		SAMPLED SI ≐ 150 MS	RECORDER TRACES AND MATCHING	C. .85 TO .95
LANGE	RANDOM STEPS RAMPS & CONT. RANDOM (FILT^{R})		SAMPLED S.I.'S FROM 125 TO 170 MS.	RECORDER TRACES AND MATCHING (INC COV^{NCE})	C. .98-.99

Chapter 3 : PRELIMINARY INVESTIGATIONS

3.1. Formulation of the Tracking Task

The considerations outlined in Section 1.4. implied a restriction on the variety of tracking tasks which could be studied. It was decided to concentrate attention on a manual compensatory task configuration, for the following reasons:-

(a) The compensatory task admits of simpler characterisation than the pursuit task, because it involves only one input to the operator. It therefore appeared more likely that the compensatory task would lead to a satisfactory model.

(b) The manual compensatory task does represent the practical situation in many applications.

(c) A given task is usually performed at least as well with pursuit tracking as it is with compensatory tracking. The latter therefore represents a more severe limitation on human performance.

(d) A good deal of previous work has been directed towards the formulation of linear continuous models by fitting them to data obtained from operators performing manual compensatory tasks. In the present case no restriction was to be placed on the form of the model which was not consistent with the accuracy of representation required. Published data would therefore provide a ready source for comparison with results derived during the course of the present study.

It was decided that investigation of the operator's performance in relation to different configurations of controlled element dynamics would probably not be very fruitful. A great deal of work had already been done in this direction, and it required lengthy experiments and numbers of trained subjects. Also, it seemed more sensible to focus attention on the inherent characteristics and limitations of the operator's performance.

With the above factors in mind, the best course appeared to be the selection of a simple controlled element, consisting of a simple gain, free of dynamics. The essential features of results derived from experiments using such a feedback sequence would probably be applicable over a fairly wide range of other experimental situations involving simple controlled elements, possessing a reasonable degree of linearity.

The most natural choice of display-operator-control configuration was to have the operator seated with a control lever grasped in his right hand, and with both display and hand movement directed horizontally, and lying in vertical planes perpendicular to the operator's line of sight. Such a direct relationship between

hand movement and display movement greatly facilitated initial task familiarisation. The operator's control movements were effected mainly by the forearm and wrist, with the humerus acting as the axis of a pivoting motion.

The physical apparatus which constituted the tracking loop is described in Appendix I.

3.2. Selection of the Statistical Structure of a Continuous Input

The experimental investigations involved studying the operator's response to both continuous and discontinuous inputs, where the term 'continuous' is meant to imply continuity in at least the first and second derivatives. The aims of the study required that continuous input signals to the operator's control loop should, at least, be 'random appearing' ; i.e. the operator should not be able to detect any regular periodicities. In addition it was desirable that such signals should possess stationary statistics ; this would facilitate analysis and comparison of results.

It was evident that there were two main classes of signals which would satisfy the above conditions:-

(1) Truly random signals, which could be considered to result from a suitable filtering operation on effectively 'white' noise.

(2) 'Synthetic' random signals, most easily obtained by filtering a sum of sinewaves, preferably not simply related in frequency or phase.

Signals of class (2) were reasonably easy to generate, but their power spectral density functions were far from smooth ; theoretically such spectra consisted of a number of Dirac delta functions. A considerable number of sinusoids was required to give a good simulation of a truly random process in the frequency domain.

Signals of class (1) were not so easy to generate, but possessed smoothly varying power spectral density functions. Their difficulty of generation arose as a result of the very small filter bandwidth required for human operator studies - usually less than 1cps. This implied that direct filtering of a conventional random noise source would yield an unacceptably low output variance, even if one could be sure that the source possessed an acceptably flat spectrum at these low frequencies.

A further requirement was that the input should possess an approximately Gaussian probability density function. This type of density function occurs in the majority of practical situations, and has the added advantage of greatly facilitating statistical analysis.

Signals of class (1) exhibit a very close approximation to a Gaussian density function, when the filtering action is a linear operation. Departures from perfection occur because voltages generated

by practical apparatus must lie within an upper and a lower bound. Signals of class (2) possess a probability density function which approaches a Gaussian form as the number of independent components is increased, assuming linear filtering. With a reasonable number of components (10 or more) the approximation is quite good.

The final choice of the class of input signal was made as a result of the following reasoning. It was recognised that the analogue model might well involve a sampling operation ; in the case of an input signal of class (2), possessing pronounced spectral peaks, the effects of spectral folding would probably be difficult to interpret. Also, the construction of a synthetic random process required considerable care , if difficulties were to be avoided in analysis. As a final consideration, the predictive features of the operator's response were of particular interest. Now the synthetic random process was, in theory, perfectly predictable. Although such a process is random appearing, it was difficult to substantiate the 'a priori' assertion that the operator would remain completely unaware of its inherent predictability, and that his response would not be affected. All these factors indicated that an input signal of class (1) would be most appropriate.

The choice of the statistical characteristics of the input was made as a consequence of the following desiderata:-

(a) Since it was desired to study the operator's predictive ability the form of the input should be such that he could improve his tracking performance substantially by means of prediction.

(b) The input spectral density should be fairly flat up to some cut-off frequency, in order to ensure a high degree of randomness of the time domain structure. This would ensure the best possible reliability and general applicability of statistical results.

(c) The best way of selecting the cut-off frequency seemed to be by experimental determination. The task was required to present a considerable challenge to the operator, but he should be able to achieve good tracking performance by dint of training.

(d) The rate of cut-off at the high frequency end of the spectrum was required to be sufficient to satisfy (a).

(e) It was not considered desirable to attenuate higher frequencies too rapidly, because of the importance such terms might assume if the representation of the operator's characteristics required a sampling operation. Moreover, it seemed reasonable to assume that such terms would be present in the more difficult of practical tracking tasks.

(f) It was necessary to attenuate at very low frequencies in order to minimise the effects of troublesome D.C. drift terms. The choice of suitable attenuation required some care, however,

because very low frequency and D.C. terms would probably be present in most practical situations. A further point to consider was that theoretical analysis would be facilitated if the input spectrum could be represented as being flat sown to D.C. It was thought satisfactory to attenuate terms below about .01 cps.. This frequency would be at least 20 or 30 times lower than any likely upper cut-off frequency.

The method of generation of the random signal corresponded to filtering of 'white' noise. It was therefore convenient to consider the statistical characteristics required of the random signal in terms of the appropriate form of the generator's filter. The simplest, and most convenient, way of achieving a reasonably smooth spectrum from D.C. to the cut-off frequency, was to cascade a number of identical exponential lags through suitable buffer amplifiers. The number of lags required was determined by a consideration of the transfer of an optimum linear least-square error (Wiener) filter, designed to predict the input signal in the absence of corrupting noise.

The method of deriving the optimum transfer for the 'noise-free' prediction by T_d secs. of a signal, whose power spectral density function is given by:-

$$\Phi_{rr}^{n}(s) = \frac{(-1)^n}{(s+a)^n.(s-a)^n} \quad ; \ (s = jw) \qquad 3.2.1.$$

is outlined in Appendix VII.

In order to gain an idea of the parameter values associated with filters corresponding to the best linear loop transfer which the operator could adopt for a particular 'n', rough numerical values were assigned to the prediction time, T_d, and filter break point, a, as follows. T_d was taken to be the prediction time necessary to offset the effect of the operator's inherent lag, due to the finite time required for perception, brain processing, and movement of the hand. In terms of step response, it corresponds to a value lying between the reaction time and the time required for the first passage through zero of the error function. A reasonable figure for T_d appeared to be around .25 to .3 secs. 'a' represented the cut-off frequency in rads/sec., of the filter which would present operators with a difficult, but feasible, task. For two lags or more, it seemed reasonable to select a value of 3 to 4, so that, approximately, $aT_d = 1$.

Substituting the numerical values given above into the theoretical optimum transfers, $G_n(s)$, gave :-

$$G_1(s) = e^{-1}$$

$$G_2(s) = e^{-1}.(2.0 + .30.s)$$

$$G_3(s) = e^{-1}.(2.5 + .60.s + .05.s^2)$$

$$G_4(s) = e^{-1}.(2.7 + .75.s + .09.s^2 + .0045.s^3) \qquad 3.2.2.$$

Inspection of the above transfers indicated that a filter composed of three cascaded, buffered lags represented a reasonable compromise between desiderata (a), (d), and (e). It was evident that a one-lag filter would give practically no scope to the operator's predictive abilities, while that afforded by a two-lag filter was somewhat restricted. A four-lag filter would give a comparatively sharp cut-off at high frequencies, without yielding any material advantages over the three lag case, since the possibility that the operator could develop a third order differentiation from a purely visual error display, seemed remote. These views were confirmed by considering the variance of the irreducible error of prediction as a percentage of the input variance ; the figures were 32% (n = 2) and 5% (n = 3).

The practical apparatus which was constructed to generate a random signal satisfying the requirements outlined above, is described in Appendix II.

3.3. Choice of Continuous Input Statistics and duration of Tracking Runs

There were three parameters of the experimental situation which it seemed best to choose by means of short preliminary investigations. These were:-

(a) The variance of the input signal.

(b) The frequency of the break point in the spectrum of the continuous input signal.

(c) The form and length of experimental tracking runs.

It was first necessary to choose an appropriate value of the gain from hand movement to display movement. The controlled element was a simple amplifier, free of dynamics. The obvious 'a priori' choice was a unity overall feedback (1 cm. of hand movement gave rise to 1 cm. of display movement in the same direction), since this corresponded to a situation to which the operator was well accustomed. The validity of this choice was confirmed by further experiments, described in Section 3.8.

The simple form of the controlled element implied a direct relationship between the variance of the input signal and the variance of the hand motion required for accurate tracking. The selection of an appropriate input variance was therefore a compromise. If the required hand motion were too small, operators found it difficult to respond accurately : Elkind (10) found that only a small increase in the relative power of the remnant was associated with input variances as low as .06 cm^2 - but his experimental set-up allowed tracking by movement of the fingers, whereas in the present

ease, tracking movements were made mainly by wrist and forearm. Conversely, too large an input variance would lead to premature operator fatigue. A value of 5 cm^2 seemed a reasonable 'a priori' choice, and this was confirmed by experiments which were simultaneously directed toward the selection of a suitable break-point frequency, as described below.

The random signal generator was initially constructed so that the filter break point could be set at five values in the range from .5 to .75 cps. The chief desiderata in the choice of the break point were that the input signal should present a challenging task, but one that allowed fairly accurate tracking (error variance less than about 25% of the input variance) after some training, and that the operator should be capable of maintaining this accurate tracking for a minimum period of about 100 secs., so that the results of individual tracking runs would show a reasonable degree of statistical reliability.

An input break point of .5 cps. was found to be too 'easy' for most operators ; there was tendency for the operator to become 'bored', and allow unnecessarily large errors to occur. A value of .75 cps. caused most operators great difficulty, and accurate tracking was possible for only a short time. It was later found that, after lengthy training some operators were capable of tracking an input possessing a break point at .69 cps. with reasonable precision. The best compromise to suit all operators seemed to be provided by a break point of .61 cps. Operators were capable of tracking this input for a period of about 3 minutes, after which fatigue was evident. The most troublesome effect of this fatigue was a 'tiredness' of the eyes (which themselves were required to perform a difficult tracking task) with a resulting increase in blink-rate and temporary loss of the displayed error signal. An experimental run length was therefore selected as representing a reasonable compromise.

It was necessary to allow operators a short period of tracking prior to the commencement of measurements, in order to avoid the effects of transients due to the initiation of tracking. The best procedure was found to be as follows. When the operator stated that he was ready, a 50 second countdown was commenced. At -20, -10, and -5 secs. the operator was warned, and at 0 secs. he was alerted by the disappearance of a small cue signal superimposed on the display; at the same time all computing circuits were activated. After 200 secs., during which silence was maintained, the computing circuits were switched into the 'hold' position, and the simultaneous reappearance of the cue signal informed the operator of the end of the run.

3.4. Selection of Discontinuous Random Inputs

It was required that the analogue model should be capable of displaying the chief characteristics of the operator's responses to discontinuous random inputs, i.e. to inputs with discontinuities in their time domain structure. Also, study of responses to such inputs was likely to be very useful in postulating the initial form of the model. Two examples of such signals were studied:-

(a) A random step function, consisting of constant segments separated by step discontinuities.

(b) A random ramp function, consisting of segments of relatively constant slope with 'corners' (points with a step discontinuity of slope) separating successive ramps.

Preliminary experiments with input (a) indicated that randomness, both in the amplitude of steps, and in their time of occurrence, was required - otherwise the operator's response tended to become precognitive. For instance, if operators were presented with steps of constant magnitude, they soon became aware of the fact. The occurrence of a step then triggered off a predetermined response - no process of prior estimation was required. This was deduced from the results of presenting a large number of steps of the same magnitude, interspersed with a few smaller steps ; the latter elicited responses which showed large overshoots, and the first part of the response was practically the same as that observed for normal sized steps.

Randomness of amplitude was achieved by forming a step function with a non-Gaussian, bimodal probability density function. A Gaussian distribution was not satisfactory because it led to a large number of undesirably small steps. Randomness of time of occurrence was achieved by forming the steps at the output of a sampler ; sampling times were determined by a pulse train with either a second, fourth, or eighth order Poisson interval distribution. The fourth Poisson was used for most experiments, because it represented a reasonable compromise between randomness and the desire to have relatively few intervals very much longer than the average.

The random ramp signal was derived from the random step signal by passing the latter through a quasi-integrator. A pure integration could not be used, because the resulting ramp process exhibited non-stationarity, with a variance increasing with time. The 'ramps' were in fact exponential decays, but with such a long time constant that the effect was hardly noticeable.

The apparatus constructed to generate the random step and ramp processes is described in Appendix III.

3.5. Probability Density Functions of Loop Variates

The results of many previous investigators, Elkind (10) in particular, indicated that the characteristics of human operators performing continuous tracking tasks could be approximated quite well by linear systems, except under conditions where the input was very difficult to follow, or where there were complicated controlled element dynamics. Now the continuous input signal possessed a Gaussian probability density function ; it therefore seemed reasonable to suppose that the error and output signals would also be Gaussian. This supposition was confirmed by the analysis of records taken during the course of experimental runs ; typical results are shown in Fig.3.5.1.

The step input was non-Gaussian, while the ramp input possessed a 'bell-shaped' probability density function, which was a reasonable approximation to a Gaussian curve ; however, its first derivative was definitely not Gaussian. The analysis of the operators' responses to these discontinuous inputs was conducted on the basis of comparison of time traces, so that difficulties of interpretation were largely circumvented.

3.6. Selection of Performance Criterion

The formulation of a performance criterion relating to the continuous input signal was guided by the following considerations:-

(a) The loop configuration was such that the three variates of chief interest were linearly dependent. It was therefore only necessary to consider any two of these quantities to describe performance adequately.

(b) The input is independent of the operator's tracking ability.

(c) The error provides the most sensitive indicator of tracking accuracy.

(d) Experiment showed that both input and error signals possessed Gaussian probability density functions.

(e) Population mean values of input and error signals were zero ; mean values calculated from reasonably long runs were close to zero.

(f) A priori, there was no reason to ascribe special significance to any particular part of an experimental run.

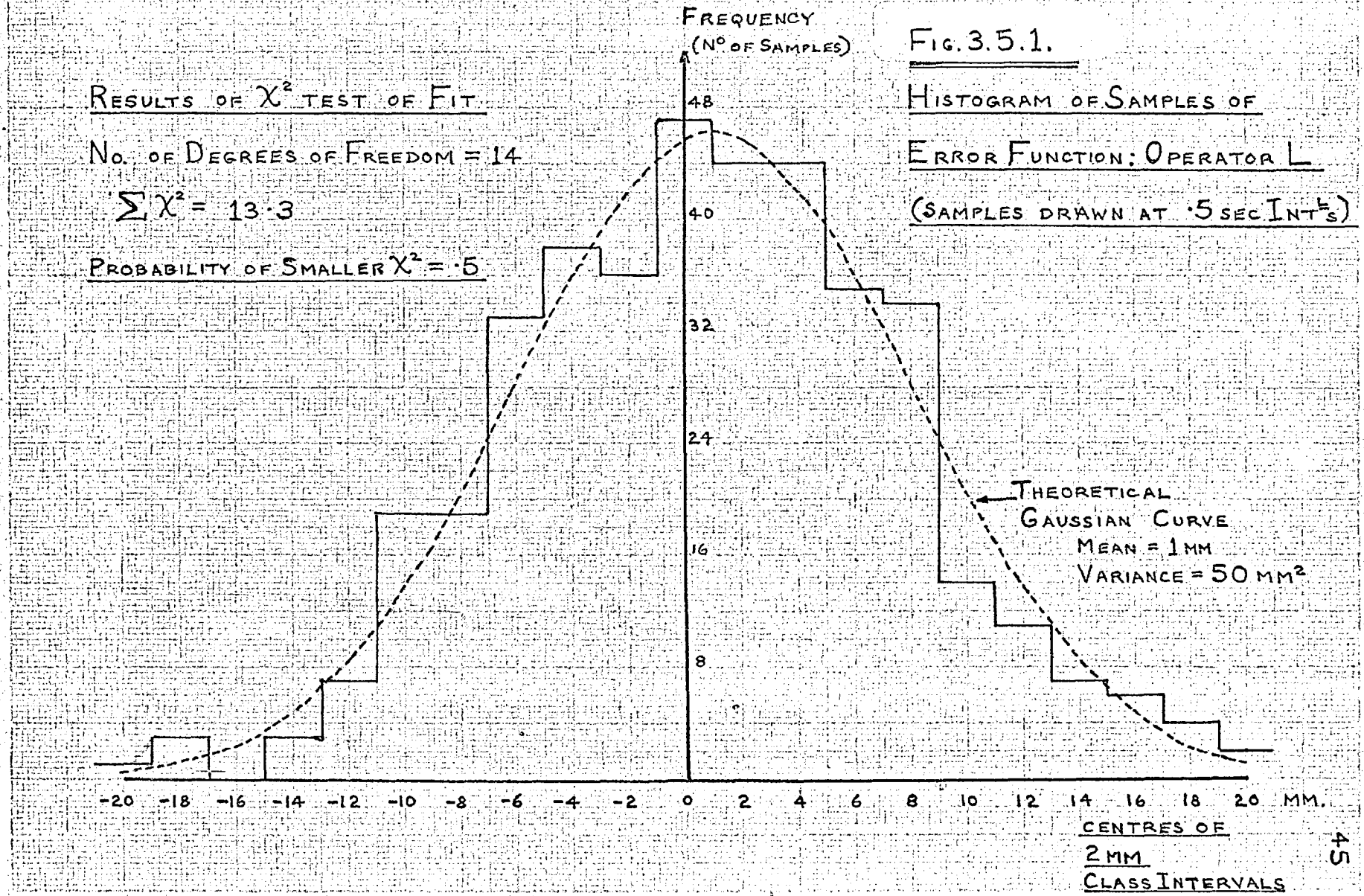


Fig. 3.5.1.

Histogram of Samples of Error Function: Operator L

(Samples drawn at ·5 sec Int^l_s)

(g) It was desired, if possible, to represent the operator in terms of a linear system. Fluctuations in, and errors of, estimation and execution would be represented as a noise voltage, additive to and not linearly correlated with output and error.

Considerations (a), (b), and (c) led immediately to a relation for the performance index, P, as follows:-

$$P = F(\text{input}, \text{error}, \text{time}) \qquad 3.6.1.$$

Consideration (f) led to the adoption of a constant weighting over time.

Now any Gaussian distribution is completely described by its mean and standard deviation, or variance.
The import of considerations (d) and (e) was, therefore, that a suitable relation for P could be written as follows:-

$$P = F(V_i, V_e) \qquad 3.6.2.$$

where V_i = variance of input
V_e = variance of error

Consideration (g) indicated that the error could be written as the sum of two components, viz.:-

Error linearly correlated with input, variance $= {}_iV_e$
Error due to operator remnant terms, variance $= {}_nV_e$

Thus, the total variance of error was given by:-

$$V_e = {}_iV_e + {}_nV_e \qquad 3.6.3.$$

Therefore consideration of variances was apparently advantageous.

It was desired that P should reflect the fact that the difficulty of effecting a given improvement in system performance is approximately inversely proportional to the difference between the theoretical maximum, and the level of performance already achieved. This implied that as $V_e \to 0$, $P \to \infty$. The simplest criterion satisfying the above requirements was given by:-

$$P = \frac{V_i}{V_e} \qquad 3.6.4.$$

This particular criterion also exhibited the useful property of being equally sensitive to changes in either component of error variance. An idea of the maximum value of the criterion index, P, which was likely to be observed in practice is provided by considering its value for the optimum filter described in Section 3.2. ; the variance of the irreducible error of prediction of this filter is about 5% of the input variance, giving P = 20.

The estimation of P required the measurement of the variance of input and of error . These variances were computed as described in Appendix IV.

Performance measurement in the case of discontinuous inputs was mainly required in order to determine when the operator had achieved a suitable level of proficiency. Chief interest was concentrated on the operator's response at points of discontinuity. The specification of a numerical performance index was difficult because these inputs possessed non-Gaussian probability density functions of amplitude or slope. A further difficulty was that it would be necessary to decide the form of an appropriate time-varying weighting function. It therefore seemed most reasonable to judge performance by visual inspection of oscilloscope traces and recordings. A check was provided by noting the subjective impressions of the operator.

3.7. Learning Phenomena and Training

The human operator exhibits a high degree of adaptability in the performance of tracking tasks. On being presented with a task to which he is completely unaccustomed, the performance of the operator is at first generally poor ; at this stage he has no knowledge of statistical dependencies present in the input signal, or of the result of a given hand movement in terms of controlled element output. Performance usually improves rapidly over the first few minutes, unless the task is a really difficult one. The observable effects of the processes of learning and adaptation arise from improved processing of sensory inputs and more accurate formulation of programmes of hand movement required to achieve desired responses at the output of the controlled element.

The temporal characteristics of improvement in performance due to learning and adaptation depend on the statistical structure of the input and the complexity of the controlled element dynamics. In general, the simpler the task, the more rapid the initial learning process. One useful result of utilising a simple controlled element in an experimental situation is that the time necessary for task familiarisation is thereby greatly reduced.

The above discussion indicates that operators must be trained before scoring runs are made - unless adaptive and learning behaviour is itself the subject of study. During the course of training the operator's performance improves, and his run-to-run variability decreases. Training may be terminated when performance has become reasonably stationary, but the interpretation of 'reasonably' will, in general, depend on the needs of the investigation and the complexity of the task. Very difficult tasks usually exhibit a slow, long-term improvement. Allowance must also be made for the possible occurrence of learning plateaux. These are evidenced by

a reasonably steady performance over quite a few successive runs, then a definite improvement over one or more successive runs,followed again by steady performance,but at a higher level.

Graphs of tracking performance over time are shown in Figs.3.7.1.& 3.7.2.,and illustrate experimental results of tracking runs carried out with three of the operators. Operator D had only two previous task familiarisation runs,with input spectrum break-point set at .5 cps. The graph shows that his performance increased from an initial value of 2 to about 6 ,during the course of 60 runs. There is a faint suggestion of a learning plateau between runs 10 and 24.

Operator F had about 10 previous runs in connection with tracking task formulation. His performance shows a moderate degree of variability,and fairly strong evidence of a learning plateau between runs 4 and 44.

Operator L had considerable tracking experience. His performance shows a fair degree of variability and a slow secular trend ; however,there is very little to indicate learning plateaux.

Run-to-run variability resultd from the combined effects of statistical fluctuations due to finite run length,and changes in the operators' health and emotional states. Also,it was difficult to eliminate outside disturbances.

The upward trend of the graphs for operators D and F was to be expected ; however,the presence of a slow secular trend in records for operator L serves to confirm the difficulty of the task. A minimum training period of 30 runs was therefore considered to be necessary.

3.8. Investigation of the effects of varying Hand-to-Display Gain

Preliminary experiments included an investigation into the effects of varying the hand motion-to-display motion gain on continuous tracking characteristics ; for this,one well trained operator was used. Throughout the course of subsequent investigations it was desired to utilise unity feedback gain,since this seemed most natural from the point of view of habituation. Any other gain would have altered the relation between visual acuity and potential accuracy of hand motion. It is well known (11,12,13,15,19) that the operator possesses the ability to vary his overall gain to adapt to changes in the controlled element,but it was desirable to obtain an idea of the extent to which his performance depended on the external loop gain,in the particular physical situation presented by the tracking task being investigated.

The experiment consisted of measuring performance with hand-to-display gains of 1,2,3,and 4 - after a short period of training in each case. Results are shown graphically in Fig.3.8.1. It may be seen that the variation of gain,within the above limits, showed that there was certainly no sharp optimum. A gain of 2 probably

FIG. 3.7.1. TRACKING PERFORMANCE VS. RUN Nº

OPERATORS D AND F

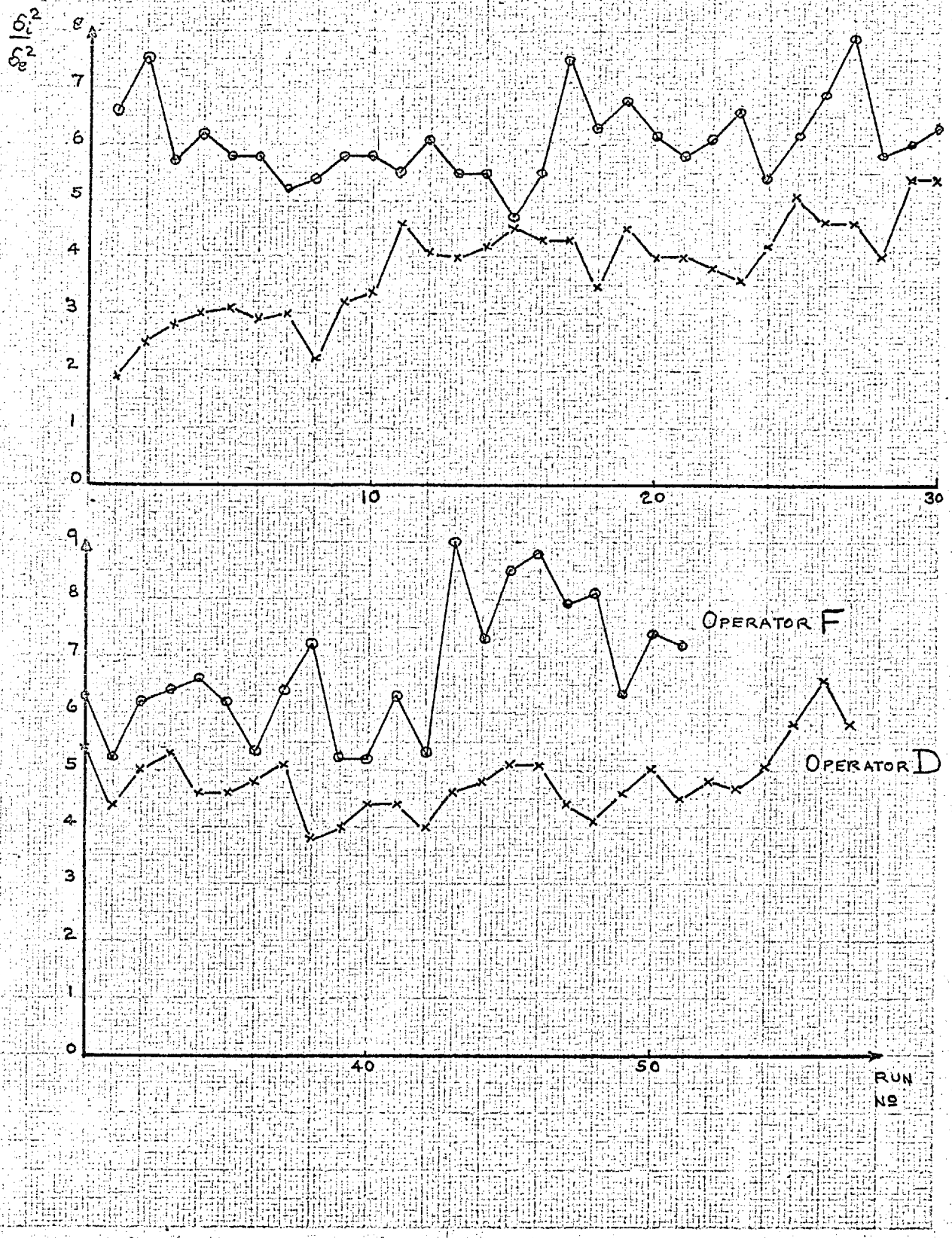


FIG. 3.7.2. TRACKING PERFORMANCE VS. RUN Nº

OPERATOR L

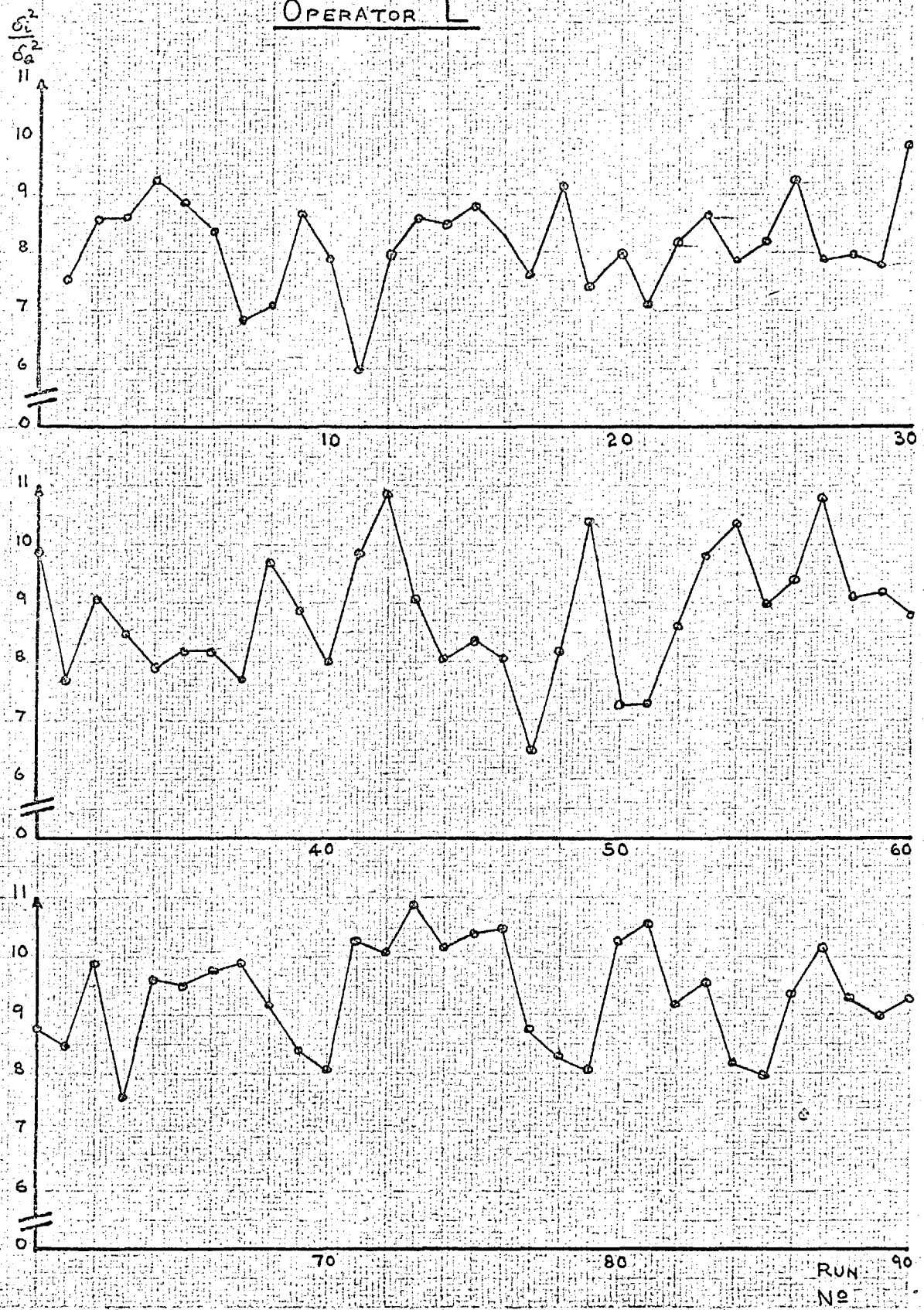
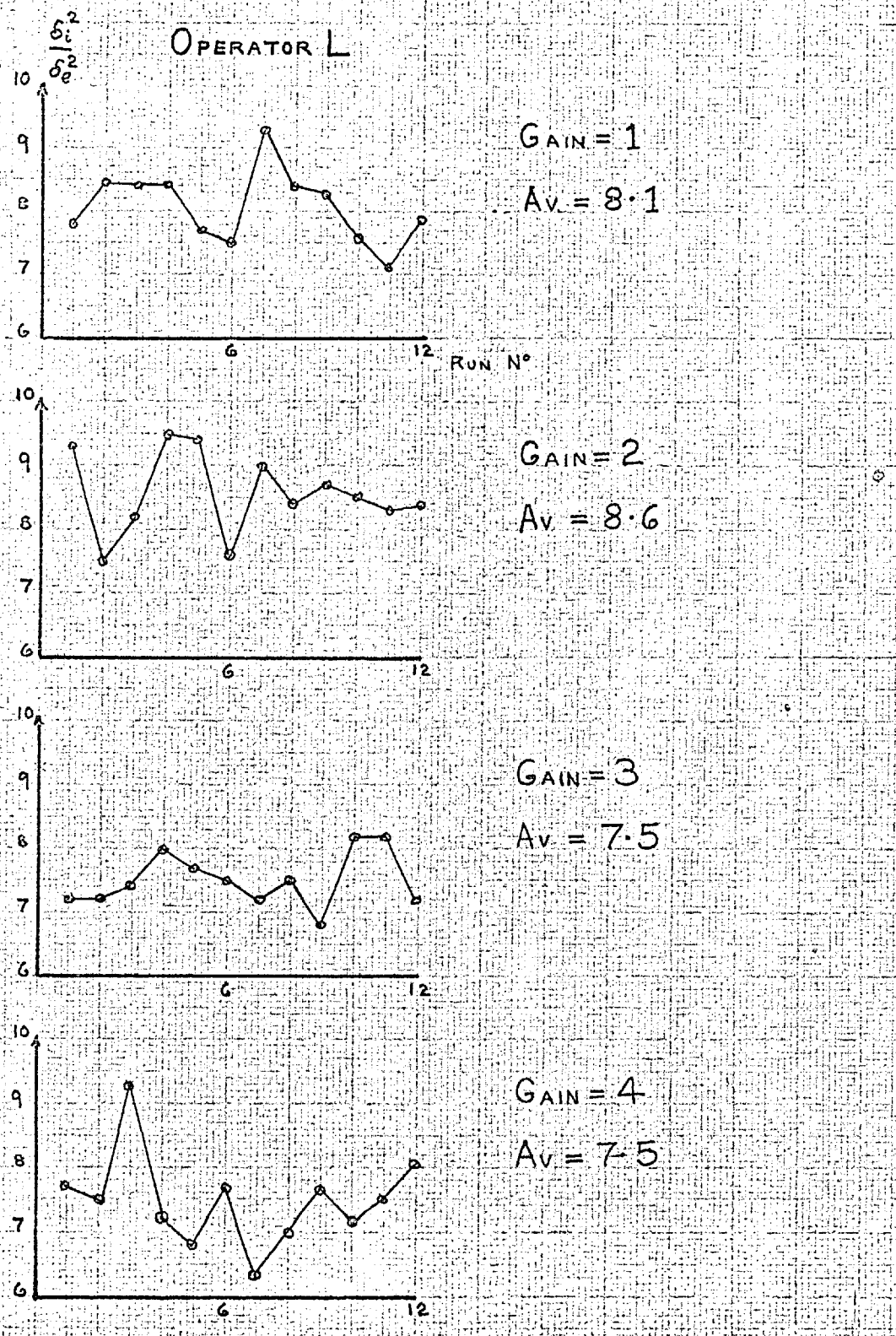


FIG. 3.8.1. TRACKING PERFORMANCE AS A FUNCTION OF HAND-TO-DISPLAY GAIN



represented about the best setting to obtain maximum performance, but the effect was marginal. This shallow maximum represented a pay-off between easier perception of small errors and increased difficulty of estimation, due to magnification of hand tremor effects, etc., as display gain was increased.

It was concluded that the choice of unity gain possessed advantages which outweighed the very slightly better performance obtainable with a gain of 2. The results also demonstrated that the operator was quite capable of varying his forward gain to maintain a performance level near to the limit permitted by the task configuration.

3.9. Investigation of the Effects of Display Nonlinearities

An investigation into the effects of nonlinearities in the error channel was conducted with two objects in view:-

(a) To investigate the effect of relatively mild nonlinearities on performance.

(b) To indicate the potential fruitfulness of further investigations into the effects of nonlinearities on the tracking characteristics of human operators.

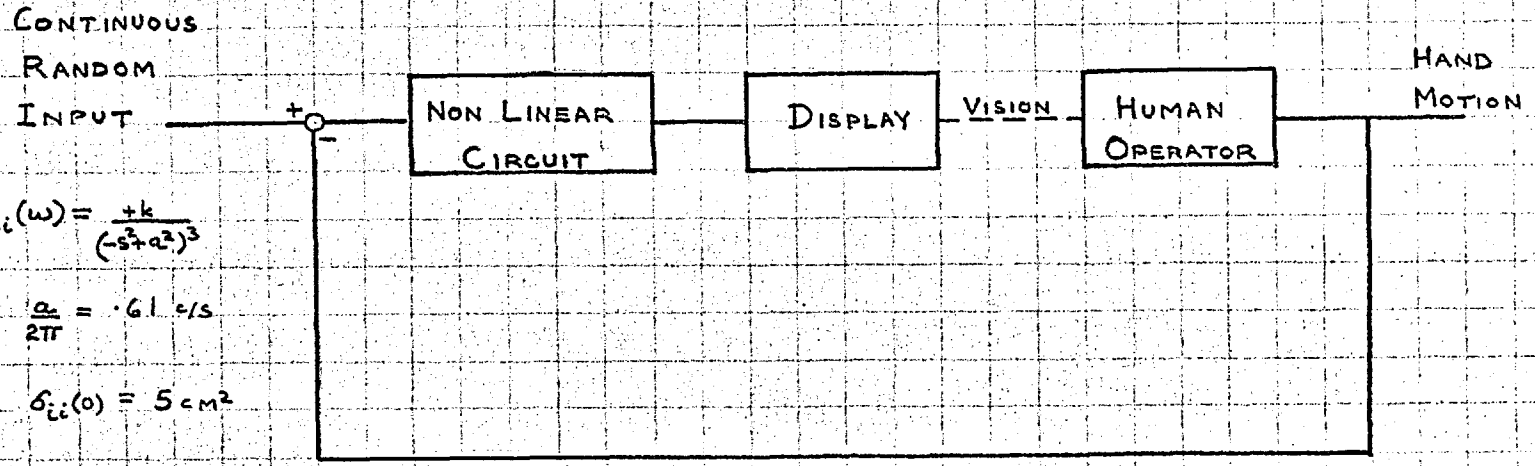
The form of the nonlinearities which were employed in the investigation was that of a cubic function approximated by three straight lines ; details are shown in Fig.3.9.1., which also shows the overall operator loop configuration. The break points at the junctions of the linear segments were placed so as to correspond with a displayed error of 1 cm., and true errors of 1 cm. and ½ cm. at the inputs of nonlinearities of type A and B respectively. An experimental check showed that these settings produced an effect which was quite noticeable on the subjective level.

Two training runs, followed by twelve scoring runs, each of 200 secs. duration, were made for each of the two forms of nonlinearity employed, using one experienced operator.

Numerical results are shown graphically in Fig.3.9.2. Their chief import was that the operator was well able to adapt his estimation procedure to allow for relatively mild forms of nonlinearity.

It was concluded that in order to be fruitful, further investigation must needs involve the study of a number of carefully selected, relatively severe nonlinearities, together with evaluation of subjects' capabilities in linear control loops. Such a series of experiments was considered to be beyond the scope of the present investigations, having regard to its likely demands in time, apparatus, numbers of subjects required, and data reduction facilities.

FIG. 3.9.1. SCHEMATIC DIAGRAM OF TRACKING LOOP WITH NON-LINEAR DISPLAY



$$\Phi_{ii}(\omega) = \frac{+k}{(-s^2+a^2)^3}$$

$$\frac{a}{2\pi} = .61 \text{ c/s}$$

$$\sigma_{ii}(0) = 5 \text{ cm}^2$$

TYPES OF DISPLAY NONLINEARITY

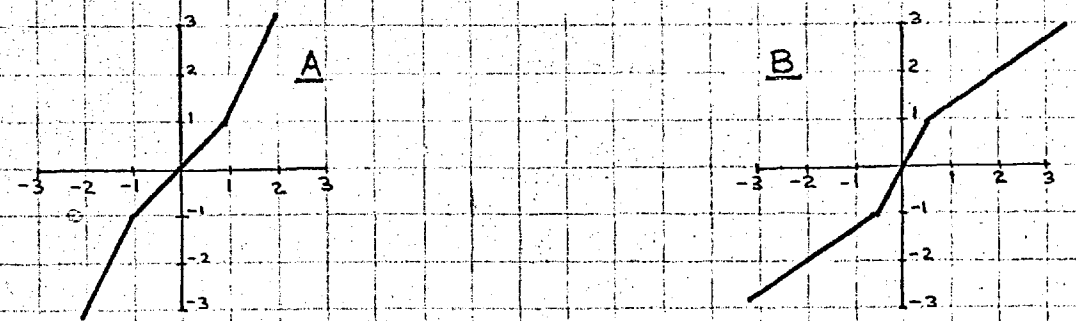
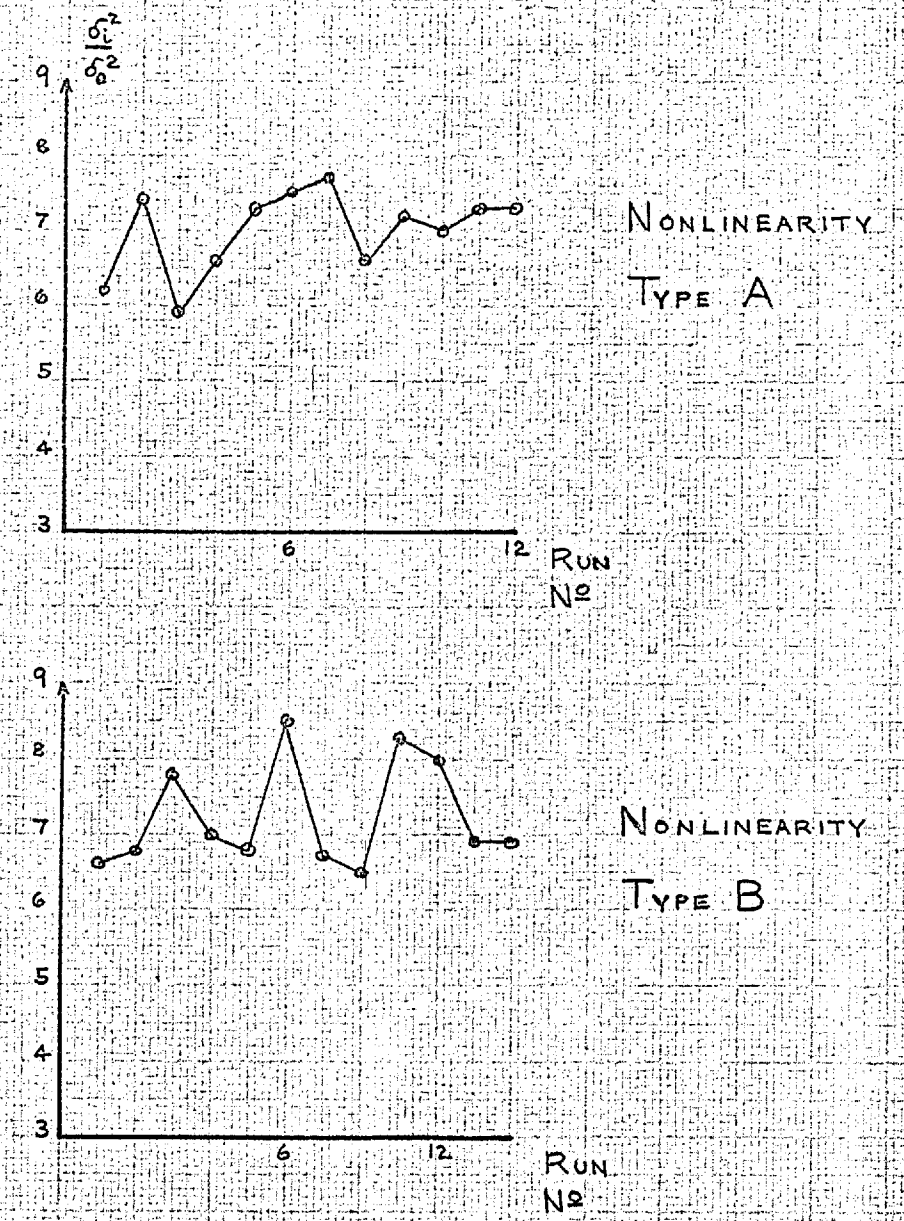


FIG. 3.9.2. TRACKING PERFORMANCE WITH TWO TYPES OF NONLINEAR DISPLAY



3.10. Summary of Results and Conclusions from Preliminary Investigations

Preliminary considerations indicated that the most fruitful area of investigation would be that of a compensatory tracking task, with a linear controlled element, free of dynamics. The most appropriate form of continuous input was a random signal, which was effectively derived by filtering a 'white' noise signal through three cascaded exponential lag circuits, separated by suitable buffer amplifiers.

Experiment indicated that the most suitable value of the input spectrum break-point frequency was .61 cps., and that an input variance implying a hand motion variance of 5 cm^2 for perfect tracking, was well suited to the experimental configuration. Under these conditions a run length of 200 secs. was found to represent a reasonable compromise between the necessity of avoiding undue operator fatigue, and the desire for statistical reliability of the results of individual tracking runs.

The Gaussian form of the probability density functions of the loop variates, verified experimentally, led to the adoption of a performance criterion equal to the ratio of the input variance to the error variance. A series of runs conducted with hand motion-to-display motion gains of 1, 2, 3, and 4, showed that, according to this criterion, there was very little practical advantage to be gained by adopting a value of gain other than unity.

Results indicated that a minimum tracking period of 30 runs was necessary for the establishment of a reasonably stable level of performance.

The study of responses indicated that the two forms of discontinuous inputs studied experimentally, viz. steps and ramps, must possess randomness of both amplitude and time structure. It was advantageous to employ a non-Gaussian probability density function of amplitude of steps, and of slopes of ramps. It was judged that the best way of estimating performance to determine when sufficient training had been given, was by visual inspection of records, and reference to operators' subjective impressions.

The results of a short investigation into the effects of nonlinear displays showed that the operator was well able to adjust his characteristics to maintain good performance in the presence of relatively mild nonlinearity. It was decided not to pursue a further course of such investigations, because of the inherent limitations of experimental facilities.

Chapter 4. EXPERIMENTS LEADING TO THE INITIAL FORMULATION OF A SAMPLED DATA MODEL

4.1. Specific Investigations Leading to the Initial Formulation of the Model

A number of investigations were conducted to elucidate the chief features of the operators' characteristics relating to a variety of tracking tasks additional to that of continuous tracking. These experiments were designed to explore the relative plausibility of the 'continuous' and 'sampling' hypotheses,and to indicate the likely structure of a model capable of displaying characteristics similar to those shown by operators engaged in tracking both continuous and discontinuous inputs.

There were three exploratory investigations:-

(a) Study of operators' characteristics under conditions of sampled display.

(b) Study of operators' responses to step inputs.

(c) Study of operators' responses to ramp inputs.

4.2. Experimental Investigation of the Effects of Sampled Error Display

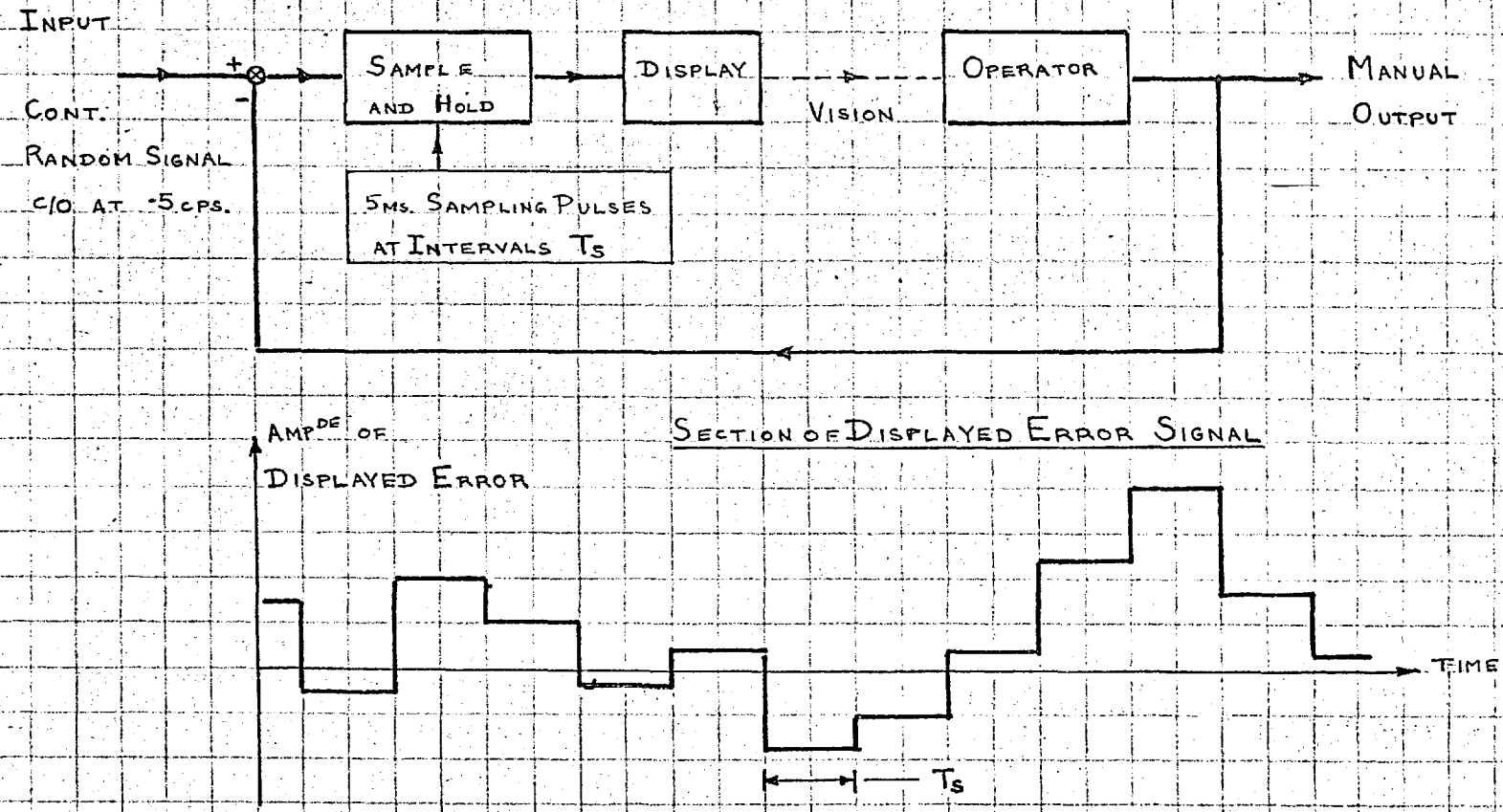
Experiments to investigate the effects of presenting the operator with a sampled error function,were carried out with the primary purpose of probing the validity of the hypothesis that the operator acts on his input in a temporally discrete fashion,rather than continuously. Experimental details were as follows:-

(a) The operator control loop was exactly the same as used in the normal tracking task (1:1 compensatory feedback),with the exception that the error channel was modified.

(b) The error signal presented to the operator was derived from the continuous error function present in the loop,by passing the latter through a sampler and zero order hold circuit,with unity D.C. transfer. The error function actually displayed to the operator therefore consisted of a series of steps,with discontinuities at sampling instants ; the sampling frequency could be varied at will. The experimental arrangement and sampled error function are illustrated in Fig.4.2.1. Circuit details are given in Appendix 1.

(c) The input to the operator control loop was a continuous random function of time,similar to that which was used for the ordinary tracking runs, but with a spectrum possessing a break-point at

Fig. 4.2.1. Arrangement for Sampled Error Display Tests and Displayed Error



.5 cps. This ensured that the tracking task would not be made too difficult as a result of the insertion of the sample and hold circuit.

(d) Four operators took part in the investigations; all had received a reasonable amount of training in the normal continuous tracking task.

(e) Tracking runs were conducted as in the normal tracking task with continuous input ; i.e. a 'warm-up' period followed by a 200 secs. duration 'scoring' period.

The aim of the experiments was to determine the relation between the operator's performance - both quantitative and qualitative - and the length of the error sampling interval. The following procedure was therefore adopted,and applied in the case of each operator. First,several runs were conducted under conditions of normal error presentation - i.e. no sampling. These served to familiarise the operator with the input,and provided a basis for comparison with the rest of the trials. The sampling interval was then set at 50 millisecs. and the operator was given several runs,until his performance ceased to show marked run-to-run improvement. Scoring runs were then made. This procedure was repeated with sampling intervals of 100,150,200,250, and 300 millisecs.,in the order given. Sampling intervals in excess of 300 millisecs were not used,since it became apparent during the course of the experiment,that the range of chief interest lay between 100 and 200 millisecs.

The criterion of performance was as used previously ; i.e. the ratio of the variance of the input to the variance of the 'true' error. This criterion was designated as P in Section 3.6. The variance of the displayed error signal was not exactly the same as that of the continuous error,because the values of the latter between sampling instants could not be considered to be drawn from precisely the same population as the values at sampling instants,due to the presence of the operator.

Curves showing the relation between P and the length of the error sampling interval,T_s,are given in Figs.4.2.2. and 4.2.3. Their general form is very similar for each operator. There is a fairly rapid fall in performance as sample interval is increased from zero to 150 millisecs.; performance then falls more slowly in the region $150ms. \leq T_s \leq 250ms.$,after which a more rapid fall is resumed. This effect may be seen quite clearly in the curves of log P vs, T_s,shown in Fig.4.2.4. These all show a definitely lower slope over a range of T_s from 150 to 250 millisecs.,indicating that a different power law is applicable in this region.

The curves of P vs.T_s could all be approximated by two straight lines, despite differences in the absolute values of P. The 'corner'between the fitted lines was found to occur in the vicinity of T_s = 150 millisecs.;this value of sampling interval also led to important subjective effects,as described below. The similarity of the forms of the curves for different operators served to

FIG. 4.2.2. CURVES OF P VS. T_S OPERATORS A AND C

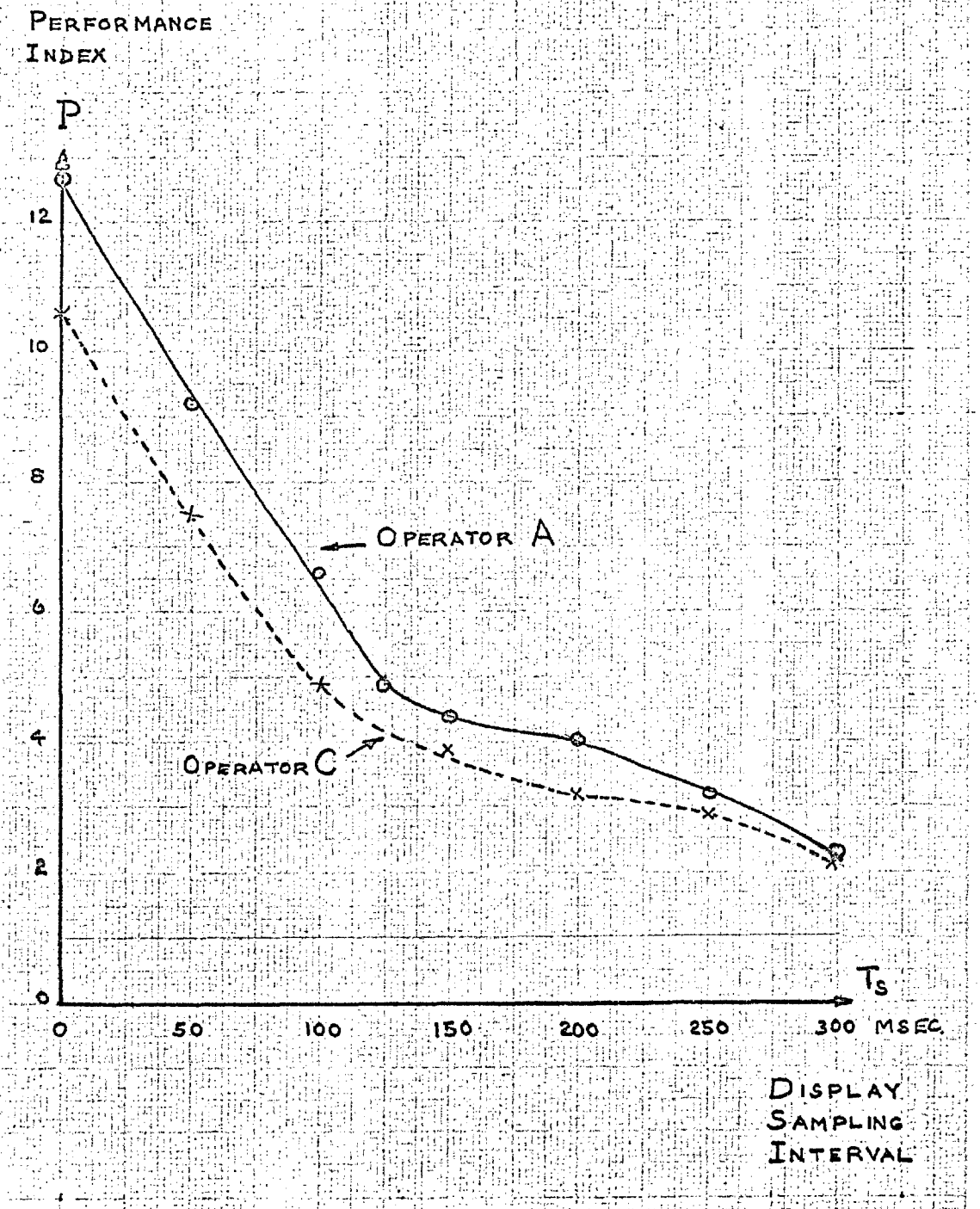


FIG. 4.2.3 CURVES OF P VS. T_S OPERATORS F AND L

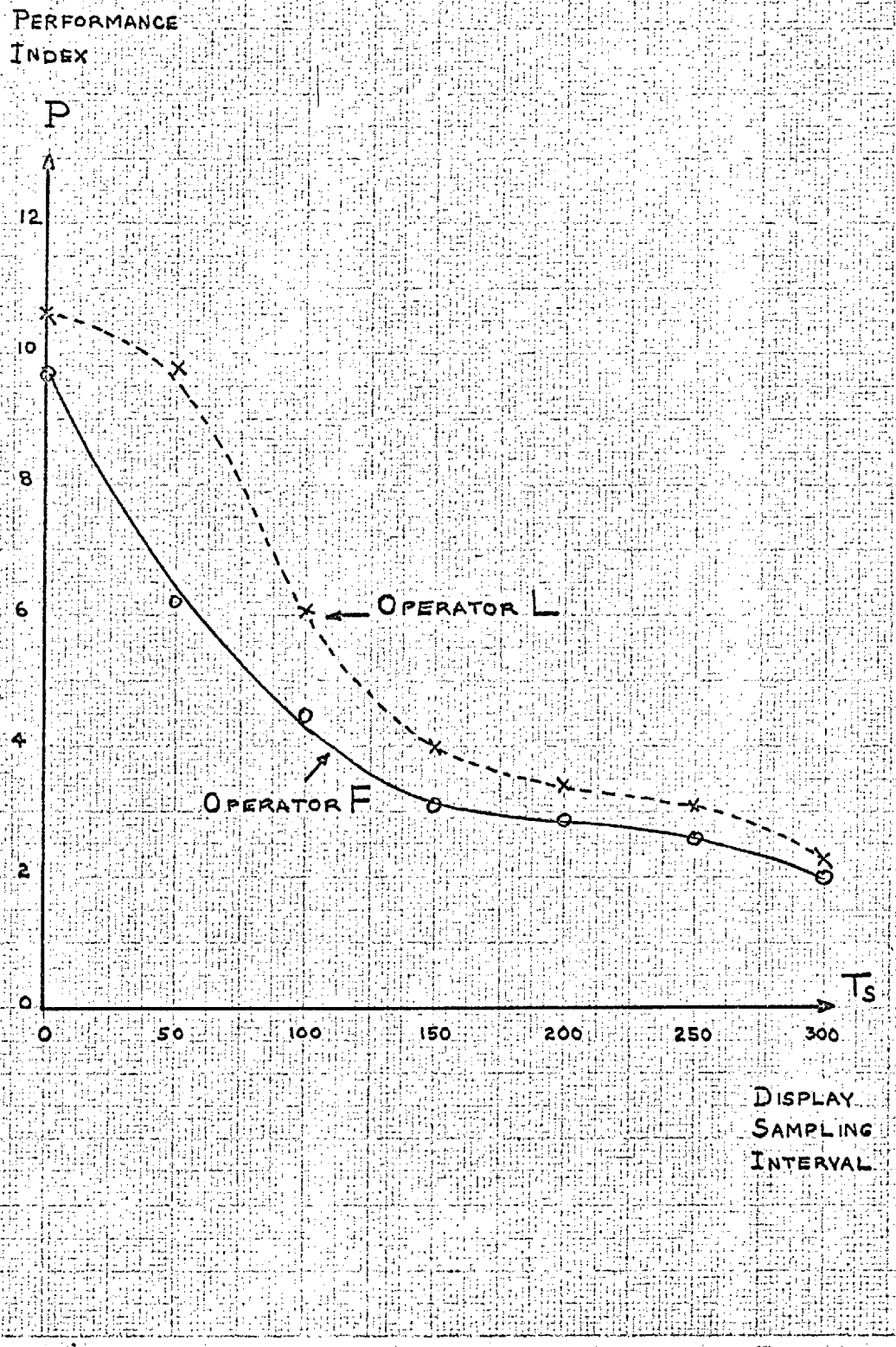
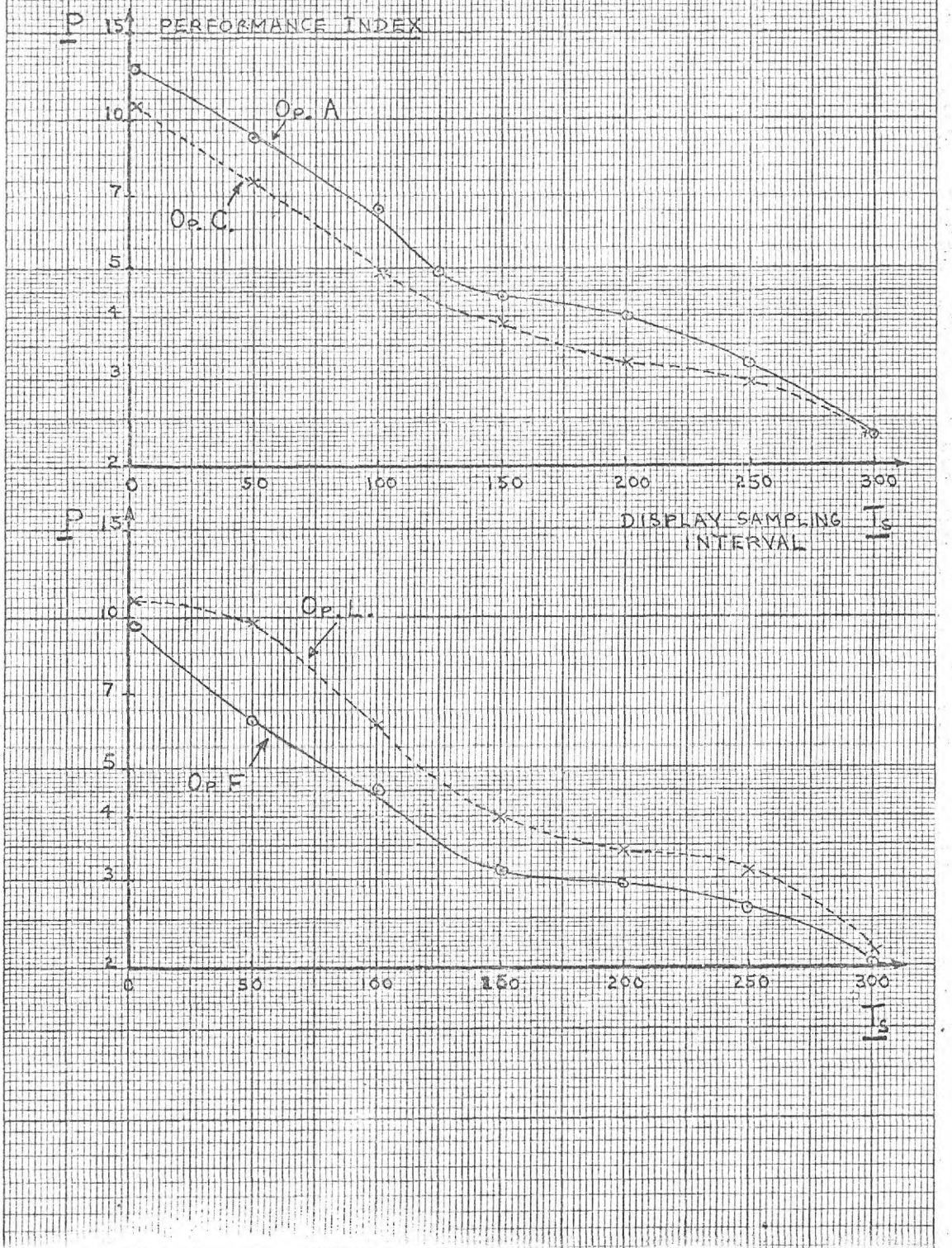


FIG. 4.2.4. CURVES OF LOG P VS. T_s.



justify the view that all the operators exhibited the same basic behaviour.

During the course of the initial investigation, operators reported that they seemed to possess the ability to 'synchronise' their actions with the displayed error function : the effect was most noticeable at T_s = 150 millisecs, but operator A found a definite effect at T_s = 100 ms., and operator F reported one at 200ms. Sampling intervals longer than 200 millisecs tended to lead to a subjective feeling of waiting for the next sample. When operating in the synchronised mode, the operator had the subjective impression that his hand was jumping about in unison with the display, and outside his direct, conscious control.

After concluding the initial series of experiments, further runs were made with various settings of T_s in the range 100 to 200ms., with the object of finding the sampling interval lengths, for each operator, at which the subjective effects of synchronisation were most noticeable. These values of T_s were found to be 125, 150, 150, 140 ms. for operators A, C, F, and L, respectively.

Further runs were conducted with sampling intervals set to these values, and recordings of sampled error and output velocity were taken, sections of which are shown in Figs.4.2.5. and 4.2.6. These show a 'velocity-triangle' type of response (vide 18) with varying degrees of emphasis, according to the operator. There is a tendency for the end-points of the triangles to occur in synchronism with the sampling points in the error curve. It may also be noted that only a few of the triangles begin or end at zero velocity.

It was evident that any hypothetical model of the operator's action must offer a satisfactory explanation of the salient features of the above results, viz.:-

The general form of the P vs. T_s curves.

The subjective effects of synchronisation.

The form of the operator's output motion.

The development of such a model is discussed in the following Sections.

4.3. Implications of the Results of Sampled Display Tests for the Hypothesis of Continuous Action

The hypothesis that the operator receives information and acts continuously in time, has been advanced by some workers (10). Others have contented themselves with describing the operator's action in terms of the best-fit quasi-linear transfer function. The forward transfer usually proposed has the general form:-

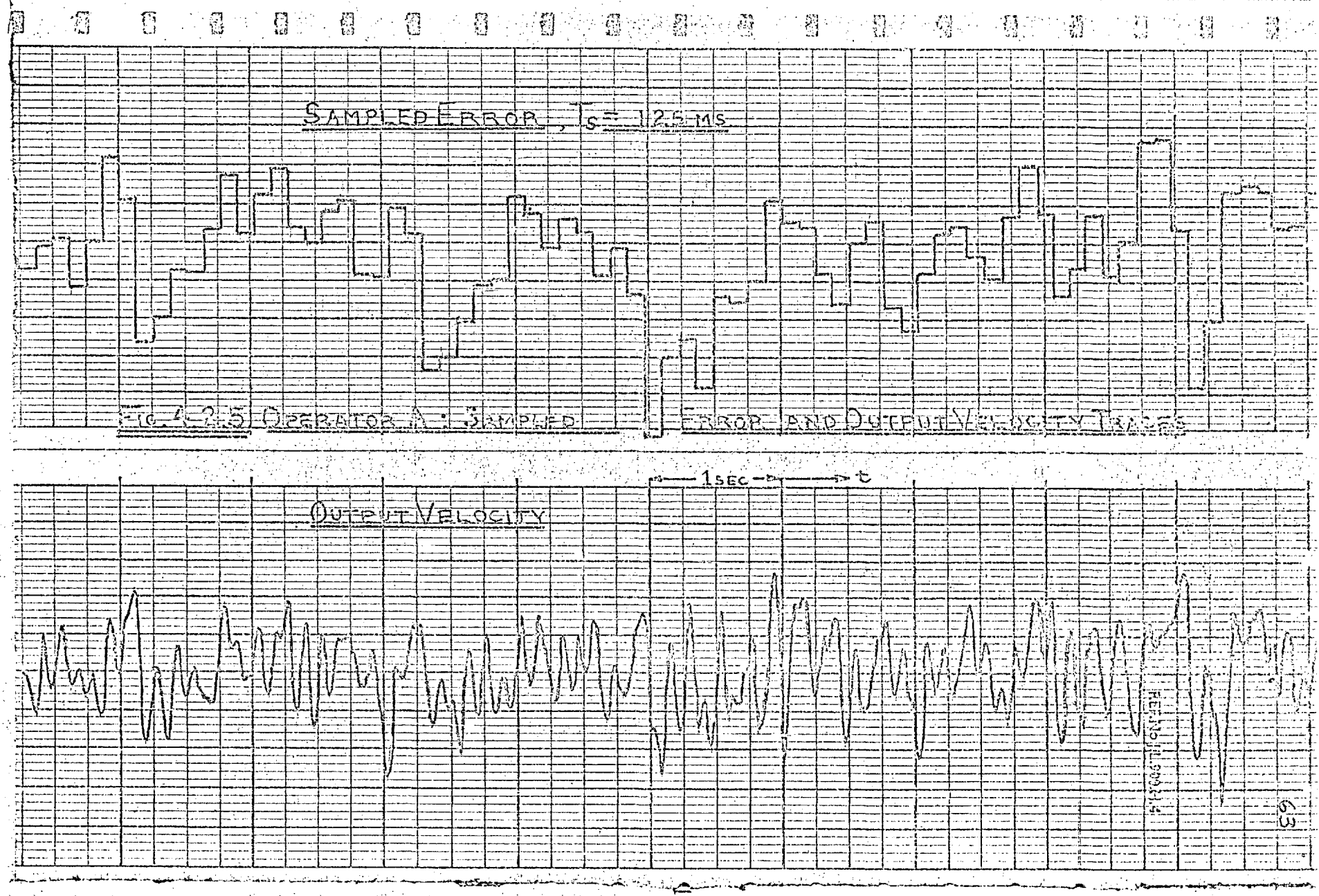


FIG. 4-2.5 OPERATOR A: SAMPLED ERROR AND OUTPUT VELOCITY TRACES

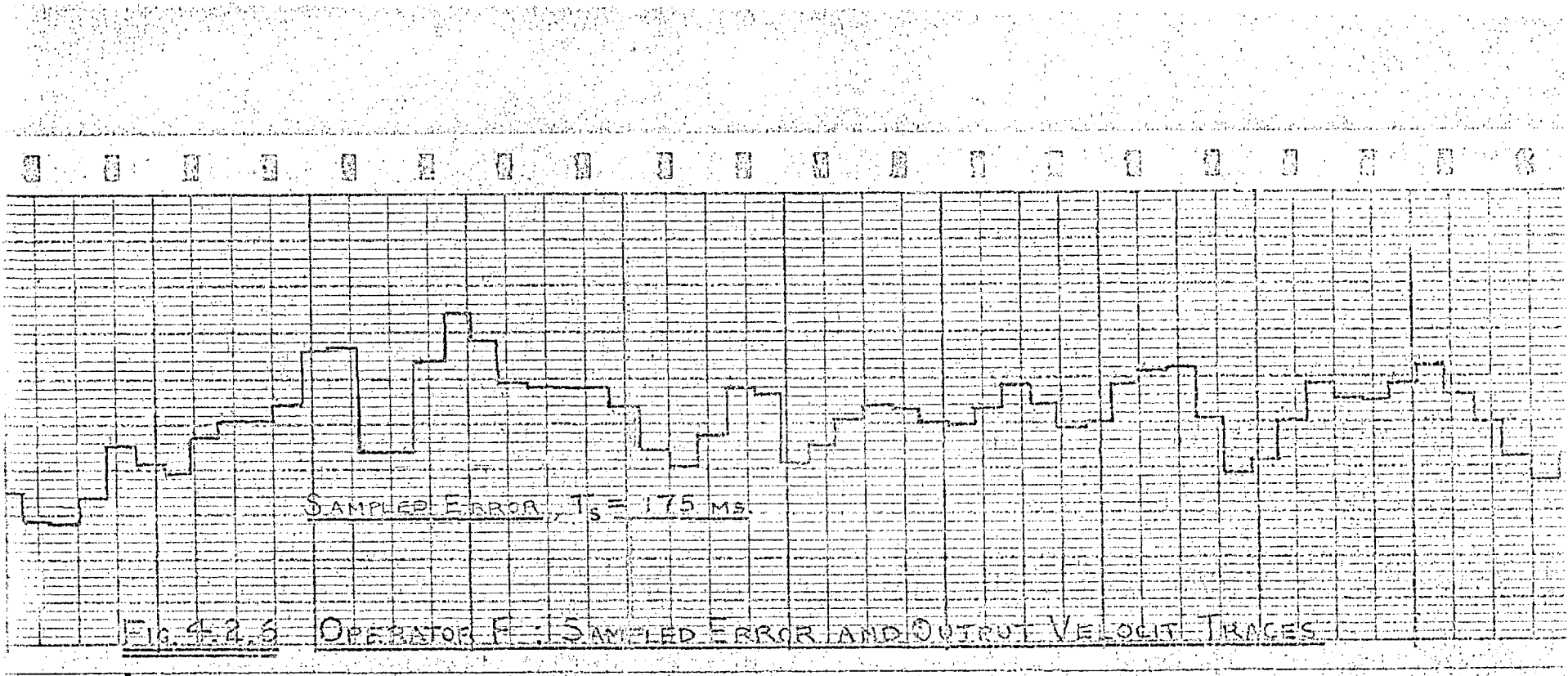
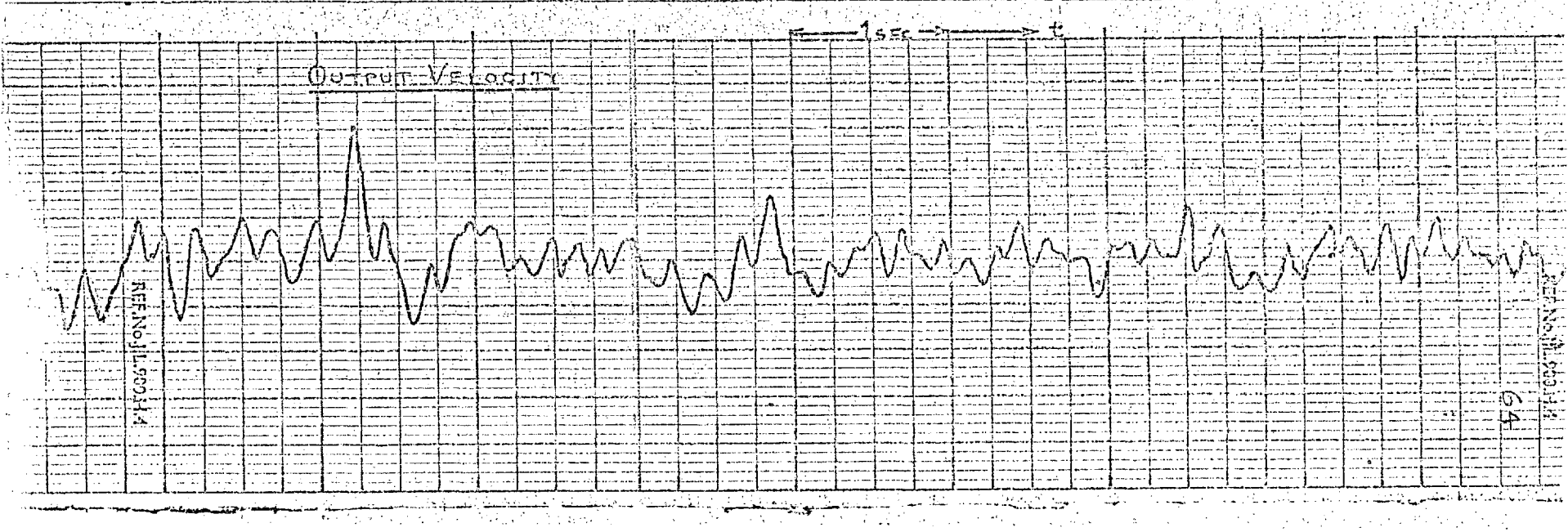


FIG. 4.2.5 OPERATOR F: SAMPLED ERROR AND OUTPUT VELOCITY TRACES



$$G(s) = \frac{k.(1 + s.T_L).e^{-s.T_d}}{(1 + s.T_1)(1 + s.T_N)} \qquad 4.3.1.$$

(see Chapter 2 and refs.11,12). The lead term is sometimes omitted, but this omission often leads to a rather low value of pure delay - usually much smaller then the delay observed in operators' step responses.

In terms of the continuous model,the experimental results may be evaluated as follows:-

(a) The action of the sample and hold circuit is to introduce a phase lag proportional to frequency. The equivalent delay is equal to $\frac{1}{2}T_s$. It also causes the effective gain to be periodic in the frequency domain,with zeros at multiples of the sampling frequency.

(b) The sampling process removes the operator's capacity for direct sensing of velocity. He is forced to estimate velocity on a differencing basis,with some smoothing and averaging to counteract the effects of sampling 'noise'.

The effect of (a) is to cause an additional delay of $\frac{1}{2}T_s$ to be added to the operator's inherent delay,and to modify the structure of the gain vs. frequency characteristics. In order to maintain performance,the operator needs to increase his reliance on prediction from velocity information. But,according to (b),the operator's capacity for sensing velocity is impaired by 'staleness' and 'noise' introduced by the differencing and filtering operations he is forced to perform. Both factors (a) and (b) would therefore result in a progressive impairment of performance as the error sampling interval is increased. This accords with experimental results.

There are,however,some points of inadequacy in the continuous hypothesis. Firstly,there is no 'a priori' reason to expect the synchronisation phenomenon. The subjective effects of synchronisation are inexplicable on the 'continuous observation and action' hypothesis. Secondly,the hypothesis offers no really satisfactory explanation for the shape of the P vs. T_s curves. One possibility is that both effects (a) and (b) are operative when T_s lies in the range 0 to 150ms.,while effect (b) is unimportant for greater values of T_s. However,one would expect a rather smooth transition.

A third point of inadequacy concerns the form of the output. The simplest continuous model proposed has a transfer similar to that of equation 4.3.1.,but with $T_L = T_1$. This transfer would lead to a series of exponentials at the output,each running for a time of one sampling interval. The more complicated model,with a lead and two lags,would give a similar response. The basic exponential form would not be greatly modified by the additional elements,which usually have fairly small time constants. The lead term would be

modified by a filtering operation,as outlined above ; its effect would be to alter the value of the output at any instant, but not to change its general form very much.

While one must allow for the fact that the continuous models so far proposed are idealised,so that one would not expect to see exact exponentials at the output,one might reasonably expect maxima and minima in the output acceleration to occur at times separated by roughly one sampling interval. The record of output velocity should therefore show a frequency of 'corners' rather lower than the error sampling frequency - occasionally one output response would lead directly into its successor. This behaviour was not noted experimentally, and acceleration peaks tended to occur at time intervals which were relatively constant,despite wide variations of T_s. Except in the case of the smallest T_s, the number of acceleration peaks was in excess of one per sampling interval. The form of the output observed experimentally is therefore not explicable in terms of the continuous models which have so far been proposed.

4.4. Implications of the Results of the Sampled Display Tests for the Sampling Hypothesis

Several investigators (e.g. Craik,6;Bekey,2;Raoult,26) have proposed that the operator may be regarded as perceiving and acting upon his input in a temporally discrete fashion,with an interval between successive samples of .3 to .5 secs. A typical model proposed to accord with this hypothesis would have the form of a sampler,followed a data hold in cascade with a linear continuous 'plant',which usually includes a pure delay (a perfect prediction has also been suggested (2),but,while allowable for the purposes of digital computation,this is of course,physically unrealisable). The hold may be omitted (26),or be of zero or first order (2). Higher orders of hold circuit are usually excluded because they incurr large penalties of phase lag. The sampler may be preceeded by a filter which includes a quasi-differentiation (practically perfect over the frequency range of interest) ; an example is the modified partial velocity hold suggested by Bekey(2).

The sampling models without the facility of a pre-filter differentiation fail to explain the experimental results. The main effect of sampling the error would be to add a phase lag proportional to sample interval,and the performance curve should show a smooth reduction,which should not be very severe - at least for values of T_s less than 200 ms.

If models with a differentiation facility are considered there is better agreement with experiment. One would expect a first rapid fall in performance as T_s is increased,followed by a more gradual rate of decline. However,the P vs.T_s curves would be expected to exhibit a relatively smooth form. Also, the model does not really

explain the synchronisation phenomena. Some effect might be expected to occur when T_s was exactly half the model sampling interval (for the models suggested, this would correspond to values of T_s in the range .15 to .25 secs.), but one would also expect some such effects to occur at other submultiples of the sampling frequency of the model. The subjective impression of 'waiting for the next sample', which occurs for T_s more than .2 sec., and which becomes quite strong as T_s approaches .3 sec., is not readily explicable in terms of the above model.

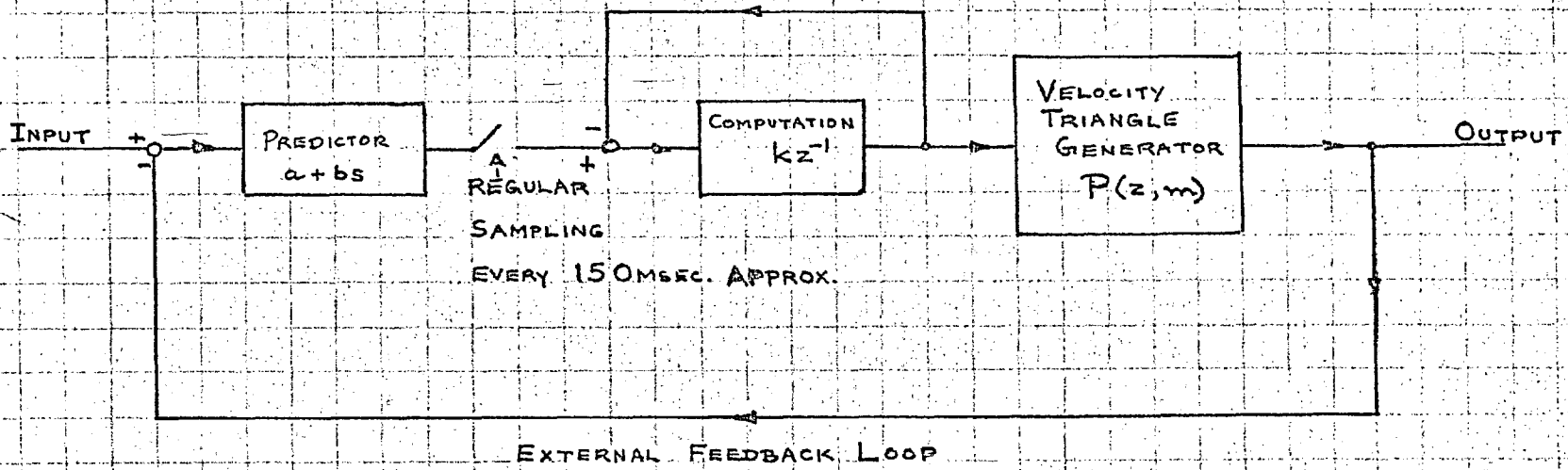
A further point which detracts from the acceptability of the above models is that they predict output motions in which peaks of acceleration occur about once every .3 to .5 sec.; even if one allows a velocity-triangle type of response, such peaks would only occur about once every .15 to .25 secs. These features of the model do not accord with experimental results.

One general point is worthy of further note. The above models all involve a value of pure delay smaller than the smallest pure delay observed in operators' responses to discontinuous inputs - some theoretical models actually involve a small prediction (2). They therefore constitute a rather abstract and restricted representation of operators' characteristics.

A sampling operation occurring at around 6 to 7 per second has been suggested, (18,4). A model first proposed by Wilde and developed by Lemay is based on the postulate that the operator samples a prefiltered error function every 150ms. or so, then forms his output in terms of velocity triangles which run between sampling instants. the structure of the model is shown in Fig.4.4.1. The prefilter contains either a pure gain (Wilde), or a gain plus a pure differentiation (Lemay) (actually a quasi-differentiation practically exact over the frequency range of interest). The Z-transfer contains a delay of one sampling instant, postulated to correspond with brain processing time, and has the form $k/(z + k)$. The value of k suggested for step inputs is approximately unity. Lemay proposes $k = 1.25$ for continuous tracking. The velocity triangle circuit generates a double-parabola position output, with zero slope at sampling instants.

The simpler form of the model (prefilter a pure gain) does not readily account for the rapid deterioration of performance shown by the first part of the P vs. T_s curve. The model possessing a prefilter containing a quasi-differentiation accounts quite well for this effect. Both models predict the synchronisation phenomenon at a value of T_s around .15 sec. One possible explanation of the effect is that the operator is able to complete programming of the next hand movement at an average time corresponding to the arrival of the next sample on the display. The subjective feeling of 'waiting for the next sample', which occurs when T_s exceeds .2 sec., is also well accounted for. The operator is able to complete his programme of response to the input well before the arrival of the next sample of error, resulting in the subjective feeling of a delay before the visual perception of the next step in the error function.

FIG. 4.4.1. EFFECTIVE STRUCTURE OF MODEL PROPOSED BY L.E MAY



$$P(z,m)\big|_{m=1} = k_1 \frac{T^2}{4} \frac{1}{z-1}$$

$$\frac{\partial}{\partial m} P(z,m)\Big|_{\substack{m=1 \\ \text{or } m=0}} = 0$$

The above reasoning also indicates a qualitative explanation of the form of the P vs. T_s curves. The initial steep fall in P occurs as direct information of error velocity is lost. For T_s more than 100 ms. the operator has to rely on differencing his input at his own sampling instants. For T_s more than about 150 ms., the operator is able to act so as to reduce the effective delay introduced by the sample and hold circuit in the error channel. P therefore falls less steeply in the range 150 to 250 ms. For T_s greater than this, there is an inevitable increase of error caused by loss of information due to the low display sampling frequency.

The output would be expected to exhibit acceleration peaks at a rate somewhat less than 2 per sampling interval, when synchronisation had occurred ; i.e. about one peak every 100 ms., on the average, allowing for occasional 'slurring over' of one velocity triangle into the next. The predicted form of the output is in general accord with experimental observation, with the important exception that the operator's output velocity does not pass through zero once per sampling interval, as may be seen from Figs.4.2.5. and 4.2.6.

4.5. Formulation of a Qualitative Model to Accord with Results of Sampled Display Experiments

The discussions presented in Sections 4.3. and 4.4. served to indicate the form of a hypothetical, qualitative model, which would accord with the experimental results of the sampled display investigations. The features of this proposed model were as follows:-

(a) There would be a sampling operation on error and derivative of error ; the average sampling interval would lie in the range of 125 to 175 ms. For convenience a regular sampling operation could be assumed, though the operator's sampling process would probably be random.

(b) There would be a computation delay to allow for brain processing time. It seemed logical to assume that this delay would be approximately equal to one sampling interval, or about 150 ms. This value would be in reasonable accord with the lowest reaction time delays observed in operators' responses to step inputs.

(c) The output would represent an idealisation of the operator's output, in terms of scalene velocity triangles, which could, as a further idealisation, be considered as a combination of an isosceles triangle and a ramp function.

The above model would account qualitatively for all the main features of experimental results. Further experiments, described in Sections 4.6. et seq., served to extend and confirm the above postulates as being useful idealisations leading to an accurate analogue model.

4.6. Experimental Study of Operators' Responses to Random Step and Ramp Inputs

The main purpose of the series of experiments involving discontinuous inputs was to derive qualitative information regarding operators' responses to step and ramp inputs, possessing random time and amplitude structure. Quantitative information was desired insofar as it aided in the formulation of a suitable model of the operators' actions.

The step tests were conducted with an input derived from the random step generator (Appendix III). There were two main lines of investigation, viz.:-

(a) It was desired to study operators' responses to fairly widely spaced step inputs, mainly from a qualitative viewpoint, and with especial reference to outputs' velocity characteristics. For the purposes of this experiment, the step generator was set to give steps with an approximate fourth Poisson interval distribution, and an average rate of about one step per 3 seconds. This arrangement resulted in reasonable freedom from very long or very short intervals but still led to an acceptable degree of randomness in times of occurrence. The amplitude probability density function of steps was bimodal.

(b) It was desired to probe the characteristics of operators' responses to closely spaced steps, with the hope that results would shed some light on the phenomenon of the 'psychological refractory period', about which there has been some conflict of opinion in the literature (Section 2.1.). Closely spaced steps would also provide a test for the sampling hypothesis. According to this, the minimum length of time between two successive reactions should correspond to the shortest possible sampling interval, whereas, on the 'continuous action' hypothesis, the length of time between two reactions should correspond to the interval between the steps. For the purposes of this experiment the step generator was set to give an approximate second Poisson distribution of intervals, with a mean rate of 1.4 steps per second. Again a bimodal amplitude probability density was employed.

The reason for utilising a non-Gaussian step height probability density function was that a Gaussian distribution entailed a preponderance of small steps, the response to which was difficult to evaluate. If the average interval were reduced to give a reasonable rate of occurrence of moderate to large steps, then the time function in between steps approached the appearance of a continuous signal. The small steps were therefore eliminated by using a nonlinear circuit, as described in Appendix III.

The step tests were carried out with a tracking loop configuration exactly the same as that used in the case of continuous inputs (1:1 compensatory feedback). Throughout all tests, the instructions given to operators were to respond as quickly as possible, with accuracy an important, but secondary, consideration. Recordings

were made after a suitable period of training, when it was judged that performance had become reasonably stable.

For the purposes of the random ramp tests, the input was formed by quasi-integrating the non-Gaussian random step function as used in step tests (a), for reasons given in Section 3.4. As in the case of the step tests, the ramp tests were carried out with a tracking loop configuration consisting of 1:1 compensatory feedback. Operators were instructed to respond so as to minimise their subjective estimate of error. The ramp input tracking task bore some resemblance to the continuous tracking task. The similarity was, however, superficial, since the random ramp process was not strictly differentiable, and possessed spectral characteristics quite different from those of the continuous input signal.

4.7. Quantitative Results of Step and Ramp Tests

The following quantities were measured from recordings of step and ramp responses:-

Reaction time .....designated..... X_1
Time from last input discontinuity ..... X_2
Magnitude of input discontinuity ..... X_3
% overshoot in position ..... X_4

where 1. Reaction time was taken as the time interval between an input discontinuity and the first change of velocity caused by the corrective response.
2. Magnitude of input discontinuity was taken as step height, or change in slope.
3. In the case of ramp inputs, the percentage overshoot in position was taken as the ratio of the - negative - height of the second extremum of the error function to the height of the first extremum, immediately following a change of slope.

The above quantities are illustrated in Fig.4.7.1.

The quantitative results of step test experiment (a) are shown in Table 4.7.I., and a typical scatter plot is shown in Fig.4.7.2. Linear multiple regression analysis failed to reveal any significant relationships (at the 5% level) between any of the variables.

Results of step test experiment (b) are shown in Fig.4.7.3. The effect of short intervals between steps was to slightly increase the second reaction time for step inputs separated by .3 to .45 secs. or so. Elsewhere there was no clear tendency, except perhaps, toward a slightly decreased reaction time for shorter intervals, down to about .15 sec. It was clear that operators could certainly respond to a second step presented within .2 sec. of the first, without a greatly

FIG. 4.7.1. ILLUSTRATION OF QUANTITATIVE MEASUREMENTS OF STEP AND RAMP RESPONSES

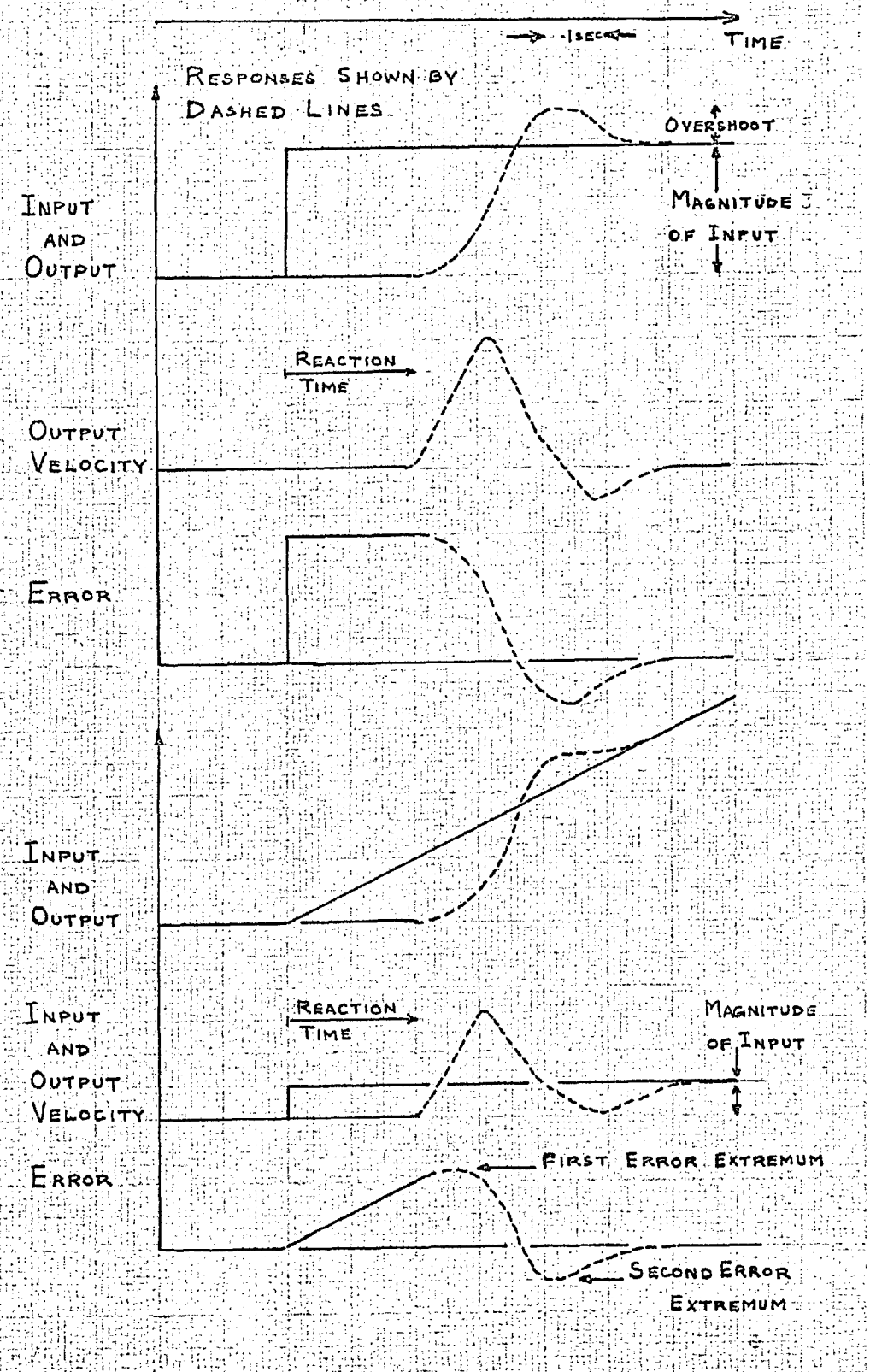


TABLE 4.7.I. QUANTITATIVE RESULTS OF STEP TESTS

QUANTITY		OP. A	OP. L
$b_{12.3}$	MS/SEC	6.5	4.1
$b_{13.2}$	MS/CM	-8.6	-7.1
$\bar{R}_{1.23}$	COEFF OF MULT. CORR^N	.29	.26
NO. OF SAMPLES		32	50
$b_{1.4}$	MS/100% OVERSHOOT	-20.3	-15.1
$\bar{R}_{1.4}$	CORR^N	-.21	-.19
$b_{4.3}$	% OVERSHOOT /CM	-5.1	-8.3
$\bar{R}_{4.3}$	CORR^N	-.21	-.26
AV. % OVERSHOOT		30	24
AV. REACTION TIME (MS)		245	235
AV. |DEVIATION|, DM, (MS)		34	31

FIG. 4.7.2. SCATTER PLOT; OVERSHOOT VS. MAGNITUDE OF STEP (OPL)

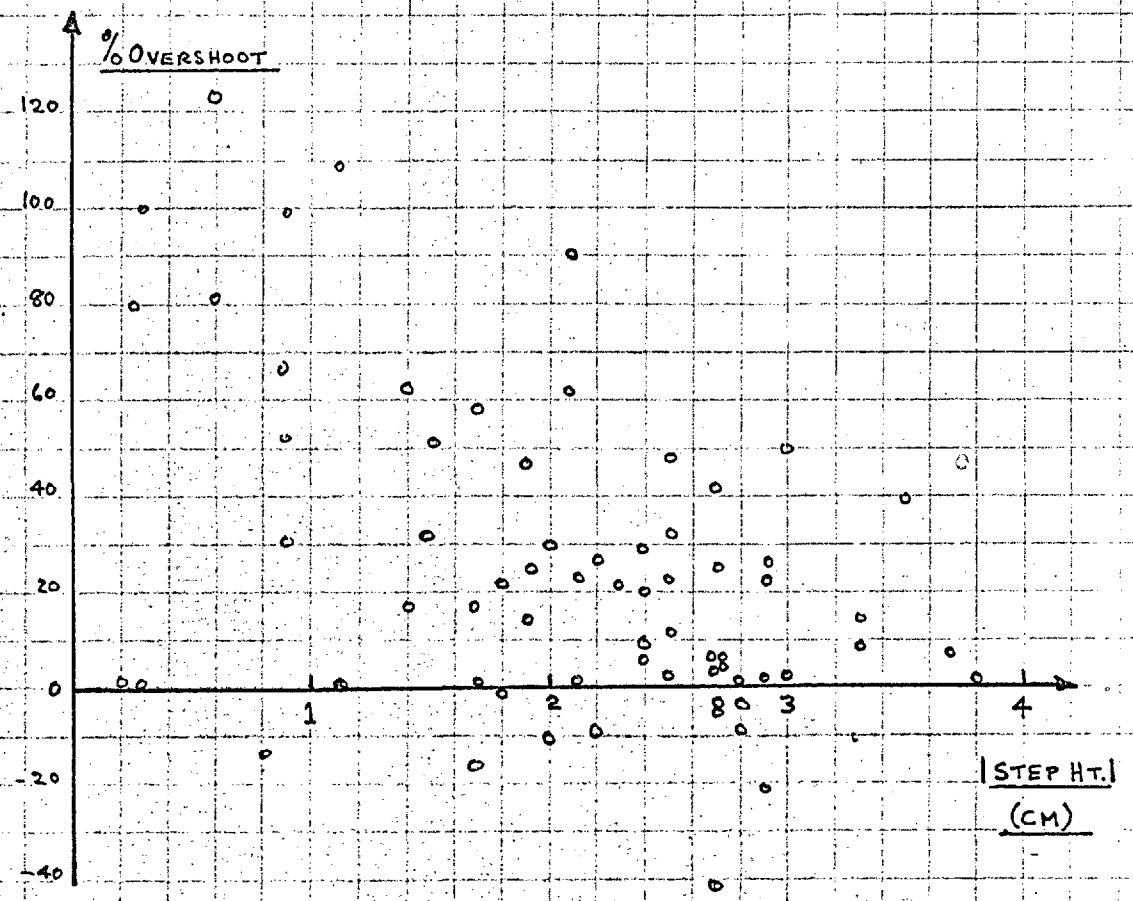
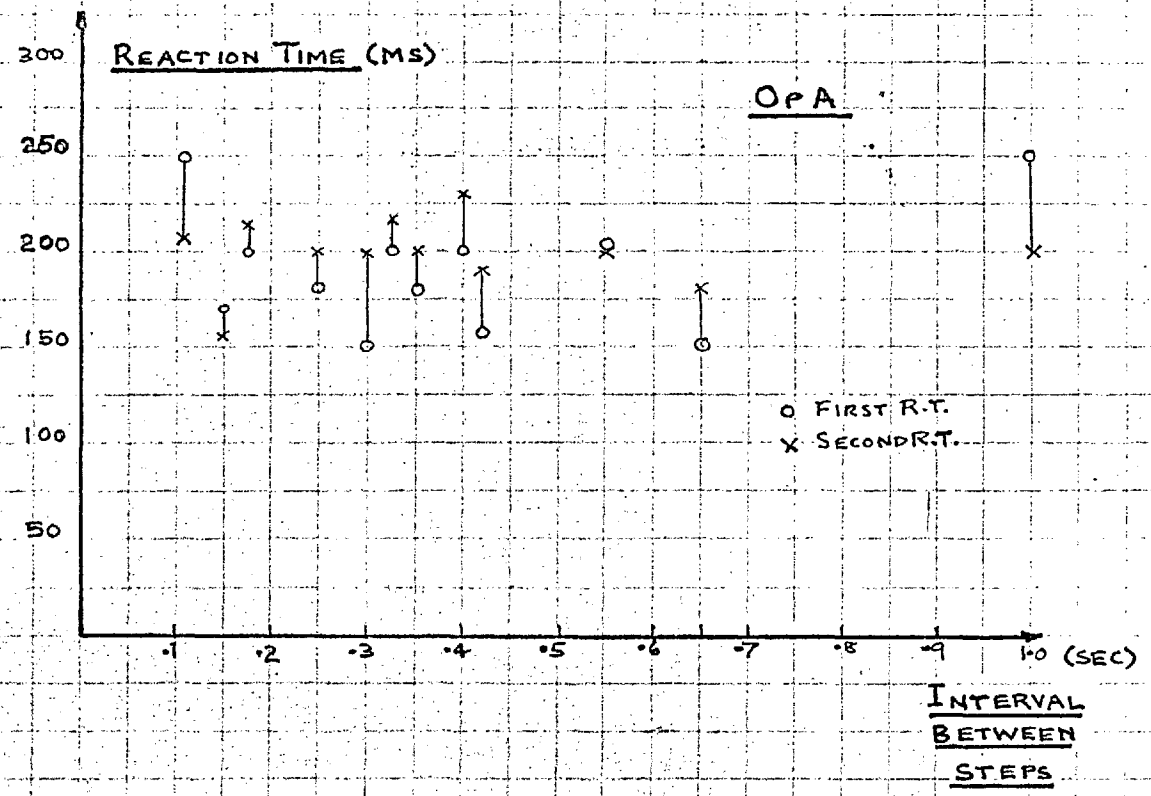
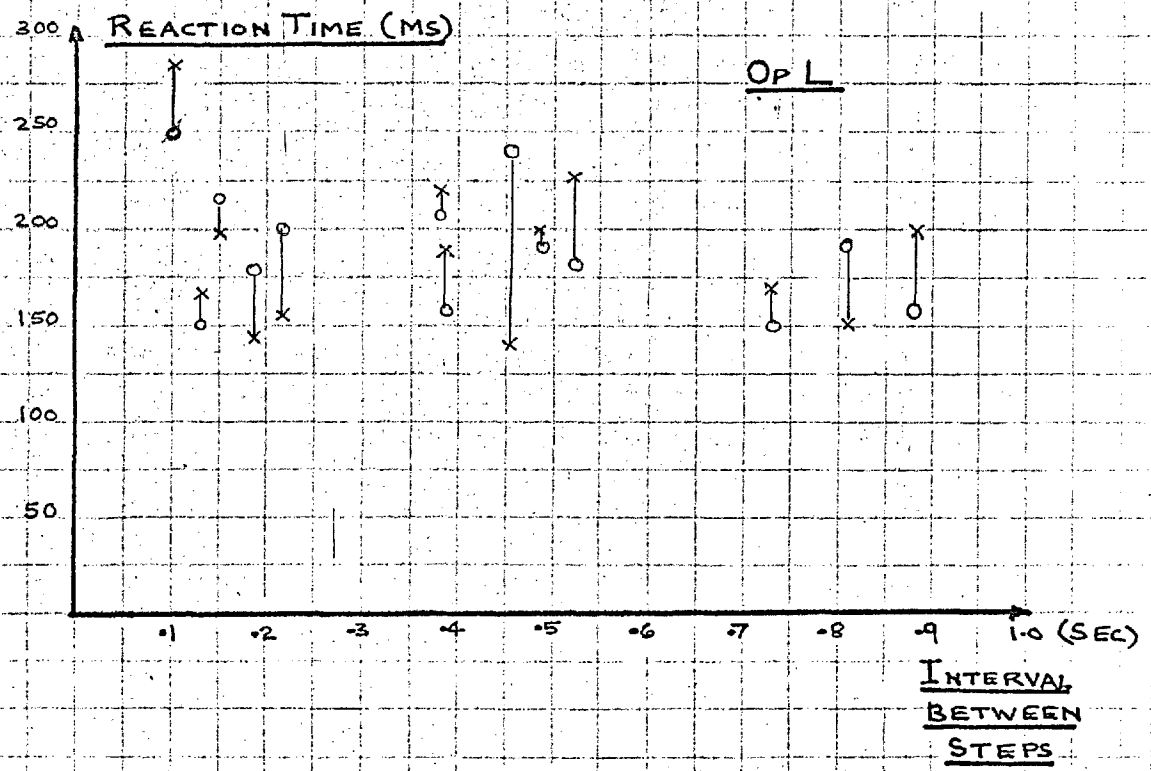


FIG. 4.7.3. SUCCESSIVE REACTION TIMES TO CLOSELY SPACED STEPS





increased second reaction time. The minimum time between successive responses to very closely spaced step inputs,was found to be around 100 ms. Some typical responses are shown in Fig.4.7.4.

The quantitative results of the random ramp experiments are shown in Table 4.7.II. Typical scatter plots are shown in Fig.4.7.5. Linear multiple regression analysis revealed no significant (at the 5% level) relations between the variables. Except in the case of reaction time vs. magnitude of change in slope,all relations were weak. Tests of significance were provided by considering the standard error of regression coefficients,and by comparing the coefficient of multiple correlation with confidence charts,as given by Ezekiel (27).

4.8. Conclusions from Quantitative Results of Step and Ramp Experiments

Neither step tests nor ramp tests revealed any significant relation between reaction time and overshoot. In the case of step inputs,it would appear that operators were following instructions,and responding 'as quickly as possible' - there was no trade-off between accuracy and reaction time,which one might otherwise expect. In the case of ramp inputs,the operator's error criterion was apparently so formulated as to lead to a similar policy of rapid response.

Partial regression coefficients indicated a tentative relation between overshoot and the magnitude of response called for, though below the level of significance. Care must be exercised in interpreting this result,because of the non-Gaussian distribution of input steps. It should be noted that the magnitude of each step was nearly independent of preceeding steps. The operator could therefore make no prediction regarding the size of the next step,except its average value. The requirement of reasonable accuracy of response constrained the operator to make an individual estimate for each step; apparently this estimate was only slightly biassed.

There was no evidence directly supporting the hypothesis of the 'psychological refractory period',as proposed in the literature. It was noted that the particular conditions of an experiment (the nature of the display,method of effecting output motion,etc.) could easily affect the result. Psychological variables must also be taken into account - in particular,the instructions given to operators,their motivation,and prior expectations,are most important. (Section 2.1.) Responses to very closely spaced steps were consistent with the hypothesis of sampling advanced in Section 4.5.The minimum time observed between successive reactions was somewhat more than 100 ms. This implied a lower limit to the length of a sampling interval of about 100 ms.

Reaction time data also have a bearing on the sampling hypothesis. If it is assumed ,as a first approximation, that sampling is regular in time,with an interval length T,and that total reaction

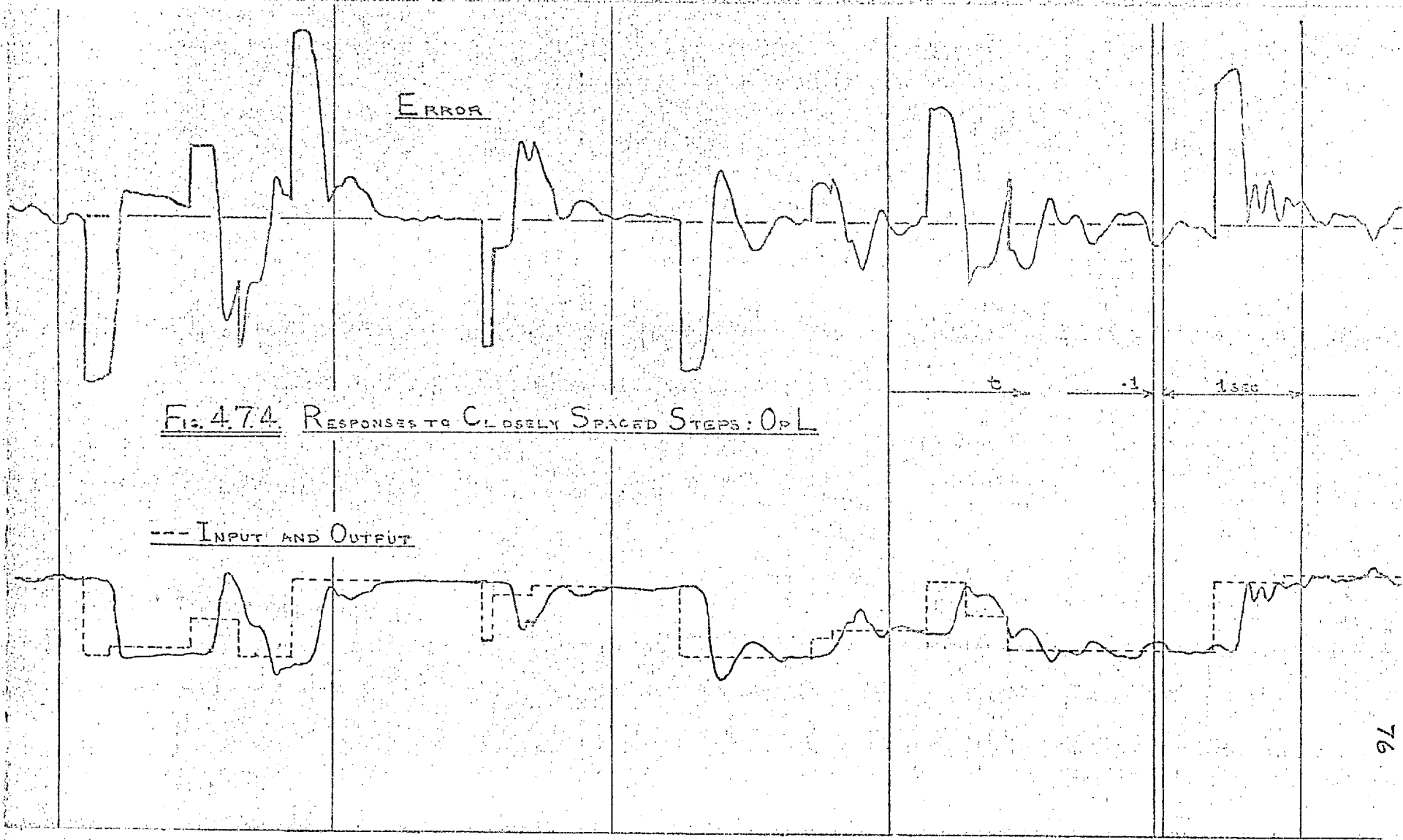


FIG. 4.7.4. RESPONSES TO CLOSELY SPACED STEPS: OPL

TABLE 4.7.II. QUANTITATIVE RESULTS OF RAMP TESTS

QUANTITY	A	F	L
$b_{12.3}$ ms/sec	5·3	-4·0	-7·3
STANDARD ERROR	11·6	6·7	7·7
$b_{13.2}$ ms/cm.sec^{-1}	-4·7	-4·14	-9·1
STANDARD ERROR	3·6	2·2	2·6
$\bar{R}_{1.23}$ COEFF OF MULT CORR^{N}	0	·25	·24
$R_{3.2}$ COEFF OF CORR^{N}	·09	·04	·1
NO. OF SAMPLES	46	43	49
AVERAGE REACTION TIME (MS)	246	256	213
AVERAGE |DEVIATION|, D_m, (MS)	36	37	27

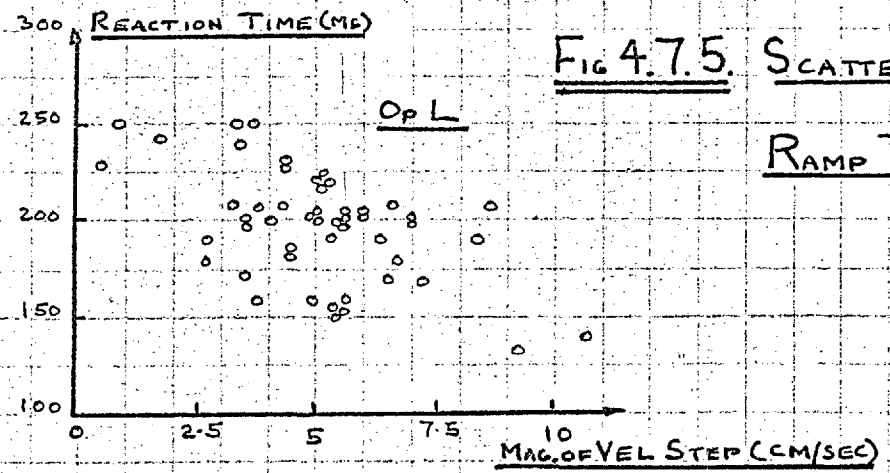
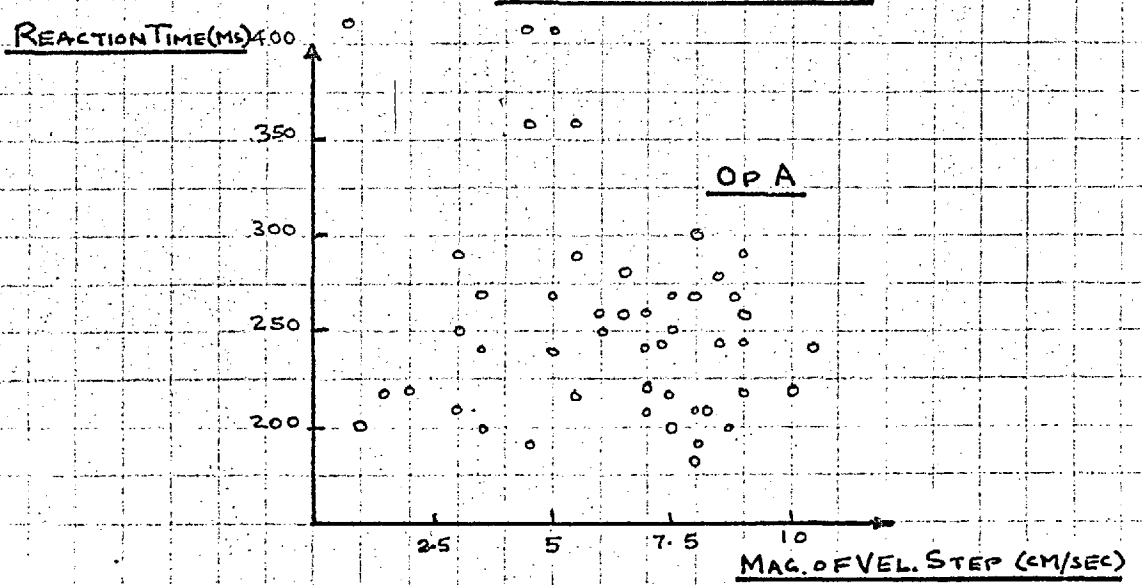


FIG 4.7.5. SCATTER PLOTS FROM RAMP TESTS



time delay, R_t, is composed of a delay of one sampling interval for computation , plus a random delay due to sampling,evenly distributed in the interval 0 to T,then the following relations may be derived:-

$$\bar{R}_t = E[R_t] = 1.5T \qquad 4.8.1.$$

$$E[R_t - \bar{R}_t] = .25T = D_m \text{ (say)} \qquad 4.8.2.$$

therefore

$$\bar{R}_t = 6.D_m \qquad 4.8.3.$$

Haynes (20),in a series of experiments on random steps of constant size,found that reaction times exhibited a somewhat skewed normal distribution,with which the results of the present investigation agree. This means that reaction times are somewhat more clustered than allowed for in the above reasoning,and so one would expect to find that experimental values of $\bar{R}_t/D_m$ were somewhat larger than 6. In fact, values found from experiment range from 6.9 to 8.3,with a mean around 7.5.

It was concluded that the hypothesis of sampling plus computation delay was consistent with experimental results. It was noted that the 'continuous action' hypothesis was not capable of any 'a priori' prediction of $\bar{R}_t/D_m$.

Linear multiple regression analysis of the ramp test data revealed a rather weak relation between reaction time and magnitude of velocity discontinuity of input. A representative value of the partial regression coefficient would be about -6 ms./cm./sec., corresponding to about 60 ms. less reaction time for the largest changes in velocity vis a vis the smallest. Although only a small degree of confidence could be placed in the validity of the relation, due to the low value of the unbiassed coefficient of multiple correlation,it was interesting to consider its implications.

The effect may be explicable in terms of a threshold in error perception ; that is ,the operator may judge small changes in error to be unimportant. The approximate value of the threshold level to accord with the results is about 1mm. or so,implying a zone 2mm. wide in which the error function fails to elicit a response. Static visual acuity would correspond with a zone width of not more than ½mm. The perceptual threshold effect has been postulated elsewhere,in both human tracking studies (11,12) and eye tracking studies (21).

An explanation is also feasible in terms of an extension of the sampling hypothesis. The supposition is made that the operator tends to shorten his sampling intervals when following a steeply sloping input ramp. Now the largest changes in velocity tend to occur at the ends of the most steeply sloping ramps ; the above hypothesis then leads directly to the dependency found experimentally. The operator would need to reduce his average sampling time by about 40 m.s. when following the steepest ramps,vis a vis the shallowest,

if the extended sampling hypothesis were to entirely account for the experimentally determined dependency. It was concluded that the observed relation may have been exaggerated by the somewhat non-Gaussian probability density functions possessed by the experimental variates, but that it could be accounted for quite satisfactorily by a combination of the effects discussed above.

4.9. Implications of the Quantitative Results of Step and Ramp Tests for the Proposed Model

The quantitative results of the experiments on the operators' responses to discontinuous inputs were in accord with the structure of the model proposed in Section 4.5. ; they were consistent with a sampling operation with a sample interval of average length about 150 ms.,plus a computation delay of one sample interval.

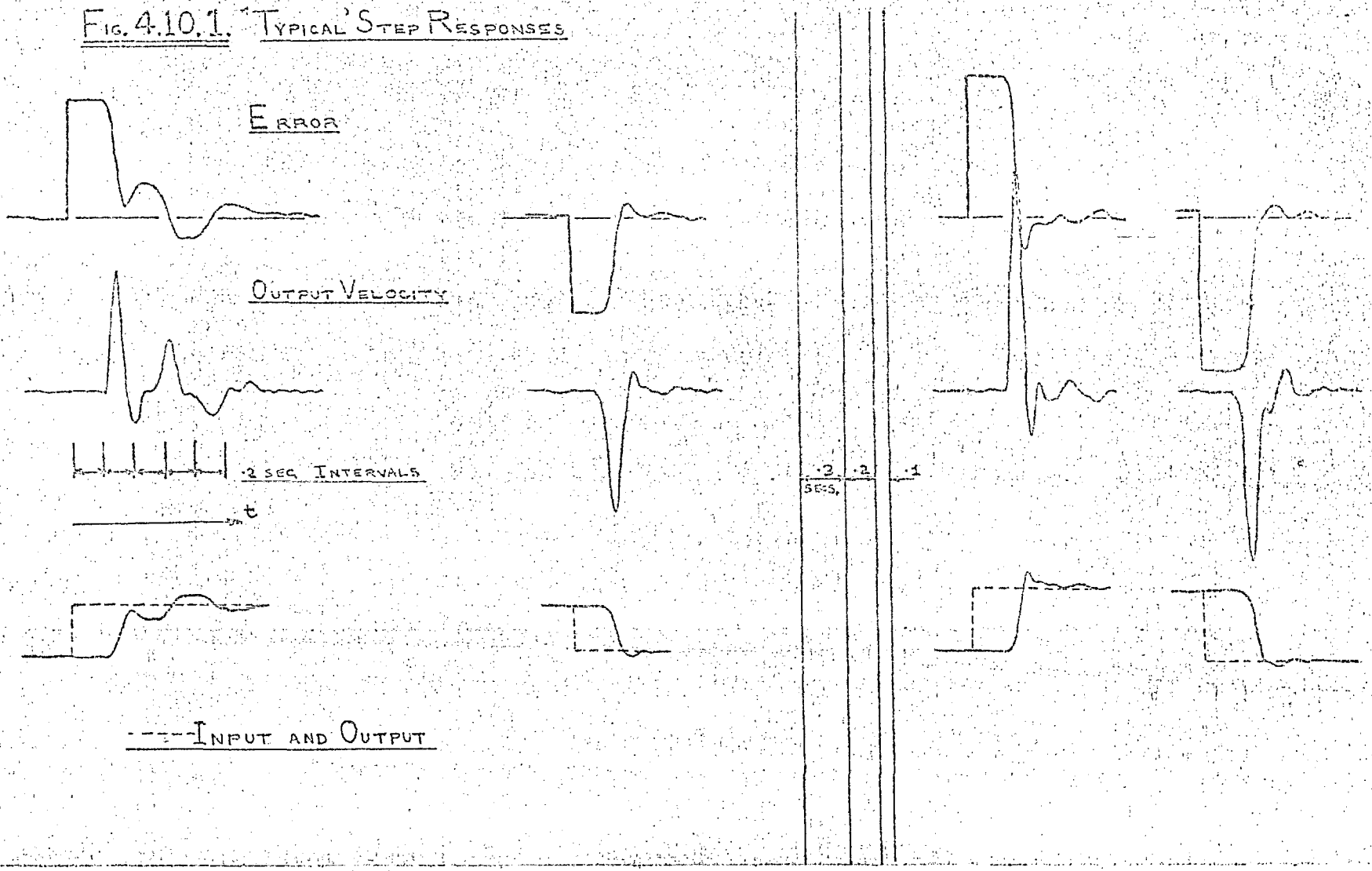
The results of linear multiple regression analysis suggested two possible additions to the model - the introduction of a small threshold effect in the error channel,and of some form of dependency between sampling interval length and input velocity,but of a rather weak nature. Unfortunately,the effect of both these operations would have been to introduce nonlinearity into the model structure. It was judged that the importance of both these effects would be very small in the continuous tracking task. It was therefore decided to hold them 'in reserve',to be utilised if a linear model proved inadequate.

4.10. Initial Formulation of a Model Structure from Qualitative Features of Step Responses

The qualitative features which should be possessed by a model capable of representing the human operator,are discussed in Sections 4.5. and 4.9. The purpose of this and the subsequent section is to describe the formulation of a set of structural interconnections between error and output motion,through a consideration of the qualitative features of operators' step and ramp responses. It was desired that the form of the model structure should be in general accord with the aims set out in Section 1.5.

Typical step responses of operators are shown in Fig.4.10.1. In considering the idealisation of these responses,reference was made to the explanation of the qualitative features of step responses offered by Wilde (18) and subsequently by Lemay (4). Observed step responses were in general agreement with the mechanism proposed by these investigators ; response could be considered to be composed of

FIG. 4.10.1. TYPICAL STEP RESPONSES



a number of quasi-parabolic segments,corresponding to one or more fairly well defined velocity triangles. It was noted that these triangles did not,in general, either begin or end at zero velocity,and that this effect was usually associated with the presence of overshoot. These 'residual' velocities were usually small in comparison with peak velocities,and were neglected by the aforementioned workers. The significance of the operator's ability to correct velocity errors was fully appreciated when ramp responses were considered,as described in Section 4.11.

For the purposes of model formulation,it seemed appropriate to consider the operator's overall output motion in terms of a combination of a position response plus a velocity response ; the desired linear form of the model would certainly permit such a superposition.Information relating to the basic position response was obtained by neglecting the small velocity errors in step responses,and by considering the operator's position output to be composed of parabolic segments,corresponding to velocity triangles. Responses idealised in this manner are shown in Figs,4.10.2a,b,&c. ; these illustrate the effects of errors of estimation and execution.

The mechanism which was proposed to account for the form of these responses was similar to that described by Lemay (4). An irregular sampling operation - average sampling interval about 150 ms. - was hypothesised,and on this basis time instants designated as A,B,C,D,and E in Figs.4.10.2. were considered to represent points at which the operator sampled his input. The sequence of actions which was considered to account for the form of these idealised responses was as follows:-

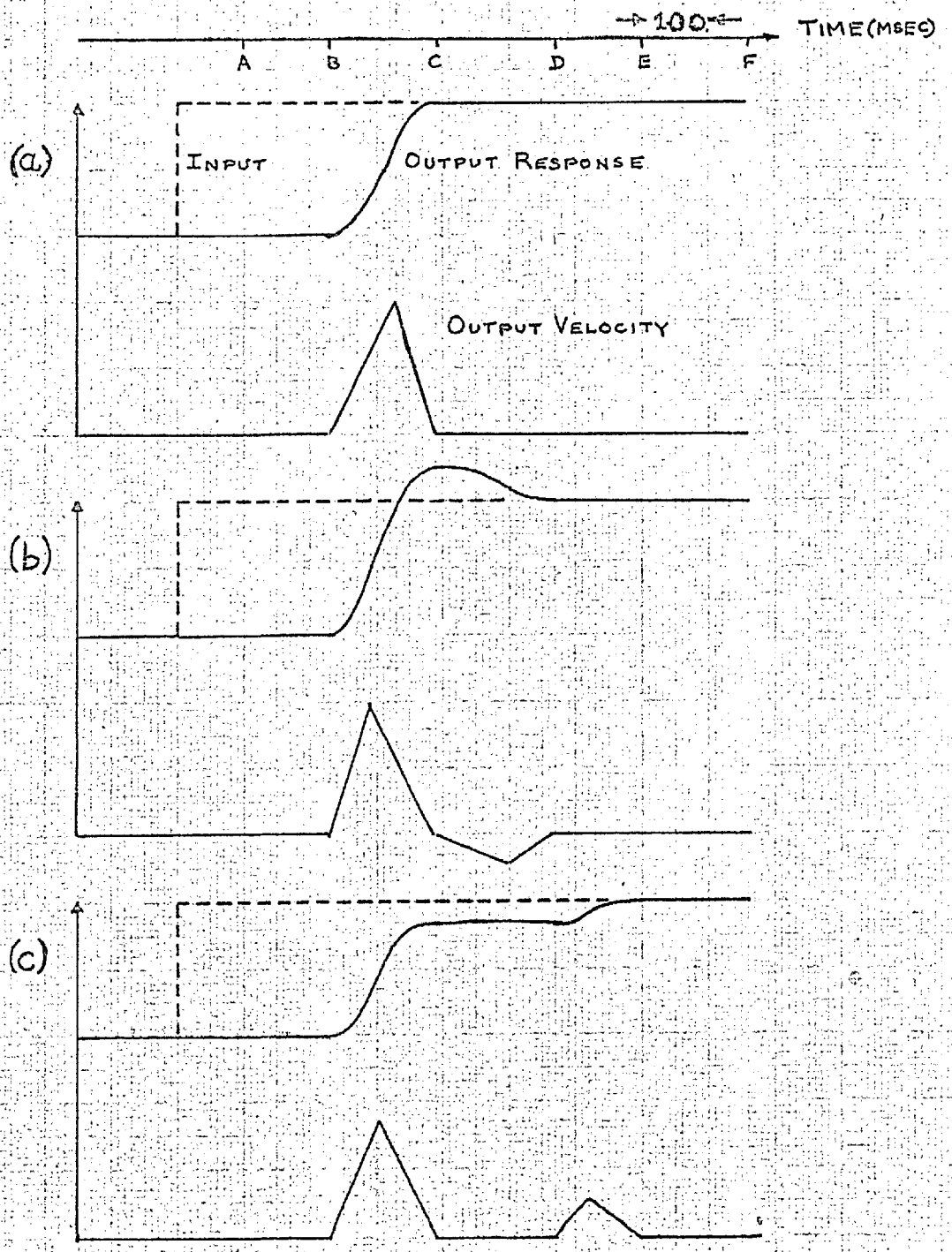
At time A the operator first observes that an input step has occurred. During the interval AB the magnitude of the error is estimated,and an appropriate programme of hand movement is formulated. This programme is executed over the interval BC,giving rise to an approximately triangular output velocity-time function. Meanwhile, the operator again samples his error at B. The estimated error is compared with the hand movement already programmed,and a further programme of hand movement is formulated,for execution in the interval CD . A similar action is pursued over subsequent time intervals,until error is finally reduced below some threshold, set by the operator in relation to his instructions.

Fig.4.10.2a. shows a case where error was estimated perfectly from the sample taken at A,and the programmed hand movement was performed without error over interval BC. No further hand movement was necessary in subsequent time intervals.

Fig.4.10.2b. shows a case of overestimation of error. The error was sampled again at B ,and the overestimate was allowed for by the formulation of a programme of reverse hand movement executed in the interval CD.

Fig.4.10.2c. shows a case of error in execution of hand movement. Error was estimated correctly from the sample at time A, but subsequently there was an error between actual hand movement and that required by the programme. The operator could not perceive this error until time C,after which he formulated a further programme of

FIG. 4.10.2. IDEALISED REPRESENTATIONS OF OPERATORS' STEP RESPONSES



movement to be executed over DE.

For the purposes of clarity, the idealised responses shown in Figs.4.10.2. constitute a separate representation of effects which usually occur in combination. Also, only inaccuracies in the initial estimation of error and execution of hand movement are shown ; usually further small errors are made while correcting these initial errors. For the purposes of illustration, these subsequent errors have been neglected.

It was thought that the tendency for inaccuracy in the initial estimation of error, leading to undershoot or overshoot, might be related to the structure of the visual system. The eye cannot move instantaneously in response to an initial step ; this implied that the fastest manual response to a step input could be obtained only if the error were estimated before any eye movement could be effected. In the experimental situation, the full width of the foveal region was covered by an image 1 cm. wide on the display. Thus, if the eye were fixated on the centre of the screen, so small a step as ½ cm. would take the retinal image to the edge of the zone of maximum resolution. The majority of steps were of a magnitude between 2 and 3 cms. ; this implied fairly serious degradation of accuracy of estimation. By comparison, estimation of error during continuous tracking of a fairly smooth input, should be very accurate. The limitations of the eye tracking system were confirmed by presenting steps random in amplitude and time of occurrence at a fairly high rate - second Poisson interval distribution with an average interval less than about ½ sec. It was found that the operator tended to 'lose' the display line at times, which demonstrated that this task was quite a difficult one as far as the visual tracking system was concerned.

In formulating a model to reproduce the basic features of the operator's step response, the following idealisations were made:-

(a) The operator was represented as sampling his input regularly in time, with a sampling interval in the vicinity of 150 ms. - a value about equal to the average length of sampling interval deduced from step responses and sampled display tests. The sampling instants were represented as being of negligible length.

(b) The operator's hand motion was represented in terms of isosceles velocity triangles.

Idealisation (a) was made in the interests of mathematical tractability, and also because it was not possible to assign a relation between sampling interval length and input characteristics with any certainty. A purely random sampling operation could have been specified, but would not have aided in direct comparison of operator and model responses to a common input function, for which purpose the model was primarily intended.

Idealisation (b) was a fairly accurate approximation to observed responses. The form of the operator's hand motion may well have arisen from the innate structure of the neuromuscular system ;

however,a second possibilty was that it represented an optimising action on the part of the operator. This might arise if muscular effort caused fatigue proportional to some power of the total muscular force required. This is discussed further in Section 9.6.

The block diagram of the model corresponding to the mode of operation described in the above paragraphs is shown in Fig,4.10.3. Here,the process of estimation of input (loop error) is represented by a sampler,and an added estimation error term, E_1. Gain g_p represents the weight attached to samples of estimated error. The process of formulation of desired response is represented by the loop containing the processing delay of one sample interval, and the feedback gain g_m , which represents the weighting of the last released programme of hand movement. The velocity triangle generating circuit gives rise to a double-parabola position output running for one sample interval, the overall magnitude of which is determined by the programme released to it at the commencement of the sampling interval. Error of execution is represented by the random term, E_2.

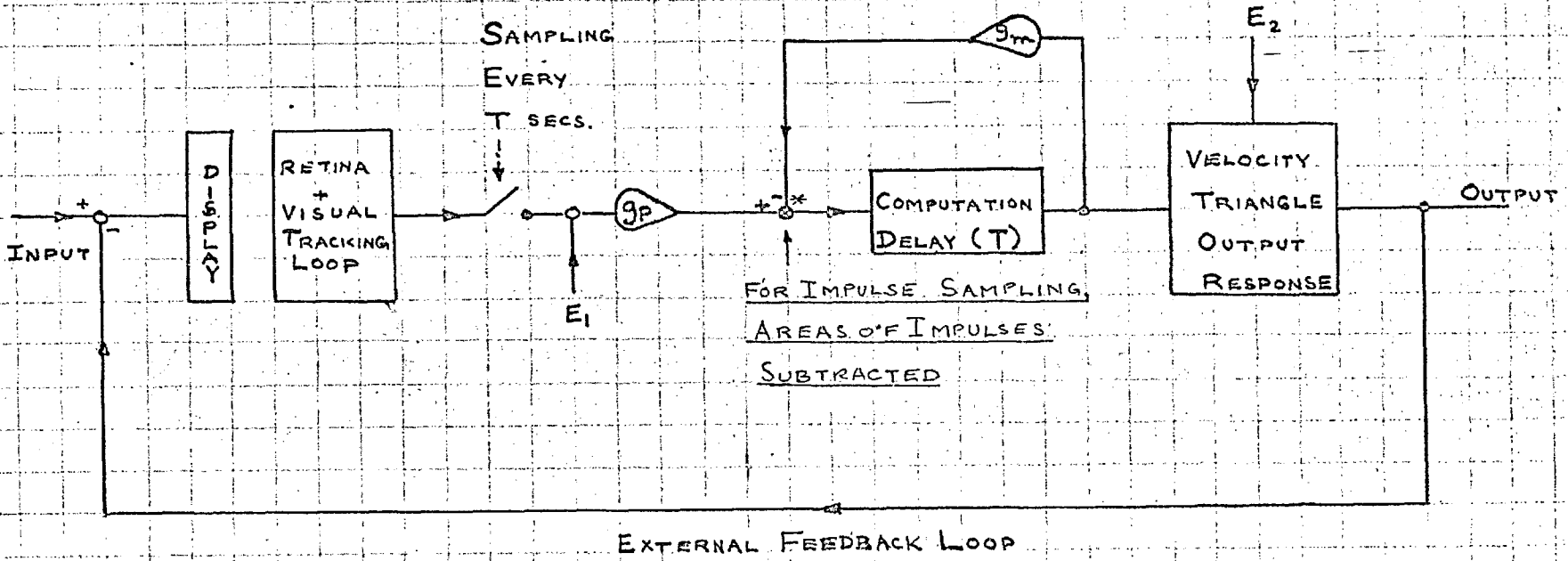
Though experimental data showed a tendency for more overshoots than undershoots,part of the effect was due to velocity error and bias in the estimation error, E_1 . There was no evidence to suggest a persistent bias in error of execution,and in any case this was quite small. Thus to simulate operators' step responses, the gain g_m would be unity, while g_p would be very nearly unity.

4.11. Further Formulation of Model Structure From Qualitative Features of Ramp Responses

The study of step responses indicated that the operator possessed some facility for correcting velocity errors. This was confirmed by the study of operators' responses to the random ramp function. It was evident that the operator possessed the ability to follow a constant input velocity by means of a well matched and relatively constant hand velocity,after an initial response whereby both hand position and velocity were matched to the input. The step response model,as derived in Section 4.10., could not explain this behaviour. It could only represent tracking by means of a number of double parabolic segments,with velocity going to zero at sampling instants. A more elaborate model structure was therefore necessary.

Ramp responses typical of those obtained experimentally are shown in Fig.4.11.1.,and in idealised form,in Figs.4.11.2a&b. In order to explain these responses in terms of the sampling hypothesis already advanced,it was necessary to consider how the operator might sense velocity. There were two possibilities : velocity could be sensed directly at sampling instants,or it could be sensed by differencing position error. The sampled display investigations (Sections 4.2.-4.5.) had already indicated that the former was more likely. Additional

FIG. 4.10.3. BLOCK DIAGRAM OF MODEL TO SIMULATE OPERATORS' STEP RESPONSES



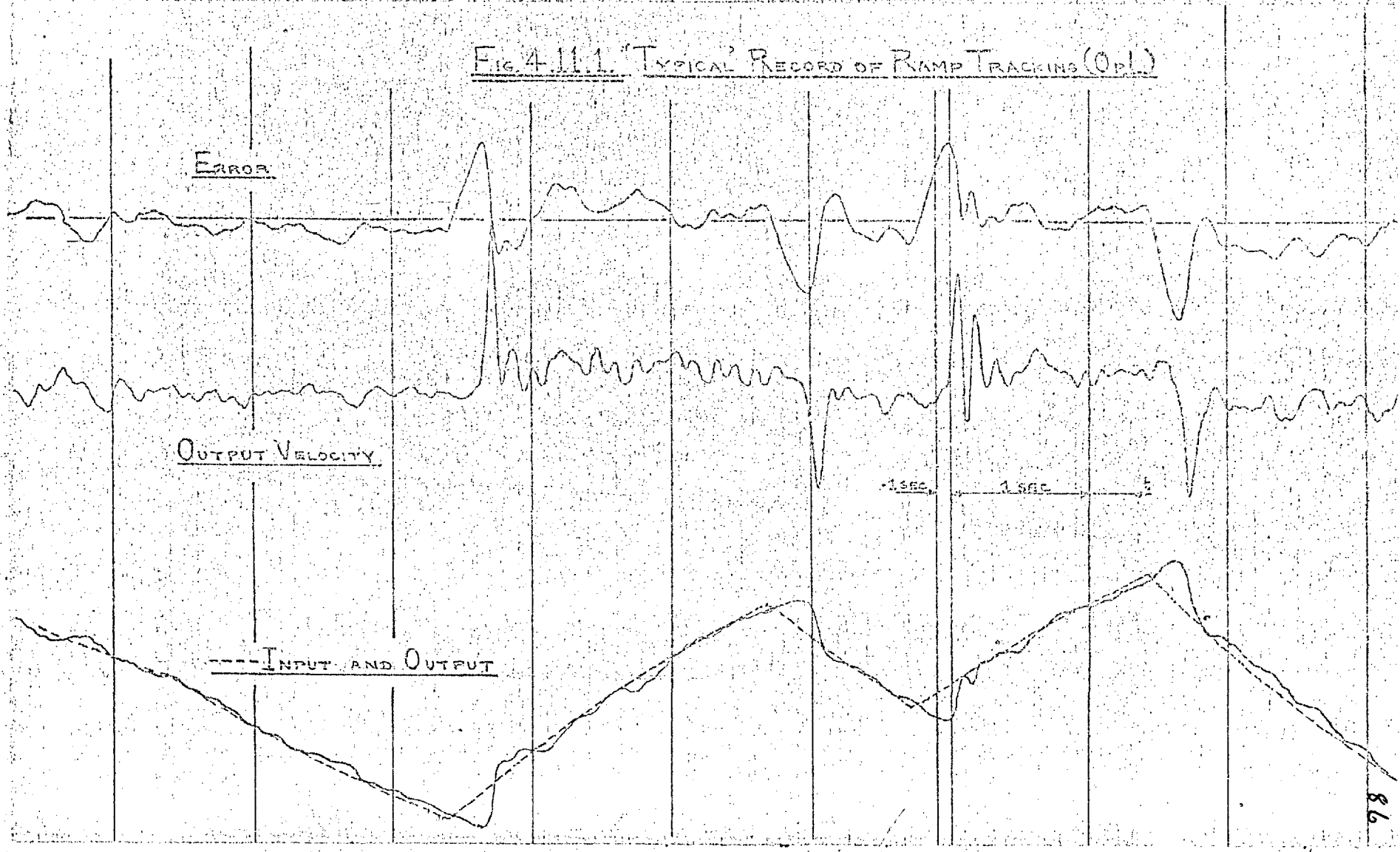
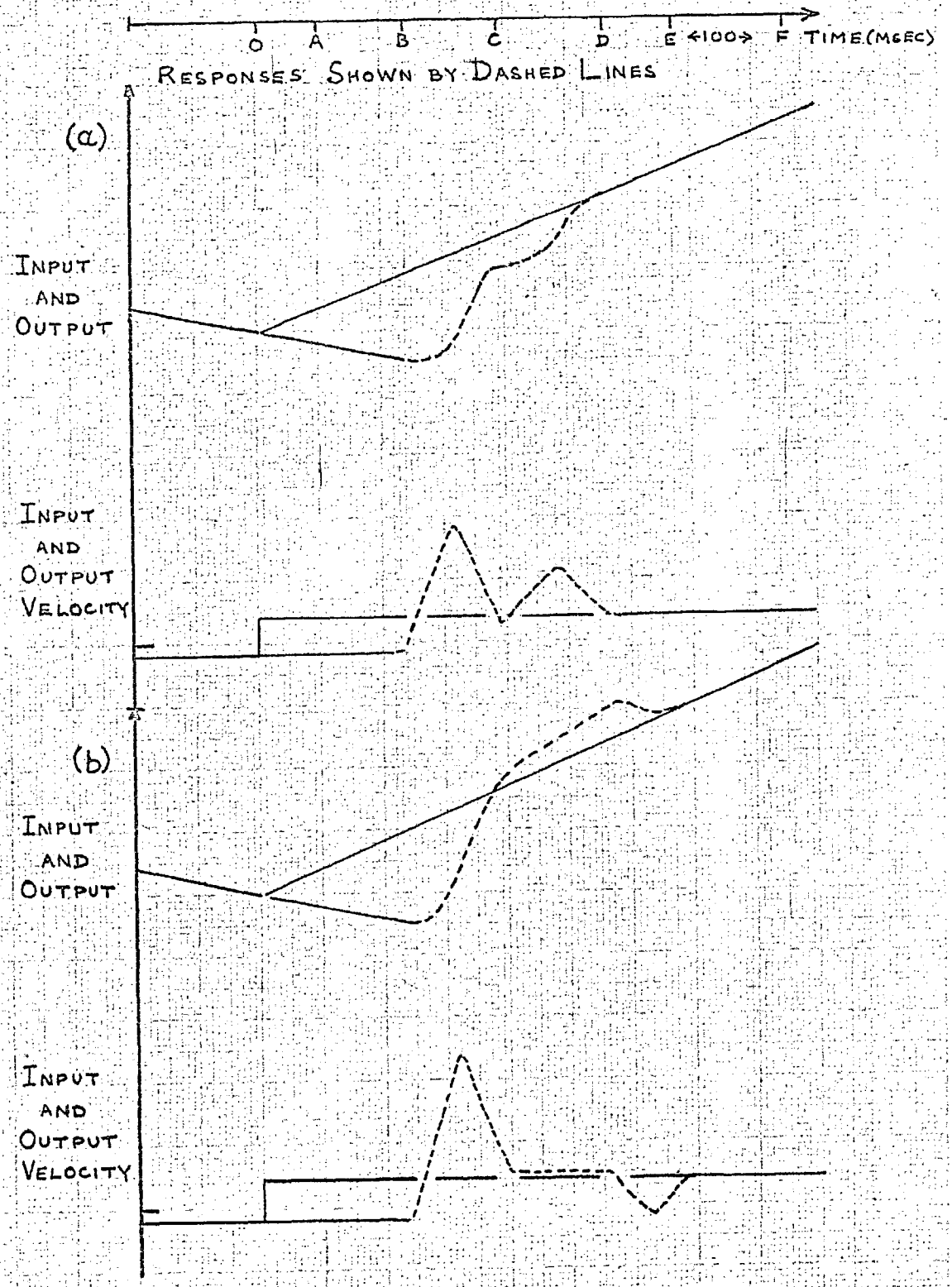


Fig. 4.11.1. 'Typical' Record of Ramp Tracking (O.L.)

FIG. 4.11.2. IDEALISED REPRESENTATION OF OPERATORS' RESPONSES TO RAMP INPUTS



information was provided by the ramp responses ,as described in the following discussions.

If velocity were sensed by differencing,then to obtain an accurate estimate of velocity error,the operator would need to take at least two samples of position error. This would lead to a minimum pure delay of two sampling intervals (about 300 ms.),allowing for computation delay,before response movement could be initiated. This was at variance with experimental data. If the operator took just one sample before formulating a response,then there would be considerable inaccuracy of response,which would also tend to be biassed. Again,this did not fit the experimental data ; operators were quite capable of effecting fairly accurate responses after a pure delay only slightly in excess of 150 ms. It was concluded that, if the operator did sense the error velocity by differencing,then this must be carried out over a very short period. As an idealisation, therefore, the operator's sensing of velocity could be represented as a sampling operation on the differentiated (practically,a sufficiently accurate quasi-differentiation) error signal. Physiological evidence does not contradict this view (Section 2.3.).

The above reasoning led to an explanation of the basic features of operators' ramp responses as follows. Referring to Figs.4.11.2a&b,times A,B,C, etc.,are considered to be sampling instants. In each figure the operator is shown as following the preceeding ramp perfectly ; there would usually be some small position error,for which he can allow quite easily in the following sequence of operations :-

(a) At time 0 a step change of input velocity occurs.

(b) At time A the operator perceives two errors - one of position,the other of velocity.

(c) In the interval AB he formulates a programme of hand movement to be released at time B , and executed in the interval BC. The programme may be considered to consist of two parts,viz. a ramp of velocity (i.e. a constant acceleration) calculated to match hand velocity to input velocity at C,plus a velocity-triangle position correction manoeuvre,calculated to match hand position to input position at C. Since the input under consideration is a ramp,the required velocity correction programme is equal to the velocity error observed at A. The required position correction programme is given by the sum of the position error at A and the predicted position error at C,arising from the velocity error at A acting over two sampling intervals,minus the position correction afforded by the velocity tracking programme.

(d) At B the position and velocity errors are again sampled. Allowance is made for the movements already programmed,and a second correction programme is formulated to be released at C. Further programmes are formulated sequentially,in an analogous manner.

In Fig.4.11.2a the operator is shown as having underestimated his velocity error. He corrected for this by a programme of hand movement executed over the interval CD. The case of an overshoot caused by error in execution is shown in Fig.4.11.2b. Here the operator could not perceive the error in hand velocity - and the resultant error in position - until movement was completed according to the initial programme ; a further two sampling intervals were required for correction. Errors of estimation and execution relating to position correction, would elicit responses as described for step inputs (Section 4.10.), which would be superimposed on the basic ramp response.

The basic block diagram of a model structure capable of a realistic representation of the above sequance of operations is shown in Fig.4.11.3. This model was developed by combining a velocity tracking loop and position prediction facility with the basic step response model. Regular sampling was assumed, as in Section 4.10. (no relation was postulated between sampling interval length and input or error velocity). A description of the action of the model is given in the following paragraphs:-

(i) Pure position errors are dealt with solely by the position correction branch, the action of which is as described for the step response model (Section 4.10.)

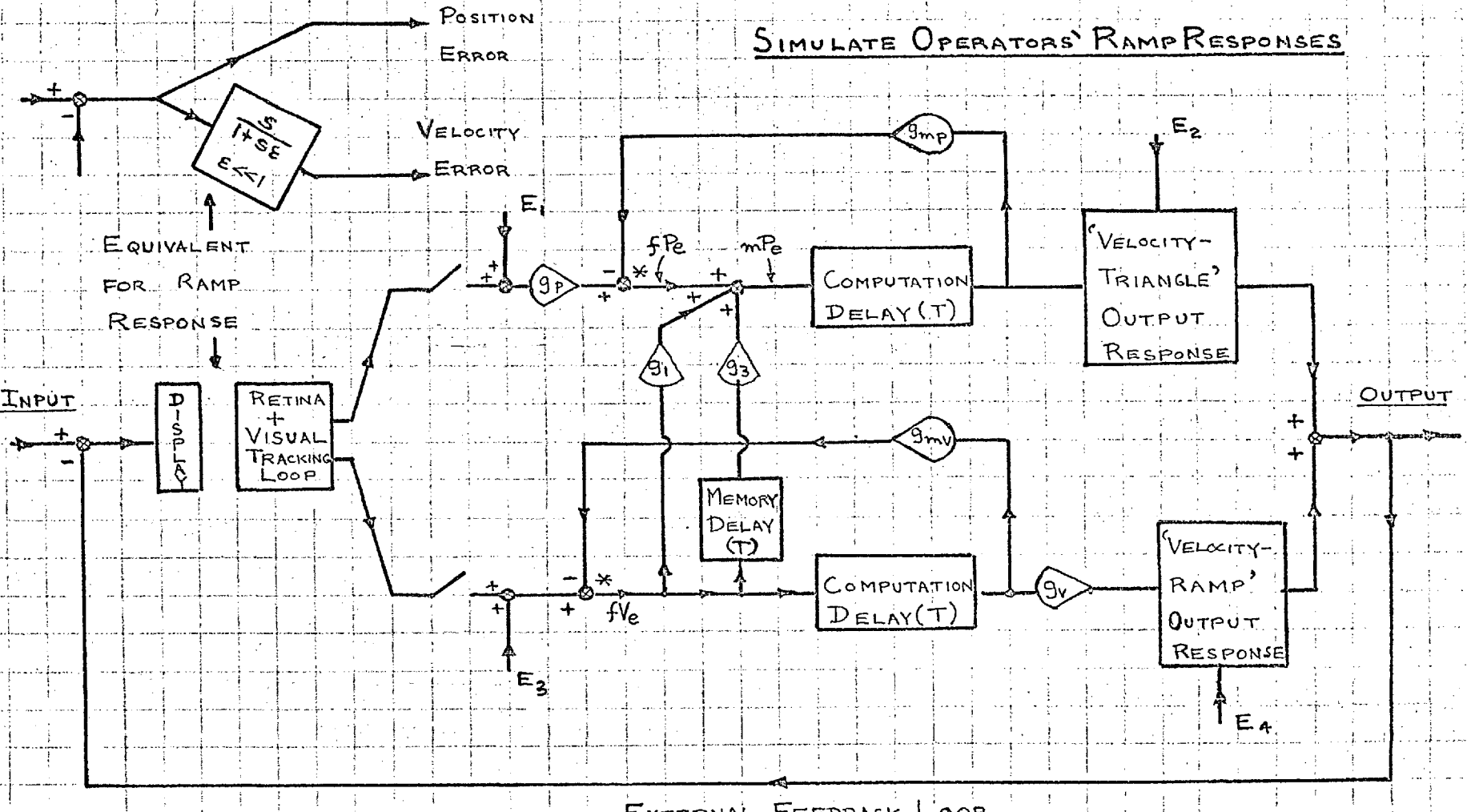
(ii) The process of estimation of velocity error is represented by the differentiator, sampler, and additive error term, E_3.

(iii) Fresh velocity error, fVe, is computed by subtracting the previously programmed velocity correction, weighted by a factor g_{mv}, from the present sample of velocity error. This computed fresh velocity error then gives rise to both a velocity correction and a position correction programme.

(iv) The velocity correction programme consists of a constant acceleration over one sample interval, leading to an alteration of hand velocity by an amount proportional to the computed fresh velocity error, fVe; the weighting is represented by the factor g_v. Error of execution of the velocity programme is represented by the additive term, E_4.

(v) The total position correction programme results from the combined effects of the observed position error and position error predicted via the velocity correction branch. The prediction process is represented by the cross-branches from the velocity error computation loop to the position error computation loop. Computed fresh velocity error, fVe, is weighted by a factor g_1 and gives rise to a position correction programme executed over the following sampling interval. The magnitude of g_1 is related to the predicted change of position error at the end of the interval;

FIG. 4.11.3. BLOCK DIAGRAM OF MODEL TO SIMULATE OPERATORS' RAMP RESPONSES



* FOR IMPULSE SAMPLING AREAS OF PULSES SUBTRACTED

i.e. position error predicted from presently observed velocity error, minus the position correction due to the concurrently formulated velocity correction programme. The cross-branch containing a delay of one sample interval and weighting factor g_3 is specified because of the need to make allowance for the position prediction programme just released, when calculating a programme for the succeeding sampling interval. Otherwise, the action of the position computation loop could give rise to spurious response. The currently released position prediction programme is proportional to the fresh velocity error computed in the previous sample interval - hence the reason for a delay in the cross-branch.

(vi) For the best possible non-overshooting response to ramp inputs, the values of the weighting factors would be as follows:-

$$g_{mp} = g_{mv} = 1 \; ; \; g_p = g_v = 1 \; ; \; g_1 = 1.5T \; ; \; g_3 = .5T$$

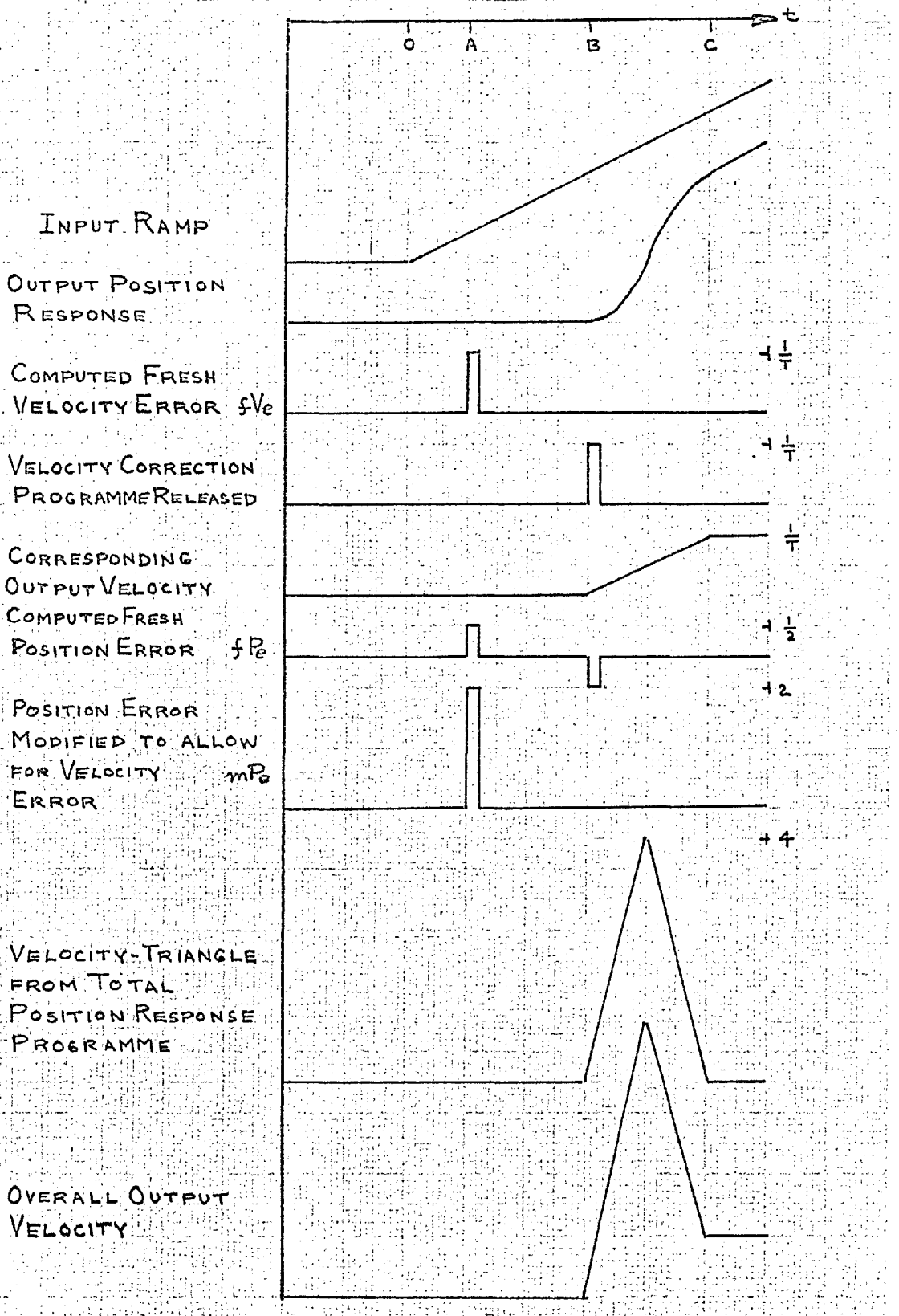
This particular choice of model parameters leads to a ramp response with a settling time of two sampling intervals, if the injected error terms are zero. This is the shortest settling time of which the model is capable.

The sequence of actions described above may be appreciated more clearly by considering the model's response to a ramp input, under conditions where the model parameters are set to values given in (vi) above, the injected error terms are assumed zero, and the input ramp has a slope of 1/T (one unit per sampling interval). Referring to Fig.4.11.4., it is assumed for convenience that the ramp input starts halfway between sampling instants, at 0. At A the model samples a position error of ½ and a velocity error of 1/T. Since there was no previous input, these values represent fresh errors, fPe and fVe respectively.

The position computation loop then programmes a velocity-triangle response, released at B, which leads to a change of hand position by ½ over the interval BC. The velocity computation loop programmes a velocity ramp released at B, which leads to a hand velocity of 1/T at C, and a change of hand position, over interval BC, of ½. Now, in the absence of position prediction, the position error at C would be 1.5 - hence the appropriate value of g_1 which, acting on the current computed fresh velocity error causes a velocity triangle giving this required motion to be added to that originally formulated by the position computation loop. Thus, programmes formulated in the interval AB lead to correct hand position and velocity at time C.

Meanwhile, velocity and position error are again sampled at B. Velocity error is unchanged, and subtraction of the velocity programme released at B leads to fVe = 0. No further velocity correction programme is required. The position error sampled at B is 1.5, but the total position programme released at B is 2.0. Since computed fresh velocity error at B is zero, this results in a value of fPe = -.5. The absence of any other action would therefore lead to an apparent, but false, modified position error, mPe, of -.5, giving erroneous response.

FIG. 4.11.4. 'SHORTEST SETTLING TIME' RESPONSE TO RAMP INPUT OF MODEL ADJUSTED AS IN (VI) ; SEE P91.



Due allowance is made for the position prediciton programme by taking note of the computed fresh velocity error which gave rise to it, i.e. that extant at time A. This operation is effected by the cross-branch containing a delay of T,plus a gain of .5T ,which feeds an input of .5 to the position computation loop at time B. This is added to the apparent fPe to yield a modified positon error,mPe,of zero,which is correct.

4.12. Evaluation of Model as Initially Formulated

As formulated in Section 4.11.,the model was theoretically capable of reproducing the main quantitative and qualitative features of operators' responses to random ramp inputs. It was also capable of reproducing operators' responses to step inputs,with the proviso that steps occurred between sampling instants. The more general case could be covered by introducing a modification,so that spurious response to a step occurring at a sampling instant was suppressed. This action could,for example,be represented by a nonlinear filtering operation immediately following the differentiator,so that a sample of velocity error was suppressed if it exceeded a preset value.

Qualitative and quantitative features of operators' tracking under conditions of sampled display could,in theory,be reproduced by the model,as modified by the addition of a variable smoothing filter to the differentiator.

The next stage in evaluating the model involved the construction of a 'real time' analogue ; this analogue was then to be used to effect a direct comparison with responses of operators, under conditions of common random ramp input. Further investigations were to be directed toward the evaluation of the model's capacity to simulate the characteristics of operators performing a continuous tracking task,and the formulation of any necessary extensions to the model structure. These aspects of experimental investigation are described in Chapter 5.

Chapter 5 : REFINEMENT OF THE SAMPLED DATA MODEL

5.1. Experimental Evaluation of the Ramp Response Model

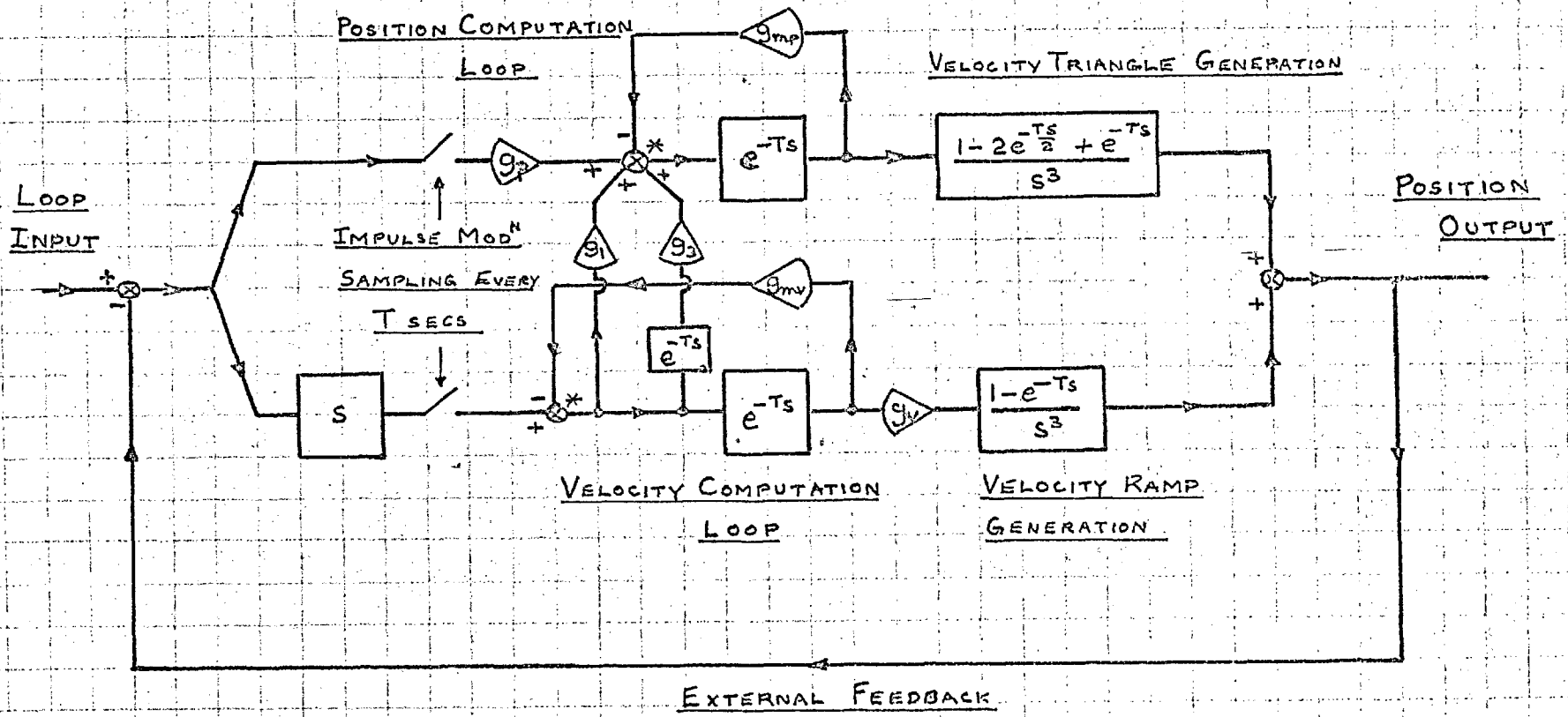
A theoretical model structure to represent operators' responses to random ramp inputs is described in Section 4.11. A 'real time' electrical analogue computer was used to simulate this model so that a direct comparison could be made between the response of model and operator, when both were subject to a common random ramp input. Fig.5.1.1. shows the component structures of the theoretical model, expressed in terms of the Laplace operator s . The various error terms were omitted from this diagram, because a direct comparison of time domain responses was required. The inclusion of such terms would have been confusing ; they would not have increased the accuracy of simulation, except in the statistical sense of, e.g., matching the shapes of autocorrelation functions. The diagram of equivalent transfers was used directly in the formulation of a practical circuit, as described in Appendix V.

The purpose of the experimental procedure was to obtain the best visual matches between operators' ramp responses and those of the model, by varying the model parameters to suit each individual operator ; three operators participated in the investigations. The model parameters were initially set so as to give the shortest settling time (Section 4.11.), then small changes were made systematically in the order $g_p, g_v, g_1, g_3, g_{mp}, g_{mv}$, and T . The gains were varied so as to achieve the best match in regard to shape of response, while overall delay was matched by alteration of T . After matching, the average discrepancies observed between operator and model responses served as the basis for evaluating the quality of the model.

It was found that the analogue model was capable of giving quite a good simulation of the operator's response immediately following a step change of slope. However, the match was not so good over the smooth portions of responses to the longer ramps. The operator followed the small curvature due to the exponential nature of the 'ramp' quite closely, and with a slight lag, whereas the model output tended to lead the ramp slightly.

The 'best match' settings of the model parameters did not show very much variation from operator to operator. These values are noted below, in terms of their effect on model response. The values of sampling interval , T, ranged from about 140 to 160 ms. ; a wide variation of T was not expected, because of its direct effect on overall model delay. The values of g_{mp} and g_{mv} were both near unity. this accorded with the a priori postulate that memory was unbiassed, and therefore weightings would be adopted so as to give about the shortest response settling time. The values of g_p

FIG. 5.1.1. RAMP RESPONSE MODEL IN TERMS OF LAPLACE TRANSFERS



* DENOTES ADDITION OF AREAS OF IMPULSES

were such as to lead to an overall position tracking gain of unity ; i.e. errors of position were corrected almost completely by one output response. The values of g_v were such as to lead to an initial small velocity overshoot. The values of g_1 were such as to give an initial overall position overshoot in the region of 10 to 20%. In all cases g_3 was found to be small - not more than .1 g_1 .

These results indicated that the basic reasoning behind the derivation of the model was sound, in that it led to satisfactory real time simulation of operators' ramp responses. It was apparent that the model could be simplified without seriously affecting accuracy of simulation, by taking $g_{mp} = g_{mv} = 1$, and neglecting g_3. The nature of the difference between smooth parts of operator and model responses gave a hint that some extensions to the model structure might be necessary, and this was confirmed by the experiments described in the following Section.

5.2. Evaluation of Ramp Response Model in Continuous Tracking Simulation

As a sequel to the ramp response investigations, it was required to evaluate the capacity of the basic ramp response analogue model (described in Section 5.1.) to simulate operators' characteristics in a continuous tracking task. Both model and operator control loops were fed with a common continuous input signal, possessing a statistical structure as described in Sections 3.2. and 3.3. The model was scaled so that both error and output voltage were comparable on a 1:1 basis with the corresponding quantities in the operator control loop.

The experimental procedure again involved a parameter variation technique to achieve a 'best fit' between model and operator error and output functions. Parameters were first adjusted to achieve a good visual fit. Then quantitative measures of fit were computed as follows. During an experimental run of 200 secs. duration, the integrated moduli of the operator error, model error, and of the difference between them , were computed. The resulting quantities were then used to calculate the operator and model error variances, and a measure of their zero lag cross-correlation, as described in Appendix IV.

The rationale for assessing the degree of fit from these measurements was as follows:-

(a) If model and operator characteristics were well matched, then the variance of their respective errors should be comparable. One would expect, a priori, that the operator's error would be somewhat greater than that of the model error variance, due to the effects of operator 'noise' not represented in the model.

(b) If the conditions outlined in (a) are satisfied, then the most sensitive measure of fit was provided by the zero lag cross-

'correlation' between errors, where the 'correlation' is defined in a particular way. With the usual definition, one could specify a very high cross-correlation between errors without defining their absolute variances, and this would not imply a good model fit. Therefore correlation is defined in terms of the variance of the difference between errors - perfect correlation then exists when this difference signal has zero variance. This corresponds to defining, in advance, a regression line between model and operator error, possessing unit slope and zero intercept, then measuring residuals from this predetermined regression line. This introduces a bias toward smaller values of correlation, according to the departure of the actual regression line from the 'preset' line. The causation of this bias is described more fully in Appendix IV, where its correction is also considered. A high value of the cross-correlation defined above automatically implies a very high cross-correlation between outputs, because of the 1:1 nature of the feedback loop.

After the initial measurement of fit, parameters were adjusted slightly in a direction which, it was thought, would lead to a better fit. A further experimental run was then performed, and quantitative fit was again computed. A number of successive runs were then carried out in a similar fashion, until no significant improvement in fit was obtained. Three operators participated in these model matching experiments.

The results of visual matching indicated that the model displayed too much phase lag when following the continuous input. Any attempt to reduce this by increasing prediction gain g_1, or velocity gain g_v, merely gave rise to undesirably low loop stability. In addition, the best value of g_3 was found to be zero, in accord with results described in Section 5.1. The best values of the gains g_{mp} and g_{mv} were again found to be near unity ; in subsequent quantitative matching runs they were kept at unity, so as to reduce the number of variables. The quantitative matching results were also disappointing. It was difficult to obtain similar error variances, and the highest error cross-correlation that could be obtained was only of the order of 0.7 ; i.e. only half the operator's error variance was matched by the model.

The explanation of the above results became clear when the statistical properties of the continuous random input were considered. The velocity tracking loop included a pure integration in the forward transfer ; the only feedback was via the external tracking loop. This implied that, once an output velocity had been established, only further samples of velocity error could change it. Now in the case of a slowly 'drooping' input velocity, as possessed by the ramp process, this gave rise to little error (though it was still noticeable - see Section 5.1.). But the autocorrelation of velocity of the continuous random signal exhibited a rapid decline (see Appendix II). Thus the velocity tracking loop output exhibited an undesirable degree of 'staleness', resulting in too large a phase lag

to achieve a good match to the operator.

The value of g_3 would be expected to be significant only in the case of an input possessing a velocity whose autocorrelation function exhibited a very slow decline over .3 to .4 secs. This condition was approximated by the random ramp process, but the continuous random signal possessed an autocorrelation of velocity which passed through zero at a lag of .4 secs. Hence the best setting of g_3 would be zero.

The conclusions drawn from this phase of the experimental investigation were that the operator was more capable of sophisticated prediction than had been allowed for in the model structure. The formulation of the appropriate extensions to allow for this prediction is discussed in the following Sections.

5.3. Extension of the Predictive Capabilities of the Model

Experiment showed that the predictive capabilities of the basic ramp response model were inadequate, and that some extension of the model structure would be necessary. It seemed logical to suppose that the operator's sophisticated prediction might be related to his ability to estimate the statistical structure of the input either directly, or in effect, by adjusting the parameters of his response so as to minimise his subjective criterion of error - in effect using a hill climbing technique. In either case, this predictive ability would lead to a more complicated model structure.

The continuous random input signal was formed by passing 'white' noise through a ~~third~~ second order lag. As described in Section 3.2., derivatives up to and including third order could theoretically be used for the purpose of prediction. In postulating the manner in which the model should be modified, the following points were considered:-

(a) The operator could probably not sense acceleration, other than by differencing velocity information. The basis of this statement was that previous investigators (Chapter 2) have found little evidence to support direct visual sensing of acceleration. Also, eye tracking studies (21) indicate that the eye tends to track a parabolic input by means of a series of saccadic jumps, separated by ramps of relatively constant velocity. Except for inputs containing only very low frequency terms, differenced velocity information would be too stale to be of much use. It would also be noisy, so that further staleness would be incurred in the necessary smoothing operations. The above remarks apply 'a fortiori' to sensing of third derivative information.

(b) The operator could estimate his hand position and velocity by means of the kinaesthetic sense, albeit with some delay and inaccuracy

(c) The operator could not perform a programme of hand motion with arbitrary precision. Knowledge of error of execution was available through the kinaesthetic sense only after some delay. Kinaesthetic reaction time experiments indicate a delay of response to sudden arm movement of the order of 100 ms.(11,24)

(d) The operator could perceive position and velocity error only with some finite error of estimation.

The points listed above indicate that only position error and velocity error could be sensed directly, but that output position and velocity could be estimated and used for the purposes of prediction. The weights which could be attached to past samples of the output functions were not arbitrary - they were subject to limitation in respect of the requirement that past errors should be suppressed. A further restriction was that the requirements of closed loop stability must be satisfied. Also, it was desired that the number of adjustable model parameters should be kept to the minimum consistent with accurate simulation. These considerations all indicated that modifications to the model should be restricted so that a minimum number of past samples were used for the purposes of prediction.

5.4. Linear Prediction of Position

Prediction of required future position correction was already inherent in the structure of the basic model, in terms of the gain parameters g_1 and g_p (considering g_3 to be zero for the continuous random input). The values of these gains were such as to give the expected required position correction two sample intervals ahead, corresponding to the time required for hand motion to be fully effective.

Position prediction from past values of position output was allowed for by incorporating velocity prediction from past samples of output position and velocity, as described in the following Sections.

5.5. Linear Prediction of Velocity from Velocity Information

Experiment indicated the necessity for prediction of velocity from present and past samples of velocity error and output. It was desired that this prediction should be represented by a linear operation. Reasoning which led to a suitable formulation of a model structure for incorporating prediction from velocity information

is described in the following paragraphs.

It was found convenient to employ the following notation in considering the form of the prediction:-

$E_v(n)$ = velocity error at time nT

${}_pE_v(n)$ = prior prediction of velocity error at time nT

$P_v(n)$ = velocity correction programme released at time nT

$C_v(n)$ = output velocity at time nT

${}_pC_v(n)$ = prior prediction of output velocity at time nT

$R_v(n)$ = input velocity at time nT

${}_pR_v(n)$ = prior prediction of input velocity at time nT

${}_rR_v(n)$ = reconstruction of input velocity at time nT

In view of the unity external feedback, the basic requirement of velocity prediction may be stated in the equation:-

$$P_v(n+2) = {}_pR_v(n+3) - {}_pC_v(n+2) \qquad 5.5.1.$$

i.e. the programme released at time (n+2)T should be that required to change the predicted existing output velocity at (n+2)T to the predicted input velocity at time (n+3)T .

Since tracking is compensatory, the operator must predict ${}_pR_v(n+3)$ either by using past samples of output velocity, or by using reconstructed past samples of input velocity. The latter method yields a better prediction. Linear formulation of ${}_pR_v(n+3)$ from reconstructed past samples of input velocity may be represented as:-

$$ {}_pR_v(n+3) = f_1 \cdot {}_rR_v(n+2) + f_2 \cdot {}_rR_v(n+1) + f_3 \cdot {}_rR_v(n) + \text{etc.} \qquad 5.5.2.$$

where the f's are constant weighting factors.

The operator may reconstruct all past ${}_rR_v$'s, except ${}_rR_v(n+2)$ by use of the relation:-

$$ {}_rR_v(n) = C_v(n) + P_v(n) \qquad 5.5.3.$$

A special reconstruction must be used for ${}_rR_v(n+2)$, because programme $P_v(n+2)$ must be released at time (n+2)T, and must therefore be formulated in the interval (n+1)T to (n+2)T. Therefore ${}_rR_v(n+2)$ must be predicted from known values at time (n+1)T.

This prediction may be accomplished as follows. First we write:-

$$ {}_pR_v(n+2) = {}_pC_v(n+2) + {}_pE_v(n+2) \qquad 5.5.4. $$

Now the operator can form ${}_pC_v(n+2)$ from kinaesthetic information, plus knowledge of the programme $P_v(n+1)$:-

$$ {}_pC_v(n+2) = C_v(n+1) + P_v(n+1) \qquad 5.5.5. $$

The predicted error, ${}_pE_v(n+2)$, must be estimated from knowledge of past error samples. Consider the error $E_v(n+1)$; ${}_pE_v(n+1)$ had been formed at time nT, and the programme $P_v(n+1)$ was formulated to correct for the predicted effect of this ${}_pE_v(n+1)$ at time (n+2)T . We may therefore write:-

$$ E_v(n+1) = {}_pE_v(n+1) + \eta(n+1) \qquad 5.5.6. $$

where $\eta(n+1)$ represents the error of prediction of $E_v(n+1)$. Since ${}_pR_v(n+2)$ has already allowed for the predicted effect of ${}_pE_v(n+1)$ at time (n+2)T , the following relation holds :-

$$ {}_pE_v(n+2) = (1+k).\eta(n+1) \qquad 5.5.7. $$

where $(1+k).\eta(n+1)$ represents the predicted effect at time (n+2)T of $\eta(n+1)$. Now $\eta(n+1)$ is a time weighted measure of the unpredicted acceleration in the interval nT to (n+1)T . Since the total effect of this acceleration in the interval nT to (n+2)T is represented by $(1+k).\eta(n+1)$, k represents the effect of the unpredicted acceleration in the interval (n+1)T to (n+2)T which is correlated with that in the interval nT to (n+1)T.

An expression for ${}_pE_v(n+2)$ may be developed in terms of past error, from equations 5.5.6. and 5.5.7., as follows :-

$$ \begin{aligned} {}_pE_v(n+2) &= (1+k).E_v(n+1) - (1+k).{}_pE_v(n+1) \\ &= (1+k).E_v(n+1) - (1+k)^2.E_v(n) + (1+k)^3.E_v(n-1) \\ &\quad - (1+k)^4.E_v(n-2) + \text{etc.} \end{aligned} \qquad 5.5.8. $$

Consider the velocity computation loop to be modified by increase of the gain of the forward sequence, so that the loop gain becomes (1+k) . Let the output of this velocity computation loop at time nT be $\nu(n)$. From the structure of the velocity computation loop, it may be seen that:-

$$ \nu(n) = (1+k)\left[E_v(n-1) - \nu(n-1)\right] \qquad 5.5.9. $$

Expansion of equation 5.5.9. shows that :-

$$\mu(n) = (1+k).E_v(n-1) - (1+k)^2.E_v(n-2) + (1+k)^3.E_v(n-3) - (1+k)^4.E_v(n-4) + \text{etc.} \qquad 5.5.10.$$

Comparing equations 5.5.8. and 5.5.10., it may be seen that the modified velocity computation loop operates on past error so that:-

$$\mu(n) = {}_pE_v(n) \qquad 5.5.11.$$

The original concept of the velocity computation loop was developed in relation to the step and ramp response models. The ramp process possesses an autocorrelation of acceleration function which is practically equivalent to a delta function, so that the appropriate value of k for this input is zero. This is in accord with experimental results. For continuous random inputs, as a function of increasing bandwidth, k varies from an initial positive value through zero at moderate values of bandwidth, becomes slightly negative, then returns to zero for the broadest bandwidth inputs. For the continuous random signal used in experimental investigations (input spectrum cut-off at .61 cps.) the corresponding value of k was around zero, with which experimental results accord.

Elkind's (10) results for very low bandwidth inputs lend support to the above reasoning. In deriving a simple, linear continuous model to represent operators' compensatory tracking behaviour for these inputs, he found it was necessary to introduce a high-frequency lag in order to preserve theoretical closed loop stability ; this lag could not be measured directly. For these low bandwidth inputs, the appropriate value of k would be positive. The velocity computation loop would have an unstable transfer, with a pole at $z = -(1+k)$, and thus would be likely to lead to an unstable forward transfer. A simple model fitted to this transfer could well appear to give an unstable closed loop transfer.

To obtain an expression for $P_v(n+2)$ we may use equations 5.5.1., 5.5.2., 5.5.4., and 5.5.11., and derive the following relation:-

$$P_v(n+2) = (f_1 - 1).{}_pC_v(n+2) + f_1.\mu(n+2) + f_2.{}_rR_v(n+1) + f_3.{}_rR_v(n) + f_4.{}_rR_v(n-1) + \text{etc.} \qquad 5.5.12.$$

where ${}_pC_v(n+2)$ may be calculated quite accurately, as indicated by equation 5.5.5. Therefore, with very little approximation, we may write :-

$$P_v(n+2) = -(1 - f_1).C_v(n+2) + f_1.\mu(n+2) + \sum_{k=2}^{N} f_k.{}_rR_v(n+3-k) \qquad 5.5.13.$$

As a first approximation, the terms involving reconstructed past inputs may be neglected ; such terms are likely to be relatively unimportant, except for very highly filtered inputs, and would in any case

be subject to inaccuracy. This approximation would also minimise the number of model parameters needed to represent prediction. The final form of the equation then becomes:-

$$P_v(n+2) = -\alpha C_v(n+2) + \beta\mu(n+2) \qquad 5.5.14.$$

where α and β should be approximately $(1 - f_1)$ and f_1 respectively ; α would probably show the largest difference, due to the neglect of the reconstructed input terms.

According to equation 5.5.14., prediction is accomplished by taking note of the autocorrelation of input acceleration to determine the velocity computation loop gain, $(1+k)$. The velocity computation loop output is then weighted by a factor β in formulating a suitable velocity correction programme. This programme is modified by taking note of the output velocity at the sampling instant immediately preceeding its execution, weighted by the factor α . This is allowable, since $C_v(n)$ can be predicted quite accurately from $C_v(n-1)$ and $P_v(n-1)$.

In the ramp response model the parameter α was zero, leading to undesirable staleness of output velocity ; by postulating a suitable value for α this staleness is reduced to the required extent. It may be noted that with a finite value of α , output velocity at sampling instants would tend exponentially to zero, even in the absence of input from the velocity computation loop, in response to a transient input to the model.

The appropriate form of the tracking loop to correspond with velocity prediction according to equation 5.5.14. is shown in Fig.5.5.1. As may be seen, the only structural modification required is the introduction of the subsidiary feedback loop from velocity output. The relation of model parameters to α , β , and k is shown in the diagram. As indicated previously, for the experimental continuous random input signal, the appropriate value of k was zero, so that $g_{mv} = 1$. The parameters g_v and g_s were evaluated in model matching trials.

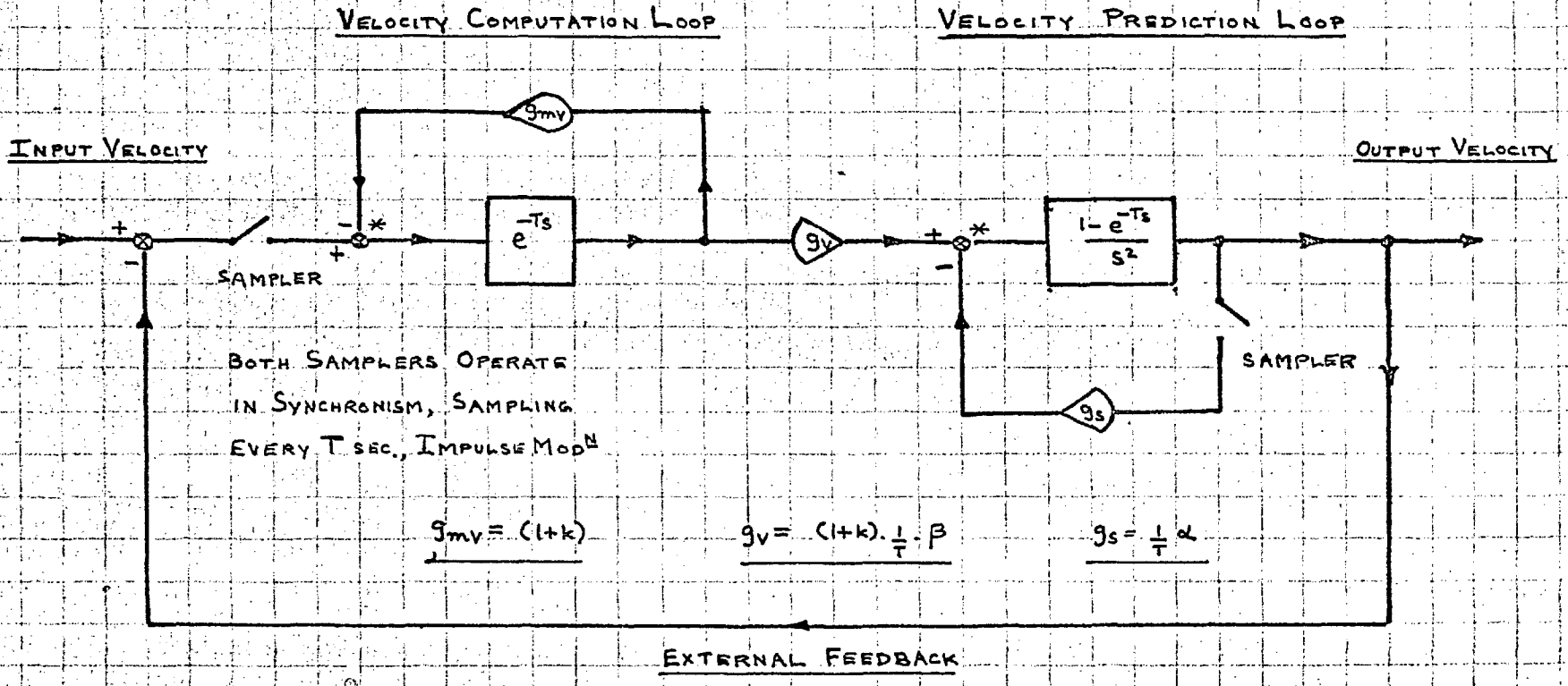
5.6. Linear Prediction of Velocity from Position Information.

The experimentally utilised continuous random signal possessed a negative cross-correlation between position and velocity for positive values of lag, as shown in Appendix II. The greatest cross-correlation occurred at a lag value of .4 sec. ; this indicated that position information could be used to improve velocity prediction. Such an improvement could arise as a result of:-

(a) The velocity prediction already proposed involved the use of only one past sample of velocity. The use of position information allowed some weight to be attached to further past samples, since position is the integral of velocity.

(b) The effect of position corrections made via velocity-triangles could be taken into account ; such an effect could not be

FIG. 5.5.1. FORM OF VELOCITY TRACKING LOOP INCORPORATING PREDICTION



* DENOTES SUBTRACTION OF AREAS OF IMPULSES

allowed for purely from past samples of output velocity.

In order to estimate input position, the operator would find it necessary to rely on kinaesthetic assessment of output position. Now consider the programme computation in the interval nT to (n+1)T. The operator can estimate his output position at time nT, $C_p(n)$, from kinaestheitc knowledge of $C_p(n-1)$ plus knowledge of $C_v(n-1)$ and programmes of hand movement released at time (n-1)T. He can then form an estimate of the input position at time nT, $R_p(n)$ by adding the position error observed at time nT, $E_p(n)$. The operators' processing of error and past velocity information to predict $R_v(n+2)$ and $R_p(n+2)$ has been considered in Sections 5.4. and 5.5., therefore the only additional information required by the operator in formulating programmes to be released at time (n+1)T is the estimate of $C_p(n)$. The appropriate modification of the model would therefore involve a cross branch running from overall position output to the velocity tracking loop.

In order to minimise the number of model parameters, it was decided to adopt a branch transfer consisting of a simple gain, subject to this proving satisfactory. The desiderata governing the choice of the point of entry of the position information to the velocity tracking loop were as follows:-

(a) There must be some processing time after each sample of position information becomes available.

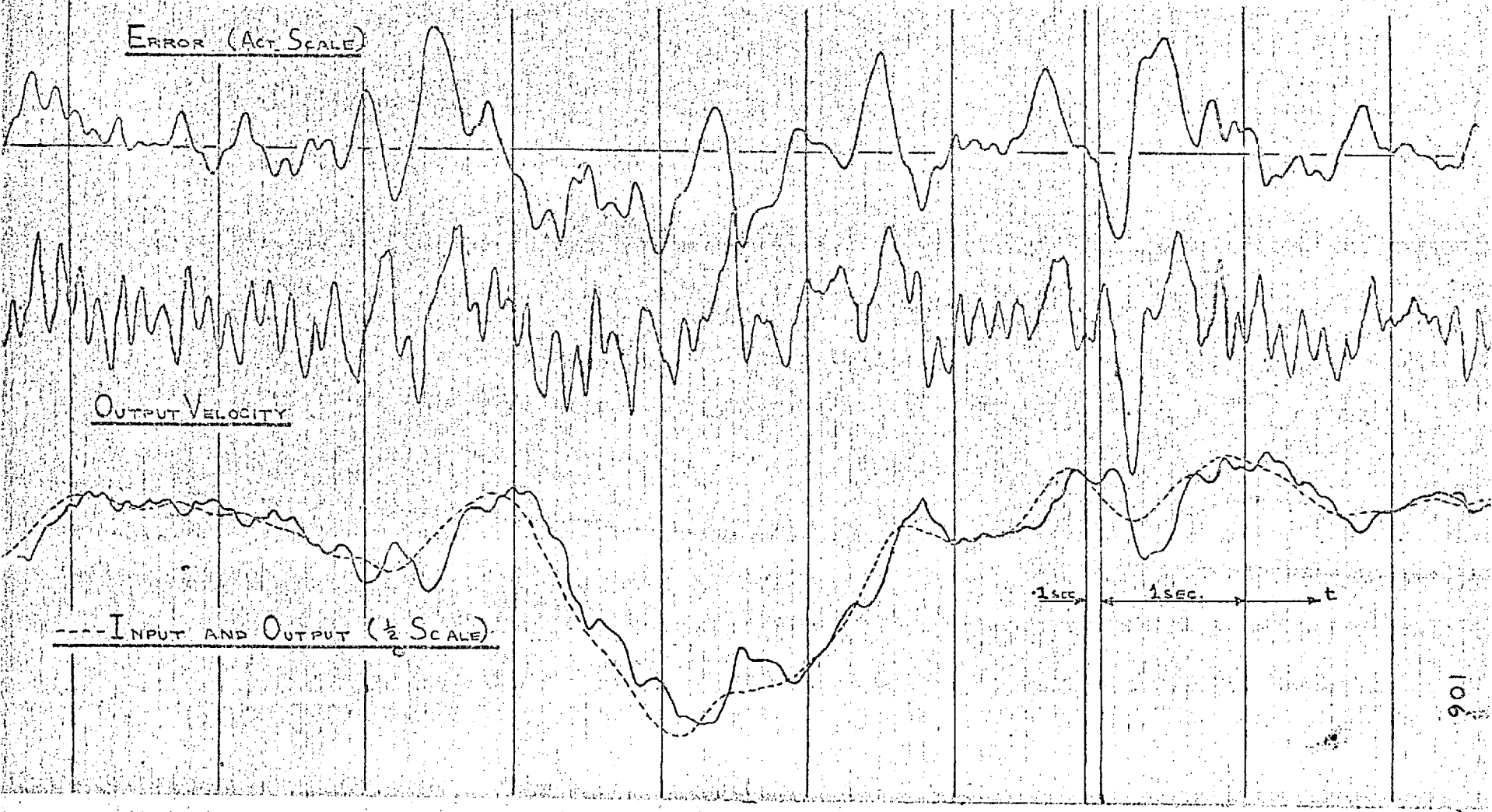
(b) The modification to the predicted velocity should give rise to a modification of position prediction from velocity.

(c) From close examination of human operator tracking records it was apparent that the operator did not allow a large standing error at points where the input remained relatively sluggish over a second or more. His output motion tended to contain one or several small excursions toward zero, followed by subsequent corrections. This behaviour is illustrated in Fig.5.6.1.

The desiderata outlined above indicated that the best scheme would be to connect the cross-branch through the sampler to the input of the velocity computation loop, and that the gain from position to velocity should be negative, to correspond with the sign of the cross-correlation between position and velocity for positive lags. In response to a steady input such a connection would give rise to a 'predictive ripple' corresponding to that described in desideratum (c) above ; a connection to the output of the loop would not give this effect. Also, if the latter connection were made position prediction would need to be realised by a further cross branch.

The chief significance of the cross-branch from sampled position output to velocity computation loop input, lay in its implication that the operator modified his presently observed velocity

FIG. 5.6.1. ILLUSTRATION OF 'PREDICTIVE RIPPLE': OPERATOR A, CONTINUOUS RANDOM INPUT



error, by subtracting from it a quantity proportional to his estimated position output. It was expected that the constant of proportionality would be relatively variable from operator to operator, and be strongly dependent on the statistical structure of the input. These prognostications were later substantiated by experimental results.

5.7. Experimental Evaluation of Model Modified to Allow for Prediction

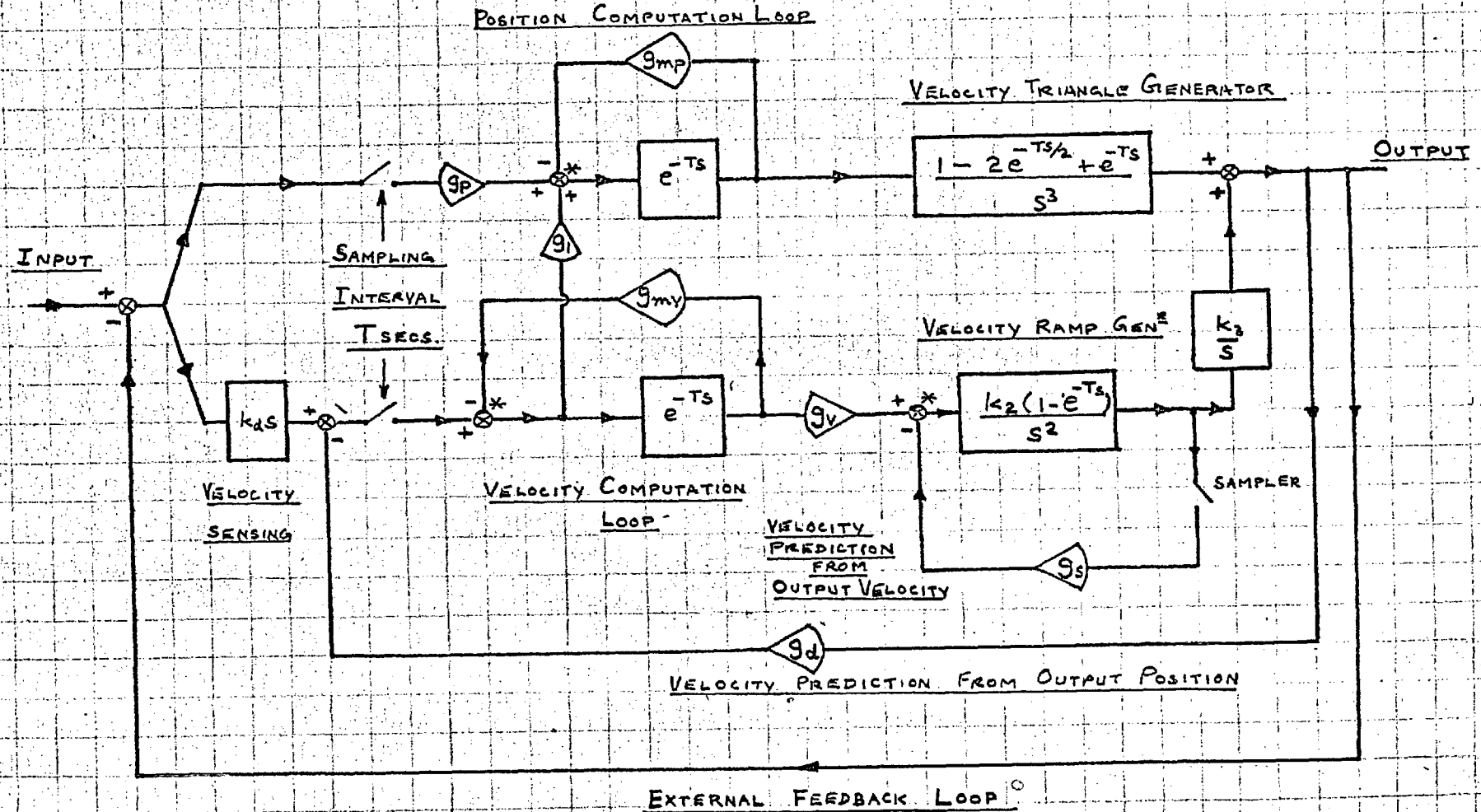
The structure of the analogue model was modified to allow for prediction as described in Sections 5.3., 5.4., 5.5., and 5.6. The corresponding theoretical model structure is shown in Fig.5.7.1. The experimental evaluation of this modified model was conducted in a similar fashion to that described in Section 5.2., using the continuous random signal with input spectrum cut-off at .61 cps. However, a more lengthy series of matching trials was conducted with the object of more thoroughly evaluating the degree to which the model was able to match the closed loop characteristics of operators. The quantitative criterion of comparison was again based on the computation of the variances of operator and model errors, and of their difference, over a run of 200 secs. duration.

The gains g_{mv} and g_{mp} were set to unity throughout, which left six variable model parameters - viz. the gains g_p, g_v, g_s, g_d, g_l, and the sample interval T. However, the initial choice of sampling interval was made on the basis of the sampled display experiments (except for operator D), and throughout subsequent parameter fitting trials T was regarded as a relatively 'fixed' parameter - changes made were small and infrequent.

Five operators participated in the series of matching trials ; each performed upwards of 30 runs. Operators A, C, F, and L had previously received quite extensive training, whereas operator D became available at a rather late date, after the conclusion of initial investigations. He was given an accelerated training programme of 20 runs, after which model fitting runs were commenced. In his case there was the subsidiary aim of trying to follow his characteristics during the later stages of training, in terms of the model parameters showing best fits over successive groups of runs.

The results of the model matching trials proved very gratifying. The qualitative fit between corresponding model and operator time traces was very good, and this was corroborated by several independent observers ; this fit is considered in more detail in Chapter 7. The quantitative results are summarised in Table 5.7.I. The best fits obtained all showed zero lag coefficients of cross-correlation between errors (defined in the special sense indicated in Section 5.2.) in excess of .8 and, in some cases, higher than .9 . The corresponding cross-correlations between outputs were greater than .98 . When considering the quality of fit represented by these figures, the following factors should be borne in mind:-

FIG. 5.7.1. STRUCTURE OF MODEL INCORPORATING PREDICTION



ALL SAMPLERS SYNCHRONOUS, FREQUENCY $\frac{1}{T}$ /SEC

* DENOTES SUBTRACTION OF AREAS OF IMPULSES

TABLE 5.7.1. SUMMARY OF QUANTITATIVE RESULTS OF MATCHING TRIALS

OPERATOR	A	C	D	F	L
OPERATOR WORST PERFORMANCE	7·9	5·6	3·8	4·7	6·0
OPERATOR PERFORMANCE BEST	12·5	8·7	6·6	9·0	11·2
OPERATOR-MODEL WORST ERROR CROSS-CORR^{N}	·59	·72	·84	·76	7·75
OPERATOR-MODEL ERROR CROSS-CORR^{N} BEST	·87	·85	·94	·91	·91
OPERATOR-MODEL WORST OUTPUT CROSS-CORR^{N}	·965	·948	·949	·966	·95
OPERATOR-MODEL OUTPUT CROSS-CORR^{N} BEST	·987	·978	·985	·986	·99
BEST FIT TRIAL Nº	28	19	58	38	92
OPERATOR PERFORMANCE	12·5	8·7	5·8	7·2	11·2
MODEL PERFORMANCE	12·5	8·4	5·8	7·2	10·4
O-M ERROR CROSS-CORR	·83	·84	·93	·91	·91
O-M OUTPUT CROSS-CORR	·987	·980	·985	·986	·990
2ND BEST FIT TRIAL Nº	22	12	44	37	95
OPERATOR PERFORMANCE	9·2	7·8	4·6	6·4	9·3
MODEL PERFORMANCE	8·3	8·9	4·6	7·2	10·4
O-M ERROR CROSS-CORR	·87	·84	·94	·88	·91
O-M OUTPUT CROSS-CORR	·984	·979	·984	·979	·989
3RD BEST FIT TRIAL No.	34	10	55	27	58
OPERATOR PERFORMANCE	10·2	6·8	5·1	7·9	9·1
MODEL PERFORMANCE	10·5	7·7	5·0	8·6	8·8
O-M ERROR CROSS-CORR	·84	·85	·90	·86	·91
O-M OUTPUT CROSS-CORR	·984	·977	·976	·986	·989

(a) The upper limit of zero lag cross-correlation between operator and model errors is set by the operator's inherent errors of estimation and execution. The variance of the operator-model error difference signal cannot be reduced below the variance due to these errors by any linear operation. Here it may be noted that Elkind (10) found that in a task similar to that presently under consideration, rather more than 1% of the output power could not be represented by a linear operation on the input. This would correspond to an upper limit of .99 in the coefficient of determination between outputs, and one might expect a typical value of the maximum attainable zero lag cross-correlation to be in the region of .99 .

(b) The characteristics of sampled data systems are non-stationary between sampling instants. Thus, if two such systems, which were identical in all respects except for the location of sampling instants, were to be compared on the same basis as model and operator, they would not show unity cross-correlation between errors. The expected cross-correlation would decrease as input bandwidth increased. Again, irregularity of sampling on the part of the operator could only serve to decrease cross-correlation.

(c) The model parameters were set before the commencement of a run ; the degree of fit obtainable therefore reflects the accuracy with which the operator's characteristics were anticipated by those of the model. Had the model been fitted after the run, an even closer representation could have been obtained.

The matching runs conducted with operator D showed that his characteristics did not change dramatically with practice, and it was possible to follow them by relatively small changes in model parameters. The most significant change was that of g_v, which increased by about 50% during the course of the trials. The implications of these results are discussed further in Section 9.6.

Consideration of the above results showed that the model was capable of very accurate simulation of those characteristics of the operator which could be represented by a linear operation on input. It was concluded that any improvement in fit which might be obtained through a greater complexity of linear operations would only be marginal. The introduction of random terms into the model, in order to represent errors made by the operator, would decrease the fit obtainable in terms of model-operator cross-correlation functions ; it could only serve to improve the comparative fit between the shapes of statistical functions between variates within the operator or model loops. The introduction of nonlinearities to improve fit was not considered to be justified, in view of the marginal nature of any likely improvement which they might effect, and the complication of theoretical analysis which would result.

The implications of the sets of values of the model parameters yielding the best experimental cross-correlations, were also considered

in terms of the model Z-transfer and Z-plane pole-zero plots. These were calculated from the results of theoretical analysis described in the following Chapter. Relevant discussion is presented in Chapters 7,8,and 9.

5.8. Summary

The model of the operator's ramp response developed as described in Chapter 4 was used as the basis of formulation of an analogue model. Direct comparison trials showed that this model was capable of simulating operators' ramp responses with very satisfactory accuracy. However,its simulation of continuous tracking was not very satisfactory. Consideration of how the operator might accomplish prediction led to the elaboration of the model structure by the addition of two subsidiary feedback loops,one from position output to velocity error,the other from velocity output to the output of the velocity computation loop. This analysis also provided a basis for selecting a value of velocity computation loop gain ; for the continuous random input used,this value was approximately unity.

The direct comparison of the more elaborate model with operators,under conditions of continuous tracking,showed that,with appropriate parameter settings,it was capable of achieving a very close match to operators' error and output functions - error cross-correlations of the order of .9 were achieved. A more elaborate comparison is presented in Chapter 7,after considering some of the implications of the model structure as revealed by Z-transform analysis presented in Chapter 6.

Chapter 6 : Z-TRANSFORM ANALYSIS OF THE SAMPLED DATA MODEL

6.1. Theoretical Analysis of the Sampled Data Model by Use of the Matrix Modified Z-Transform

Fig.5.7.1. shows the structure of the model derived as a result of the consideration of prediction presented in Sections 5.3. to 5.6. The corresponding diagram of pulse transfers is shown in Fig.6.1.1. Here the transfers are shown in terms of modified Z-transforms with notation as used by Jury (30). z represents the operator e^{Ts}, and m is a parameter lying between 0 and 1, so that times lying between sampling instants may be represented by the format $(n+1-m)T$. The values of transforms for m=0 and m=1 should, in general, be considered as the result of a limiting operation. The sampling process has been represented in terms of impulse modulation ; this involves some idealisation, because the analogue model could not sample instantaneously- in fact the sampling relay remained closed for 5 ms. The error of approximation is extremely small, however, due to the low-pass nature of the circuits following the samplers - effectively a triple integration. (The experimental realisation of the analogue model is described in Appendix V). The effect of the velocity tracking loop output on hand position is shown as a separate branch from the velocity predictor loop. This arrangement is necessary in order to avoid complications due to lack of a sampler between velocity and position outputs.

The representation of the velocity-triangle circuit requires the specification of two ranges of m, because there is a discontinuity of slope at $m = \frac{1}{2}$. The action of the circuit is illustrated and analysed in Appendix V. Its overall transfer may be represented as:-

$$P(z,m) = \tfrac{1}{4}T^2 \cdot \frac{1}{z(z-1)} \cdot \Lambda(z,m) \qquad 6.1.1.$$

where the function $\Lambda(z,m)$ assumes the appropriate form, according to whether $0 \leq m < \frac{1}{2}$ or $\frac{1}{2} \leq m < 1$. Also, the form of $\Lambda(z,m)$ is such that:-

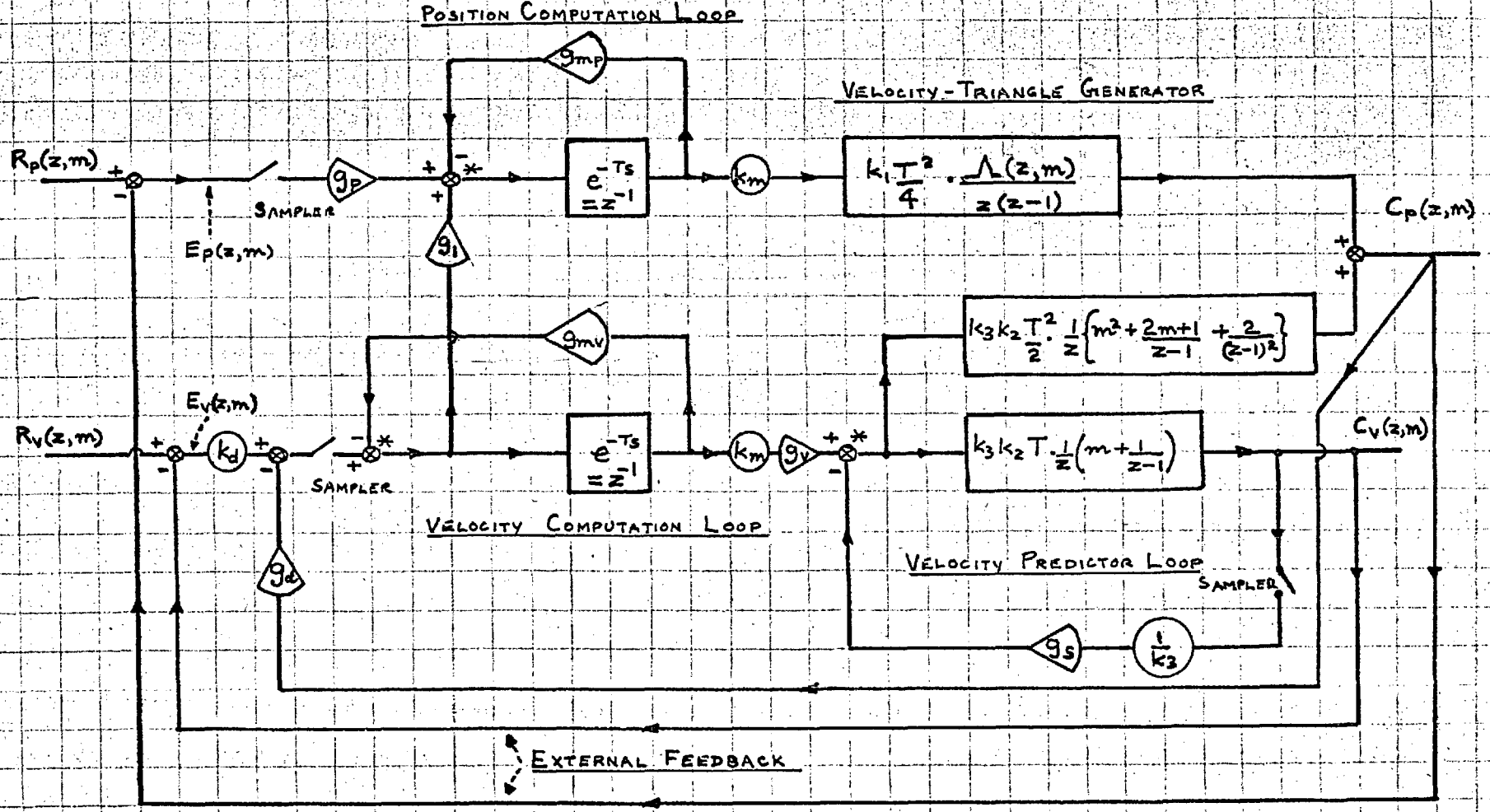
$$\lim_{m \to 0} \frac{\partial \Lambda(z,m)}{\partial m} = \lim_{m \to 0} 4m(z-1) = 0 \qquad 6.1.2.$$

$$\lim_{m \to 1} \frac{\partial \Lambda(z,m)}{\partial m} = \lim_{m \to 1} 4(1-m)(z-1) = 0 \qquad 6.1.3.$$

These equations signify that the output velocity is zero at sampling instants.

The analysis of the overall model was guided by the following reasoning. Reference to Fig.6.1.1. shows that the continuous error function at the input to the model can only elicit output response via the sample sequences derived from it and its derivatives.

FIG. 6.1.1. DIAGRAM OF EFFECTIVE PULSE TRANSFERS FOR MODEL INCORPORATING PREDICTION



*DENOTES SUBTRACTION OF AREAS OF IMPULSES

ALL SAMPLERS SYNCHRONOUS ; FREQUENCY $\frac{1}{T}$ /SEC

Also, the model's internal feedback from output and derivative of output is such that each feedback path effectively contains a sampler. Therefore the modified Z-transforms of the output will, in general, depend on the simple Z-transforms of both error and output and their derivatives, each associated with an appropriate modified Z-transfer function. This action is illustrated in Fig.6.1.2. which also displays the notation employed in the following analysis. With this notation the following equations may be written :-

$$C_p(z,m) = a_{pp}(z,m).E_p(z) + a_{pv}(z,m).E_v(z) + b_{pp}(z,m).C_p(z) + b_{pv}(z,m).C_v(z) \qquad 6.1.4.$$

$$C_v(z,m) = a_{vp}(z,m).E_p(z) + a_{vv}(z,m).E_v(z) + b_{vp}(z,m).C_p(z) + b_{vv}(z,m).C_v(z) \qquad 6.1.5.$$

where the form of the a's and b's may be determined through consideration of the pulse transfer diagram.

One may define a vector space in terms of the complex basis vectors e_p and e_v, such that the components of the vectors $\underline{R}(z,m)$, $\underline{E}(z,m)$, and $\underline{C}(z,m)$ are respectively :-

$R_p(z,m).e_p$ and $R_v(z,m).e_v$; $E_p(z,m).e_p$ and $E_v(z,m).e_v$;

$C_p(z,m).e_p$ and $C_v(z,m).e_v$. (symbols defined in Fig,6.1.2.)

By considering the linearity of the Z-transformation it may be shown that these vectors satisfy the requirements for a linear vector space. Therefore we may put equations 6.1.4. and 6.1.5. directly into matrix vector form, and with obvious notation write:-

$$\underline{C}(z,m) = \left[\underline{A}(z,m)\right].\underline{E}(z) + \left[\underline{B}(z,m)\right].\underline{C}(z) \qquad 6.1.6.$$

where the matrices $\left[\underline{A}(z,m)\right]$ and $\left[\underline{B}(z,m)\right]$ represent arrays of modified Z-transforms ; i.e. linear operators.

First, a relation between $\underline{C}(z)$ and $\underline{E}(z)$ may be derived by letting $m \to 1$ in equation 6.1.6., using the relation:-

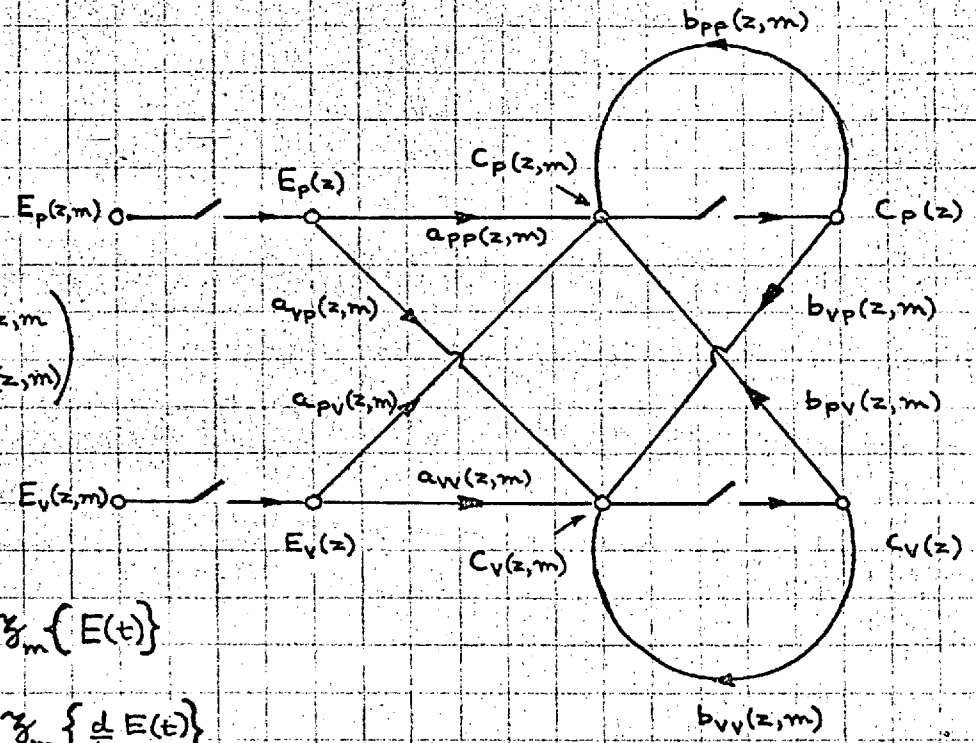
$$\lim_{m \to 1} \underline{C}(z,m) = \underline{C}(z) \qquad 6.1.7.$$

which holds if sudden jumps at sampling instants are excluded ; such jumps are not possible in the model output. Applying this relation gives:-

$$\underline{C}(z) = \left[\underline{A}(z,1)\right].\underline{E}(z) + \left[\underline{B}(z,1)\right].\underline{C}(z) \qquad 6.1.8.$$

from which we may derive the following equation:-

FIG. 6.1.2. REPRESENTATION OF THE ACTION OF THE MODEL IN TERMS OF MODIFIED Z-TRANSFERS



$$\underline{E}(z,m) = \begin{pmatrix} E_p(z,m \\ E_v(z,m) \end{pmatrix} \qquad \underline{C}(z,m) = \begin{pmatrix} C_p(z,m) \\ C_v(z,m) \end{pmatrix}$$

$$E_p(z,m) = \mathcal{Z}_m\{E(t)\} \qquad C_p(z,m) = \mathcal{Z}_m\{C(t)\}$$

$$E_v(z,m) = \mathcal{Z}_m\left\{\frac{d}{dt}E(t)\right\} \qquad C_v(z,m) = \mathcal{Z}_m\left\{\frac{d}{dt}C(t)\right\}$$

$E(t)$ = CONTINUOUS ERROR FUNCTION $\qquad$ $C(t)$ = CONTINUOUS OUTPUT FUNCTION

ALSO, FOR CONTINUOUS LOOP INPUT $R(t)$, WE DEFINE $\underline{R}(z,m) = \begin{pmatrix} R_p(z,m) \\ R_v(z,m) \end{pmatrix} = \begin{pmatrix} \mathcal{Z}_m\{R(t)\} \\ \mathcal{Z}_m\left\{\frac{d}{dt}R(t)\right\} \end{pmatrix}$

$$[I - B(z,1)].\underline{C}(z) = [A(z,1)].\underline{E}(z) \qquad 6.1.9.$$

where I is the identity matrix. Provided that $[A(z,1)]$ and $[I - B(z,1)]$ are non-singular, the following expressions may be written:-

$$\underline{C}(z) = [I - B(z,1)]^{-1}.[A(z,1)].\underline{E}(z) \qquad 6.1.10.$$

and

$$\underline{E}(z) = [A(z,1)]^{-1}.[I - B(z,1)].\underline{C}(z) \qquad 6.1.11.$$

Substituting back from equations 6.1.9. and 6.1.10. into equation 6.1.6. results in the following expressions :-

$$\underline{C}(z,m) = [A(z,m)].\underline{E}(z) + [B(z,m)][I - B(z,1)]^{-1}.[A(z,1)].\underline{E}(z) \qquad 6.1.12.$$

and

$$\underline{C}(z,m) = [A(z,m)][A(z,1)]^{-1}.[I - B(z,1)].\underline{C}(z) + [B(z,m)].\underline{C}(z) \qquad 6.1.13.$$

Equations 6.1.12. and 6.1.13. indicate that $\underline{C}(z,m)$ can be derived directly from $\underline{C}(z)$ or $\underline{E}(z)$, so long as $[A(z,1)]$ and $[I - B(z,1)]$ are not simultaneously singular. Such singularity would only occur under conditions where the effective structure of the model is drastically changed.

Furthermore, from a consideration of the output response at sampling instants, it may be seen that 'hidden oscillations' (i.e. oscillations which are not sensible at sampling instants) cannot occur within the system. This statement may be substantiated as follows. 'Hidden oscillations' would involve the simultaneous satisfaction of the following conditions :-

$$C_p(nT,m) = 0 \quad ; m = 0 \text{ or } 1 \qquad 6.1.14a.$$

$$C_p(nT,m) \neq 0 \quad ; 0<m<1 \qquad 6.1.14b.$$

$$C_v(nT,m) = 0 \quad ; m = 0 \text{ or } 1 \qquad 6.1.14c.$$

$$C_v(nT,m) \neq 0 \quad ; 0<m<1 \qquad 6.1.14d.$$

Now the velocity-triangle generating circuit cannot satisfy requirements a, b, and d simultaneously. Moreover, the velocity tracking system can only change velocity monotonically in any one sampling interval ; therefore it is unable to satisfy conditions c and d simultaneously. Therefore the model output cannot satisfy all of the above conditions simultaneously.

The above reasoning indicates that it is sufficient to consider $\underline{C}(z)$ to determine stability . The condition for this is that the elements of $\underline{C}(z)$ have no poles outside the unit circle $|z| = 1$.

It is impossible for $\underline{E}(z)$ to have poles other than those of $\underline{C}(z)$ and $\underline{R}(z)$, since it is formed from a linear combination of the latter quantities.

6.2. Derivation of the Model Z-transfers

The forward (open-loop) modified Z-transfer of the model is given by equations 6.1.12. and, at sampling instants, 6.1.10. The former equation may be written in the form:-

$$\underline{C}(z,m) = \frac{1}{h(z)} . [H(z,m)] . \underline{E}(z) \qquad 6.2.1.$$

where $h(z)$ is the L.C.D. of the elements of the matrix $[I-B(z,1)]^{-1}[A(z,1)]$ The common poles of the forward transfer are given by the roots of $h(z)$, and the zeros are given by the elements of $H(z,m)$. These may be determined by direct inspection of the pulse transfers diagram, Fig.6.1.1., and substitution for the elements of $A(z,m)$ and $B(z,m)$ in the relation:-

$$\frac{1}{h(z)} . [H(z,m)] = [A(z,m)] + [B(z,m)][I - B(z,1)]^{-1} . [A(z,1)] \qquad 6.2.2.$$

To obtain the response at sampling instants only, we let $m \to 1$ and obtain:-

$$\frac{1}{h(z)} . [H(z,1)] = [I - B(z,1)]^{-1} . [A(z,1)] \qquad 6.2.3.$$

To derive the closed loop input-output Z-transfer matrix, we first note that the external feedback loop has a transfer of -1, and therefore:-

$$\underline{E}(z,m) = \underline{R}(z,m) - \underline{C}(z,m) \; ; 0 \leqslant m \leqslant 1 \qquad 6.2.4.$$

Combining equations 6.2.4. (with $m = 1$) and 6.1.9. we derive:-

$$[I - B(z,1)] . \underline{C}(z) = [A(z,1)] . \{\underline{R}(z) - \underline{C}(z)\} \qquad 6.2.5.$$

whence :-

$$\underline{C}(z) = [I + A(z,1) - B(z,1)]^{-1} . [A(z,1)] . \underline{R}(z) \qquad 6.2.6.$$

subject to the matrix $[I + A(z,1) - B(z,1)]$ being non-singular. $\underline{C}(z,m)$ may now be derived from equations 6.2.6. and 6.1.13.

Thus we may write the following :-

$$\underline{C}(z,m) = \frac{1}{q(z)}.[Q(z,m)].\underline{R}(z) \qquad 6.2.7.$$

where $q(z)$ is the L.C.D. of the matrix $[I + A(z,1) - B(z,1)]^{-1}.[A(z,1)]$ and:-

$$\frac{1}{q(z)}.[Q(z,m)] = \left\{[A(z,m)][A(z,1)]^{-1}.[I - B(z,1)] + [B(z,m)]\right\} \times [I + A(z,1) - B(z,1)]^{-1}.[A(z,1)] \qquad 6.2.8.$$

The common poles of the overall closed loop transfer are given by the roots of $q(z)$, and the zeros are given by the elements of $[Q(z,m)]$ These may be determined as described for the open loop transfer.

The closed loop transfer at sampling instants is given by:-

$$\frac{1}{q(z)}.[Q(z,1)] = [I + A(z,1) - B(z,1)]^{-1}.[A(z,1)] \qquad 6.2.9.$$

The closed loop input-error Z-transfer matrix may be obtained by combining equations 6.2.4. and 6.2.7., which gives:-

$$\underline{E}(z,m) = \underline{R}(z,m) - \frac{1}{q(z)}.[Q(z,m)].\underline{R}(z) \qquad 6.2.10$$

The input-error transfer at sampling instants may be obtained from equations 6.2.4. and 6.2.9. :-

$$\underline{E}(z) = \frac{1}{q(z)}.[q(z).I - Q(z,1)].\underline{R}(z) \qquad 6.2.11.$$

Equations 6.2.7. and 6.2.11. show that $\underline{E}(z)$ and $\underline{C}(z)$ have the same set of common poles, as would be expected from the form of the feedback.

Closed loop instability can only occur if $q(z)$ has roots with moduli greater than unity. The roots of this polynomial may be determined conveniently by use of the Newton-Raphson iteration or of Newton sums. These methods are described in Appendix VI.
In order to determine only whether the closed loop transfer is stable it is not absolutely necessary to factorise $q(z)$. Stability may be investigated by the companion matrix method or through the bilinear transformation $z = (w + 1)/(w - 1)$, and subsequent application of Routh's criterion. Still another alternative is the construction and examination of Sturm sequences.

The procedure for evaluating the model Z-transfer matrices was first to evaluate the elements of matrices $[A(z,m)]$ and $[B(z,m)]$ and thence to calculate out the matrix inversions and products giving $h(z)$, $[H(z,1)]$, $q(z)$, $[Q(z,1)]$; the elements of matrices $[H(z,m)]$ and $[Q(z,m)]$ were not utilised, because they were rather intractable and yielded little further information, since the output response during sampling intervals was completely determined by the response

vector at sampling instants.

Performing the above calculations showed that both $q(z)$ and $h(z)$ involved powers of z up to the fourth. The elements of the associated matrices each contained no more than two zeros. Thus,e.g., an expression for position output involved one set of four poles plus two sets of two zeros. Explicit expressions for the transfers in terms of model parameters are given in Appendix VI.

6.3. Calculation of the Closed Loop Frequency Domain Characteristics of the Model

It was desired to calculate the frequency domain characteristics of the model with the objects of comparing these with results obtained by other investigators,and of gaining an insight into the nature of the model's frequency response. The theoretical basis for this calculation is given in Appendix XI.

The most accurate calculation of frequency response involves taking account of the response of the sampled data system over the whole sampling interval. This is obtained either by putting $s = jw$ in the Laplace transform of the continuous output,or by considering the modified Z-transform ,as in the following expression :-

$$G(jw) = \int_0^1 G(z,m)\Big|_{z=e^{jwT}} . e^{jwT} . e^{-jwmT} . dm \qquad 6.3.1.$$

This corresponds to representing the continuous output as the resultant of a sum of infinitesimal sinusoidal components.

The most accurate way of calculating the model's overall frequency response was through the utilisation of equation 6.3.1., as extended to cover the case where $G(jw)$ and $G(z,m)$ are matrix transfers. An equivalent method was to construct the matrix Laplace transfer corresponding to the matrix modified Z-transfer $\frac{1}{q(z)} . [Q(z,m)]$ and thence to derive the frequency response directly. Unfortunately,both procedures were extremely cumbersome.
The form of the output during sampling intervals suggested that very good accuracy could be obtained by considering only that sinusoid which was coincident with the position output at sampling instants. Experiment and calculation served to confirm this view.

The output of the position and velocity tracking circuits was considered in terms of the magnitude and phase of:-

$$\frac{\text{Sinusoid fitted to position output over sampling interval}}{\text{Sinusoid fitted to position output at sampling instants}} = \frac{S_a}{S_f}$$

In the case of the velocity-triangle position response this expression gives:-

$$\frac{S_a}{S_f} = \left| \frac{(\sin wT/4)^2 . (\sin wT/2)}{(wT/4)^2 . (wT/2)} \right| \angle 0 \qquad 6.3.2.$$

This expression shows that the error of approximation is a small overestimate in magnitude at frequencies of interest - about 1 dB. at 1.7 cps. Phase error is zero.

In the case 66 the position output due to the velocity tracking circuit, the following expression holds:-

$$\frac{S_a}{S_f} = \left| \frac{(\sin wT/2)^2 . (\tan wT/2)}{(wT/2)^3} \right| \angle 0 \qquad 6.3.3.$$

This expression shows that the error of approximation is a very small underestimate at frequencies below 2 cps. ; at 2.5cps the error is about 2 dB. The rapid increase of error as $wT \rightarrow \pi$ results from the presence of a zero at $z = -1$ in the simple z-transfer of the circuit. The ripple component at 1/2T cps. is not shown by the values of the output at sampling instants. The phase error is zero at all frequencies.

Experiment showed that the overall position output of the model could be closely represented by considering only the response at sampling instants, for frequencies below about 3 cps. Accuracy deteriorated at higher frequencies ; the model introduced ripple components at frequencies of 1/2T and 1/T cps., so that representation in terms of frequency response tended to become a rather inadequate description of model characteristics. The importance of these ripple components is discussed in Section 6.4.

Consider the model's position response at sampling instants to a complex sinusoidal input, of the form e^{jwT} .
We have:-

$$\underline{R}(z) = \frac{z}{z - e^{jwT}} \cdot \left| \frac{1}{jw} \right| \qquad 6.3.4.$$

and

$$\underline{C}(z) = \frac{z}{z - e^{jwT}} \cdot \frac{1}{q(z)} \cdot [Q(z,1)] \cdot \left(\frac{1}{jw} \right) \qquad 6.3.5.$$

whence:-

$$C_p(z) = \frac{z}{z - e^{jwT}} \cdot \frac{1}{q(z)} \cdot [q_{pp}(z,1) + jw.q_{pv}(z,1)] \qquad 6.3.6.$$

so that, by virtue of the theory presented in Appendix XI, we may derive the following expression for the apparent closed loop frequency response of the model from position input (to the loop) to position output :-

$$C_p(jw) = \frac{1}{q(e^{jwT})} . [q_{pp}(e^{jwT},1) + jw.q_{pv}(e^{jwT},1)] \qquad 6.3.7.$$

This response is the equivalent of that derived by other workers in connection with simple linear continuous models.

The equivalent forward, or open loop, response, $E_p(jw)$, is given by considering the relation between the sinusoids fitted (at sampling instants) to the position error and position output sequences. Due to the unity negative feedback, a relation of the following form holds:-

$$E_p(jw) = 1 - C_p(jw) \qquad 6.3.8.$$

For reasons presented above, the spectral density of the continuous model output may be approximated by considering the spectral density of the infinite sum of infinitesimal sinusoids fitted to the position output at sampling instants. This spectral density is given by:-

$$\Phi_{cc}(w) = C_p(-jw).C_p(jw).\Phi_{ii}(w) \qquad 6.3.9.$$

where $\Phi_{ii}(w)$ is the spectral density of the loop input.

A similar expression holds for the approximate error spectral density.

The above methods of calculating frequency responses and spectral densities are extremely convenient, in that they involve only substition into relatively simple functions of z. They do suffer from some inaccuracy, in that they ignore response between sampling instants ; however, as indicated above, this inaccuracy is very small at frequencies of interest. The convenience of the methods led to their use in the calculation of results presented in Chapters 7 and 8.

A more exact method of calculation of the position output spectral density is provided by considering the leading element in the matrix of modified pulse spectral densities given by:-

$$\left[\Phi_{oo}(z,m)\right] = \frac{1}{q(z^{-1}).q(z)} . \left[Q(z^{-1},m)\right]\left[\Phi_{ii}(z)\right]\left[Q(z,m)\right]^T \qquad 6.3.10$$

where $\left[\Phi_{ii}(z)\right]$ is the input spectral density matrix in pulsed form. T = transpose

Unfortunately, this method is cumbersome to the point of intractability and was not used for the purposes of calculation. If the parameter m is set equal to 1 in equation 6.3.10., the resultant expression gives a result exactly equivalent to equation 6.3.9.

6.4. Calculation of the Forward Sequence Frequency Domain Characteristics of the Model

Care is needed in defining the forward transfer of the model, since the relation between the components of the error vector, $E_p(z)$ and $E_v(z)$, differs according to whether the feedback loop is open or closed.

To obtain the open loop frequency response one may postulate a complex sinusoidal error signal, such that :-

$$\underline{E}(z) = \frac{z}{z - e^{jwT}} \cdot \begin{pmatrix} 1 \\ jw \end{pmatrix} \qquad 6.4.1.$$

Then, from equation 6.2.1., and by analogy with equation 6.3.7., the relative magnitude and phase of the sinusoid fitted to the position output at sampling instants is given by :-

$$C_p^f(jw) = \frac{1}{h(e^{jwT})} \cdot \left[h_{pp}(e^{jwT},1) + jw.h_{pv}(e^{jwT},1) \right] \qquad 6.4.2.$$

which is the frequency response defined from position error to position output, under conditions where the feedback loop is open.

An alternative definition of frequency response from position error to position output may be obtained by considering the feedback loop to be closed and the input vector to be defined as in equation 6.3.4. ; i.e. a complex sinusoidal function defined from $R_p(t) = e^{jwt}$. The sinusoids fitted to $E_p(nT)$ and $C_p(nT)$ are related in magnitude and phase to the input sinusoid, and to each other, as indicated in equations 6.3.7. and 6.3.8. ; i.e. by $E_p(jw)$ and $C_p(jw)$. Then the required closed loop equivalent forward frequency response, $F_p(jw)$, is given by :-

$$F_p(jw) = C_p(jw) \; / \; E_p(jw) \qquad 6.4.3.$$

The direct expression for $F_p(jw)$ is, from equations 6.3.7. and 6,3.8. :-

$$F_p(jw) = \frac{q_{pp}(e^{jwT},1) + jw.q_{pv}(e^{jwT},1)}{q(e^{jwT}) - q_{pp}(e^{jwT},1) - jw.q_{pv}(e^{jwT},1)} \qquad 6.4.4.$$

Comparison of equations 6.4.2. and 6.4.4. clearly shows the difference between the two definitions of forward transfer. Previous workers have always measured $F_p(jw)$; calculated plots of this function are given in Chapter 7.

The difference between the two definitions of frequency response arises because of the fact that ,when operating under closed loop conditions the relation :-

$$E_v(z=e^{jwT}) = jw.E_p(z=e^{jwT}) \qquad 6.4.5.$$

which is true for open loop operation,does not hold. The difference arises through the effect of ripple at the model output,which does not affect $C_p(nT)$,but does affect $C_v(nT)$ and hence $E_v(nT)$.

To exemplify the effects of ripple,the steady state response to a step input may be considered. Under closed loop conditions,we have,from equations 6.2.11. and 6.3.4. :-

$$\underline{E}(z) = \frac{z}{(z-1).q(z)} . \left[q(z).I - Q(z,1)\right] . \begin{pmatrix}1\\0\end{pmatrix} \qquad 6.4.6.$$

The final value theorem may now be applied to obtain :-

$$\lim_{n\to\infty} \underline{E}(nT) = \lim_{z\to 1} (z-1).\underline{E}(z)$$

$$= \frac{1}{q(1)} . \left[q(1).I - Q(1,1)\right] . \begin{pmatrix}1\\0\end{pmatrix} \qquad 6.4.7.$$

The above equation shows that there is a steady state velocity error at sampling instants,given by :-

$$\lim_{n\to\infty} E_v(nT) = - q_{vp}(1,1) \,/\, q(1) \qquad 6.4.8.$$

The form of the output between sampling instants could be derived by using equation 6.1.12. However,this method is extremely cumbersome. An alternative,and more enlightening way of calculating the form of the output between sampling instants is to use knowledge of the form of the output directly. From equation 6.2.7. (with m = 1) we may derive:-

$$\lim_{n\to\infty} \underline{C}(nT) = \frac{1}{q(1)} . \begin{pmatrix} q_{pp}(1,1) \\ q_{vp}(1,1) \end{pmatrix} \qquad 6.4.9.$$

This shows that output velocity at sampling instants is steady,but not zero. Therefore there can be no change of output velocity during the sampling interval due to the action of velocity tracking.Output position is also steady at sampling instants,therefore the programmed change of position must exactly cancel the change of position due to the steady velocity tracking circuit output,at sampling instants.

If the output of the velocity tracking circuit is v ,then the change of position which would occur over one sampling interval, in the absence of position tracking output, would be vT. Therefore the position tracking output over one sampling interval must be -vT. Thus,at time mT after the commencement of the sampling interval, we have:-

$$\begin{aligned} D_p(m) &= vT.(m - 2m^2) && 0 \leq m < \tfrac{1}{2} \\ &= vT.(m - 1 + 2.\overline{1 - m}^2) && \tfrac{1}{2} \leq m < 1 \end{aligned} \qquad 6.4.10.$$

where $D_p(m) = C_p(nT,m) - C_p(nT)$

The above equations show that the output between sampling instants is a 'double-parabolic' ripple, which may be closely approximated by a sinewave of frequency 1/T cps. and amplitude vT/8 . From equation 6.4.9. it may be seen that the relative amplitude of this position output ripple to the absolute magnitude of the position output at sampling instants is given by:-

$$\frac{\underset{m}{\text{Max}}\left|D_p(m)\right|}{\left|C_p(nT)\right|} = \frac{T}{8}.\left|\frac{q_{vp}(1,1)}{q_{pp}(1,1)}\right| \qquad 6.4.11.$$

Substitution of typical values of model parameters in the R.H.S. of equation 6.4.11. indicates that the relative amplitude of ripple is about ½% ; even relative to the error ,the ripple only represents about 8% in terms of amplitude and .6% in terms of power.

Though,as shown above,the relative amplitude of the ripple is small,it nevertheless has an important effect on response. To exemplify the effect of ripple on the steady state response to a step input,consider the case where the feedback loop is opened,and the model is fed with a steady position error signal,having zero velocity,but otherwise exactly equal to that generated within the closed loop,at sampling intervals. The steady output response would then be only about half of that generated in the closed loop configuration.

The results obtained in this Section indicate that care is needed in defining the forward response of the operator. According to the model,different responses are obtained,depending on whether the equivalent open loop (forward) response,or the actual open loop response from position input (error signal) to position output is considered. The difference arises because of the presence of ripple in the output and the separate sensing of both position and velocity errors at sampling instants. Previous workers have effectively measured the equivalent open loop response ; one of the aims of the present study was to measure the actual open loop response directly.

The example of the model's response to a step input shows that the power of the ripple component is quite small when compared with that of the output or the error functions, yet it accounts for nearly half the amplitude of the steady output response. These considerations indicate that the customary dismissal of those higher frequency terms in the operator's output not linearly correlated with the loop input, as 'remnant' or 'noise', without effect on his low frequency performance, warrants more careful examination. The simulation of, and effects of this ripple component are considered further in Chapters 7 and 8.

Chapter 7 : COMPARISONS OF THE CLOSED LOOP CHARACTERISTICS OF OPERATORS WITH THOSE OF FITTED MODELS

7.1. Introduction

Chapter 5 has described the evolution of a sampled data model incorporating prediction, capable of closely matching the tracking characteristics of human operators. The purpose of the present chapter is to present further evaluation of the quality of match afforded by this model, under conditions of closed loop operation.

First, actual time traces of operator and model response to common random ramp and continuous random inputs are considered. The implications of best-fit model parameter settings are discussed in terms of Z-plane pole-zero plots.

Next, covariance functions computed from records of 100 secs. duration, of operator and model input and error functions, are considered.

Finally, computed closed loop spectral characteristics of the model are compared with operators' spectral characteristics, as derived by other workers.

7.2. Real Time Comparison of Operators' and Models' Responses to Common Random Ramp Inputs

The purpose of fitting the model, as developed to incorporate prediction, to operators' responses to random ramp inputs, was primarily to demonstrate the capability of the model to match operators' responses over a wide range of inputs. These responses also showed the effects of ripple somewhat more clearly than similar responses to a continuous random input.

Actual time traces of operators' ramp responses, and of the corresponding fitted model responses, are shown in Figs.7.2.1.,2.,& 3. Inspection of these time traces shows the good degree of fit between the qualitative features of operator and model error and output functions. The output velocity functions also fit quite well, except in the case of operator F, where the fit is moderately good; this appears to be the result of his applying a deliberate 'filtering' action to hand motion, which also increases his delay in response to the onset of a ramp. Finally, it may be noted that there is considerable similarity between the 'ripple' in the model and operator outputs.

Differences between operator and model responses would be expected to arise through errors of estimation and execution, and random sampling on the part of the operator. Also, the model is idealised, and gives sharp output velocity triangles; the operator cannot achieve a step change of output acceleration, so that there is

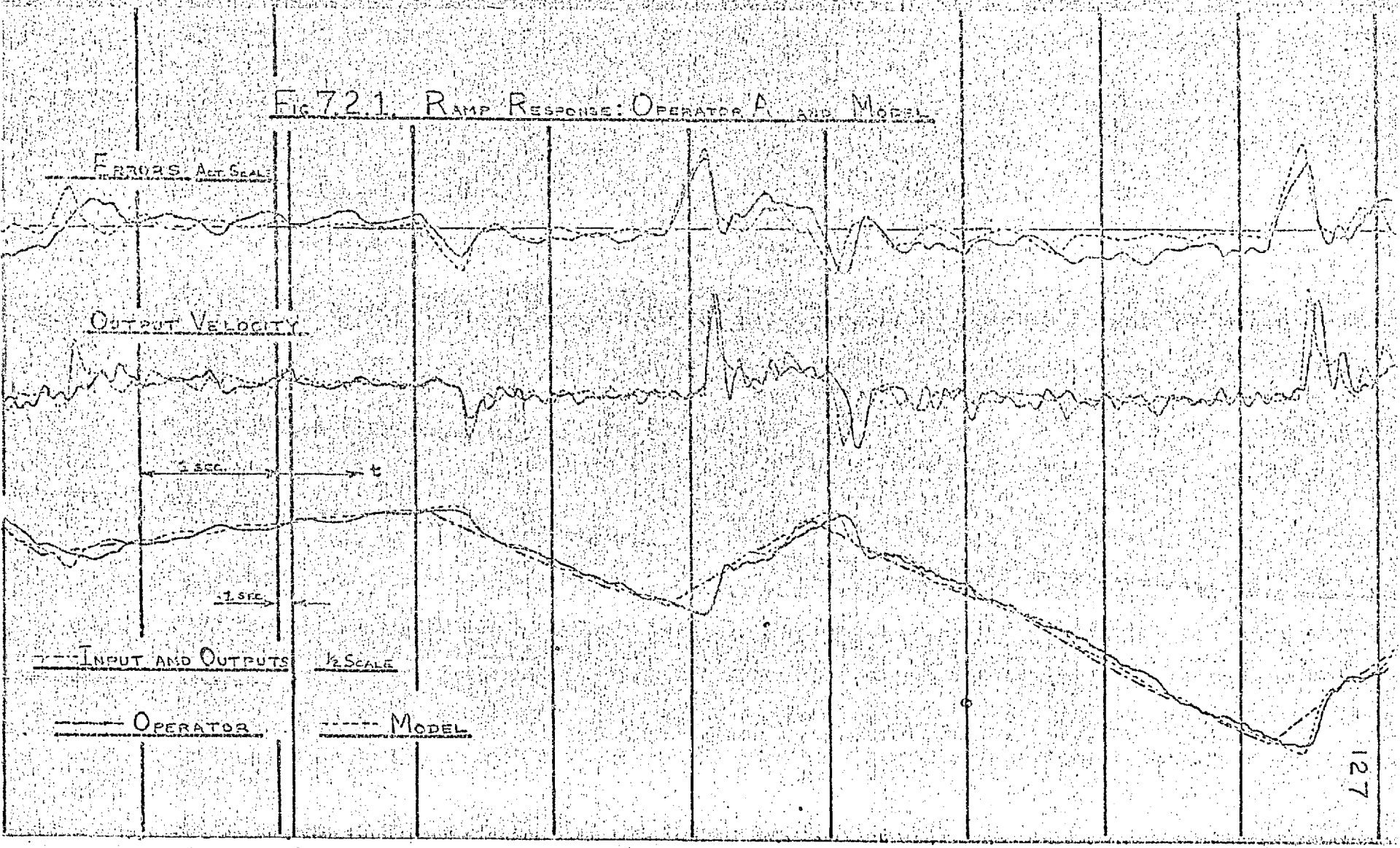


Fig. 7.2.1. Ramp Response: Operator A and Model

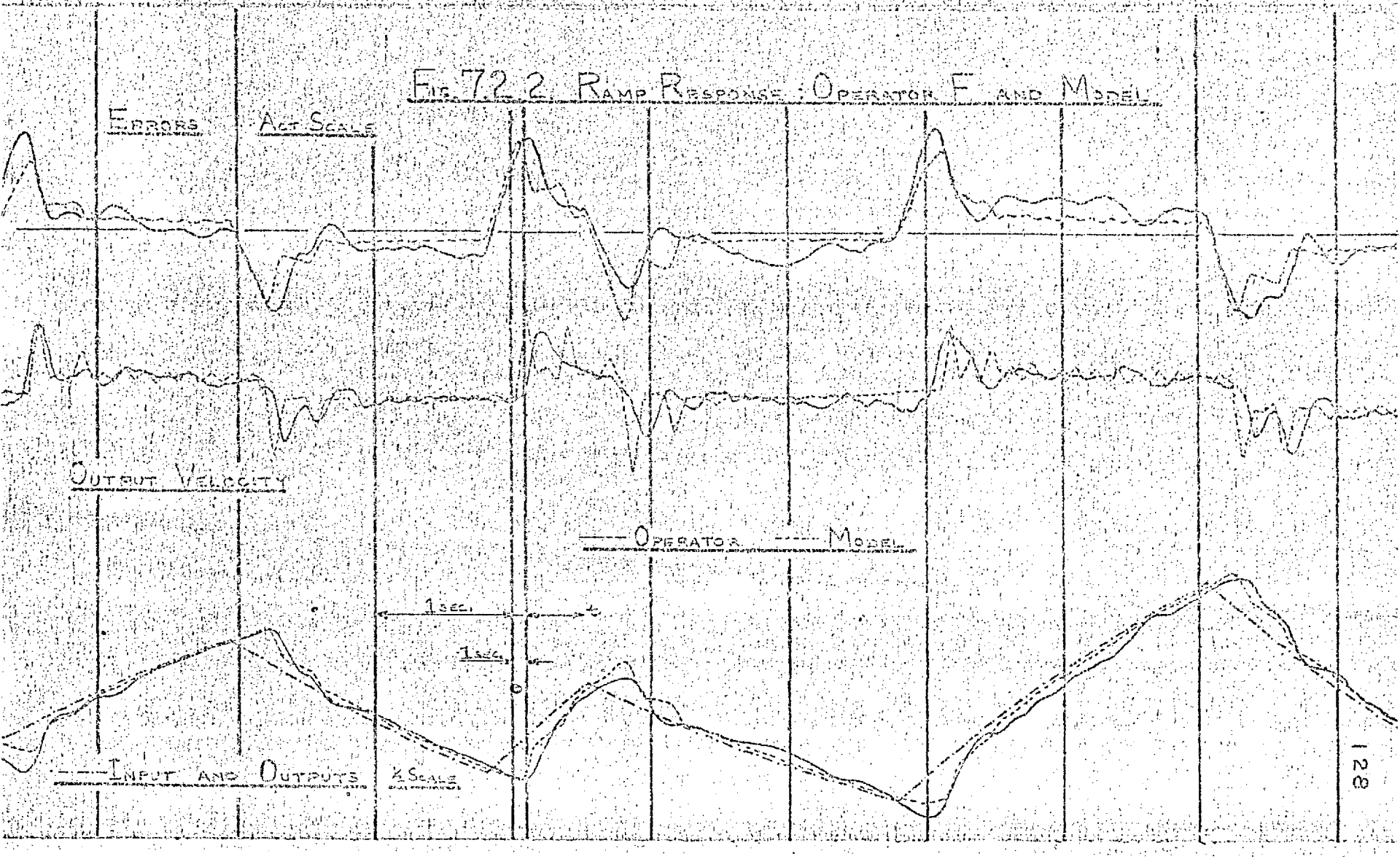


Fig. 7.2.2 Ramp Response: Operator F and Model
Errors
Act Scale
Output Velocity
Operator
Model
1 sec.
1 sec.
Input and Outputs
½ Scale

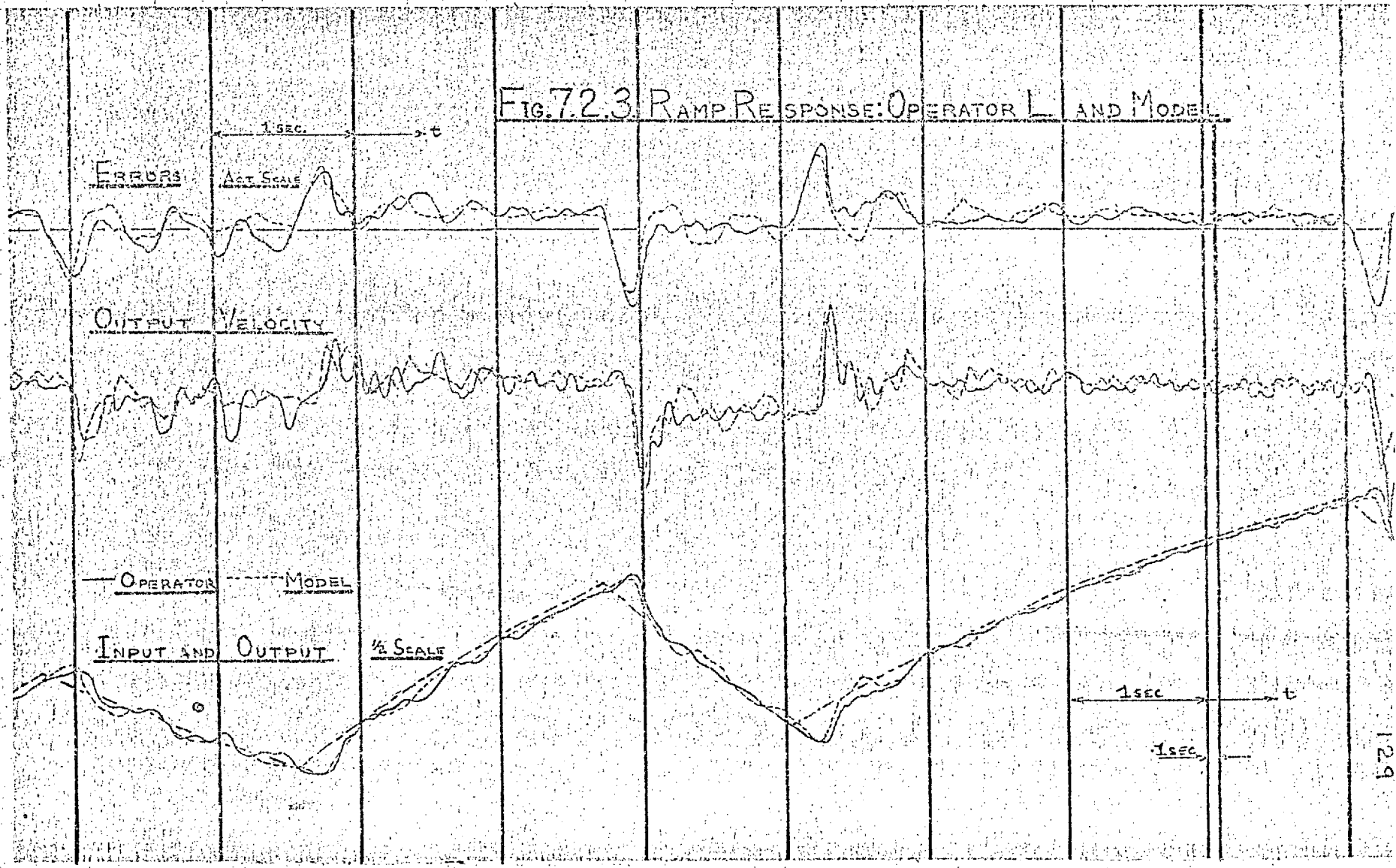


Fig. 7.2.3 Ramp Response: Operator L and Model

an inevitable filtering action tending to blur the basic velocity-triangle.

The values of best-fit model parameters were much as expected, considering the nature of the random ramp input (model parameters are shown in Fig.5.7.1.). Thus g_d was zero and g_s was relatively small. The overall closed loop Z-transfers of the best-fit models are shown in Fig.7.2.4., in terms of Z-plane pole-zero plots. These plots illustrate the well-damped nature of the model response. All poles lie inside the circle $z = .54$, except in the case of operator L. Here, poles with $z = .8$ are cancelled almost exactly by zeros of $q_{pp}(z,1)$, and less effectively by zeros of $q_{pv}(z,1)$; hence the main effect is to give a moderately damped oscillatory response of velocity, at a frequency of about 1/4T cps.- about 1.8 cps.

7.3. Consideration of Continuous Tracking Response of Operators and Model

The quantitative comparison of operator and model response to a common continuous random signal input has been described in Section 5.7. Simultaneous recordings were made of operator and model error, output position, and output velocity, with the model parameters set to the appropriate values showing the best quantitative fit during matching trials. These parameters' values are given in Table 7.3.I. Recordings were obtained for all five operators, and typical sections of these are shown in Figs.7.3.1-5.

It may be seen that the qualitative fit of error and output functions is very good in all cases. In particular, the model produces a realistic simulation of the ripple in the corresponding operator functions, both as regards frequency and amplitude. This could not be obtained from linear continuous models, as proposed in the literature, unless special injected 'noise' functions were specified. These noise functions would have to be 'tailored' to fit each case, and would be based on a 'brute-force' matching technique. Therefore they could not easily be related to the innate structure of the operator.

A more searching test is provided by the comparison of output velocity traces. Here again the simulation is generally very good. There is a 'rounding' of the operators' velocity traces, due to the finite rate of change of acceleration imposed by mechanical components in the arm ; however, the model as proposed is only capable of an idealised representation of the operator's output velocity, in terms of straight line segments. In the case of operator D the simulation is only fair ; the operator's output velocity varies much more smoothly than that of the model. Operator F shows the same tendency, but to a smaller extent. Inspection of the traces shows that the effect could well be explained in terms of a 'filtering' action

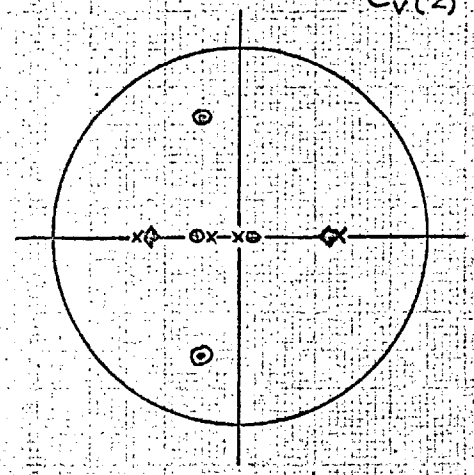
FIG. 7.2.4.

OVERALL CLOSED-LOOP Z-TRANSFER OF MODEL WITH PARAMETERS SET FOR BEST MATCH TO RAMP INPUT RESPONSES

TRANSFER: $C_p(z) = \frac{1}{q(z)}\left[q_{pp}(z,1)\,R_p(z) + q_{pv}(z,1)\,R_v(z)\right]$

$C_v(z) = \frac{1}{q(z)}\left[-2m_3m_5T^3R_p(z) + q_{vv}(z,1)\,R_v(z)\right]$

CIRCLES REPRESENT $|z| = 1$



OPERATOR A

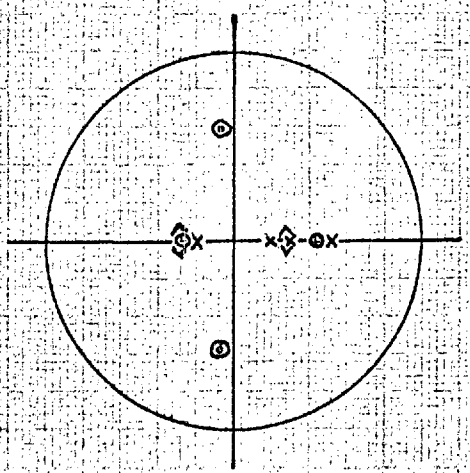
LEGEND

× POLES OF $1/q(z)$

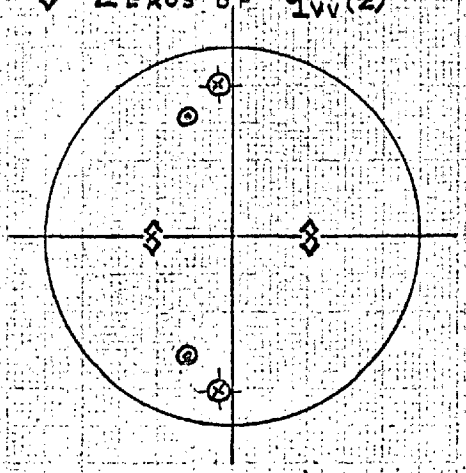
⊕ ZEROS OF $q_{pp}(z,1)$

⊙ ZEROS OF $q_{pv}(z,1)$

◇ ZEROS OF $q_{vv}(z)$



OPERATOR F



OPERATOR L

TABLE 7.3.I. VALUES OF 'BEST-FIT' MODELS' PARAMETERS

(FOR MODEL STRUCTURE AS SHOWN IN FIG 6.1.1.)

VALUES OF PARAMETERS FOR CONT TRACKING	OPERATOR				
	A	C	D	F	L
g_p	·627	·535	·412	·446	·630
g_i	·474	·403	·252	·278	·400
g_v	·343	·294	·197	·245	·340
g_s	·467	·467	·374	·467	·42
g_d	·1	·05	·4	·2	·1
T (MS)	133	150	166	150	143
VALUES OF PARAMETERS FOR RAMP TRACKING					
g_p	·59			·59	·59
g_i	·264			·31	·403
g_v	·39			·294	·662
g_s	·141			·141	·141
g_d	0			0	0
T	133			150	143

FIXED PARAMETERS :— $g_{m_p} = 1 = g_{m_v}$; $k_m = ·89$; $k_d = ·216$

$k_1 = 310$; $k_2 = 8·3$; $k_3 = 10^0$

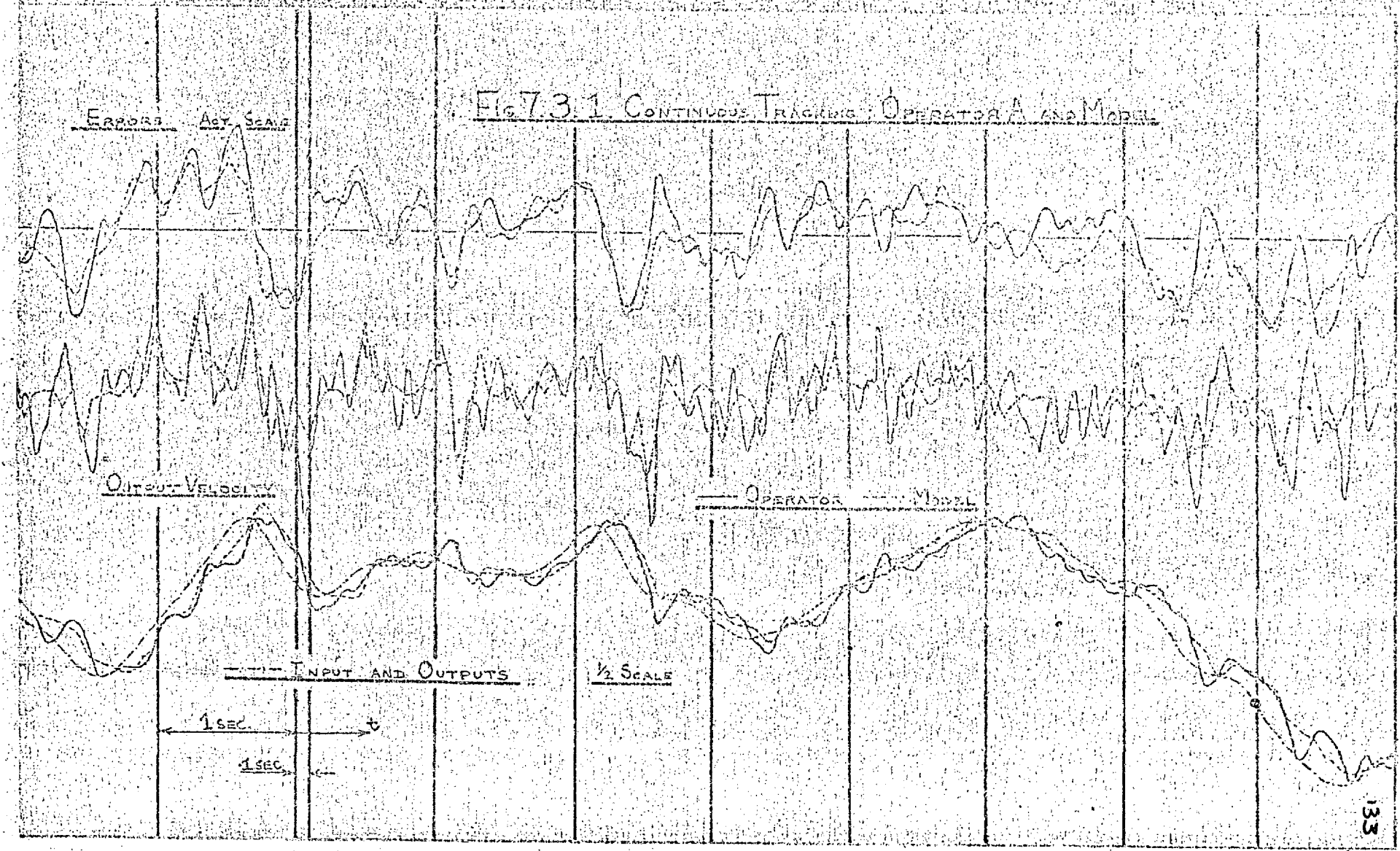


Fig 7.3.1 Continuous Tracking Operator A and Model
Errors Act. Scale
Output Velocity
Operator
Model
Input and Outputs
½ Scale
1 sec.
t
1 sec.

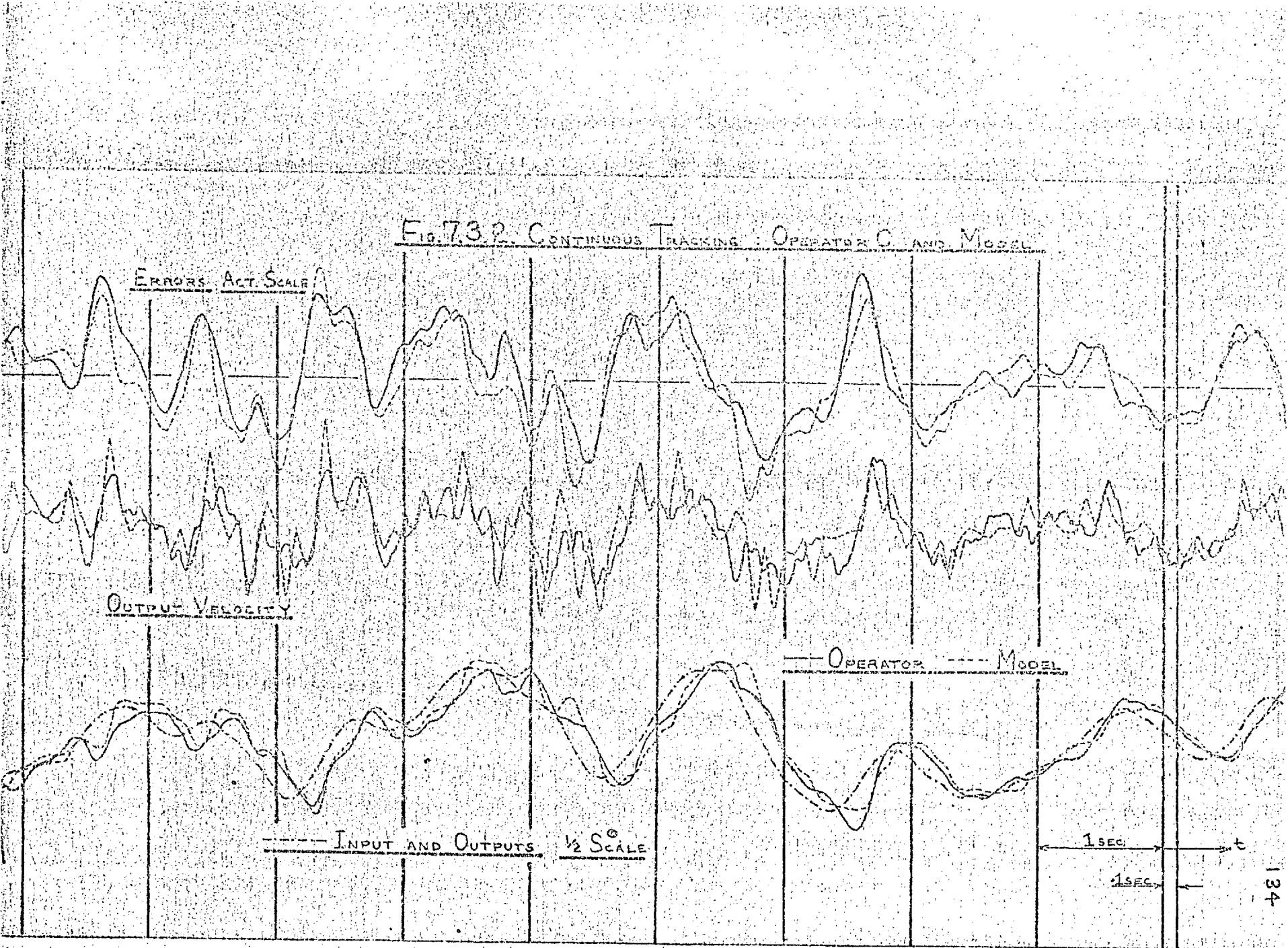


Fig. 7.3 P. Continuous Tracking : Operator C and Model
Errors Act. Scale
Output Velocity
Operator
Model
Input and Outputs ½ Scale
1 sec.
t
1 sec.

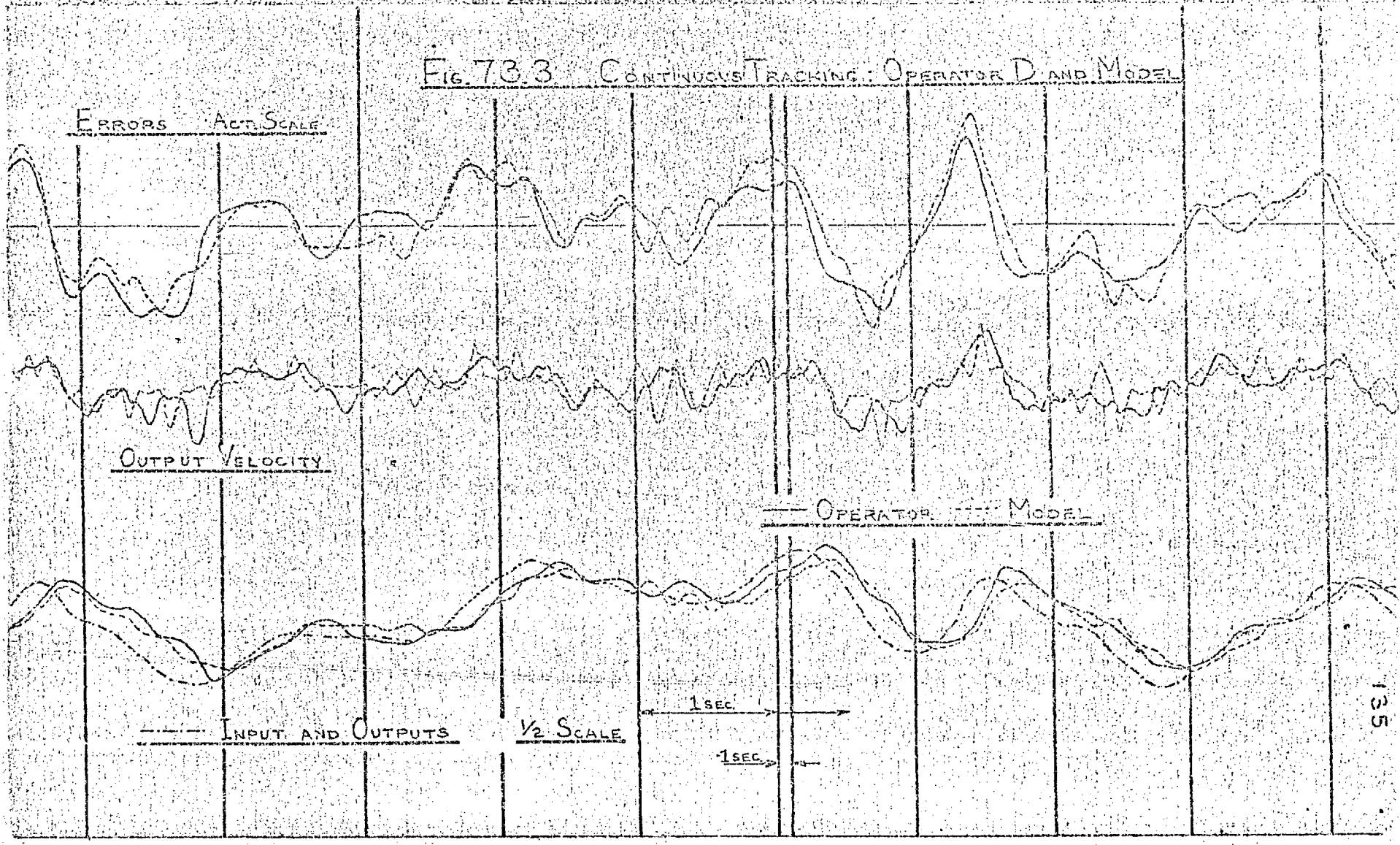
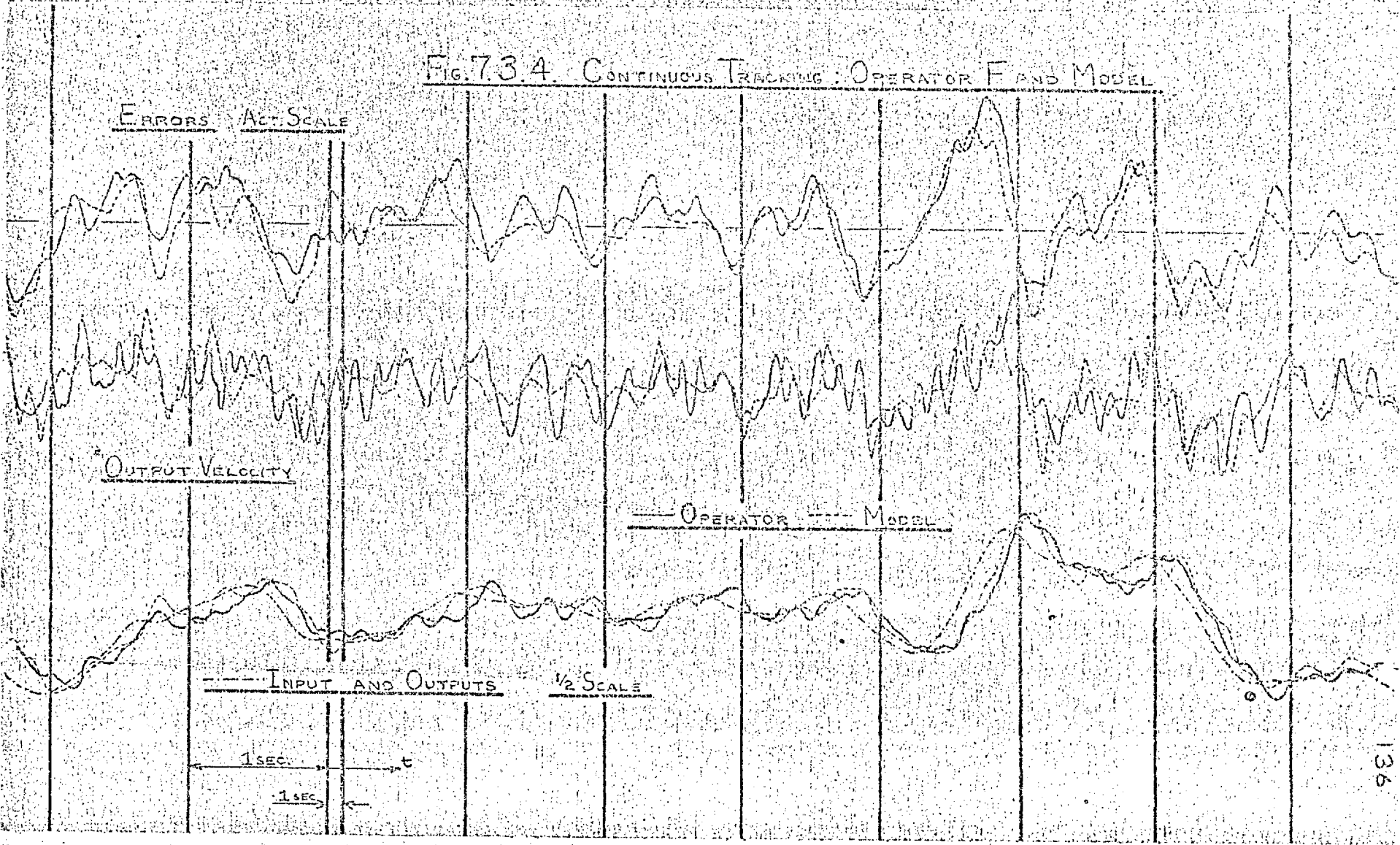


FIG. 7.3.3 CONTINUOUS TRACKING: OPERATOR D AND MODEL
ERRORS ACTL. SCALE
OUTPUT VELOCITY
OPERATOR
MODEL
INPUT AND OUTPUTS
½ SCALE
1 SEC.
1 SEC.

FIG. 7.3.4. CONTINUOUS TRACKING: OPERATOR F AND MODEL



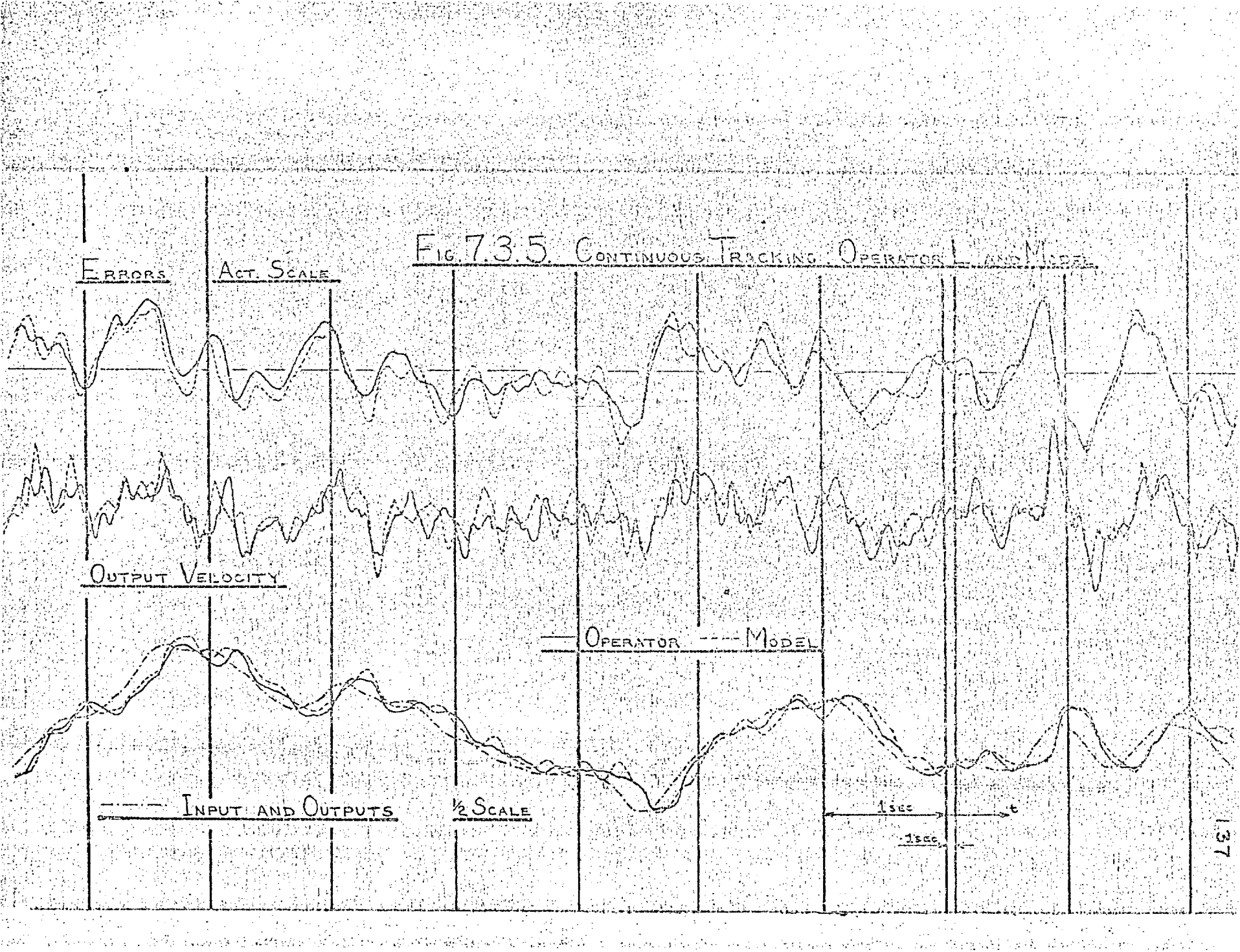


Fig. 7.3.5. Continuous Tracking Operator and Model
Errors
Act. Scale
Output Velocity
Operator
Model
Input and Outputs
½ Scale
1 sec
t
.1 sec

through the neuromuscular system. The fact that very good cross-correlation between operator and model errors was nevertheless achieved lends support to this view.

Z-plane pole-zero plots of the best-fit model transfers are shown in Fig.7.3.6. It may be seen that these transfers respresent a generally well damped closed-loop response. Poles and zeros occur well within the Z-plane unit circle,and those in the right half plane tend toward mutual cancellation.

The values of best-fit model parameters possessed some interesting implications. The lowest sensitivities to velocity information were displayed by operator D. Now this operator also filtered his hand motion most noticeably. This filtering action would tend to corrupt the velocity output at sampling instants,as a result of residual velocity from previously formulated position correction responses,so that velocity error sensing would be impaired. One would therefore expect less weight to be given to samples of velocity error,and that to compensate for the loss of information for the purposes of prediction,more weight would be given to past samples of reconstructed input position. One would expect a parallel in the case of operator F ; as his filtering action was less pronounced,the best-fit model should possess parameter values lying between those of operator D and the other operators. Observed parameter values are in agreement with this reasoning. In particular operator D showed much the greatest value of g_d - the weight attached to position output.

7.4. Comparison of Operator and Model Covariance Functions Calculated from Records of Continuous Tracking

The comparison of real-time traces was augmented by the comparison of covariance functions ,digitally computed from records obtained as described in Section 7.3. The requirement governing the choice of record length and data sampling frequency was that of minimising computation time,while retaining reasonable statistical reliability and covariance function detail.

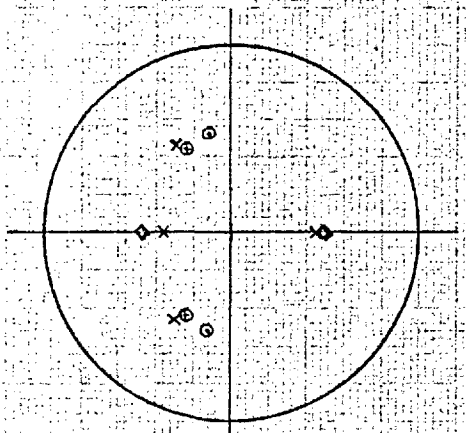
The record length was determined by considering the continuous random signal input function,as this contained the largest relative power at very low frequencies. The theoretical autocorrelation of the input signal is shown in Appendix II. It may be seen that the theoretical correlation is less than .2 at a lag of 1 sec.,and is thereafter quite small. It was considered that about 100 effectively independent samples would yield sufficient accuracy for the purposes of the investigation ; therefore a record length of 100 secs.would be quite suitable.

The data sampling frequency was determined by reference to the error functions,since they contained the largest relative power

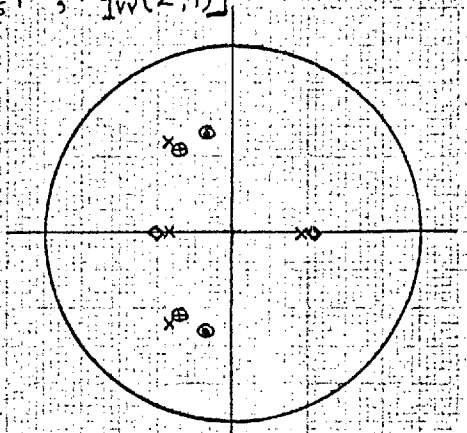
Fig. 7.3.6.

Overall Closed-Loop Z-Transfer of Model with Parameters set for Best Match to Continuous Tracking Response

Transfer: $$\underline{C}(z) = \frac{1}{q(z)} \cdot \begin{bmatrix} q_{pp}(z,1) & ; & q_{pv}(z,1) \\ -2m_3m_5T^3 & ; & q_{vv}(z,1) \end{bmatrix} \cdot \underline{R}(z)$$

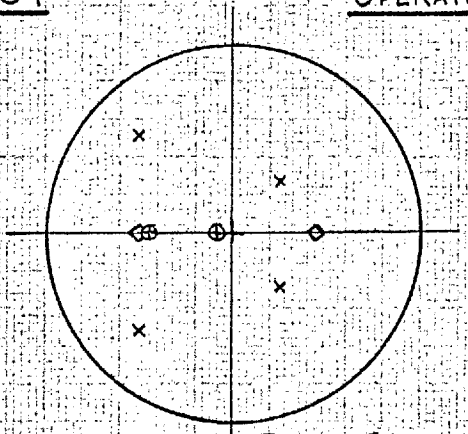


Operator A | Trial 34



Operator C | Trial 19

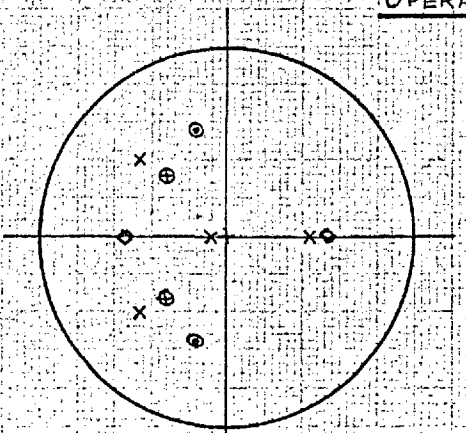
Circles are $|z|=1$



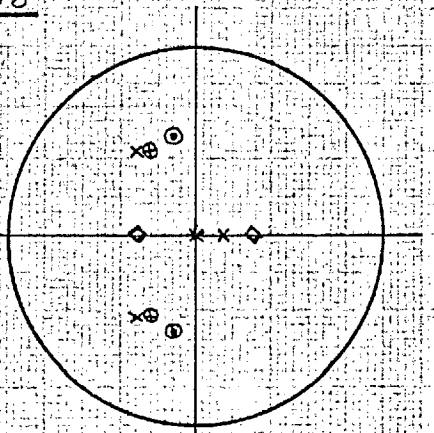
Operator D | Trial 58

Legend

- × Poles of $1/q(z)$
- ⊕ Zeros of $q_{pp}(z,1)$
- ⊙ Zeros of $q_{pv}(z,1)$
- ◇ Zeros of $q_{vv}(z,1)$



Operator F | Trial 38



Operator L | Trial 92

at high frequencies. A pilot estimation of the error spectrum obtained from a record of operator C ,who was thought most likely to generate high frequency remnant terms,showed that there was very little power beyond about 2cps.;this was in accord with results reported in the literature. On this basis,a data sampling frequency of 10/sec. was considered to represent a reasonable compromise between retention of covariance function detail,and cost and time required for data processing. Visual inspection of records confirmed the suitability of this choice.

A maximum lag of 5 secs. was chosen on the basis that :-

(a) It was desired to minimise computation effort,consistent with not discarding material information.

(b) All correlations of interest were likely to be very small at lags in excess of 5 secs. - especially so because a genuine random input was used. The assertion might not have been so valid in the case of a pseudo random input composed of only a few sinewaves.

(c) The amount of computation effort increases nearly in proportion to the number of lags required. As the time lag is increased, however,the reliability of the result decreases. A maximum lag between 5 and 10% of the record length is a generally accepted compromise.

In terms of spectral analysis,the above record length,lag and sampling frequency correspond to estimation of a reasonably flat spectrum with a resolution of .2 cps.,up to a maximum frequency of about 3 cps.,with a 90% spread of about 3 dB. (28)

In computing the sampled covariance functions the unbiassed estimator :-

$$\sigma_{ab}(k.\Delta t) = \frac{1}{N-k} \cdot \sum_{n=1}^{n=N-k} a(n.\Delta t).b(\overline{n+k}.\Delta t) \qquad 7.4.1.$$

(where Δt = record sampling interval and N = total number of samples) was employed. Although this estimator exhibits increasing variance as k is increased,the effect was small enough to be neglected in the present case - the maximum lag was given by $k = N/20$.

D.C. level terms were rejected,as implied in the calculation of covariance functions,for the following reasons:-

(1) Limitations of data logging and reading made it extremely difficult to determine D.C. levels accurately.

(2) The long term mean of the input was zero.

(3) Experiment indicated that the operator estimated the mean position of his arm, in order to generate kinaesthetic information. As the

control stick was free-moving, it provided no direct information as to hand position. The experimental procedure consisted of conducting a number of runs with the stick null position (the stick position corresponding to zero feedback voltage) differing by 1 cm. or so either side of its normal setting. The effect on operators' performances was completely insignificant - indeed some operators were not aware that any change had been made in experimental conditions. A displacement of 2 cms. was noticeable, but its effect on performance was only marginally significant. The average values indicated during the course of computation were not more than 1/10 th. of this.

The quantities computed directly were as follows:-

$\delta_{ii}(.1k), 0 \leq k \leq 50$, input autocovariance

$\delta_{ee}(.1k), 0 \leq k \leq 50$, error autocovariance

$\delta_{ie}(.1k), -50 \leq k \leq 50$, input-error covariance

The quantities derived from these were,

$\delta_{eo}(.1k), -50 \leq k \leq 50$, error-output covariance

$\delta_{io}(.1k), -50 \leq k \leq 50$, input-output covariance

$\delta_{oo}(.1k), 0 \leq k \leq 50$, input autocovariance

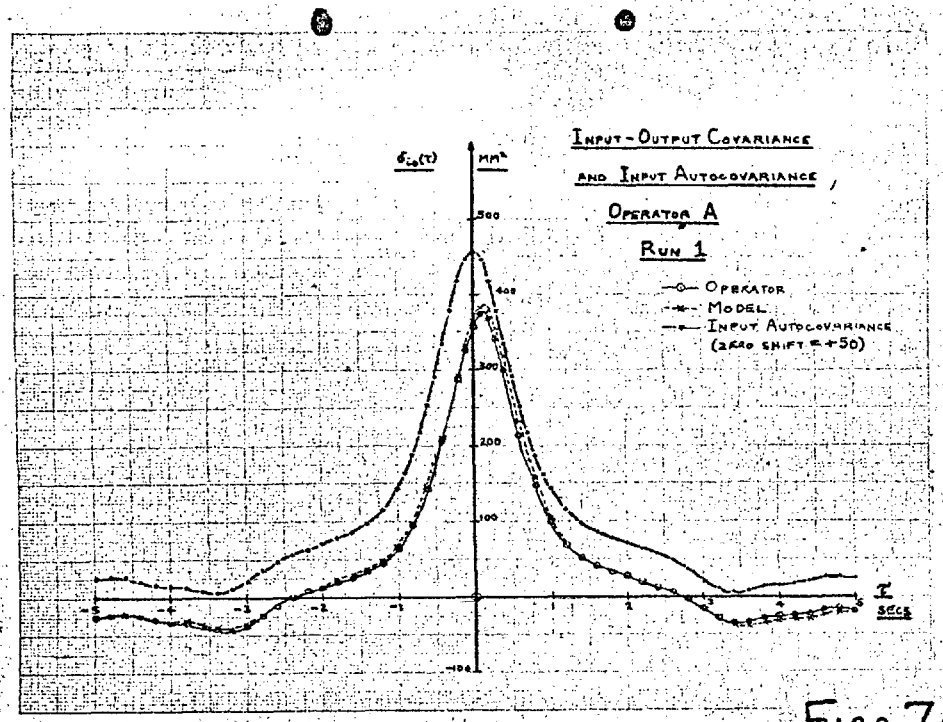
These derived covariances were computed from the following relations:-

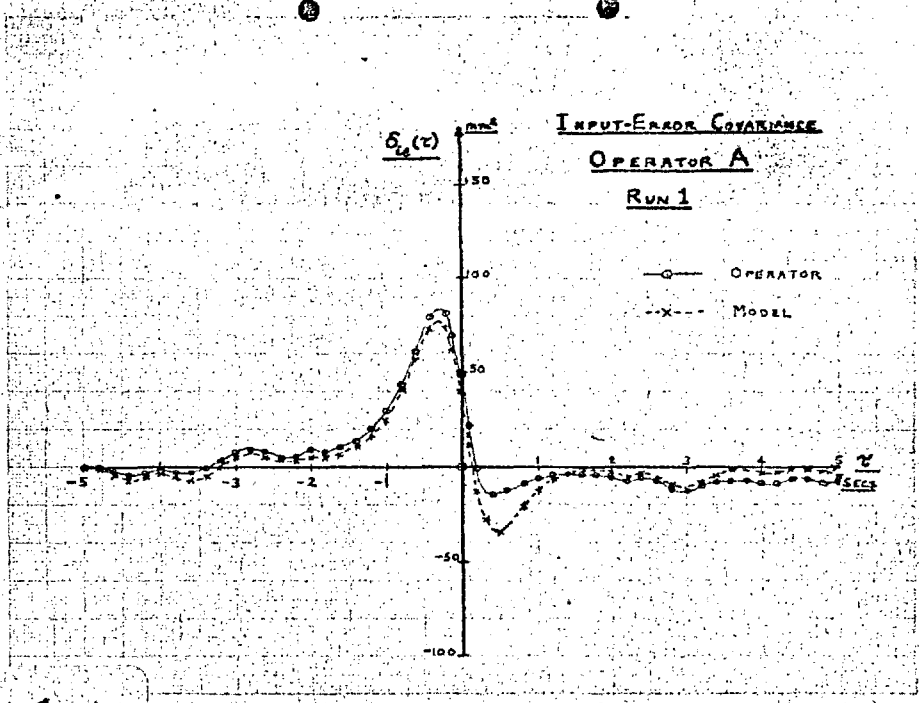
$$\delta_{eo}(.1k) = \delta_{ie}(-.1k) - \delta_{ee}(.1k) \qquad 7.4.2.$$

$$\delta_{io}(.1k) = \delta_{ii}(.1k) - \delta_{ie}(.1k) \qquad 7.4.3.$$

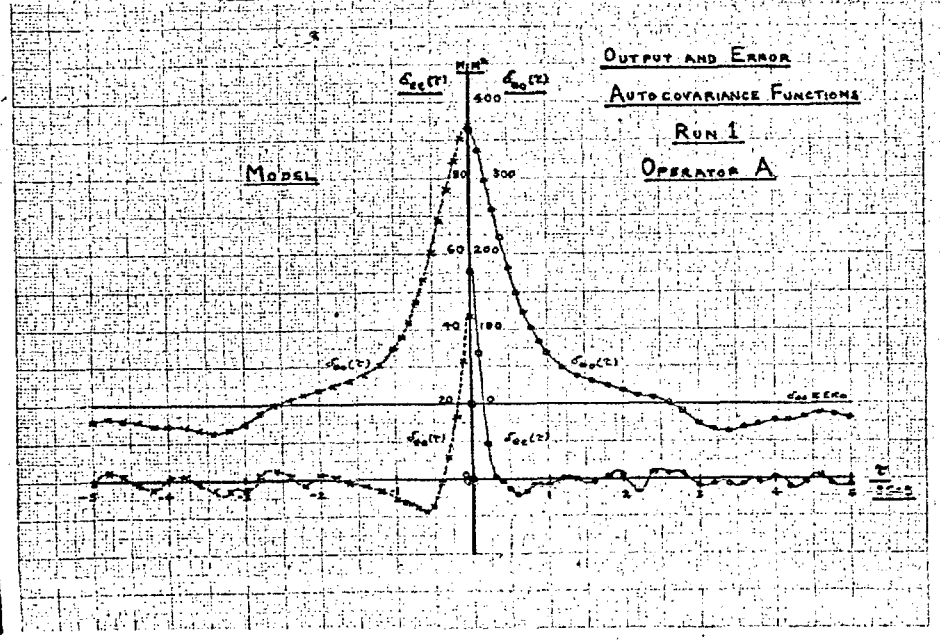
$$\delta_{oo}(.1k) = \delta_{ii}(.1k) - \delta_{ie}(.1k) - \delta_{ie}(-.1k) + \delta_{ee}(.1k) \qquad 7.4.4.$$

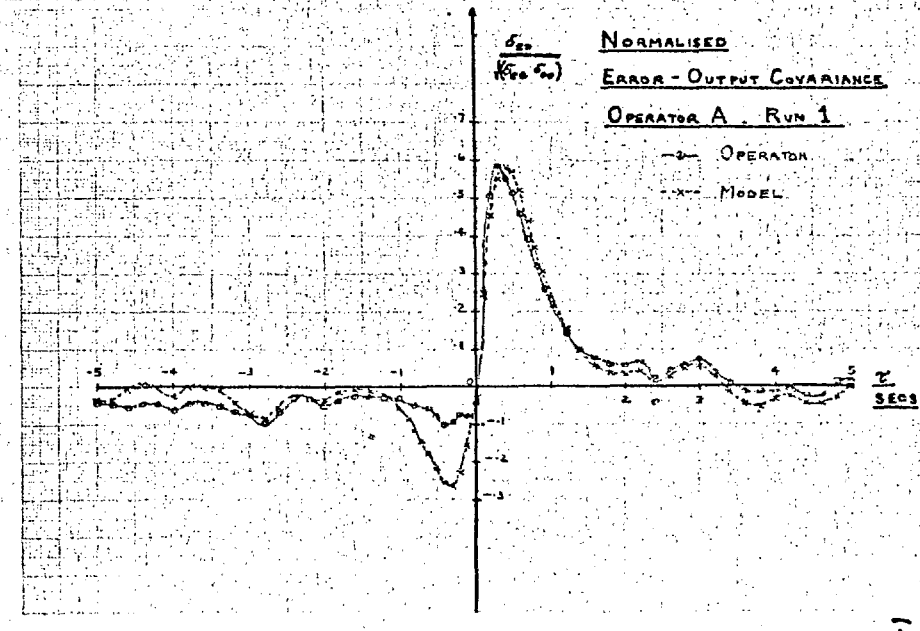
Graphs of all the above covariance functions are shown in Figs.7.4.1-20, in which corresponding operator and model covariances have been plotted together, so as to facilitate a direct comparison. In principle one may calculate the approximate variance of a particular estimate of covariance, and the degree to which two estimates of covariance for different lags are correlated(29). However, such calculation is unwieldy and unnecessary, because both model and operator were fed with the same input function. If the model were a perfect simulator of the operator's response, then all covariance functions should be practically identical. Differences would arise as a result of the operator's remnant term (the part of the operator's output not linearly correlated with the input), and by

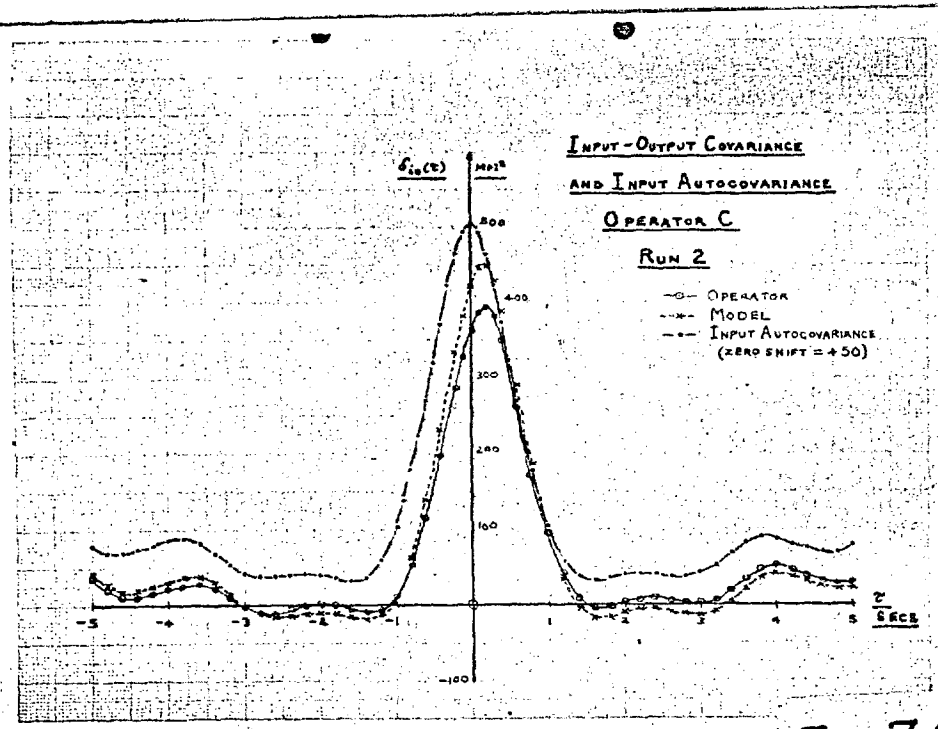

Input-Output Covariance
and Input Autocovariance
Operator A
Run 1
Operator
Model
Input Autocovariance
(zero shift = +50)
$\delta_{io}(\tau)$
mm²
τ
secs

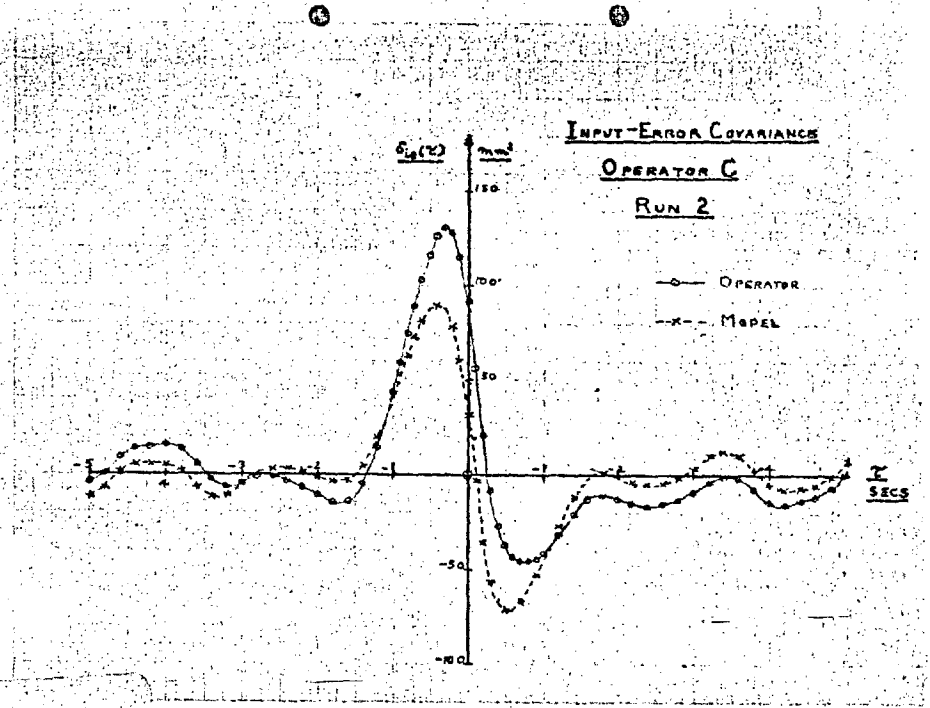

Input-Error Covariance
Operator A
Run 1
Operator
Model
$\delta_{ie}(\tau)$
mm²
τ
secs

Figs. 7.4.1-4.

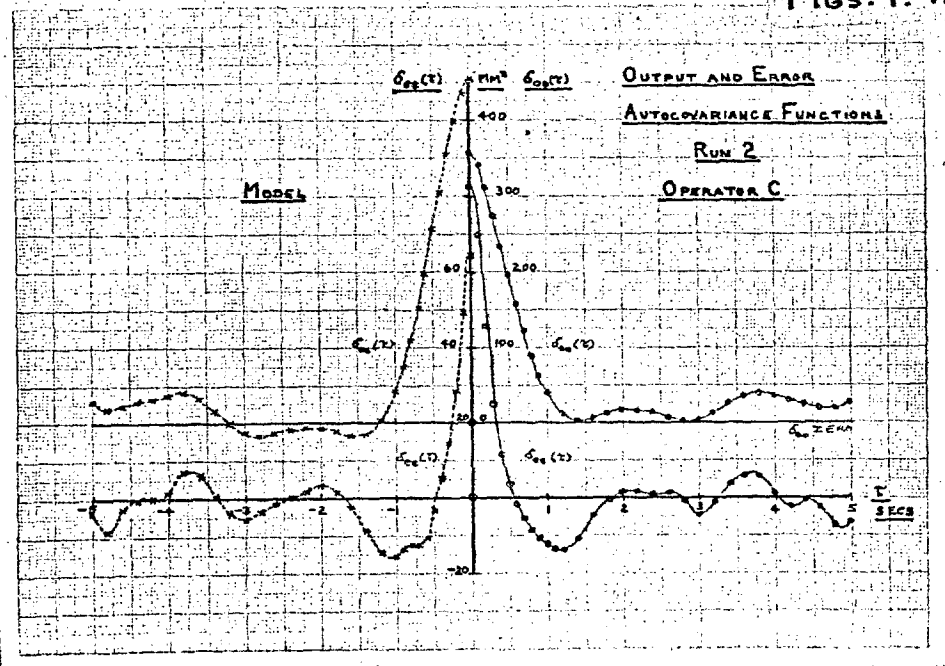

Output and Error
Autocovariance Functions
Run 1
Operator A
Model
$\delta_{oo}(\tau)$
$\delta_{ee}(\tau)$
mm²
τ
secs

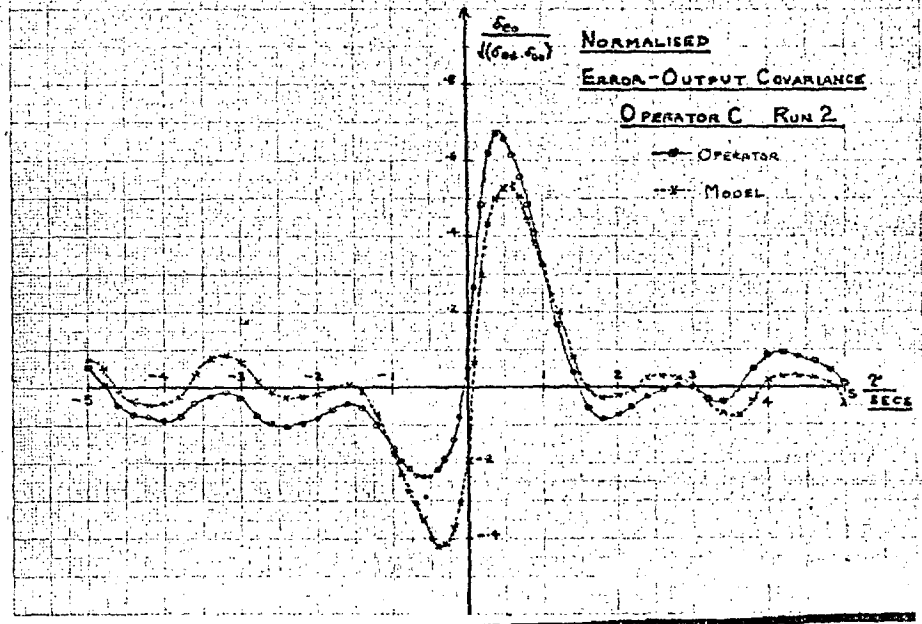

Normalised
Error - Output Covariance
Operator A . Run 1
Operator
Model
$\frac{\delta_{eo}}{\sqrt{\delta_{oo}\,\delta_{ee}}}$
τ
secs

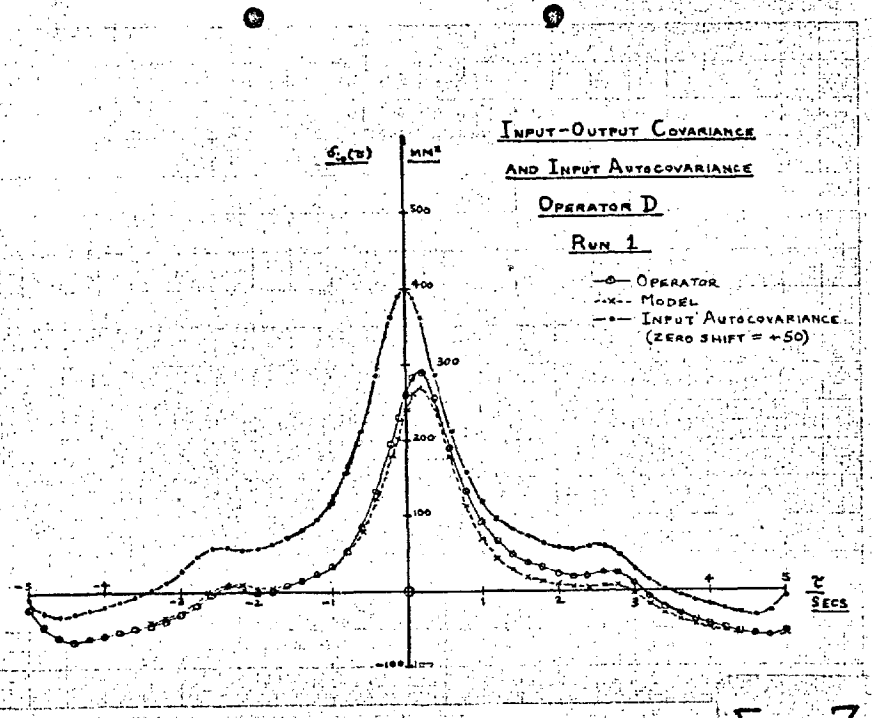

Input-Output Covariance
and Input Autocovariance
Operator C
Run 2
Operator
Model
Input Autocovariance
(zero shift = 450)
τ secs

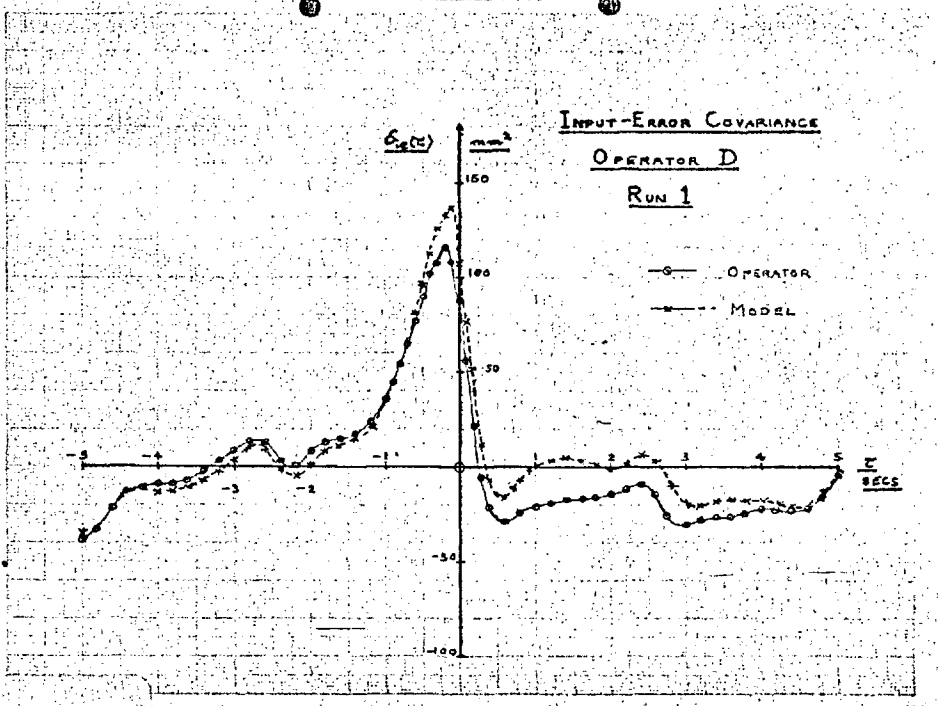

Input-Error Covariance
Operator C
Run 2
Operator
Model
τ secs

Figs. 7.4.5.-8.

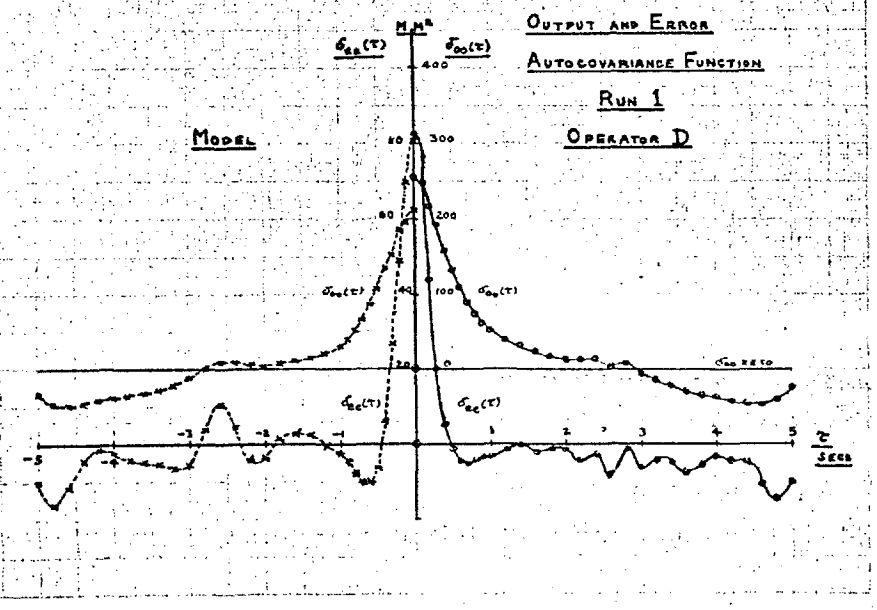

Output and Error
Autocovariance Functions
Run 2
Operator C
Model
τ secs

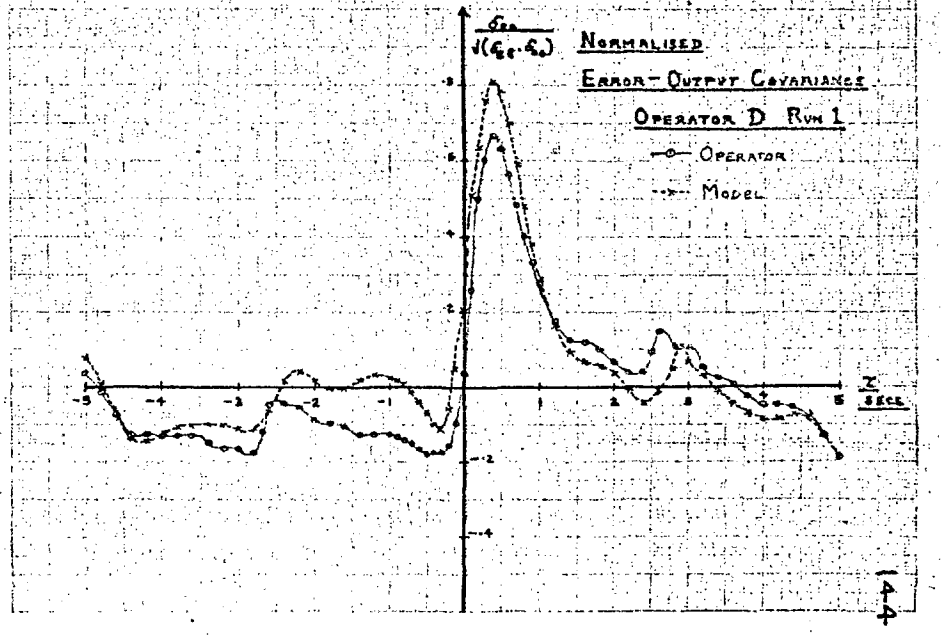

Normalised
Error-Output Covariance
Operator C Run 2
Operator
Model
τ secs

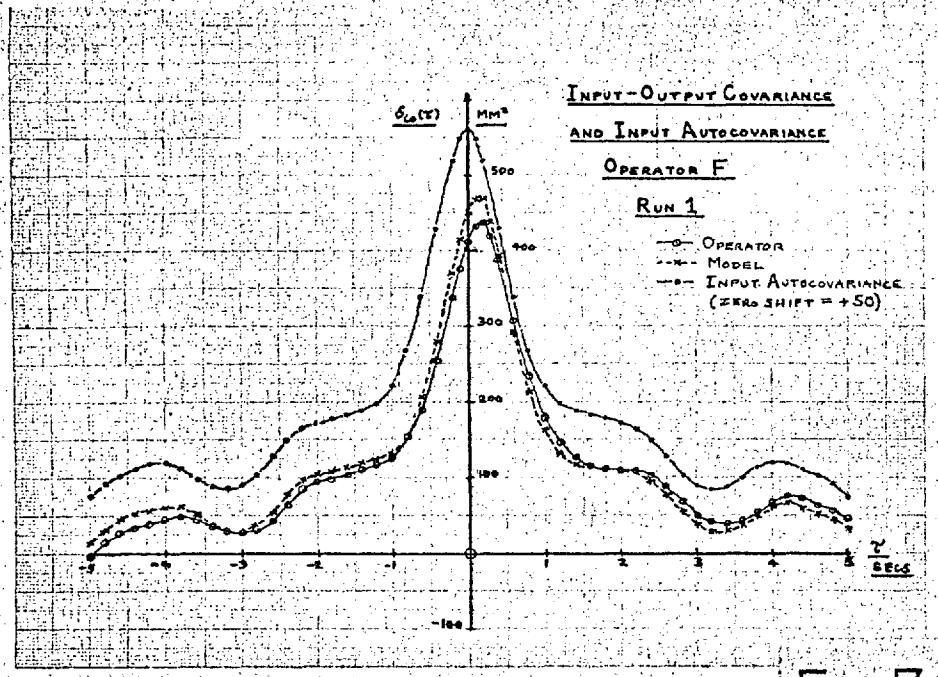

Input-Output Covariance
and Input Autocovariance
Operator D
Run 1
Operator
Model
Input Autocovariance
(zero shift = +50)
mm²
τ secs

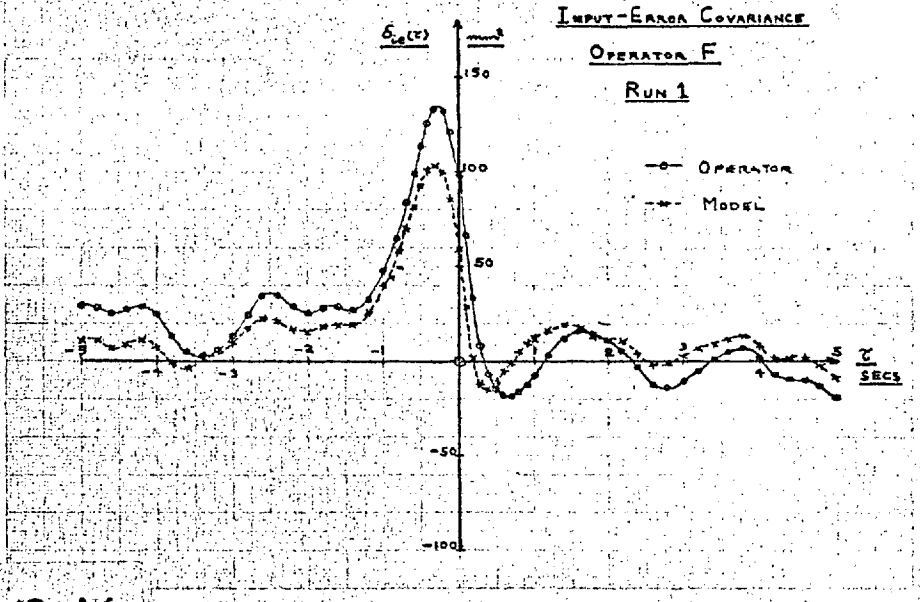

Input-Error Covariance
Operator D
Run 1
Operator
Model
mm²
τ secs

Figs. 7.4.9.–12.

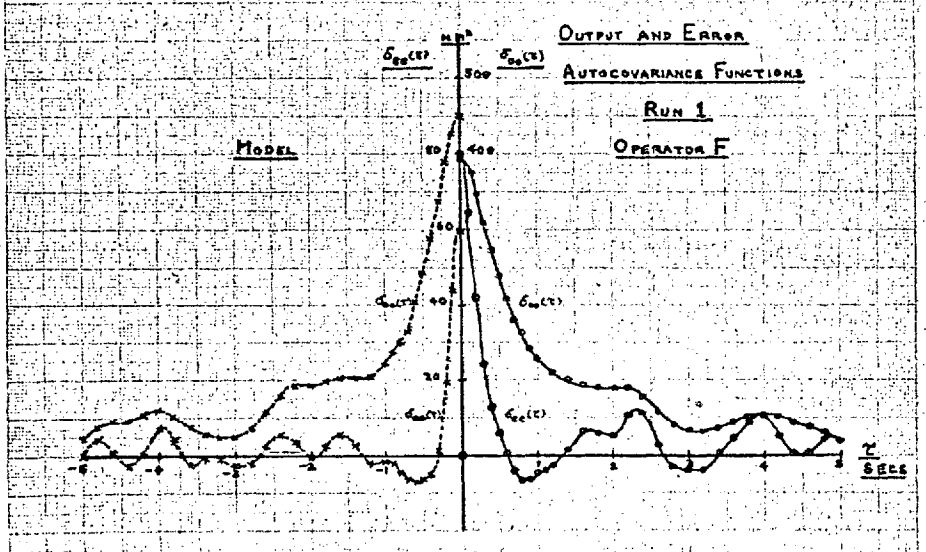

Output and Error
Autocovariance Function
Run 1
Operator D
Model
M.M²
τ secs

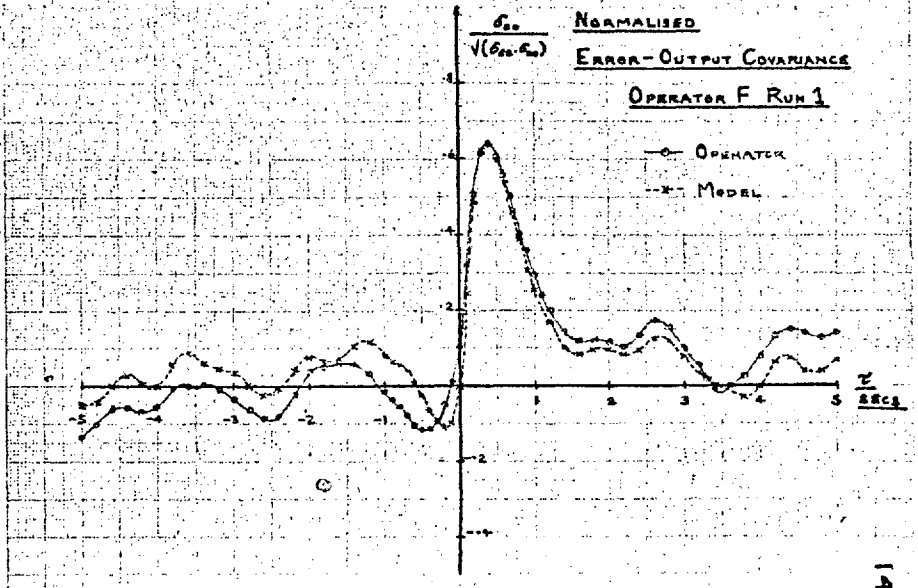

Normalised
Error-Output Covariance
Operator D Run 1
Operator
Model
τ secs

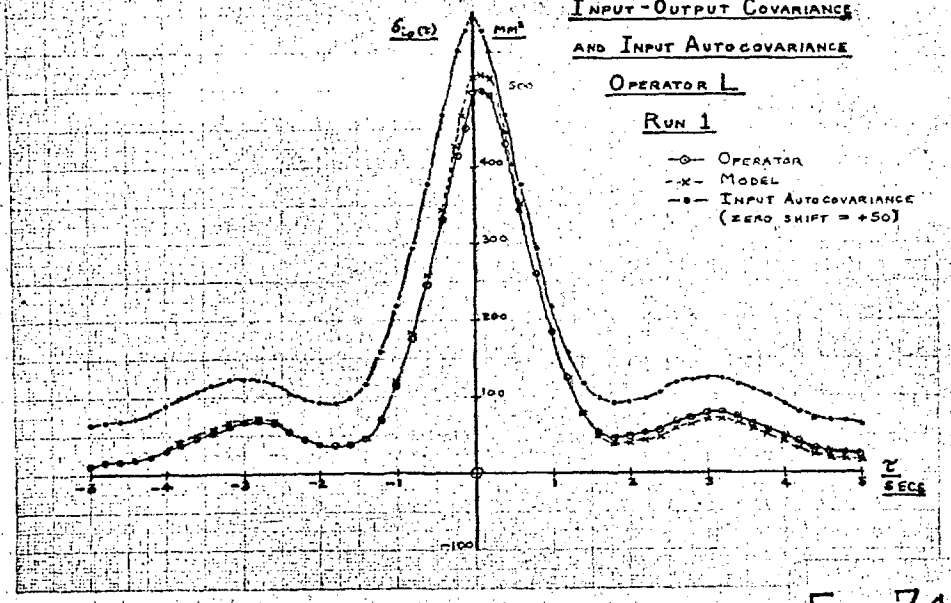

Input-Output Covariance
and Input Autocovariance
Operator F
Run 1
Operator
Model
Input Autocovariance
(zero shift = +50)
τ secs

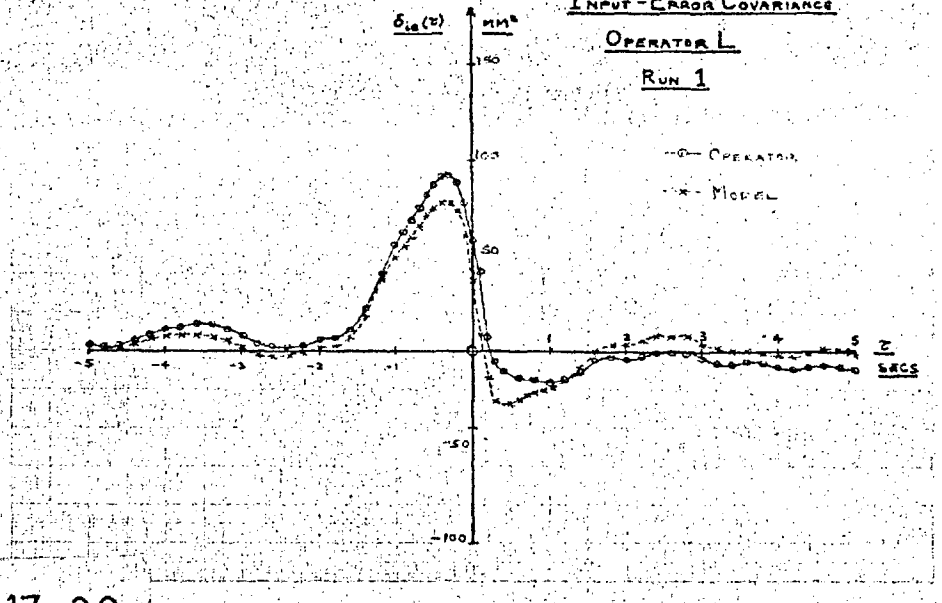

Input-Error Covariance
Operator F
Run 1
Operator
Model
τ secs

Figs. 7.4.13-16.

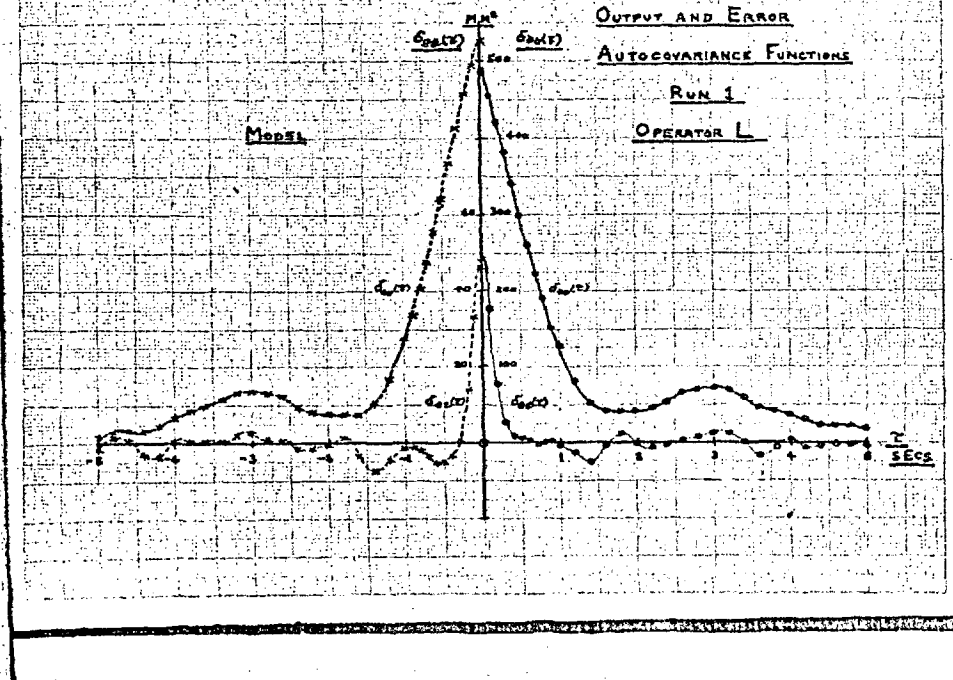

Output and Error
Autocovariance Functions
Run 1
Operator F
Model
τ secs

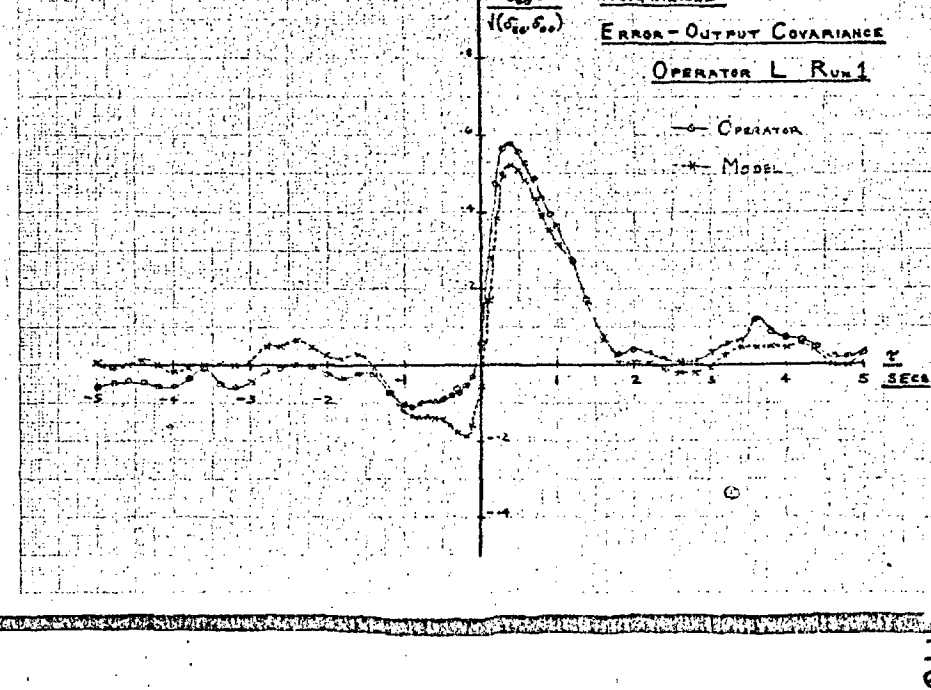

Normalised
Error-Output Covariance
Operator F Run 1
Operator
Model
τ secs


Input-Output Covariance
and Input Autocovariance
Operator L
Run 1
Operator
Model
Input Autocovariance
(zero shift = +50)
mm²
τ secs


Input-Error Covariance
Operator L
Run 1
Operator
Model
mm²
τ secs

Figs. 7.4.17-20.


Output and Error
Autocovariance Functions
Run 1
Operator L
Model
mm²
τ secs


Normalised
Error-Output Covariance
Operator L Run 1
Operator
Model
τ secs

departure from the best possible linear simulation on the part of the model.

Comparison between operator and model covariance functions reveals generally very good matches, especially in regard to shape. Matches between input-output covariances were very good to excellent ; those between input-error covariances were very good. In the latter case there was some tendency for the model to show a somewhat more pronounced negative peak at small positive values of lag. It was noted that this was associated with slightly too much amplitude in the input-output covariance function at similar values of lag.

The fits between output autocovariance functions were good as regards shape, but zero-lag variances showed as much as 25% discrepancy in the case of operators C and D. This may be explained by the fact that output variance is very sensitive to the overall loop gain, and therefore any mismatch of the model is emphasised.

The degree of match shown by the error autocovariance functions could be described as reasonable to good. These functions were most sensitive to operator remnant and model mismatch, because their variance was much smaller than either the input or output variances - by a factor of the order of 8. Thus operator remnant could easily represent 10 - 20% of the total error variance. One would therefore expect the shapes of the operator and model autocovariance functions to differ quite materially. It was noted that the model autocovariances tended to show a slightly more pronounced first minimum than did those of the operators, but the form of the 'tails' (lags in excess of 1 sec.) was very similar.

The error-output covariance functions are related to the corresponding equivalent forward transfer weighting functions, each convolved with the appropriate error autocovariance function. Even if the model were a perfect match to the linear portion of the operator's transfer, the remnant terms recirculating around the operator control loop would give rise to spurious covariance, the effect of which would be greatest at small negative values of lag (see Appendix VIII).

The covariance functions were normalised to give error-output cross-correlation functions (relating the derived zero-mean error and output variates) in order to facilitate a direct comparison of shapes. As may be seen, these cross-correlations show very good matches for positive values of lag, and a maximum of discrepancy for lags in the range 0 to -2 secs. Since the model and operator outputs were highly correlated, this confirms that a substantial part of the differences between the error autocovariance functions could be explained in terms of operator remnant. It also demonstrates that the equivalent forward transfers of operator and model were well matched.

The model exhibited a tendency toward showing rather more pronounced negative peaks in the $\delta_{ie}(\tau)$ and $\delta_{ee}(\tau)$ covariance functions. This could be partly explained in terms of the operators' random sampling, which would lead to smoother and rather less pronounced peaks than the regular sampling operation of the model would give. However, the effect also arose as a result of

mismatch of the model ; slightly different parameter settings could largely account for the observed discrepancy. This point is discussed in the next Section.

7.5. Fitting Model and Operator Covariance Functions

It was desired to determine how much of the observed discrepancies between operator and model characteristics could be attributed to model mismatch. It was therefore necessary to evaluate the effects of changes of the model parameters on the form of the model covariance functions, so that a new set of parameters could be formulated to improve the model's fit to operator covariance data. It was most convenient to consider fitting the input-error covariance functions, since these were not affected by the operator's remnant, and were sensitive to mismatch.

The basis of the fitting procedure was as follows. From equation 6.2.11. one may derive :-

$$\left[\Phi_{ie}(z)\right] = \left[\Phi_{ii}(z)\right].\left[I - \frac{1}{q(z)}.\,Q(z,1)\right]^{T} \qquad 7.5.1.$$

T = transpose

where $\left[\Phi_{ie}(z)\right]$ denotes a matrix of pulse spectral densities derived by Z-transforming the matrix of sampled covariance functions,

$$\text{Cov}\begin{bmatrix}\text{Input-error} & ; \text{ Input-error velocity}\\ \text{Input velocity-error} & ; \text{ Input velocity-error velocity}\end{bmatrix}$$

$\Phi_{ii}(z)$ is similarly defined.

Manipulation of equation 7.5.1. shows that:-

$$\Phi_{ie}(z) = \left\{1 - \frac{q_{pp}(z,1)}{q(z)}\right\}.\Phi_{pp}(z) - \left\{\frac{q_{pv}(z,1)}{q(z)}\right\}.\Phi_{pv}(z) \qquad 7.5.2.$$

where -

$\Phi_{ie}(z)$ is the leading term of the corresponding matrix of pulse spectral densities

and $\Phi_{pp}(z) = \sum_{n=-\infty}^{\infty} \delta_{ii}(nT).z^{-n}$

and $$\Phi_{pv}(z) = \sum_{n=-\infty}^{\infty} \delta_{ii'}(nT).z^{-n} \quad \text{where } \delta_{ii'}(nT) = \frac{\partial}{\partial \tau}\delta_{ii}(\tau)\Big|_{\tau=nT}$$

The above relations for $\Phi_{pp}(z)$ and $\Phi_{pv}(z)$ indicate the convenience of working with pulse spectral densities ; they may be derived merely by sampling the appropriate continuous covariance function, and then associating each sample with the corresponding power of z. In the present case the continuous covariance functions were obtained by visual interpolation of the digitally computed covariances. Thus $\delta_{ii}(nT)$ and $\delta_{ii'}(nT)$ could be obtained from the computed input covariance, and the model transfers could be calculated from knowledge of the model parameters. Hence $\delta_{ie}(nT)$ could be calculated and compared with the carresponding operator function.

The error autocovariance function could then be calculated by using the relation, derived from equation 7.5.1., :-

$$\left[\Phi_{ee}(z)\right] = \left[\Phi_{ei}(z)\right].\left[I - \frac{1}{q(z)}.Q(z,1)\right]^{T} \qquad 7.5.3.$$

in a manner similar to that described above.

Using the above relations, the effect of changes in the model parameters was evaluated in terms of the input-error covariance function. When an improved fit had been obtained, the corresponding error autocovariance function was calculated. The application of this procedure showed that covariance function fits could be improved by modifying the model parameters as follows:-

Operator A ; g_p smaller, g_1 larger.

Operator C ; g_p smaller, g_1 larger.

Operator D ; g_d smaller.

Operator F ; g_1 smaller.

Operator L ; g_p smaller.

In all cases the required changes in parameter values were relatively small. The improvement in fit was marginal, except in the case of operator C, where there was most initial model mismatch. However, the improved fits did exhibit a reduction in the discrepancy of the negative peaks in the $\delta_{ie}(\tau)$ and $\delta_{ee}(\tau)$ functions.

The effect of incorporating an extra model parameter, in the form of an additional sample for velocity prediction (i.e. elaborating the structure of the model), was also investigated. The fitting procedure was made difficult by the complication of the model transfer (five poles and two sets of three zeros, for position output). The results were disappointing, and if anything, poorer fits were obtained. The fit of the model was already very good, and it was

concluded that the improvement of fit by elaborating the structure of the model in the above fashion was likely to prove fruitless.

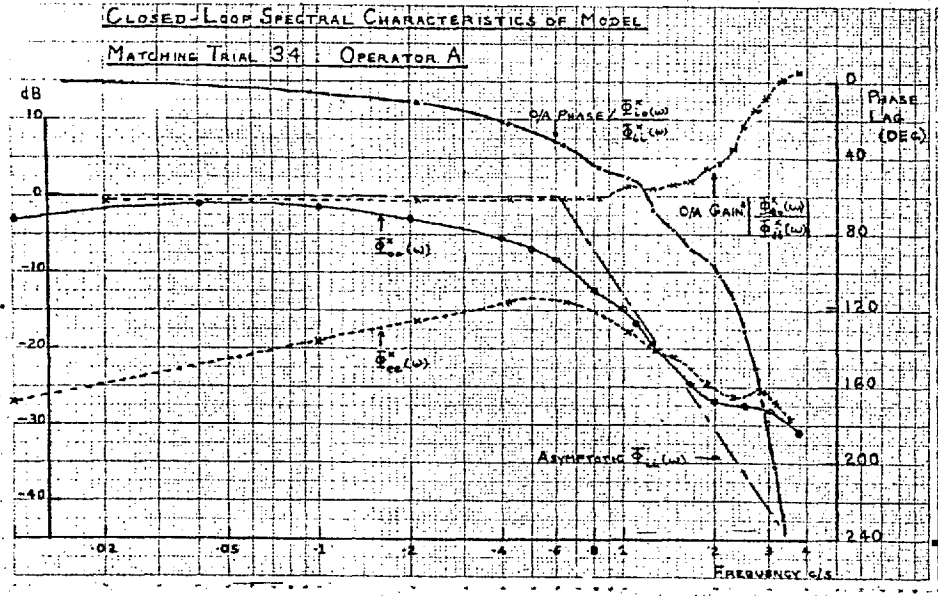
7.6. Consideration of the Calculated Spectral Characteristics of the Model

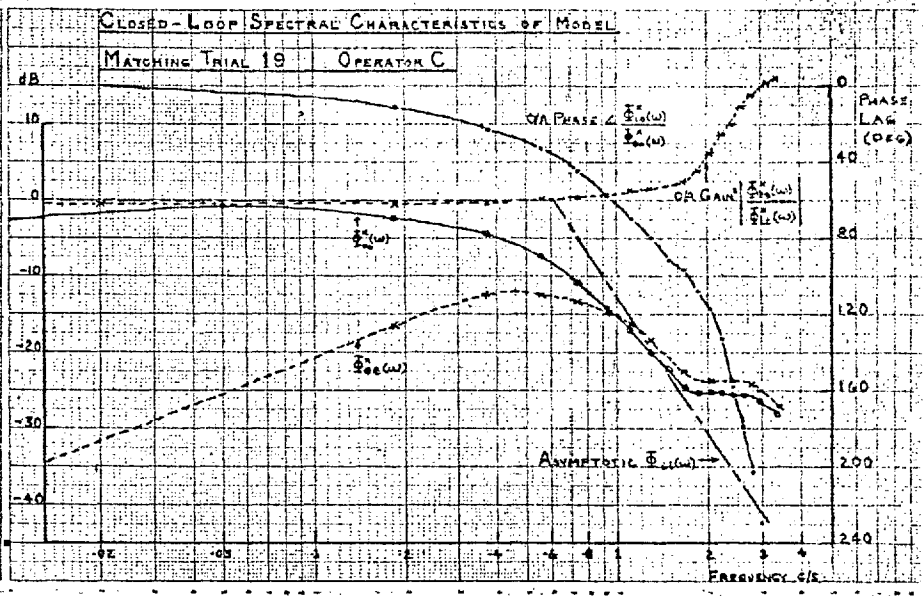
The frequency domain characteristics of the model were calculated by considering the position input, error, and output at sampling instants, in terms of the appropriate Z-transfers. Then the results derived in Section 6.3. were used to find the equivalent spectral characteristics. These are plotted in terms of magnitude and phase in Figs.7.6.1-10. Here the spectral functions, such as $\underline{\Phi}_{oo}(w)$, relate to the leading elements in the corresponding spectral density matrices. They have been normalised with respect to the 'white' noise function which, when appropriately filtered, would generate the continuous random input signal. These spectra may be thought of as representing spectra in an equivalent linear continuous model configuration, if frequencies higher than 1/2T cps. (where T = model sampling interval) are neglected.

The most directly comparable data are provided in published work due to Elkind (10) and to Bekey (2). The most apposite of Elkind's results are shown in Figs.7.6.11&12., and of Bekey's in Figs.7.6.13&14. The plots of calculated spectra demonstrate the quite pronounced variations in 'best-fit' model characteristics from operator to operator. However, it may be noted that there are certain features common not only to all the calculated spectra, but also to the results derived by Elkind and Bekey.

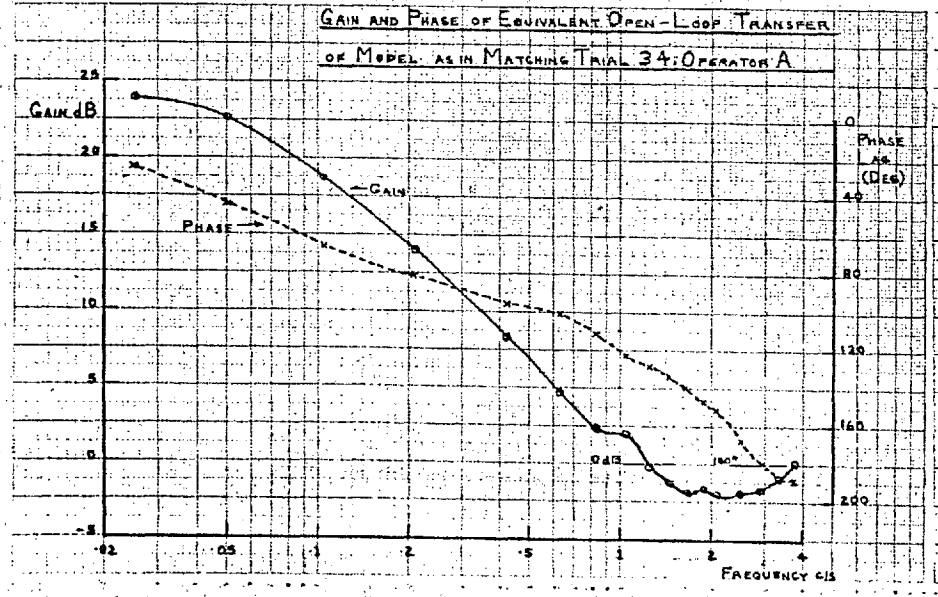
It may be noted that there is a general rise of overall closed loop gain with frequency. Associated with this rise is the form of the $\Phi_{oo}(w)$ curves, which tend to follow the $\Phi_{ii}(w)$ curves at low frequencies, but exhibit a pronounced 'flattening out' at frequencies greater than about 1.5 cps. Moreover, Bekey states in his thesis (2), "Both man and model show a tendency to magnify input distrubances when these occur near the sampling frequency of the model". Now this would indicate a maximum of gain at around 3 cps. (a typical sampling frequency of Bekey's model), with which the calculated overall loop gain of the present model is in very good agreement.

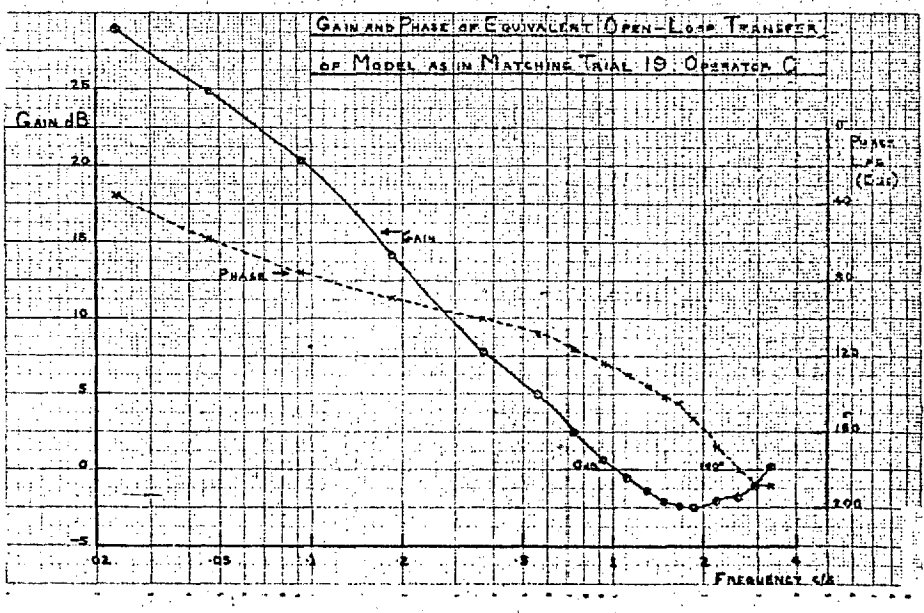
The curves of caculated closed loop phase lag are very similar in form to those given by Elkind. Although the absolute values tend to be somewhat different, the curves are practically parallel. They show that the operator is well able to compensate for his inherent lag at low frequencies, when following a reasonably smooth filtered input, but that the compensation steadily deteriorates at higher frequencies, until the slope of the phase characteristic corresponds to practically the full time lag required for response to a discrete input.

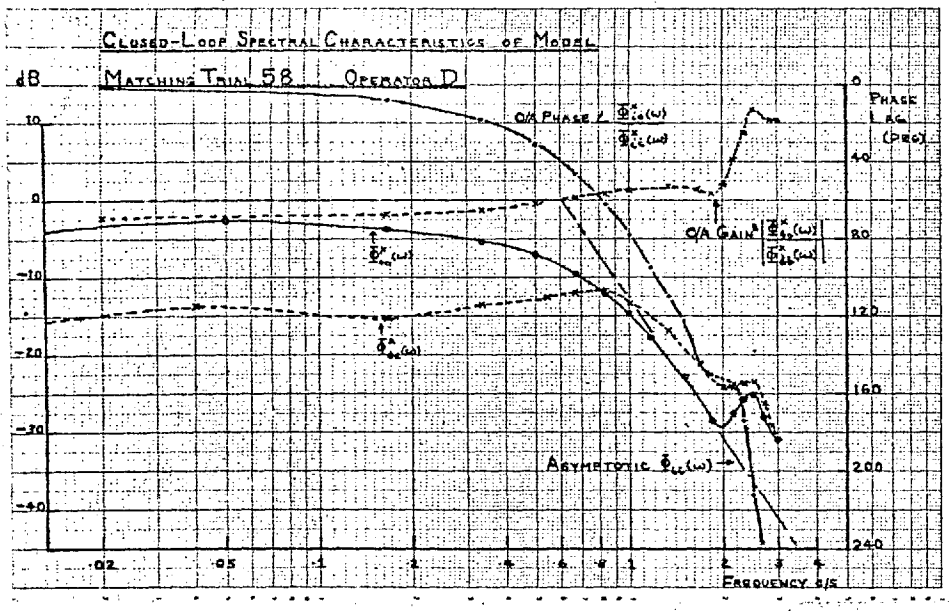

Closed-Loop Spectral Characteristics of Model
Matching Trial 34 Operator A
dB
Phase Lag (deg)
O/A Phase
O/A Gain
Asymptotic
Frequency c/s

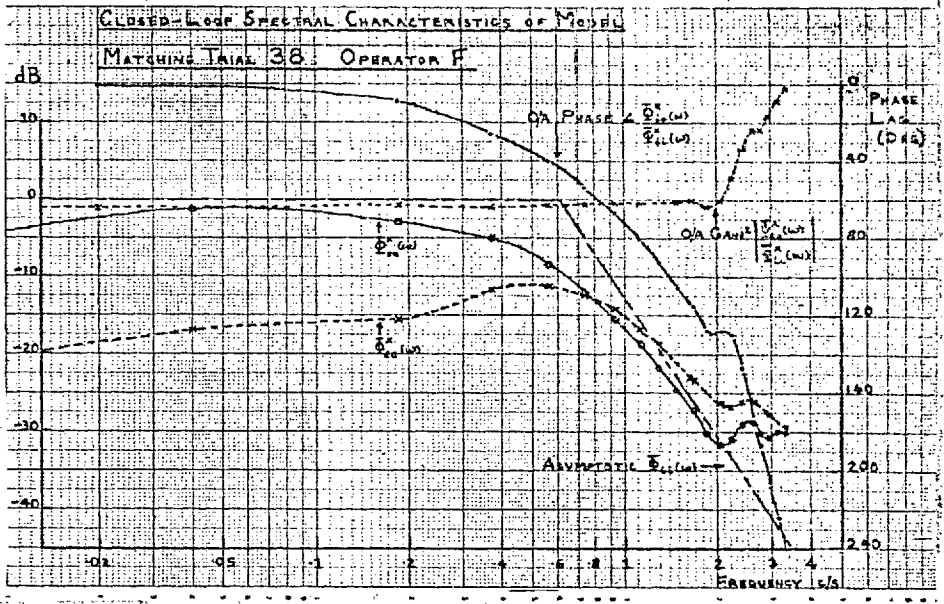

Closed-Loop Spectral Characteristics of Model
Matching Trial 19 Operator C
dB
Phase Lag (deg)
O/A Phase
O/A Gain
Asymptotic
Frequency c/s

Figs. 7.6.1-4.

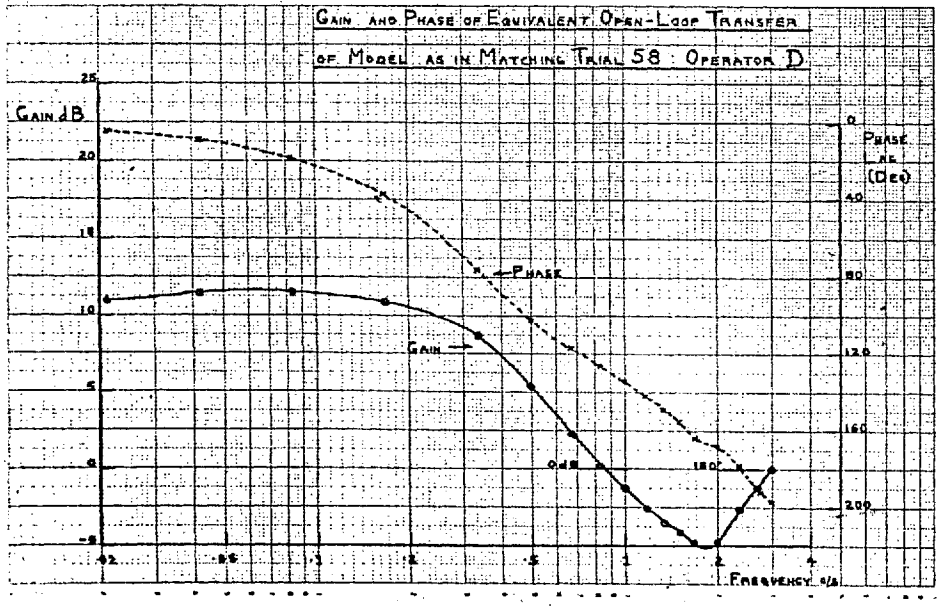

Gain and Phase of Equivalent Open-Loop Transfer
of Model as in Matching Trial 34: Operator A
Gain dB
Phase Lag (Deg)
Gain
Phase
0dB
180°
Frequency c/s

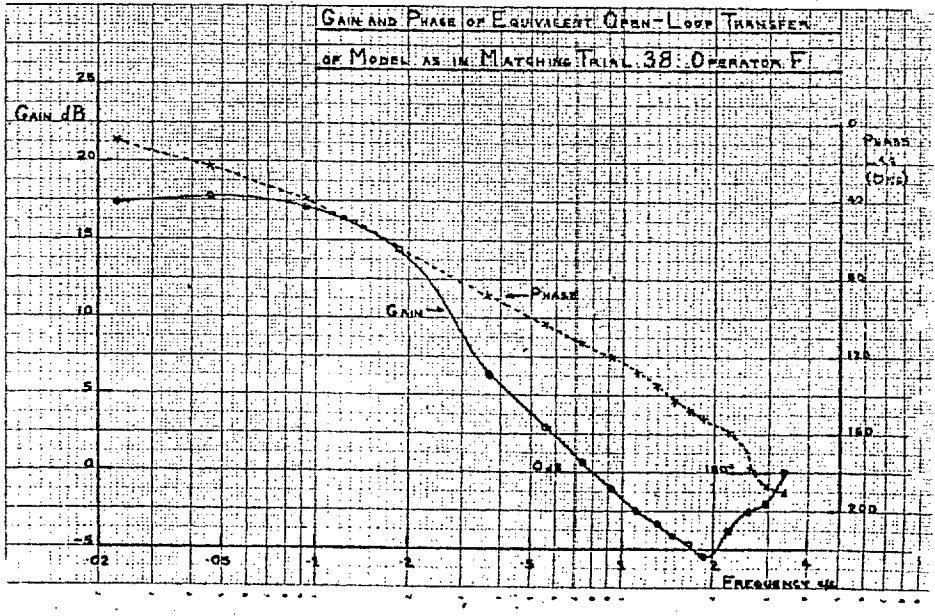

Gain and Phase of Equivalent Open-Loop Transfer
of Model as in Matching Trial 19: Operator C
Gain dB
Phase Lag (Deg)
Gain
Phase
0dB
180°
Frequency c/s

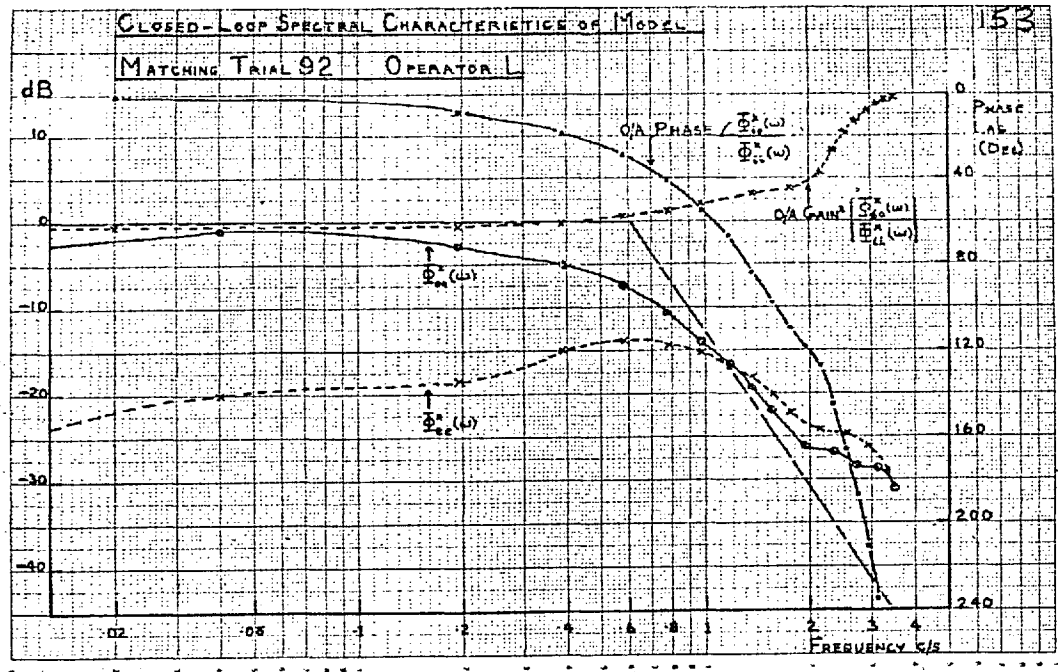

Closed-Loop Spectral Characteristics of Model
Matching Trial 58 Operator D
dB
O/A Phase
O/A Gain²
Asymptotic
Phase Lag (Deg)
Frequency c/s


Closed-Loop Spectral Characteristics of Model
Matching Trial 38 Operator F
dB
O/A Phase
O/A Gain²
Asymptotic
Phase Lag (Deg)
Frequency c/s

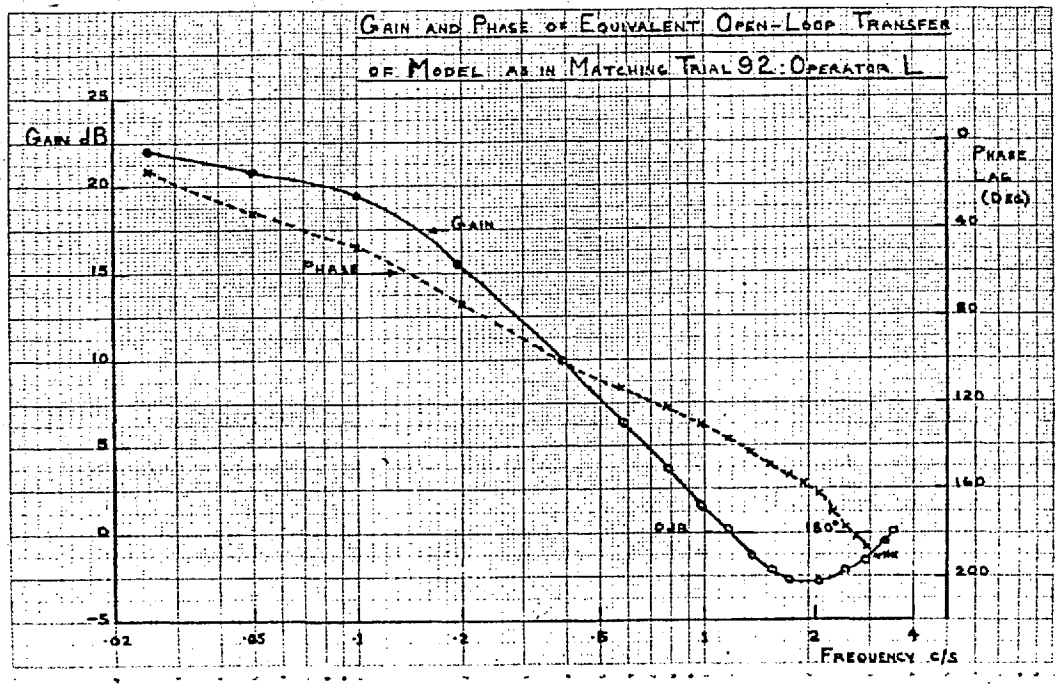
Figs. 7.6.5-8

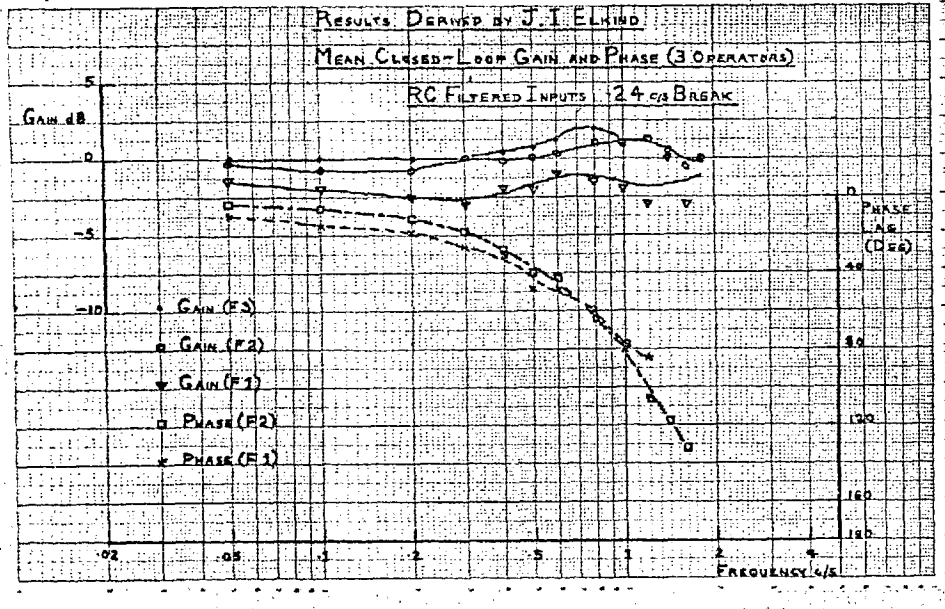

Gain and Phase of Equivalent Open-Loop Transfer
of Model as in Matching Trial 58 Operator D
Gain dB
Phase
Gain
0 dB
180°
Phase Lag (Deg)
Frequency c/s

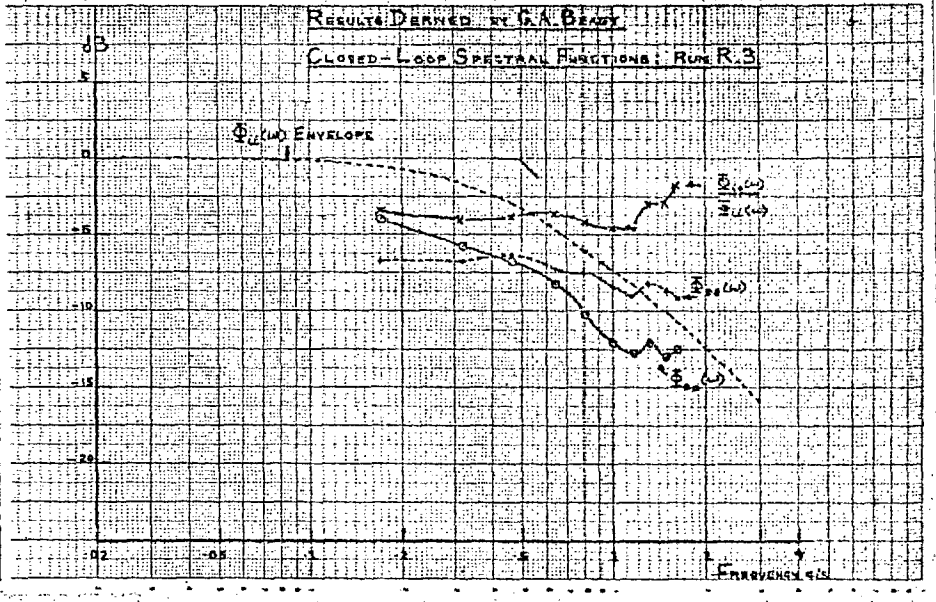

Gain and Phase of Equivalent Open-Loop Transfer
of Model as in Matching Trial 38 Operator F
Gain dB
Phase
Gain
0 dB
180°
Phase Lag (Deg)
Frequency c/s


CLOSED-LOOP SPECTRAL CHARACTERISTICS OF MODEL
MATCHING TRIAL 92 OPERATOR L
dB
10
0
-10
-20
-30
-40
O/A PHASE ∠ Φ*xo(ω) / Φ*xx(ω)
O/A GAIN² |Φ*xo(ω) / Φ*xL(ω)|
Φ*xx(ω)
Φ*xε(ω)
0
40
80
120
160
200
240
PHASE LAG (DEG)
.02
.05
.1
.2
.4
.6
.8
1
2
3
4
FREQUENCY c/s

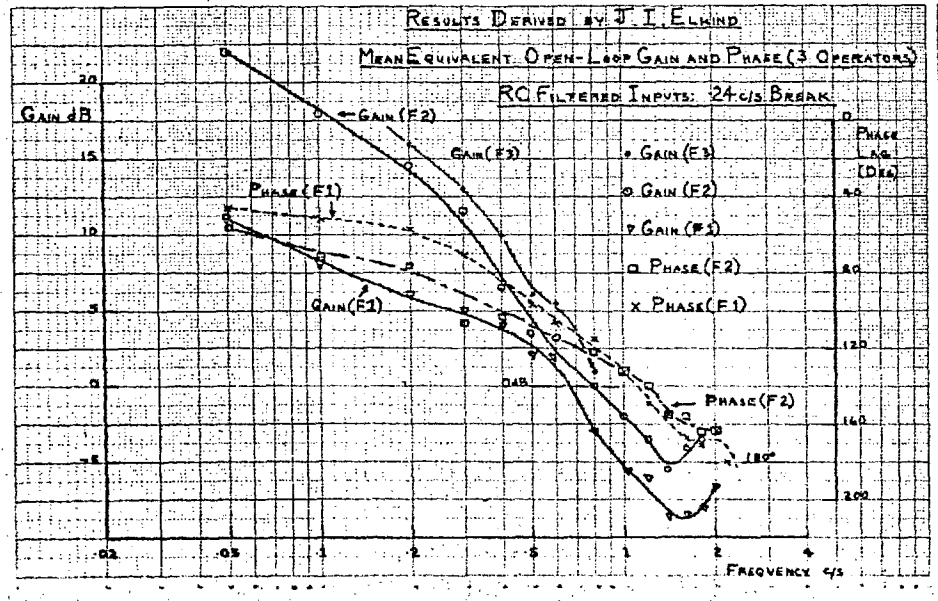
FIGS. 7.6.9-10.

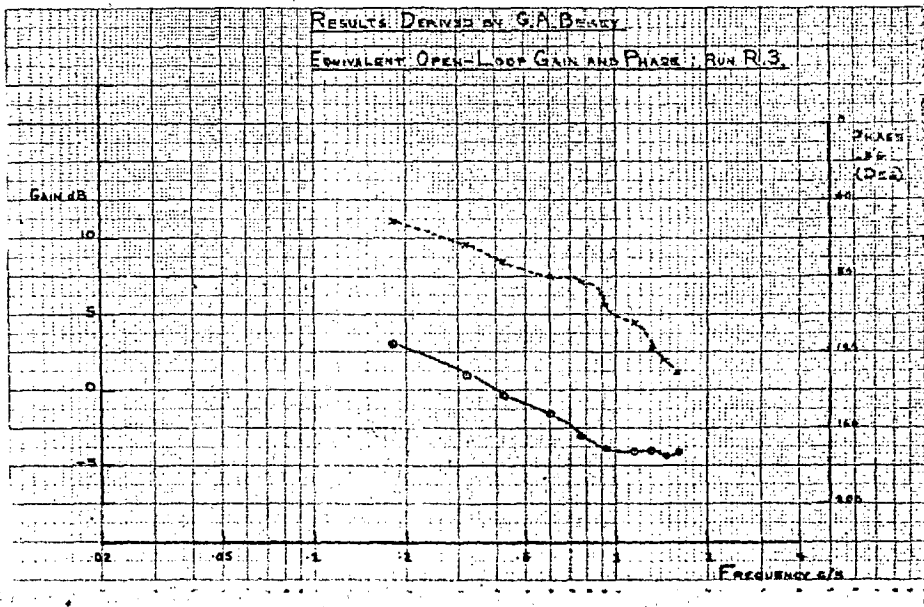

GAIN AND PHASE OF EQUIVALENT OPEN-LOOP TRANSFER OF MODEL AS IN MATCHING TRIAL 92: OPERATOR L
GAIN dB
25
20
15
10
5
0
-5
GAIN
PHASE
0 dB
180°
0
40
80
120
160
200
PHASE LAG (DEG)
.02
.05
.1
.2
.5
1
2
4
FREQUENCY c/s


Results Derived by J. I. Elkind
Mean Closed-Loop Gain and Phase (3 Operators)
RC Filtered Inputs: 24 c/s Break
Gain dB
Phase Lag (Deg)
Gain (F3)
Gain (F2)
Gain (F1)
Phase (F2)
Phase (F1)
Frequency c/s


Results Derived by G. A. Beaty
Closed-Loop Spectral Functions: Run R.3
dB
Frequency c/s

Figs. 7.6.11.–14.


Results Derived by J. I. Elkind
Mean Equivalent Open-Loop Gain and Phase (3 Operators)
RC Filtered Inputs: 24 c/s Break
Gain dB
Phase Lag (Deg)
Gain (F2)
Gain (F3)
Gain (F1)
Phase (F1)
Phase (F2)
180°
Frequency c/s


Results Derived by G. A. Beaty
Equivalent Open-Loop Gain and Phase: Run R.3
Gain dB
Frequency c/s

The form of the equivalent open loop gain curves is remarkably similar to those given by Elkind, especially in regard to the upturn of gain which occurs near the high frequency end of the characteristics, at frequencies of about 1.5 to 2 cps. Elkind's gain curves tend to exhibit a greater initial rate of fall, and a somewhat lower minimum. Unfortunately, Bekey's curves do not include any data beyond 1.6 cps., but the form of the curve illustrated is consistent with an upturn at about this frequency.

The curves of equivalent open loop phase are also quite similar ; here the model tends to exhibit rather less low frequency phase shift. The calculated gain and phase margins are about 2 dB. and 20° smaller than those given by Elkind's data. Bekey's data show a phase margin of 110°, which is considerably larger than those shown by other data ; it was not possible to determine the gain margin because of the limited frequency range of the data.

The above discussion illustrates the similarity of the calculated spectral characteristics of the best-fit models to the published results of the two most directly comparable studies. This similarity is made all the more significant by recognising that the input functions used by Elkind and by Bekey differed considerably from each other, and from that used in the present study. The differing statistical characteristics of the input signals could well explain such discrepancies as were observed. This point will be considered further in Chapter 9, where the spectral functions are considered from the point of view of optimisation theory.

7.7. Summary

The sampled data model incorporating prediction, as described in Chapter 5 , was evaluated by comparison of its responses with those of operators, both in regard to actual time traces, and to covariance functions computed from these.

Time traces relating both to a random ramp and to a continuous random signal input were examined. With suitable parameter settings it was found that the model gave an excellent match to operators' position output and error functions, and also gave a realistic simulation of 'ripple' in the operators' outputs. Comparison of output velocity traces showed that there was very good simulation, except in the case of operator D. This operator tended to display a 'filtered' form of hand motion, to which the sharp velocity-triangles of the model could only give a fair approximation. Pole-zero plots of Z-transfers calculated from the parameter values associated with the best-fit models, showed that all the operators displayed very stable closed loop response, with a tendency for zeros to cancel poles in the left half Z-plane.

Covariance functions, digitally computed from records of

continuous tracking showed close matching between corresponding operator and model functions. Excellent matches were shown by input-error and input-output covariances. The match of output covariances was good as regards shape, but absolute values of output variance showed some discrepancy, largely attributable to model mismatch. The match shown by error autocovariances was only moderately good. These functions were most sensitive to model mismatch and operator remnant. Examination of error-output cross-correlations showed close correspondence between operator and model functions, except in the range of negative lags from 0 to -2 secs. This could be ascribed to spurious correlation due to operator remnant, and indicated that a large part of the observed discrepancy of error autocovariances could be attributed to the effects of remnant. Consideration of the model's Z-transfer showed that slightly better fits of covariance data could be obtained by small alterations of the model parameters ; i.e. some of the observed differences between covariance functions were due to model mismatch. Investigation indicated the probable fruitlessness of extension of the model structure to obtain even better fits to operators' covariance functions.

Calculated spectral characteristics of best-fit models reflected the differences and similarities between operators. Their chief features were in agreement with results published by Elkind and by Bekey, particularly as regards an overall loop gain increasing with frequency, and an equivalent open loop gain showing an upturn in the region of 2 cps. Phase curves were also in good agreement, and showed that the operator was able to compensate for his inherent pure delay, and thus retain relatively small overall loop phase shift, at low frequencies. It was noted that such discrepancies as were observed between calculated and published data could probably be largely explained by the use of different forms of input function. This point is considered further in Chapter 9.

Chapter 8 : QUALITATIVE AND QUANTITATIVE RESULTS OF OPEN LOOP TESTS

8.1. Motivations for Conducting Open Loop Tests

The present chapter describes experiments conducted with a view to directly evaluating the forward,or open loop,transfer which the operator adopts when tracking a continuous random input. This entailed presenting the operator with a synthetic error signal, derived from the model,while at the same time his external feedback path was interrupted. The motivations for conducting this investigation were :-

(1) The interruption of external feedback prevented the recirculation of operator remnant. If useful open loop responses could be obtained,then this would provide an opportunity for estimating uncorrupted statistical functions (see Appendix VIII) and hence, direct evaluation of the operator's forward transfer.

(2) The presentation of model error as a synthetic error signal to the operator would provide a severe test of the quality of the model,as described below. Also,the features of the operators' open loop responses could be compared with those which might be predicted by considering the form of the model's open loop transfer. Since the model had been formulated by comparing closed loop tracking,this would provide a valuable further test of model quality.

8.2. Method of Conducting Open Loop Experiments

It was required that the open loop test data should represent, as far as possible,the forward transfer usually adopted by the operator when tracking a continuous random input with characteristics as described previously. It was necessary to conduct the test during the course of what seemed to the operator to be a normal tracking run. The experimental procedure was as described below.

The operator was given a count-down and commenced the tracking run in the usual manner (see Section 4.3.) About 50 secs. after the beginning of the run,the error switching circuit was activated (for description see Appendix I). At the appropriate time this circuit switched out the operator's error channel,and simultaneously switched the model error signal onto the operator's display. The exact moment of switching was arranged so as to give minimum disturbance to the operator. Switching only occurred when :-

(a) The difference between the operator and model error signals was zero.

(b) There was fairly close parallelism of operator and model error signals.

After a period of 6 secs. the reverse process of switching occurred. This period was chosen by experiment as being a satisfactory compromise between the desire for a long period of open loop operation to obtain statistical reliability of results, and the fact that the operator tended to sense the change in tracking conditions at about this time. The run was then completed in the normal manner.

Operators A, D, and F participated in the tests, and it proved possible to obtain 5 useful sections of record in each case. Records were rejected if the operator reported disturbance, or his tracking performance during the rest of the run was abnormal ; i.e. his normal tracking performance was affected by the presence of a period of open loop operation. For this reason it proved necessary for the operator to remain ignorant of those runs during which an open loop test was to be conducted, and tests were carried out over a period of three weeks. Preliminary tests carried out by the investigator confirmed that foreknowledge of the occurrence of a test was likely to invalidate the utility of results.

8.3. Consideration of Qualitative Results of Open Loop Tests

The actual form of the time traces obtained during open loop operation is typified by the responses shown in Figs.8.3.1-4. It may be noted that, when operating in an open loop configuration, the operator tended to display a somewhat more oscillatory output motion and to exhibit an overall drift. However, if allowance is made for these effects, the form of the operator's output motion appears to be similar to that of the model. Those tests where operators showed undue disturbance were characterised by rather large output oscillations of very variable frequency, but still had some features in common with the model's response.

The subjective effects of open loop operation, which eventually caused the operator to become aware that a change had been made in the tracking configuration were twofold :-

(1) The synthetic error signal, though appearing generally similar to the operator's own signal, seemed to possess points of unusually high velocity. This phenomenon was most noticeable to operator A who interpreted it in terms of a change of input spectrum. It was less marked in the cases of operators D and F.

(2) The operator tended to notice his unusually oscillatory hand motion Again, the effect was noted most strongly by operator A.

These subjective effects imposed an upper limit on the period of open

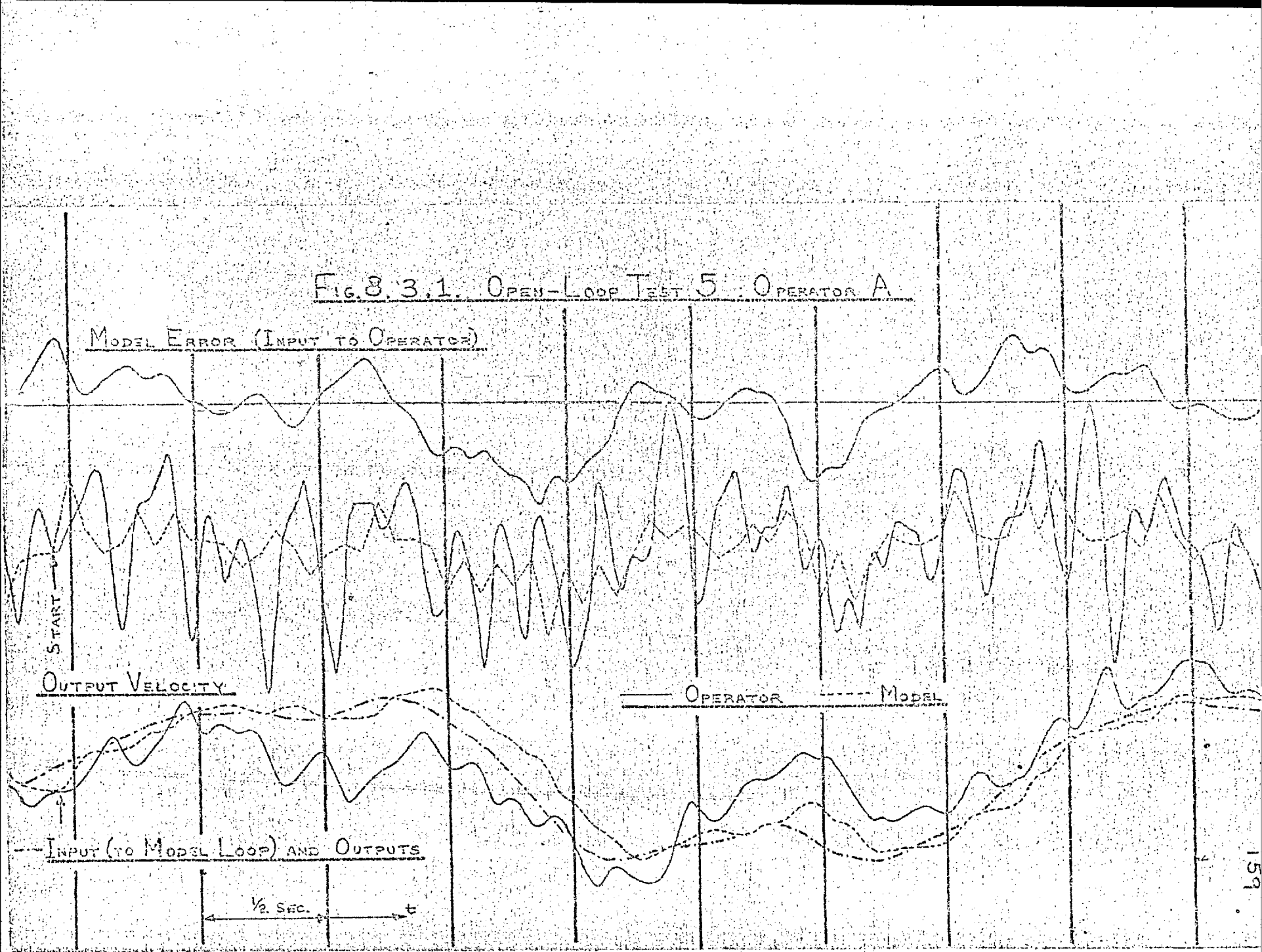


Fig. 8.3.1. Open-Loop Test 5 : Operator A
Model Error (Input to Operator)
Start
Output Velocity
Operator
Model
Input (to Model Loop) and Outputs
½ Sec.
t

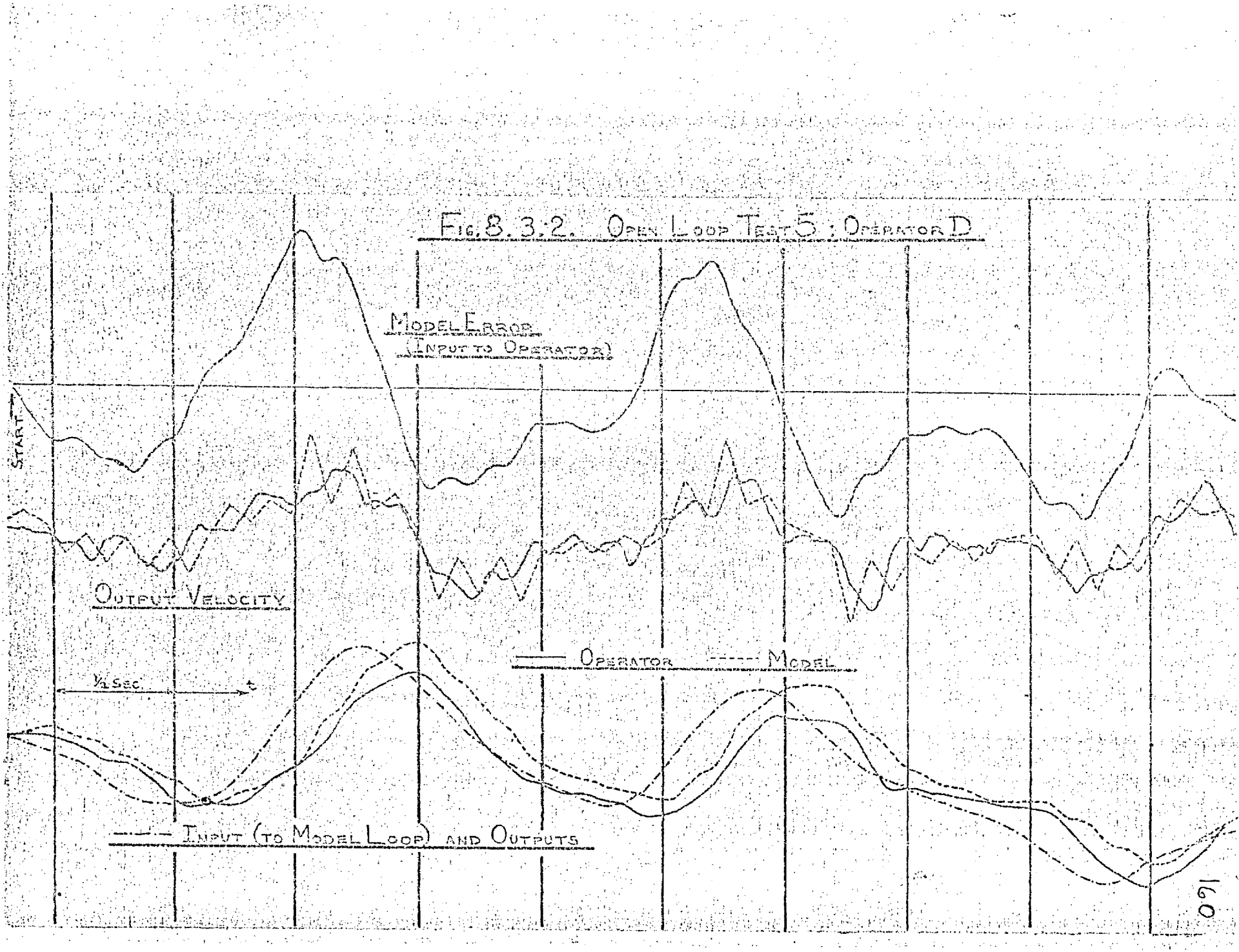


FIG. 8.3.2. OPEN LOOP TEST 5 : OPERATOR D
MODEL ERROR
(INPUT TO OPERATOR)
START
OUTPUT VELOCITY
OPERATOR
MODEL
½ SEC
t
INPUT (TO MODEL LOOP) AND OUTPUTS

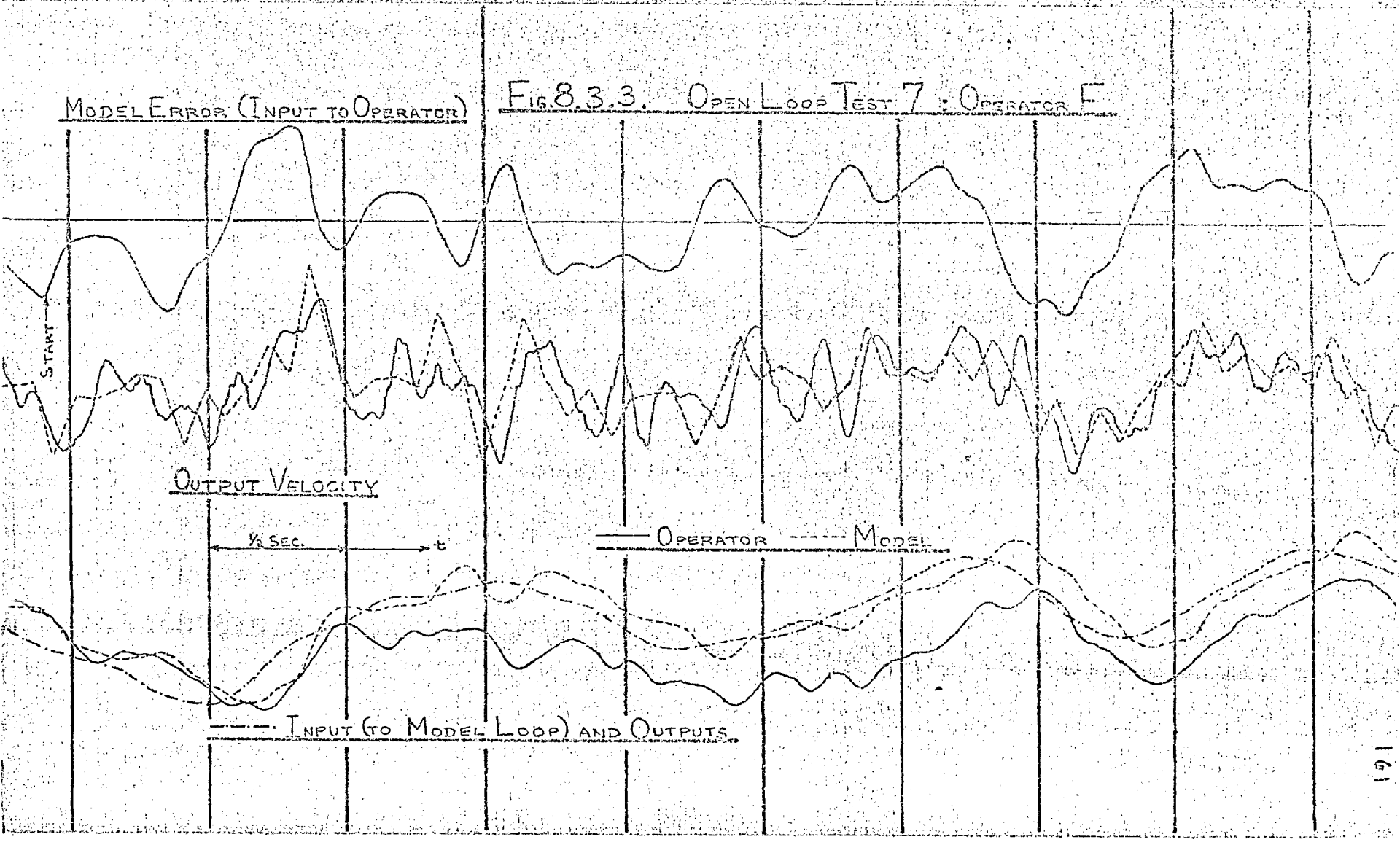


FIG. 8.3.3. OPEN LOOP TEST 7 : OPERATOR F

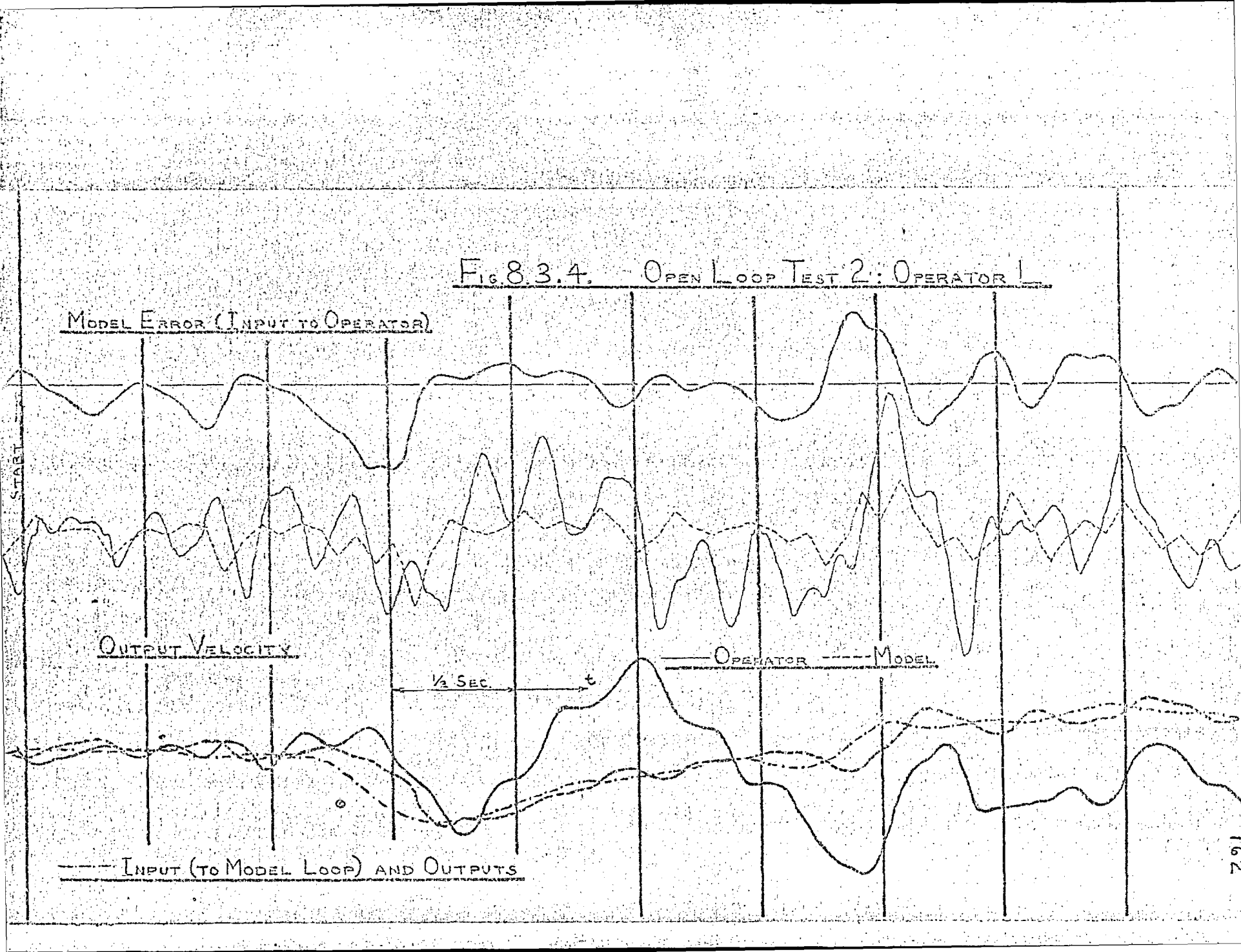


Fig 8.3.4. Open Loop Test 2: Operator 1
Model Error (Input to Operator)
START
Output Velocity
½ Sec.
t
Operator
Model
Input (to Model Loop) and Outputs

loop operation,and also on the number of tests which could be performed. After about seven or eight tests,operators' normal tracking tended to deteriorate,and they tended to inject more noise than usual - this was probably part of a probing action to test loop gain.

The above results could be explained quite well on the basis of the model,which had been developed in connection with closed loop operation. Thus the subjective effect due to the injection of the synthetic error signal was due to the lack of coincidence of model and operator sampling instants. Peaks of model velocity error occurred between sampling instants,as a result of the operation of the position correction velocity-triangle circuit ; the operator would not perceive similar peaks in his own error. The form of recordings of operator and model error (see Section 7.3.) suggested no other basis on which the operator might readily distinguish between synthetic and actual errors.

The fact that the operator was able to perceive his unusually oscillatory hand motion lent weight to the hypothesis that he utilised kinaesthetic feedback in normal closed loop tracking. The oscillatory form of the operator's output may be explained in terms of two effects,viz.,his perception of rather high apparent velocity errors,and the form of his forward transfer,as indicated by that of the model.

The forward Z-transfer of the model is shown in terms of Z-plane pole-zero plots in Fig,8.3.5. Salient features common to these transfers are (a) the presence of a pole near $z = 1$,and (b) the presence of complex conjugate poles lying practically directly on the unit circle near $z = -1$. (a) would give rise to an effective quasi-integration,so that one would expect that differences arising between operator and model responses would tend to be cumulative, which would account for the drift observed in practice - though part of this drift is probably due to drift in information from kinaesthetic feedback. (b) would give rise to a neutrally stable response,so that there would be a tendency towards oscillation at near the half sampling frequency - i.e. about 3 cps. In the case of the operator ,considerable frequency modulation would be expected, due to random spacing of sampling instants. Such an oscillatory effect may be observed most clearly in records of operators A and L (the latter showed too much disturbance to yield quantitatively useful results) ; operators D and F tended to adopt a rather smoother, 'filtered' form of hand motion in normal tracking,and so show the effect much less clearly.

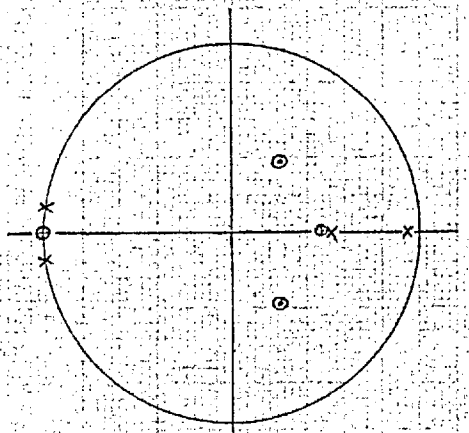
Operators' open loop responses differed from each other in a manner which accorded with the difference of their closed loop responses. The tendency for operators F and D to show less oscillatory open loop responses could also arise as a result of two further factors :-

(1) Models matching operators F and D had longer sampling intervals (150 and 166 ms. respectively) than those matching A and L. The reasoning of Section 5.5. would indicate that their respective best-fit models might have need of velocity computation loop gains

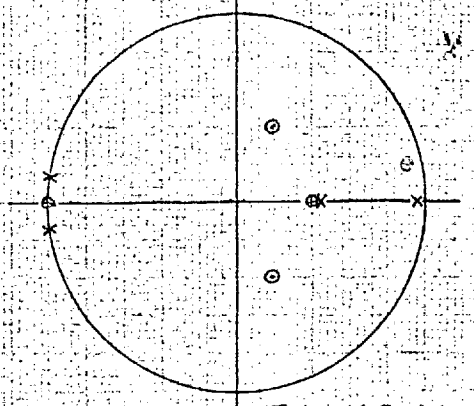
FIG. 8.3.5.

OVERALL OPEN-LOOP (FORWARD) Z-TRANSFER OF MODEL WITH PARAMETERS SET FOR BEST MATCH TO CONTINUOUS TRACKING RESPONSE

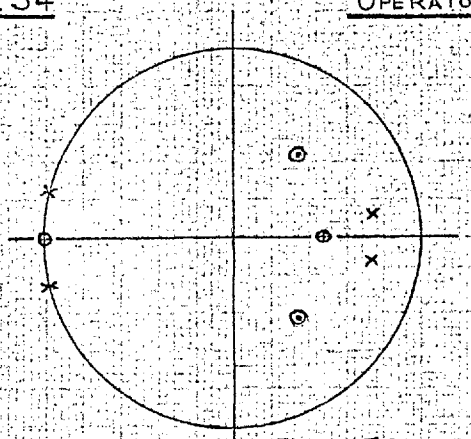
TRANSFER: $\underline{C}(z) = \frac{1}{h(z)} \cdot \begin{bmatrix} h_{pp}(z,1) & ; & h_{pv}(z,1) \\ -2m_3m_5T^3 & ; & Tm_2(z^2-1) \end{bmatrix} \cdot \underline{E}(z)$



OPERATOR A TRIAL 34



OPERATOR C TRIAL 19



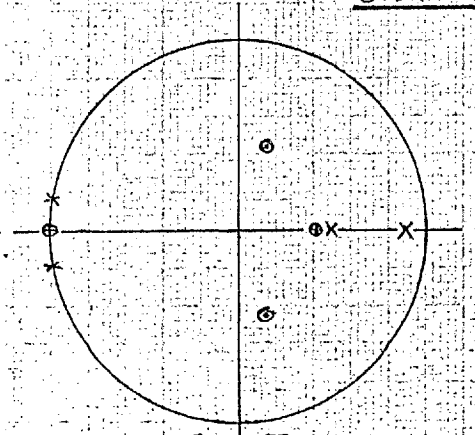
OPERATOR D TRIAL 58

LEGEND

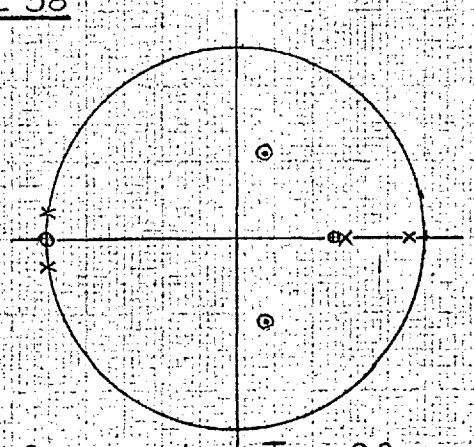
× POLES OF $1/h(z)$

⊕ ZEROS OF $h_{pp}(z,1)$

⊙ ZEROS OF $h_{pv}(z,1)$



OPERATOR F TRIAL 38



OPERATOR L TRIAL 92

slightly less than unity. This would complicate the models' transfers, adding two further poles, and while it would have little effect on the closed loop transfer, the poles near $z = -1$ in the open loop transfer would be moved inside the unit circle.

(2) The hypothesis of instantaneous sampling of position and velocity errors, which was incorporated into the model structure, was an idealisation. The operator might either adopt a differencing operation over a short time, or he might measure the time required for the error signal to change by a given amount. The latter hypothesis seemed plausible, in view of the structure of the retina and the apparently 'digital' nature of brain processing. In either case the precise time taken to sense a given velocity would vary from operator to operator. In addition, the more slowly varying error functions of operators D and F would be likely to lead to a longer average estimation time. Thus these operators would be less likely to sense peaks of velocity ; they would tend to estimate average velocity over a period as long as 50 ms. This latter figure is the likely upper limit to the velocity estimation time indicated by the results of the sampled display tests (Section 4.2), which showed that operators' performances were impaired by display sampling at frequencies as high as 20/sec.

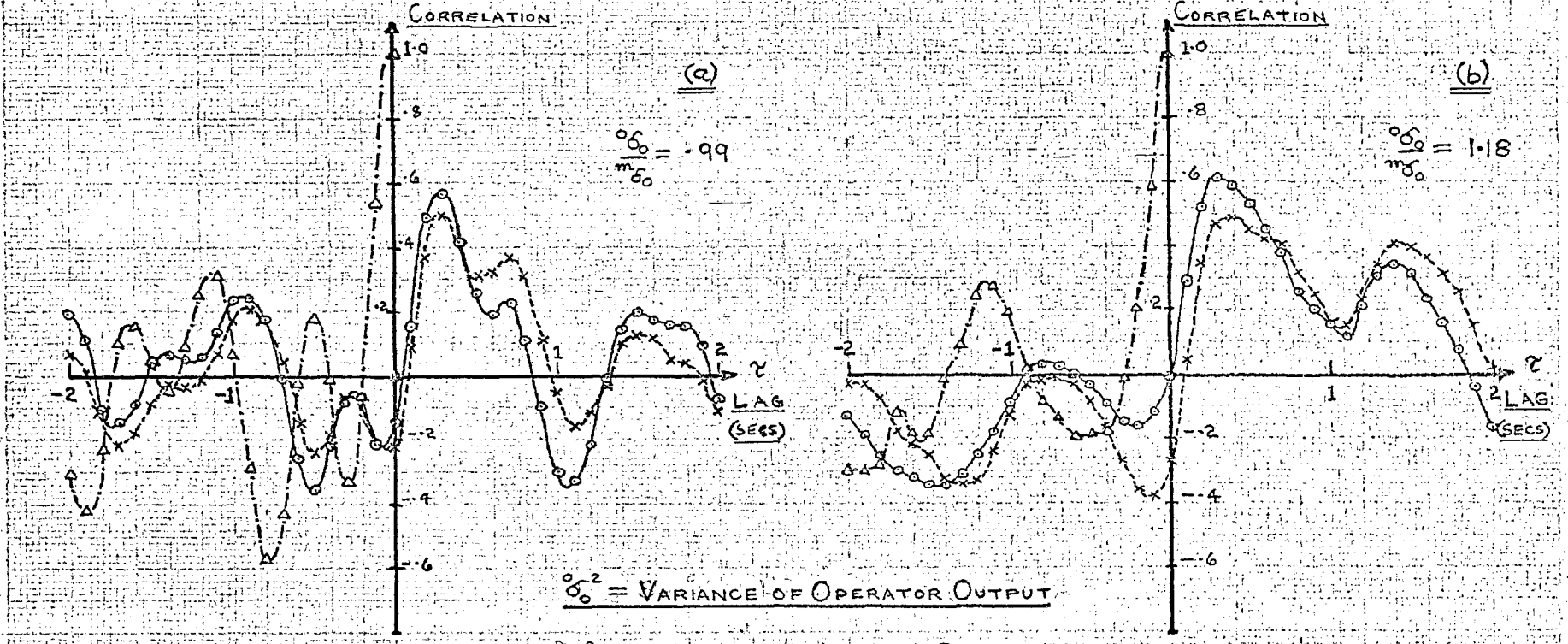
8.4. Consideration of Quantitative Results of Open Loop Tests

In order to extend and quantify the comparison between model and operator forward transfers, correlation functions were computed from records of open loop tests. These records were sampled at 100 ms. intervals, and the raw data thus obtained were subjected to a regression analysis in order to remove D.C. levels and linear trends. This process was necessary because only six seconds of raw data were available from each recording. The refinement of data to give approximate stationarity was therefore a prerequisite for computation of correlation functions through time averaging ; sufficient data were not available for computation of ensemble averages. The removal of D.C. and linear trend corresponds, in terms of spectral estimation, with the proposition that it is impossible to make useful estimates of power density at frequencies lower than $1/T_n$ cps. from a record of length T_n .

Estimates of sampled correlation functions were computed by using the unbiassed estimator as given in equation 7.4.1. A maximum lag of 2 secs. was considered to represent a reasonable compromise between accuracy and the desire to fully extract the information latent in the experimental data. Error autocorrelation and error-output cross-correlation functions are shown in Figs. 8.4.1-3., where the curves connecting data points represent a visual interpolation to facilitate comprehension of the overall form of the functions.

FIG. 8.4.1. CORRELATION FUNCTIONS COMPUTED FROM OPEN-LOOP TEST RECORDINGS

OPERATOR A



$\sigma_{\delta_o}^2$ = VARIANCE OF OPERATOR OUTPUT

$m\sigma_{\delta_o}^2$ = VARIANCE OF MODEL OUTPUT

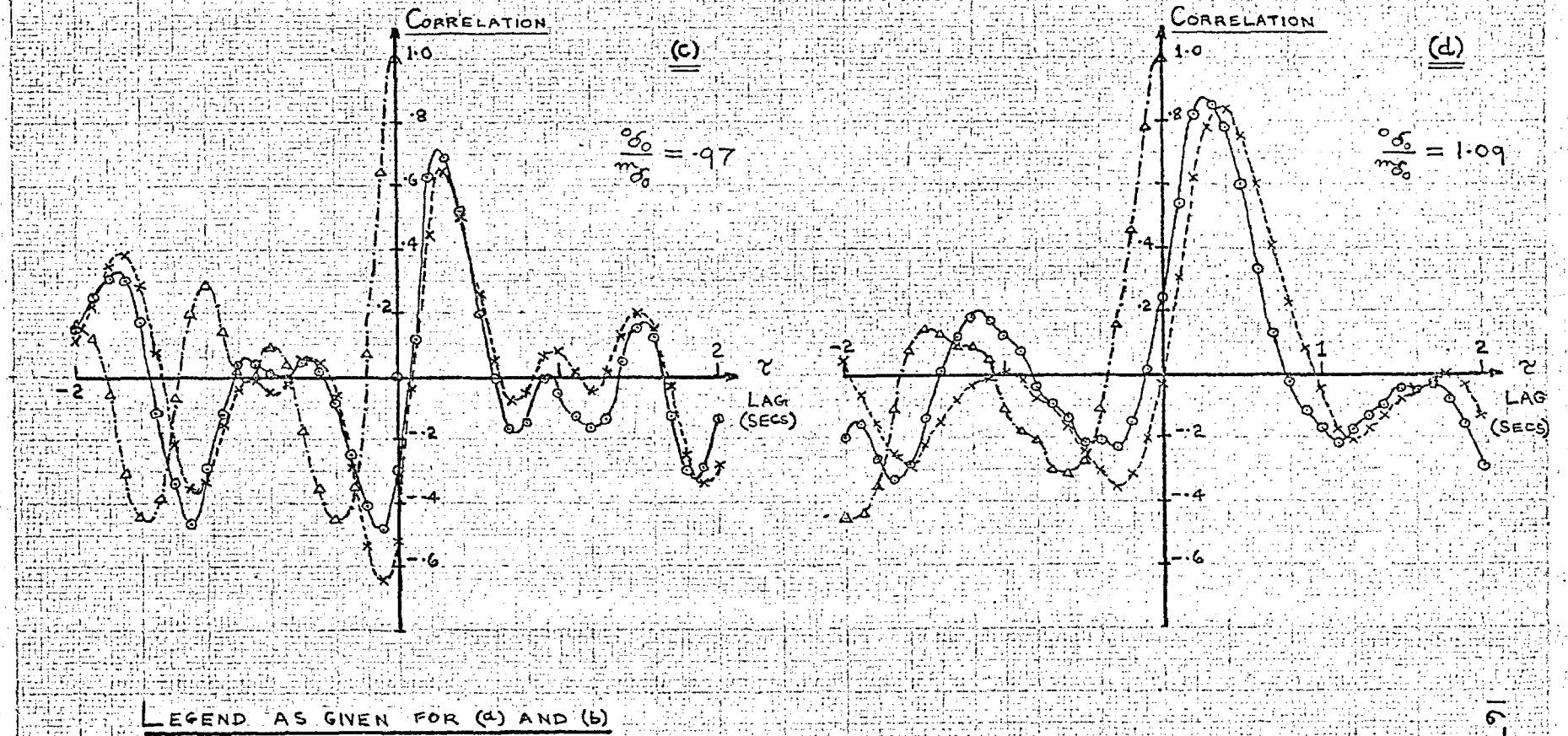
—·—△—·— ERROR AUTOCORRELATION

—o— ERROR-OPERATOR OUTPUT CROSS-CORRELATION

---x--- ERROR-MODEL OUTPUT CROSS-CORRELATION

FIG. 8.4.1. (CONT)

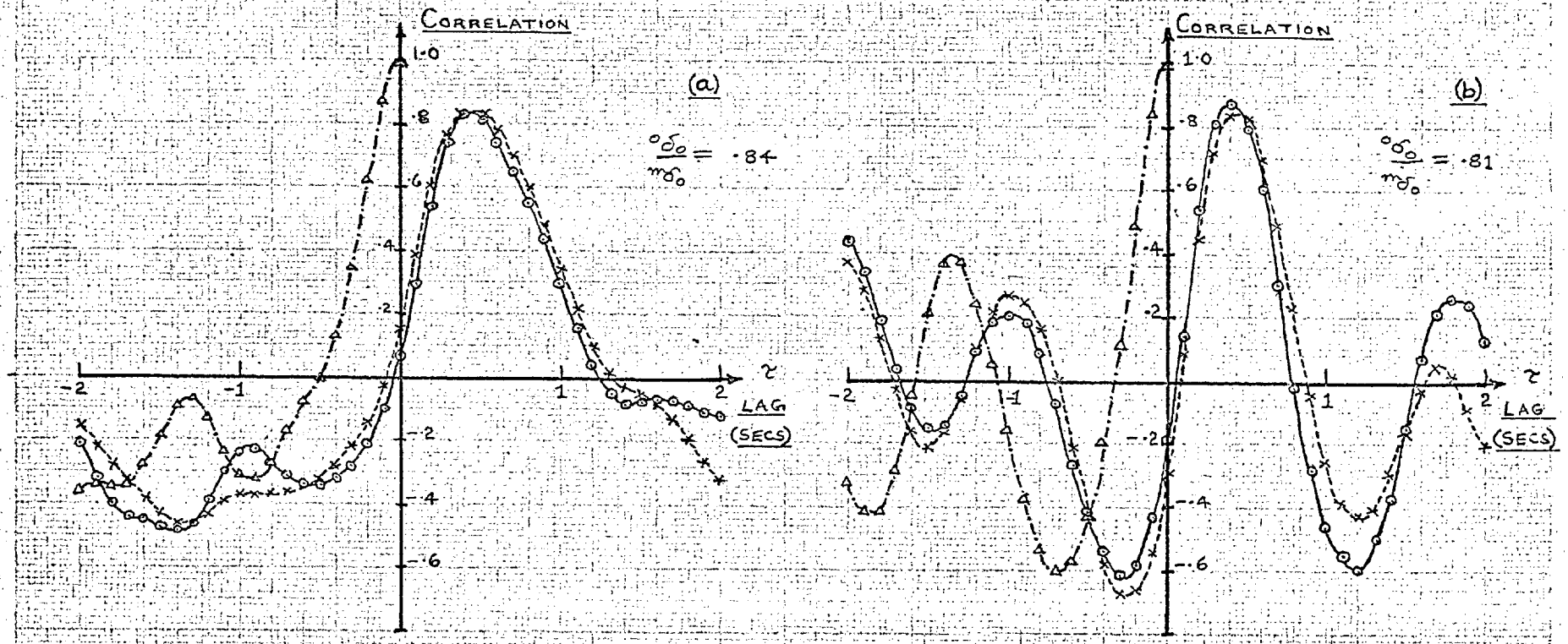
OPERATOR A



LEGEND AS GIVEN FOR (a) AND (b)

FIG. 8.4.2. CORRELATION FUNCTIONS COMPUTED FROM OPEN-LOOP TEST RECORDINGS

OPERATOR D



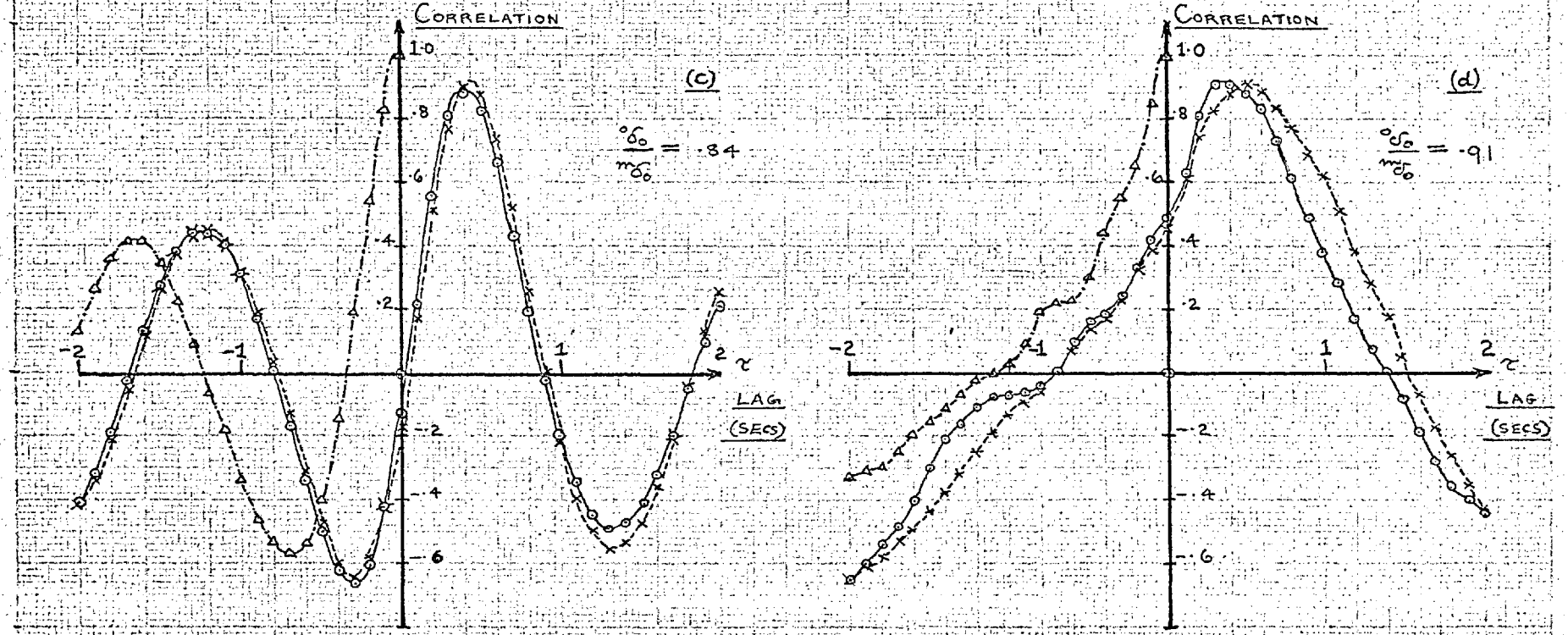
—·—Δ—·— ERROR AUTOCORRELATION

—o— ERROR-OPERATOR OUTPUT CROSS-CORRELATION

--×-- ERROR-MODEL OUTPUT CROSS-CORRELATION

FIG. 8.4.2. (CONT)

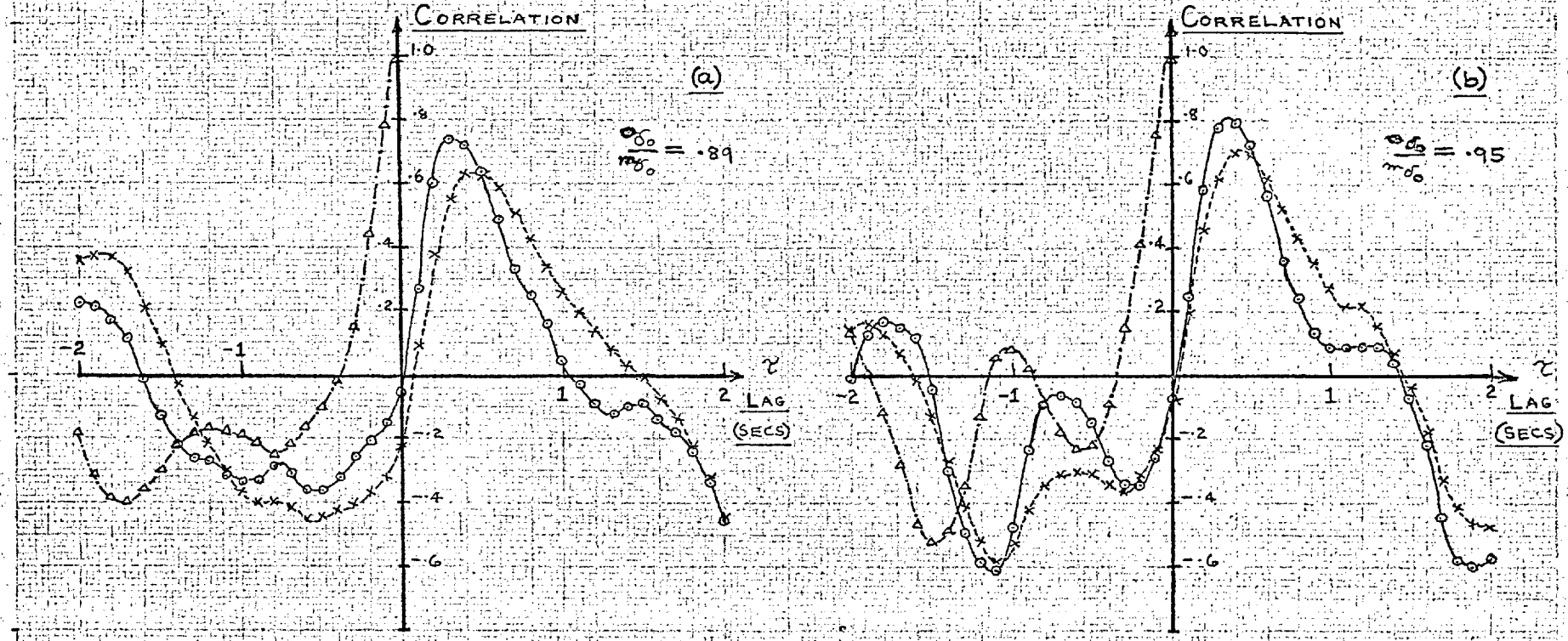
OPERATOR D



LEGEND AS FOR (a) AND (b)

FIG. 8.4.3. CORRELATION FUNCTIONS COMPUTED FROM OPEN-LOOP TEST RECORDINGS

OPERATOR F



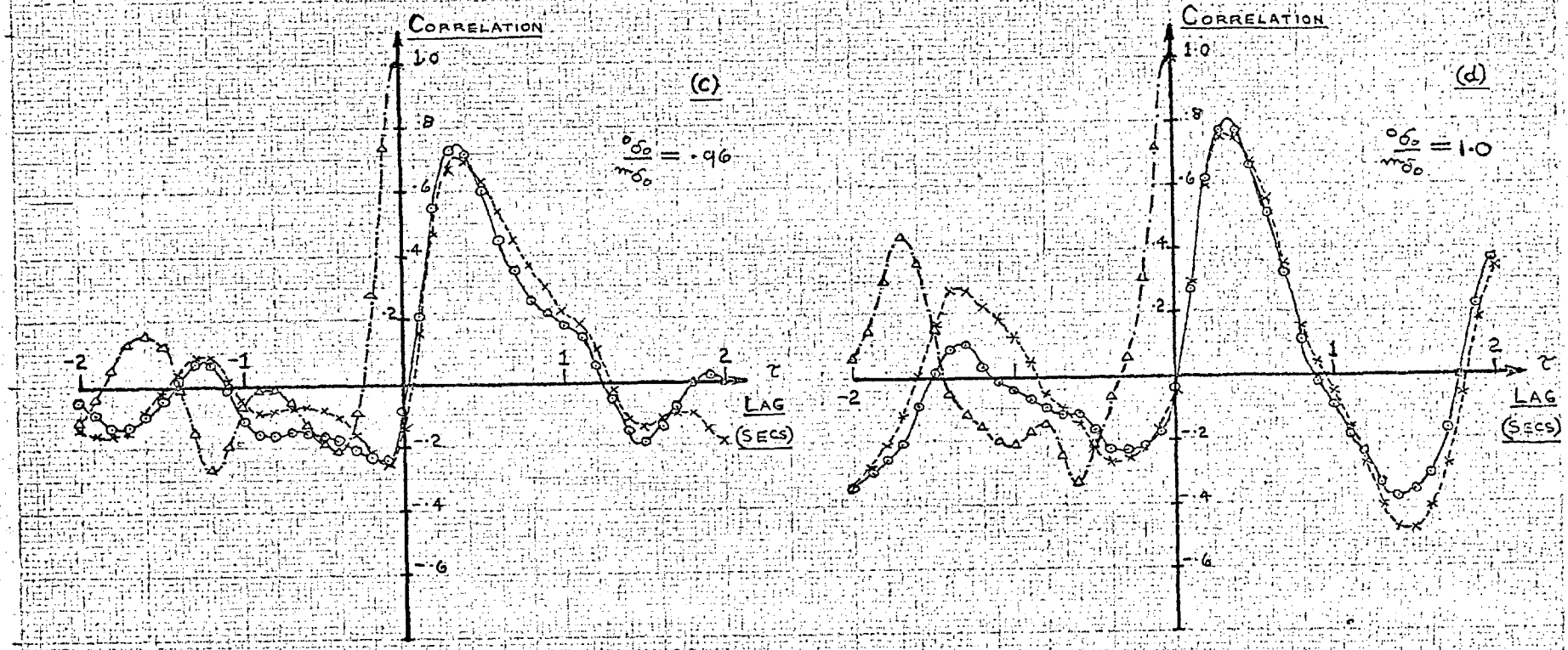
---Δ--- ERROR AUTOCORRELATION

—o— ERROR-OPERATOR OUTPUT CROSS-CORRELATION

---x-- ERROR-MODEL OUTPUT CROSS-CORRELATION

FIG. 8.4.3. (CONT)

OPERATOR F



LEGEND AS FOR (a) AND (b)

The above correlations represent characteristics averaged over the sampling interval, for both operator and model. It may be seen that the model and operator error-output cross-correlation functions exhibit a general similarity of form, but also some quite noticeable differences of detail. If model and operator were to sample synchronously, then one could expect the error-output cross-correlation functions to be very closely comparable, provided the model constituted a close representation of the operator. However, the conditions of the experiment were such that the operator and model could not be expected to sample in synchronism. The equivalent open loop (EOL) and actual open loop (AOL) transfers of the model were not one and the same thing, so even if the model were a perfect representation of the operator, one could expect their respective open loop correlation functions to exhibit some noticeable differences; i.e. functions obtained while the operator was open-looped, would differ.

In considering the degree to which the open loop correlation functions might be expected to differ, the following points were evident. The operator's sampling between model sampling instants implies that he would tend to sense the 'ripple' in the model output (see Section 8.3). Sensing of this ripple would lead to the following effects at the output:-

(a) Divergence of D.C. level and trend.

(b) Oscillatory response.

(c) Differences between shapes of operator and model response.

Factors (a) and (b) above would be practically eliminated by the process of correlation computation. Factor (c) would be minimised by taking correlations, because the ripple between model sampling instants is not highly correlated to the position error at sampling instants ; the ripple is more highly correlated with the velocity error at sampling instants, though the operator's random sampling, and random variations in delay would tend to minimise the effect on correlation functions.

The above arguments indicate that one could expect at least a moderate amount of divergence between the respective error-output cross-correlation functions of operator and model, obtained while the operator was in the open loop configuration. Experimental results concur with this prognosis.

Further consideration is given to the quantitative results of the open loop tests in terms of the frequency domain, in Section 8.7. In order to effect this, simple models were fitted to the operators' effective open loop transfers, as described in the following two Sections.

8.5. The Fitting of Models to the Open Loop Test Data

The open loop test data were further evaluated by fitting simple models to the operators' forward transfers through direct use of error and output functions. The first step in the analysis was to obtain representative covariance functions by taking an ensemble average over five individual records, for each of the three operators from whom useful records were obtained. These averaged covariance functions are illustrated in Figs.8.5.1-3., where the lines joining data points result from visual interpolation.

At first sight, the most promising method of analysis seemed to be to transform the covariances (which were in a suitable form, D.C. and trend having been removed from the raw data) so as to give the error spectrum, $\Phi_{ee}(w)$, and the error-output cross-spectrum, $\Phi_{eo}(w)$; these spectra could then be used to estimate the frequency response function, $G(jw)$, according to :-

$$G(jw) = \Phi_{eo}(w) / \Phi_{ee}(w) \qquad 8.5.1.$$

A suitable simple form of model could then be proposed, and its parameters fitted to $G(jw)$ in the spectral domain. The method of fitting could either be visual, or by regression.

While appearing attractive, in the present case the above procedure unfortunately suffered from a number of drawbacks :-

(1) The open loop test data was obtained from records only six seconds long, which would indicate a maximum usable covariance lag of .6 secs., according to generally accepted criteria (28). Even if lags up to 1 sec. were used, this would only give a frequency resolution of about 1 cps. Since the data were sampled at 10/sec., corresponding to a folding frequency of 5 cps., it would only be possible to estimate three or four samples of the spectral functions. The accuracy of these estimates was likely to be poor.

(2) The process of division of spectral functions would tend to accentuate errors of estimation.

(3) A review of pertinent literature (10,11,12) indicated that the simplest models should contain at least two parameters, and exhibit a break frequency less than 1 cps. Thus the accuracy of parameter estimation would be very poor.

(4) It would be difficult to estimate the degree of fit in terms of time domain functions, with any accuracy.

(5) A suitable programme for digital computation of spectral estimates was not readily available, and it was not feasible to develop one, having regard to restriction of available computer time.

The above difficulties caused by inaccuracy of spectral analysis

FIG. 8.5.1. OPEN-LOOP COVARIANCE FUNCTIONS (AVERAGE OF 5 RECORDS)

OPERATOR A

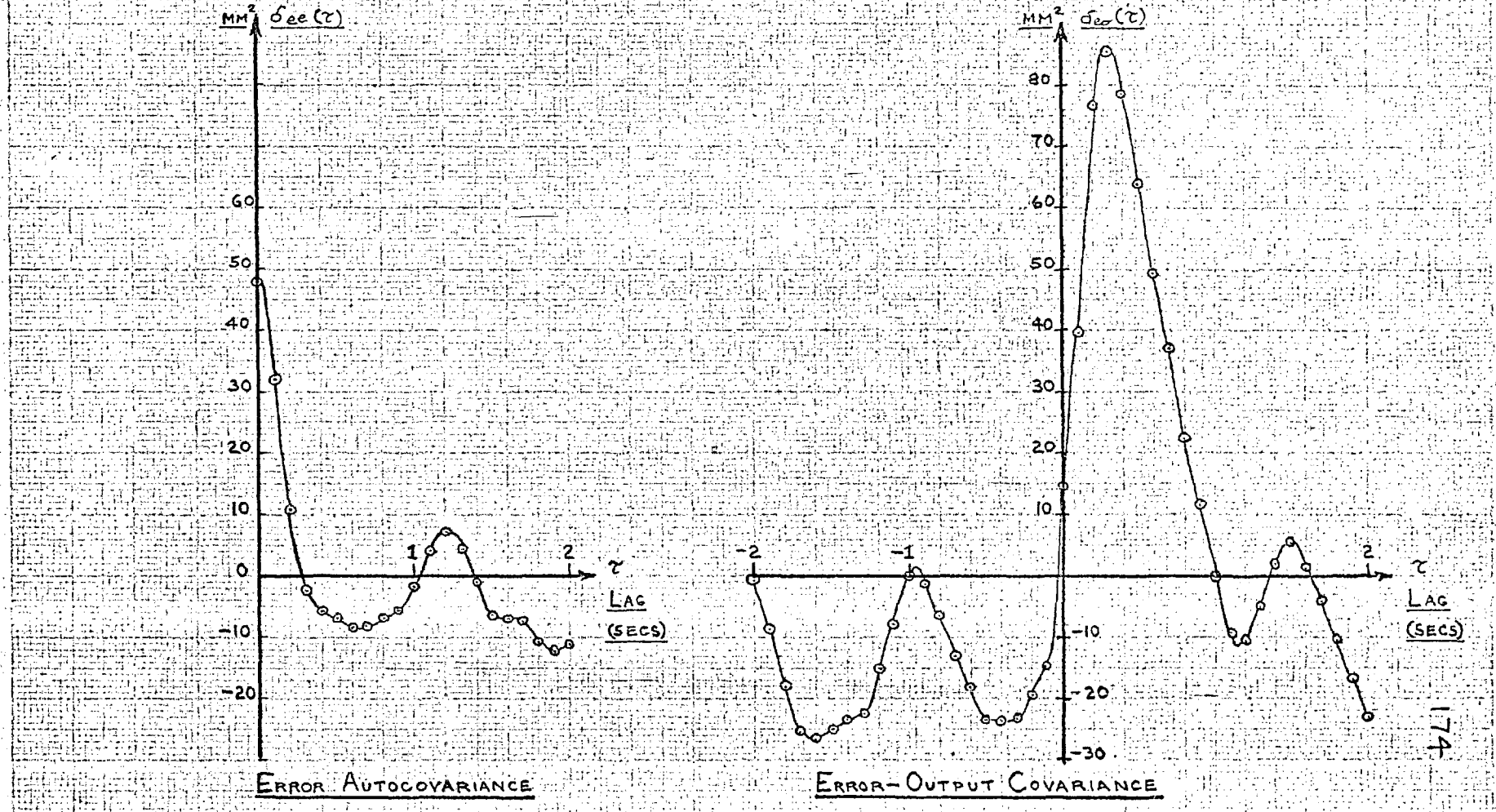


FIG. 8.5.2. OPEN-LOOP COVARIANCE FUNCTIONS (AVERAGE OF 5 RECORDS)

OPERATOR D

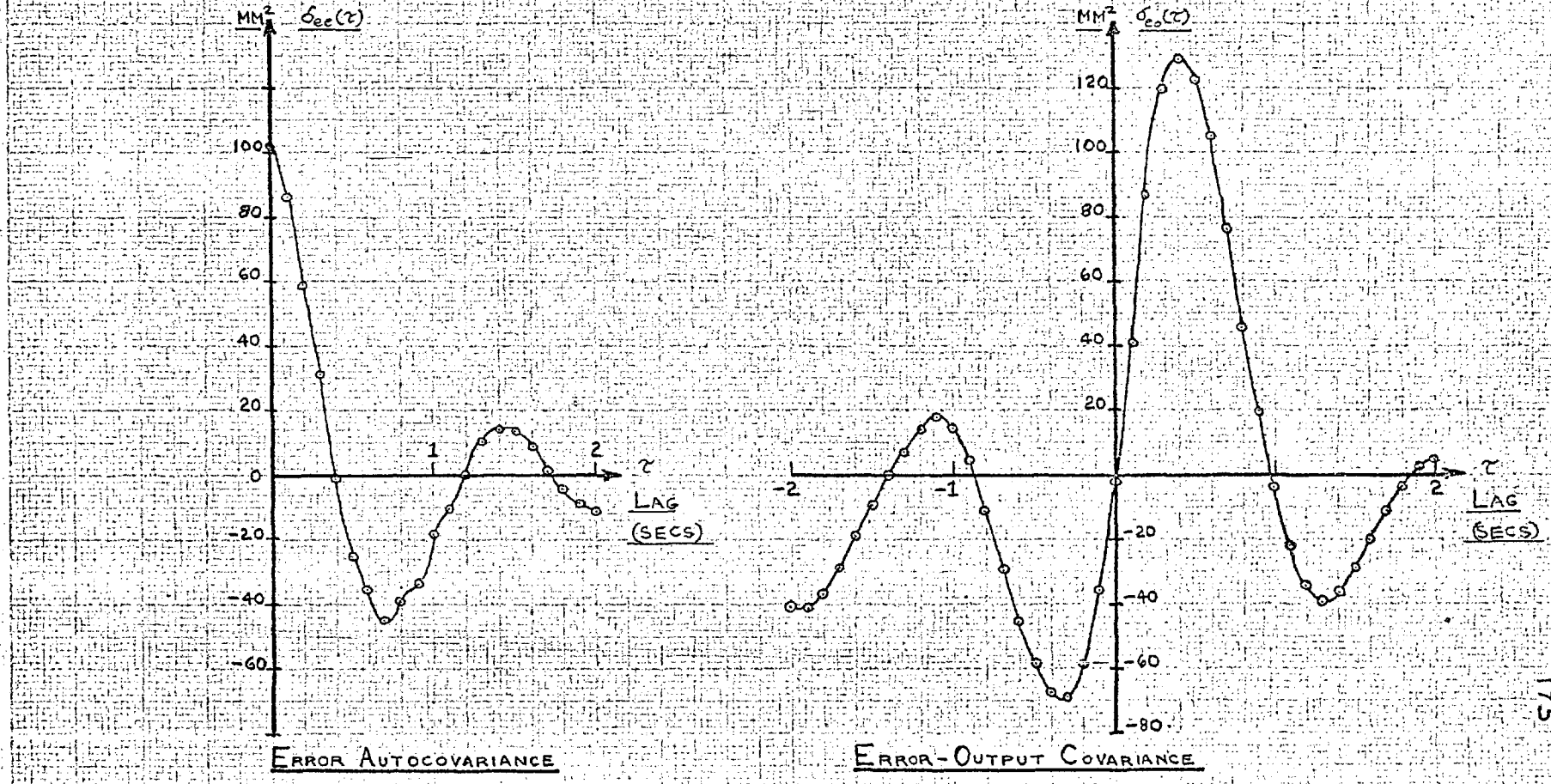
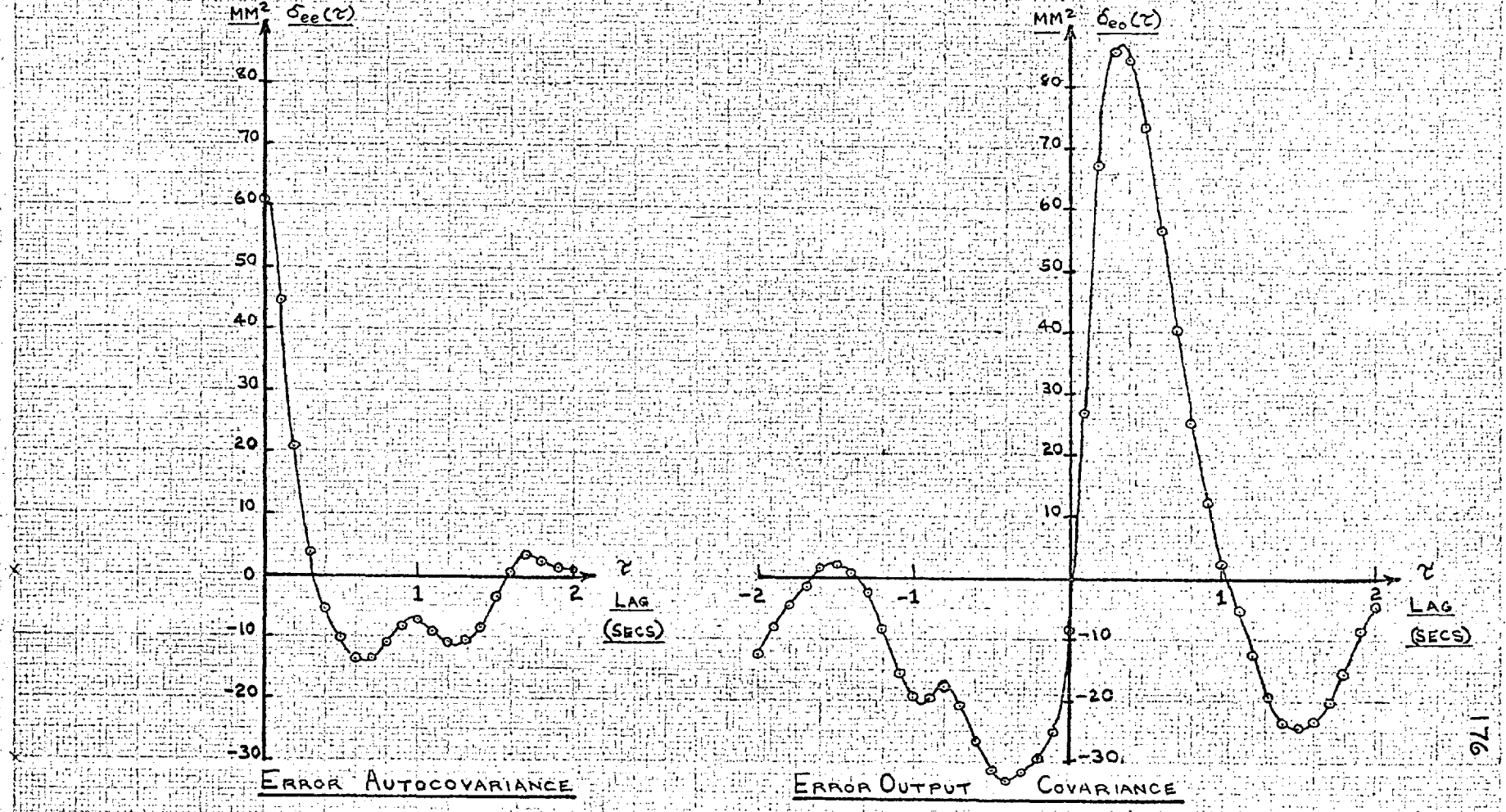


FIG. 8.5.3. OPEN-LOOP COVARIANCE FUNCTIONS (AVERAGE OF 5 RECORDS)

OPERATOR F



arise because the process of spectral estimation inevitably discards much information present in the original time traces and correlation functions. The mechanism by which this arises has been described by Elkind & Green (16),and may be appreciated by reference to Appendix IX.

Consideration of the above facts led to the fitting of model parameters by direct use of correlation function data. The method used was essentially one of multiple regression analysis,and is derived from results presented by Bartlett (29,p259). The method entails the use of sampled data,which was convenient in the present case. The method is based on the use of the equation,written in terms of Z-transforms:-

$$\bar{\Phi}_{eo}(z) = G(z).\ \bar{\Phi}_{ee}(z) \qquad 8.5.2.$$

where G(z) represents of the fitted model,(Either a sampled model,or a linear continuous model with a fictitious sampling and data reconstruction operation,can be proposed) and $\bar{\Phi}_{eo}(z)$ and $\bar{\Phi}_{ee}(z)$ represent the Z-transforms of the error-output cross-correlation and error autocorrelation functions respectively. Details of the method are described in Appendix IX.

It is difficult to determine the probable error in the values of the regression coefficients derived by the above method of analysis. Even if the data were perfectly accurate,and handled with perfect accuracy (both practical impossibilities) the method of using the data implies that the number of effectively independent samples is smaller than the number which would be deduced through the usual formula. The best indicator of likely accuracy is the coefficient of determination,provided that the number of regression parameters is small compared with the number of data points. One could also use a Monte Carlo method to determine the statistics of the regression coefficients. Unfortunately,the data were insufficiently extensive in the present case. The application of the above regression method in fitting simple models to experimental data is described in the following Section.

8.6. Models Fitted to the Open Loop Test Covariance Data

Two simple continuous models were fitted to the open loop test covariance data :-

(a) $$G_1(s) = k.\frac{b.e^{-(n+m)\Upsilon.s}}{(s + b)}$$

(b) $$G_2(s) = k.\frac{b(s + a)}{a(s + b)}.e^{-(n+m)\Upsilon.s}$$

where n is an integer, m = 0 or ½ ,and Υ = .1 sec.(data sampling interval)

The form of (a) is as suggested by Elkind (10),while (b) is an extension to allow for the operator's sensing of input velocity. In fitting (b) it was also possible to evaluate the parameters of a sampled model,sampling at 10/sec. This model was of interest because it has been suggested by Wiener (23) that the operator might sample at a frequency similar to that of the α-rhythm observed in e.e.g. records of a subject at rest.

The simplest continuous model,(a),was fitted by the method of direct estimation,as outlined in Appendix IX ; values of m and n were specified in advance. Visual interpolation and estimation of $\frac{\partial}{\partial\tau}\delta_{eo}(\tau)$ was employed. The values of model parameters derived from the regression analysis are shown in Table 8.6.I.

The more complicated continuous model,(b), was fitted by forming an equivalent hypothetical sampled model,and fitting the Z-transfer,$G_h(z)$,of this model to the sampled covariance data ; the basis of the method is described by Jury (30). The first step was to introduce a fictitious sampler into the model input,and to interpose a suitable data reconstruction circuit,transfer Q(s) ,between the sampler and continuous model transfer. The following relation was then evident :-

$$G_h(z) = \mathcal{Z}\text{-}\{Q(s).G_2(s)\} \qquad 8.6.1.$$

The form of the data reconstruction circuit which was employed is given by :-

$$Q(s) = \frac{z(1 - z^{-1})^2}{\Upsilon.s^2} \quad ; z = e^{\Upsilon s} \qquad 8.6.2.$$

which represents a first order hold,plus a pure prediction of Υ . This function was chosen because it was the simplest reasonable data reconstruction,and avoided troubles due to transcendental relations between regression coefficients. It implied a data reconstruction on the error autocovariance which represented the function accurately at data points,and interpolated by means of straight lines joining these points. The appropriate form of the transfer $G_h(z)$ was given by :-

$$G_h(z) = \mathcal{Z}\text{-}\left\{z\frac{(1 - z^{-1})^2}{\Upsilon s^2}.\frac{b}{a}.k.\frac{(s + a)}{(s + b)}.e^{-(n+m)\Upsilon.s}\right\} \qquad 8.6.3.$$

TABLE 8.6.I

Parameter Values of Simple Continuous Model

Fitted to Open-Loop Test Covariance Data

Form of continuous model:-

$$G_1(s) = k.b.\frac{e^{-(n+m)T.s}}{(s + b)}$$

Operator	Fitted Values of Model Parameters				Coeff. of Determin-ation
	k	b	k dB.	Break Freq. c/s	
A *	3.54	2.9	11.0	.46	.63
	3.10	3.6	9.8	.57	.70
D	2.74	2.1	8.8	.33	.73
F	3.32	2.5	10.4	.40	.75

All model delays are .15 sec., except *, .1 sec.

Whence :-

$$G_h(z) = \frac{(z.a_1 + a_o).z^{-(n+m)}}{(z - b_1)} \qquad 8.6.4.$$

The values of the three parameters were then derived as described in Appendix IX, after specifying values of m and n. From these the values of a, b, and k were calculated.

The above fitting procedure also served to determine the parameters of a sampled model, consisting of a sampler operating at 10/sec., followed by a first order hold feeding directly into a continuous transfer of the form :-

$$C(s) = G_2(s).e^{-\tau s} \qquad 8.6.5.$$

The parameters of both continuous (k,a,b) and sampled data (a_o,a_1,b_1) model transfers are shown in Table 8.6.II.

It may be noted that the addition of a lead term to the basic, single lag, continuous model led to much better model fits, with coefficients of determination in excess of .9 . With similar values of pure delay, the fitted model transfers, $G_2(s)$, showed lower values of lag break frequency than $G_1(s)$, and higher D.C. gains. Too much emphasis should not be placed on individual parameter values, although the high coefficient of determination indicates a fairly low spread. Average values of lead and lag break frequencies were 2.7 and .28 cps. respectively.

Objection might be raised to the form of $G_2(s)$, because it implied a step response exhibiting a sudden jump. However, the values of the model parameters indicate that this initial jump would be relatively small, and could be eliminated by postulating a further lag, with a break at higher frequency than that of the lead. Also, it should be remembered that the model was derived from data relating to a continuous input, and is only an idealised representation of a very complex structure.

The fitting of models involving the determination of more than three parameters was not attempted for the following reasons:-

(1) As Table 8.6.II. shows, the coefficient of determination obtainable with a three parameter model was very good.

(2) It was considered that there were insufficient data points to justify more complicated models. Also, the addition of one extra parameter reduced the number of usable data points by one.

(3) Parameters deduced for more complicated models would be subject to inaccuracy due to build up of rounding errors, which would rapidly become more severe as the number of model parameters increased.

TABLE 8.6.II

Parameter Values of Continuous Model

Fitted to Open-Loop Test Covariance Data

Through Equivalent Discrete Model

Continuous Model:- $G_2(s) = k.\frac{b(s + a)}{a(s + b)}.e^{-(n+m)T.s}$

Discrete Model:- $G_h(z) = \frac{(a_o + z.a_1)}{(z - b_1)}.z^{-(n+m)}$

Operator		A	D	F
Values of Parameters Fitted at		Pure Delay 100 msec.	Pure Delay 150 msec.	Pure Delay 200 msec.
a_o		.111	.170	-.042
a_1		.173	.305	.720
b_1		.885	.825	.815
k		5.3	2.7	3.7
a		25.0	7.5	17.9
b		1.18	1.92	2.05
Break Freq. c/s	LAG	.19	.31	.33
	LEAD	4.0	1.2	2.85
Coeff. of Determination		.95	.91	.97

The parameters of the simple models derived above were subject to inaccuracy due to three principal factors :-

(a) The method of calculation involved finding regression coefficients from correlated data,and these would be subject to larger statistical fluctuations on account of this.

(b) The removal of D.C. and trend from the raw data,together with the short open loop record length,led to a loss of information at low frequencies.

(c) Inaccuracies in data processing.

Nevertheless,the values of parameters as calculated from the data are in general accord with results obtained both in this study and by other workers (10,11). The implications of these fitted model parameters are considered in the next Section.

8.7. Implications of Simple Models Fitted to Open Loop Covariance Data

It is most convenient to consider the implications of the simple fitted models (SFM's) described in the previous Section,in terms of frequency domain characteristics. The frequency response functions of the SFM's are illustrated in Figs.8.7.1-3.,which also show the calculated actual open loop frequency response (AOL) of the best-fit models,as described in Chapter 6. The corresponding equivalent open loop responses (EOL) of these best-fit models are shown in Figs.7.6.2.,6.,& 8.

Comparisons of the magnitude curves show that :-

(a) At very low frequencies the SFM curves are closer to the AOL curves,except in the case of operator D,where they lie between the AOL and EOL curves.

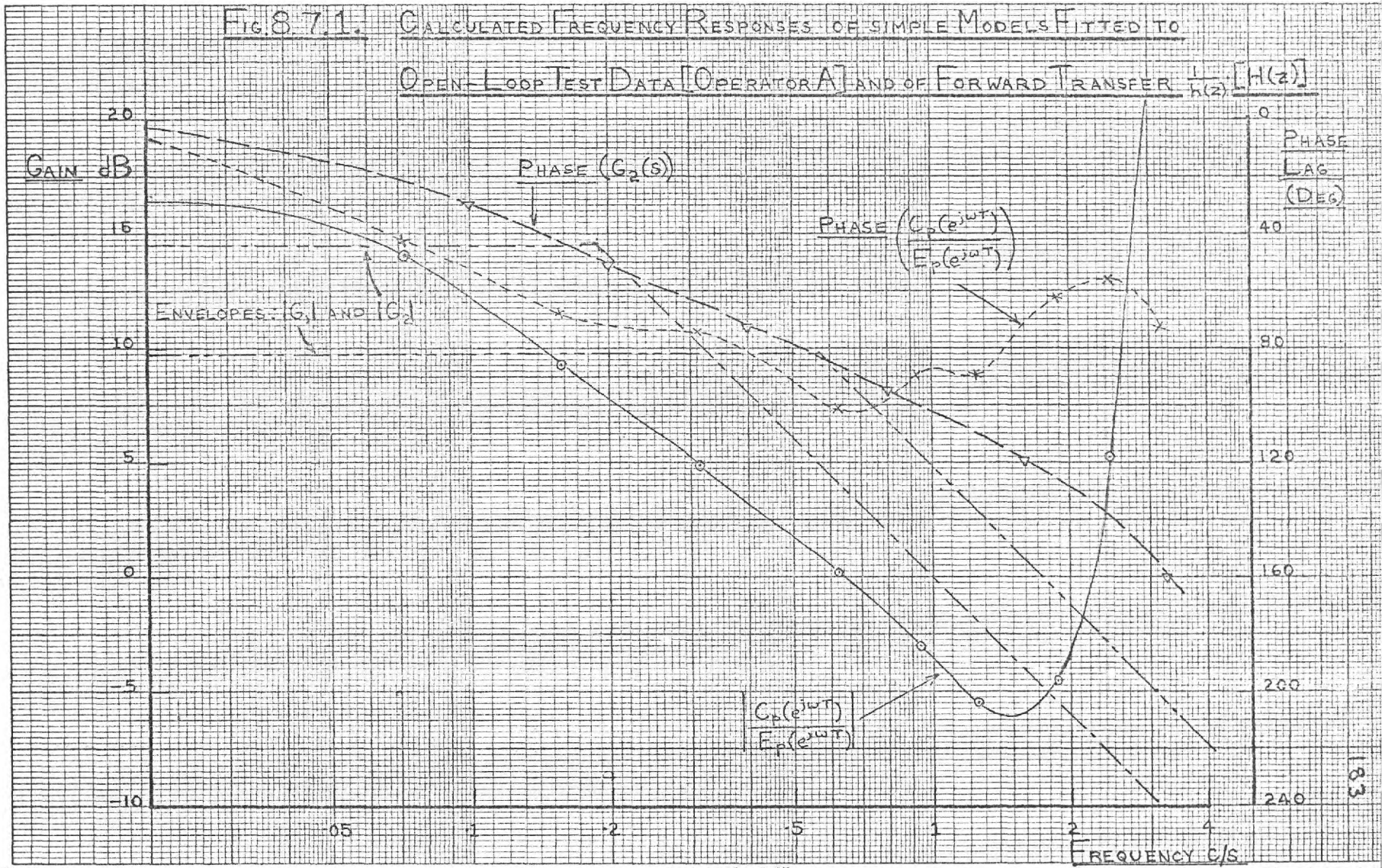
(b) In the higher frequency range (above ½ cps.) the SFM curves fit the EOL curves more closely,though exhibiting less gain,except in the case of operator F .

The SFM phase curves show good agreement to the AOL curves at frequencies below about 1½ cps.,and better agreement with the EOL curves at higher frequencies. The SFM's thus displayed a mixture of characteristics of both AOL and EOL transfers.

The general form of the SFM curves,and their relation to the AOL and EOL curves may be accounted for as follows :-

(1) A priori,one might expect the SFM curves to fit the AOL curves more closely. However,the AOL curve relates to the model's actual

FIG. 8.7.1. CALCULATED FREQUENCY RESPONSES OF SIMPLE MODELS FITTED TO OPEN-LOOP TEST DATA [OPERATOR A] AND OF FORWARD TRANSFER $\frac{1}{h(z)}$. $[H(z)]$



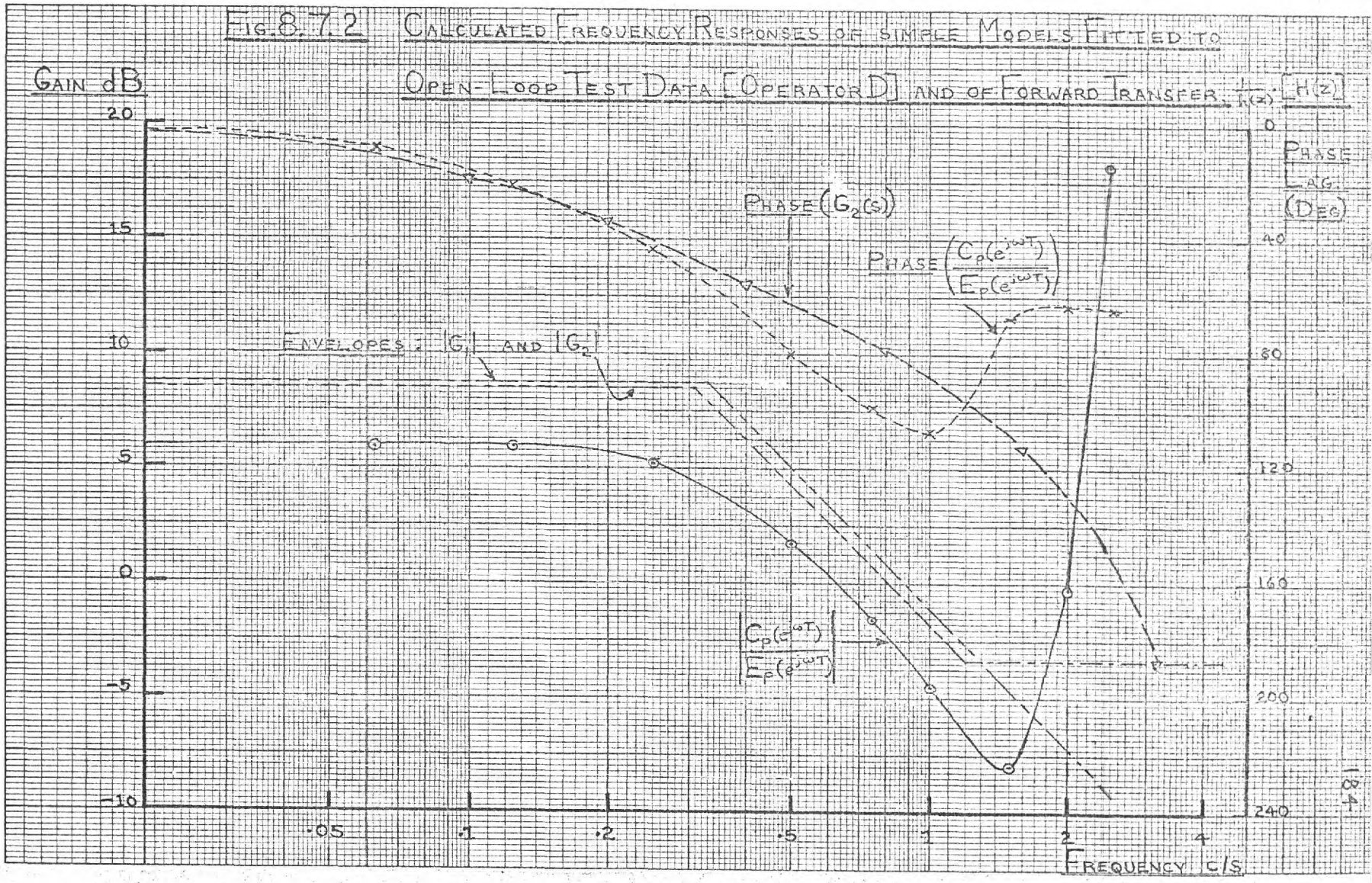


FIG. 8.7.2 CALCULATED FREQUENCY RESPONSES OF SIMPLE MODELS FITTED TO OPEN-LOOP TEST DATA [OPERATOR D] AND OF FORWARD TRANSFER, $\frac{1}{A(z)}$ [H(z)]
GAIN dB
20
15
10
5
0
-5
-10
PHASE LAG (DEG)
0
40
80
120
160
200
240
PHASE $(G_2(s))$
PHASE $\left(\frac{C_p(e^{j\omega T})}{E_p(e^{j\omega T})}\right)$
ENVELOPES: $|G_1|$ AND $|G_2|$
$\left|\frac{C_p(e^{j\omega T})}{E_p(e^{j\omega T})}\right|$
.05
.1
.2
.5
1
2
4
FREQUENCY C/S

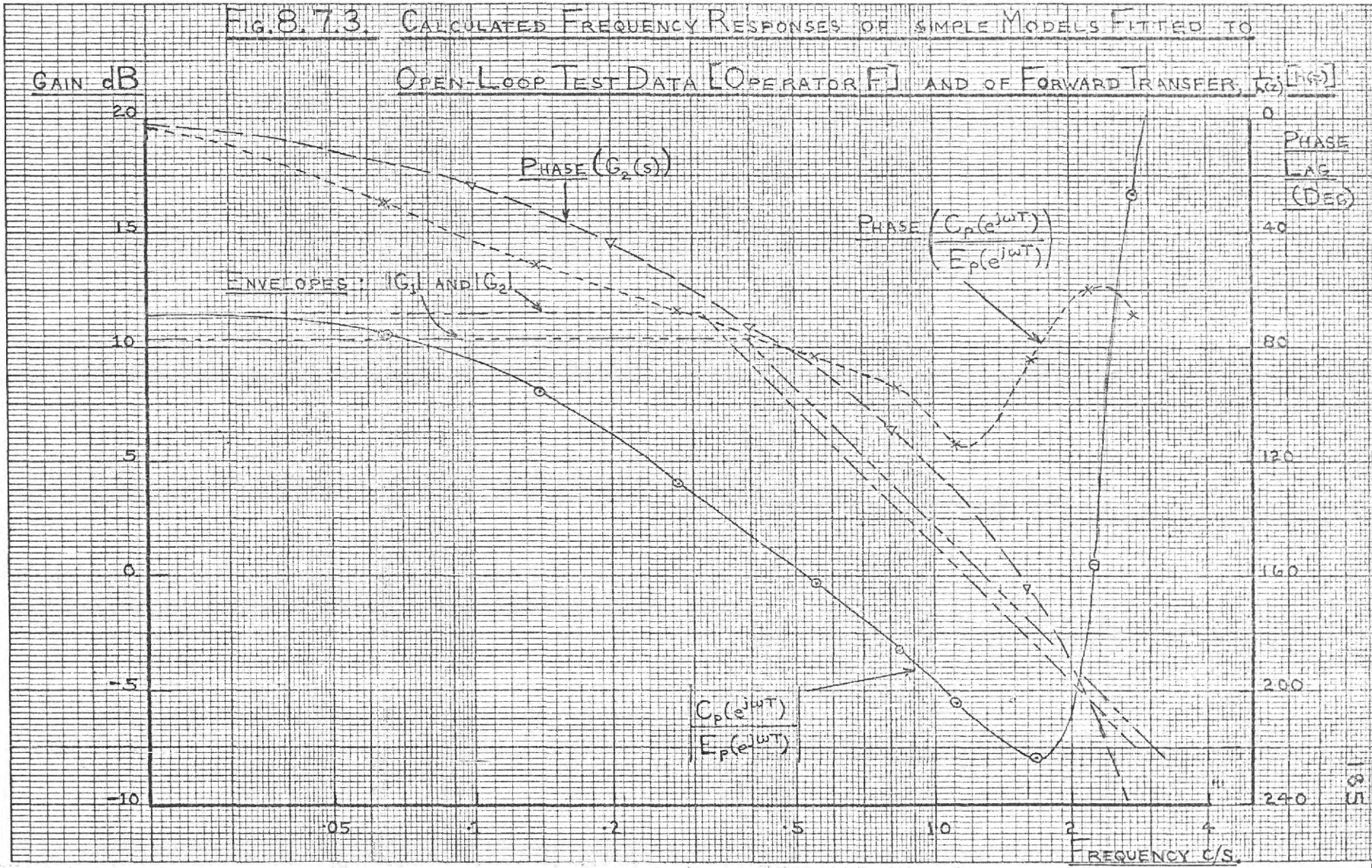


FIG. 8. 7.3 CALCULATED FREQUENCY RESPONSES OF SIMPLE MODELS FITTED TO OPEN-LOOP TEST DATA [OPERATOR F] AND OF FORWARD TRANSFER, $\frac{1}{\mu(z)}[\mu(z)]$
GAIN dB
20
15
10
5
0
-5
-10
PHASE $(G_2(s))$
ENVELOPES $|G_1|$ AND $|G_2|$
PHASE $\left(\frac{C_p(e^{j\omega T})}{E_p(e^{j\omega T})}\right)$
$\left|\frac{C_p(e^{j\omega T})}{E_p(e^{j\omega T})}\right|$
PHASE LAG (DEG)
0
40
80
120
160
200
240
$-\frac{\pi}{\omega}$
.05
.1
.2
.5
1.0
2
4
FREQUENCY C/S.

open loop response to a pure sinewave at the input,whereas the operator was presented with a signal containing ripple. Hence,one could alternatively argue that the signal perceived by the operator should give rise to a practical EOL transfer,with some bias due to the effects of ripple. Therefore it seems reasonable that the operator should exhibit aspects of both transfers.

(2) The SFM curves show rather too little phase shift and gain at low frequencies. This effect may be explained as being due to loss of low frequency information,as a result of the short available record length,and necessary data processing.

(3) As described in Section 8.4.,there was a tendency toward obtaining error-output covariance functions roughly similar to those of the model. This would tend to emphasise aspects of the EOL transfer.

(4) The model represented an idealisation,in that it incorporated instantaneous sampling of velocity. A finite time would be required for such sampling,with effects as indicated in Section 8.3.

(5) Despite the greatest care in conducting open loop tests,it was not possible to avoid causing some disturbance to the operator. Experimental results suggested that he tended to 'filter' large velocity errors due to perception of ripple, i.e. to compute over a number of observed velocity errors. This action would tend to emphasise EOL characteristics. A further point was that those runs in which operators displayed transfer characteristics most similar to AOL responses were more likely to be rejected as showing severe operator disturbance. In this connection,it may be noted that valid records could not be obtained from operator L. It seemed reasonable to hypothesise that those operators most likely to display a 'filtering' action were those who normally tended to filter their hand motion; i.e. operators D and F. These operators did,in fact, show least disturbance during open loop operation.

(6) The SFM's were of such form that they could not give more than a general representation of transfers in the frequency domain. They would thus tend to give an apparently closer fit to the EOL curves, which were of simpler form than the AOL curves.

(7) The model constitutes an idealised and simplified sequence of operations to represent a far more complex action,involving optimisation and self-organisation phenomena (see Chapter 9). These properties of the operator's response would have an effect on open loop responses which could not be predicted by reference to the model,since 'open looping' the operator represents a gross change in his external environment.

Thus,the form of the open loop responses,as revealed by the SFM's,may be accounted for fairly completely,and tends to substantiate

the basic reasoning lying behind the formulation of the original sampled data model. At the same time they served to reveal the complex nature of perception, and to emphasise the extent to which instantaneous position and velocity sampling represents an idealisation of visual sensing of error.

The sampled data model which was derived during the course of calculating the parameters of $G_2(s)$ was of interest, because it has been suggested that the operator samples his input at 10/sec.; this frequency is suggested to accord with that of the α-rhythm observed in the e.e.g. records of human subjects. This sampled model would give a step response starting at a finite, but small, velocity. Its frequency domain characteristics would be those of $G_2(s)$, multiplied by the frequency transfer of the first order hold - an even function of period 10 cps. The inadequacies of the model are that it predicts too low a delay in initiation of response to a step or ramp input (the data indicates a delay of 100 ms.) ; also, the predicted spread in reaction times to these discrete inputs is rather low.

8.8. Summary

A series of tests were carried out in which the operator was presented with a synthetic error signal, derived from the appropriate best-fit model, while at the same time his feedback loop was broken. Their purpose was to test the quality of the model, and at the same time derive error and output data which were not subject to corruption by a common remnant noise term, so that models might be fitted directly to the operators' forward transfers.

It was found that the maximum time over which useful tests could be conducted without undue operator disturbance was 6 secs. During tests operators displayed rather oscillatory hand motion and drift effects. It was concluded that these observed effects arose from lack of coincidence of operators' sampling instants with those of the model. There was subjective perception of apparent large velocity errors and of oscillatory hand motion, which agreed with the hypotheses embodied in the model structure. Also, Z-plane pole-zero plots of the model's actual open loop transfer showed features in agreement with observed effects. Results indicated that the operator might take up to 50 ms. to sample his position and velocity error, although a much shorter time was more probable.

A comparison was made of error-output covariance functions computed from records after removal of D.C. and trend. These showed a good deal of similarity between operator and model, but differences were also noticeable. These were explained in terms of the difference between the EOL and AOL transfers of the model, noting the fact that selection of data and computation gave a bias toward EOL characteristics.

It was not feasible to fit simple models to covariance data via the spectral domain. However, simple linear continuous models were

fitted by a regression technique. Values of the parameters of such models were in general accord with those reported in the literature. It was found that a lag-lead model gave a much better fit to the data, in terms of the coefficient of determination, than a simple lag. Fitting of models involving more than three parameters was not justified by the extent of the data.

The frequency response of the SFM's showed aspects EOL and AOL transfers . The somewhat closer fit to the EOL transfer was explained by several factors,including bias in analysis,disturbance of operators' normal mode of operation,differences of sampling,the simpler form of the model EOL transfer, and inadequacy of representation. It was concluded that the open loop test data were consistent with the hypotheses embodied in the formulation of the sampled data model with prediction,as described in Chapters 4 and 5, so long as the idealisations embodied in the hypothesis of instantaneous sampling and the comparative simplicity of the model were taken into account.

A simple sampled data model,with a sampling interval of .1 sec., was also fitted to the experimental covariance data. While it gave a good fit to the data at sampling instants,this model was not consistent with data concerning the operator's response to discrete inputs.

Chapter 9 : FURTHER ASPECTS OF HUMAN TRACKING BEHAVIOUR

9.1. Randomness of Sampling

When describing the formulation of the sampled data model (Chapter 5),it was explained that the hypothesis of a deterministic sampling action was an idealisation of the operator's characteristics. It was not feasible to directly determine the distribution of sampling intervals at that time. However,the excellent performance of the model,and the similarity of simultaneous recordings of operator and model output velocity gave added plausibility to the hypothesis of action embodied in the model. On the basis of this,the operator's sampling 'instants' are revealed by the quasi-triangular segments in his output velocity.

Histograms of the length of operators' sampling intervals, deduced from output velocity traces,are shown in Fig.9.1.1. These graphs can only be regarded as indicative of the true probability distributions,because,apart from sampling error,they are subject to error and bias resulting from the difficulty of measuring the correct lengths of output velocity triangles. It was difficult,at times,to determine where one triangle finished and the next started,and also to distinguish between their midpoints and endpoints.

The above histograms may be used to calculate the probability density function of the operator's delay to step inputs (i.e. the time before any response movement can be observed) ; details of theoretical analysis are outlined in Appendix X. The calculation is bsed on approximating the histograms of sampling intervals by a continuous probability density function,and making the following assumptions :-

(1) The mechanism of step response is as described in Section 4.10.

(2) Input steps occur randomly at intervals very much longer than the average operator sampling interval.

(3) Successive sampling interval lengths are effectively uncorrelated.

Assumption (3) seemed reasonable 'a priori',and was consistent with the form of the experimental data,though it cannot be considered to be justified by them.

The fairly gross approximation that the probability density function of operators' sampling intervals,p(I), was square,with a range from 100 to 200 ms.,led to a form of $\tau(L)$,(the probability density function of delay,L, to a step input) as shown in Fig.9.1.2. This diagram also shows the form of $\tau(L)$ calculated for a more refined parabolic approximation to p(I). As may be seen there is not much difference in overall form of the $\tau(L)$ curves,despite the different p(I) .

Bearing in mind the assumptions made,both the above curves show remarkable similarity to curves observed by Haynes (20),under

Fig. 9.1.1. Histograms of Length of Operators' Output Velocity Triangles

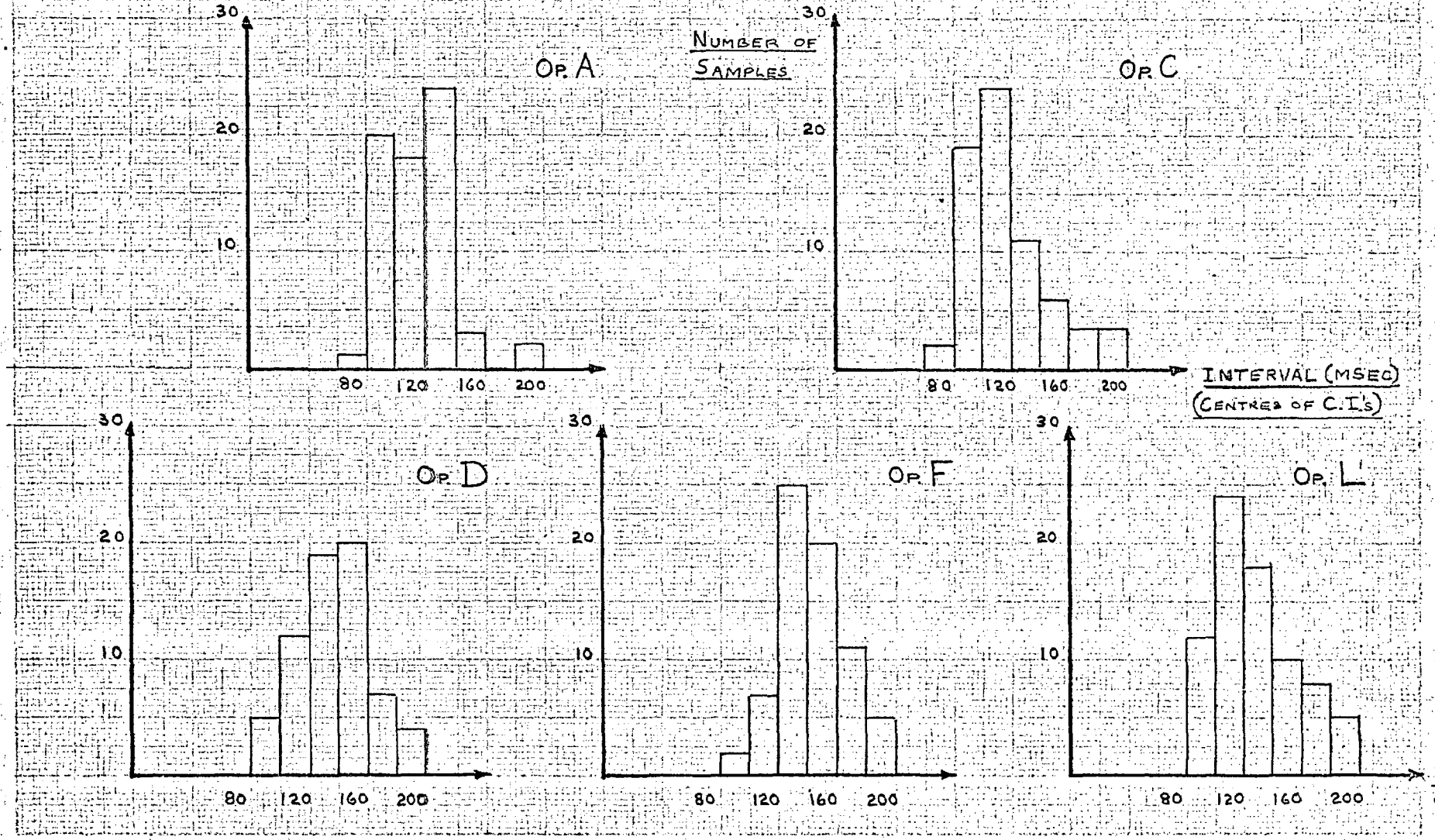
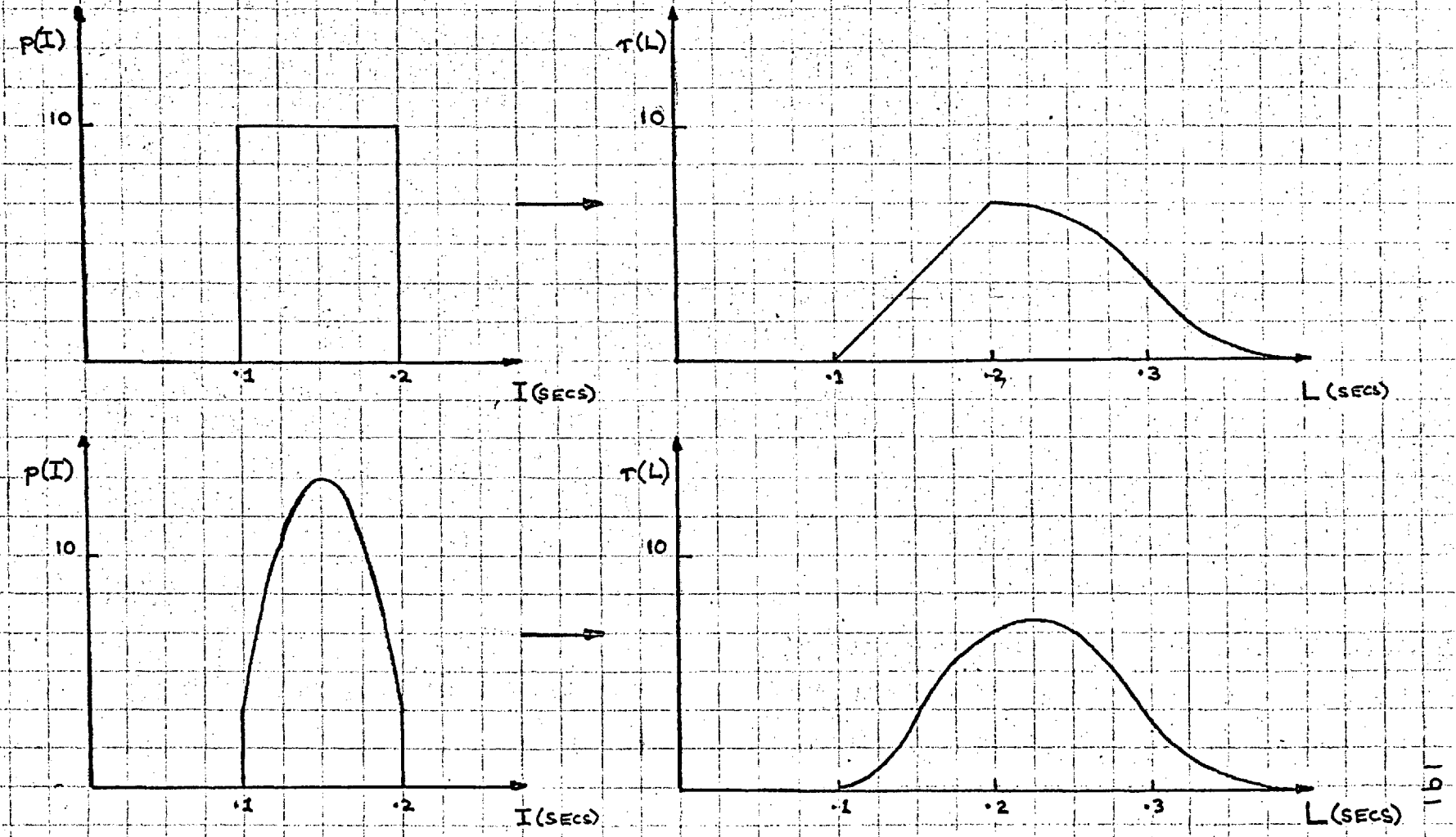


FIG. 9.1.2. SAMPLING INTERVAL (I) AND STEP RESPONSE DELAY (L) PROBABILITY DENSITY FUNCTIONS



experimental conditions in accord with assumption (2) above. Haynes fitted the curves by means of a Gaussian curve, but also commented that they could probably have been fitted quite well by a somewhat skewed normal, or even a lognormal, curve. These considerations lend a good deal of weight to the hypothesis of sampling action proposed in Chapter 4.

It has been proposed that the operator should tend to sample at a rate of about 10/sec. (23), to accord with the frequency of the α-rhythm observed in e.e.g. records of human subjects. The above data do not seem to accord with this hypothesis. However, they may be related to the α-rhythm frequency if the hypothesis is made that the operator takes an average time of 100 ms. to compute a programme of hand movement. This programme is then released for transmission via the spinal cord to the muscles. While the programme is en route, the brain is occupied in assessing input information. Thus, the time taken to 'read in' information would correspond approximately to that necessary for conduction of the programme of hand movement to the muscles (actually a muscle-spinal root feedback system). According to data presented by Walsh (22), the time lag between release of a programme and commencement of muscle activity would range from about 25 to 35 ms., made up of transmission time of approximately 20 to 30 ms. plus muscle refractory period of 2-5 ms. A few milliseconds should probably be added to the transmission time, to allow for neuron excitation time at synapses. It is difficult to arrive at a precise figure.

Predictions according to the above hypothesis agree not only with step response delay data, but also with experimental results described in previous Chapters. It predicts an average sampling interval ranging from about 125 to 145 ms., including a sampling operation occupying from about 20 to 45 ms. As explained earlier, and indicated by experimental results - especially in Chapter 8 - the hypothesis of instantaneous sampling, on which the model structure was based, represented an idealisation. The sampling intervals of best-fit models ranged from 133 to 166 ms. ; the latter figure was probably biassed toward the high side as a result of operator D 's 'filtered' hand motion. Also, sampling interval was not treated as a readily variable parameter, so that some mismatch might reasonably be expected. It was therefore concluded that the weight of evidence tended to support the form of sampling hypothesis advanced in the above paragraph.

9.2. Consideration of Model Structure in Relation to Compensatory and Pursuit Tracking

During the present course of studies attention was directed towards developing a model for compensatory tracking. It is interesting to note that the form of this model has implications which accord well with the results of other workers in regard to the relation between compensatory and pursuit tracking.

Elkind (10) found that, while operators exhibited similar behaviour in both pursuit and compensatory tracking situations, they tracked a given input better in the pursuit configuration. Tracking was better mainly because there was less operator remnant, and the operator's prediction was better. These results can be explained very satisfactorily in terms of the model structure described in Chapter 5. It will be recalled that, in considering how the operator might achieve prediction, it was postulated that he attempted to reconstruct the loop input function. Owing to the nature of compensatory tracking, and the effects of operator noise, the extent of this prediction was somewhat restricted. In the pursuit configuration the operator would have direct knowledge of the input function. Input reconstruction could therefore be replaced by direct observation, thus leading to more accurate prediction. Also, the effects of operator noise on prediction would be less severe, so that the amount of this noise recirculating around the control loop would be appreciably reduced.

Krendel & McRuer have proposed a "Successive Organisation of Perception" (SOP) model to describe the process of learning and adaptation. According to this, the operator tracking under conditions of compensatory display is eventually able to reorganise his functioning so as to approach the characteristics of a pursuit configuration. With a perfectly predictable input, he may finally achieve a 'monitored open loop' type of response.

The model structure reflects the above organisation from compensatory toward pursuit tracking, in that it was derived by postulating that the operator attempts to reconstruct the loop input. The model was derived by considering an input which was only statistically predictable. Therefore, in the form presented, it does not possess the capability of representing monitored open loop behaviour. However, a modification to include more past samples for formulating prediction could give it the capacity to generate this type of response.

The structure of the model, modified as above, and the random sampling hypothesis, are capable of explaining observed characteristics of precognitive tracking, and, in particular, of sinewave tracking. Thus, one would expect random variations in overall lead and lag, accompanied by occasional loss of synchronism caused by the occurrence of relatively improbable combinations of sampling intervals. Resynchronisation would then take some time, corresponding to the number of past samples used for prediction. Just such effects are observed in experimental records.

9.3. Relation of the Form of the Model's Response to Remnant Data

The form of the operator's remnant spectrum was not obtained in the present study, but excellent results concerning this are presented by Elkind (10). He found that compensatory tracking noise possessed a fairly flat spectrum, with a tendency toward a peak in the range 1 to 2 cps., while at higher frequencies there was moderately rapid attenuation.

To investigate the remnant in terms of the noise which the model might be expected to generate, it is theoretically necessary to specify the spectral density matrices of the operator's estimation and execution noises ; this is unfortunately not possible. However, one may consider the model's response to disturbances in the error channel, $\underline{E}_d$, and in output motion, $\underline{C}_d$. Considering output response to these disturbances, one may derive :-

$$\underline{C}(z) = \frac{1}{q(z)} . \left[Q(z,1) \right] . \underline{E}_d(z) \qquad 9.3.1.$$

and

$$\underline{C}(z) = \frac{1}{q(z)} . \left[q(z) . I - Q(z,1) \right] . \underline{C}_d(z) \qquad 9.3.2.$$

These equations show that the model's response to disturbance in the error and in the output channel is given, respectively, by the output and by the error response to loop input. The total model output response is therefore a combination of these two contributory responses. Thus, the form of the model's response to individual disturbances is an oscillation at around 2 cps., damping out fairly rapidly (as indicated by pole-zero diagrams for overall transfer). Random sampling action would tend to flatten the spectrum, as indicated by the interval histograms shown in Section 9.1. If the hypothesis of fairly independent successive disturbances at sampling instants is satisfied, one might reasonably expect an output spectrum very similar to that observed by Elkind.

Unfortunately, Elkind's measurements did not extend to frequencies high enough to show ripple effects. However, it is dubious whether these would have been shown, because their power would have been very small, and spread over a considerable range of frequencies by randomness of sampling. Randomness of sampling may also be partly responsible for the causation of output disturbances, to the extent that the operator cannot predict the length of the next sampling interval. The operator's response, programmed over this interval, could then only be formulated in terms of the response required over the predicted length of interval, so that some error would be incurred.

9.4. Relation of Operators' Responses to Impulse Response Optimisation

The characteristics of the continuous random input to the operator were decided (as described in Secion 3.2.) by considering Wiener filtering in relation to the desire to allow the operator to make use of his ability to predict. The suitability of this approach may be seen by considering the response of trained operators in relation to the response of an optimum linear filter.

It is apposite to note the possible effect of instructions to the operator. In the present case,the operator was motivated to attempt to reduce error to zero at all times. It was apparent that the operator's subjective error criterion would be the major factor shaping his formulation of an optimum form of response. An hypothesis concerning the basis of this subjective criterion was constructed from the following considerations. Now,the loop input was a zero-mean Gaussian signal,and experiment showed that the operator's error signal also possessed a close approximation to a zero-mean Gaussain probability density function. The most important attribute of such a signal is the variance,and it seemed plausible that the operator effectively attempted to minimise this ; minimisation of most other reasonable criteria of error would simultaneously result in minimisation of variance. It was noted that in a system with a Gaussian input, Gaussian disturbance or noise,and a linear controlled process,and where there is no particular reason to adopt a particular form of error criterion,the optimum controller is a linear controller (31). All these conditions were satisfied in the present case,and it therefore seemed reasonable to propose that the mean square error criterion was a realistic representation of that actually adopted by the operator. If alternative instructions had been given to the operator,e.g. "Do not allow the error to exceed a threshold level of 2 cm.", it would be reasonable to expect the operator to have adopted a nonlinear form of response,since in this case the optimum linear controller would not have been a linear optimum controller....

The form of the optimum predictor according to the mean square error criterion,is that of a continuous Wiener filter,as described in Appendix VII. The form of the overall optimum transfer is given by the combination of an optimum predictor with a pure delay corresponding to the prediction time. It is most convenient to compare the spectral characteristics of the operators' transfers,as revealed by the best-fit models (Figs.7.6.1-10) with those of the optimum transfer,shown in Fig.9.4.1. These optimum transfer characteristics have been calculated for a pure delay of .26 sec., which represents an idealisation of the operator's step response. It may be seen that the fit of both overall loop transfer and of equivalent forward transfer characteristics is quite impressive, especially when the following points are taken into account :-

(1) The fixed element in the operator's response (brain processing delay plus neuromuscular lag) has been idealised,and represented by a pure delay in calculating the corresponding optimum filter

16. . . . REQUENCY ESPONSES OF PTIMUM IENE REDICTOR PLUS RE LAY

FOR CONTINUOUS RANDOM SIGNAL INPUT (TRIPLE BREAK AT .61 cps)

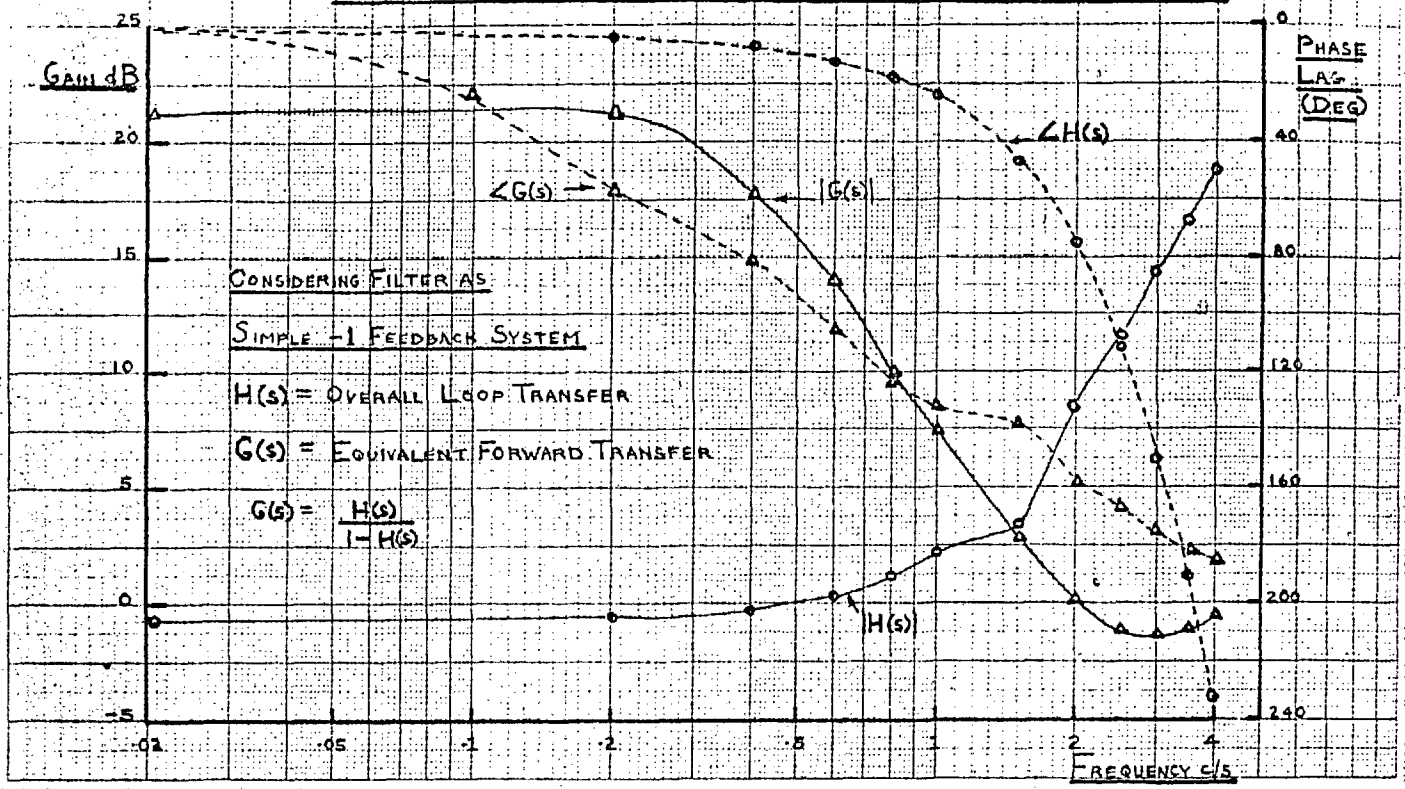
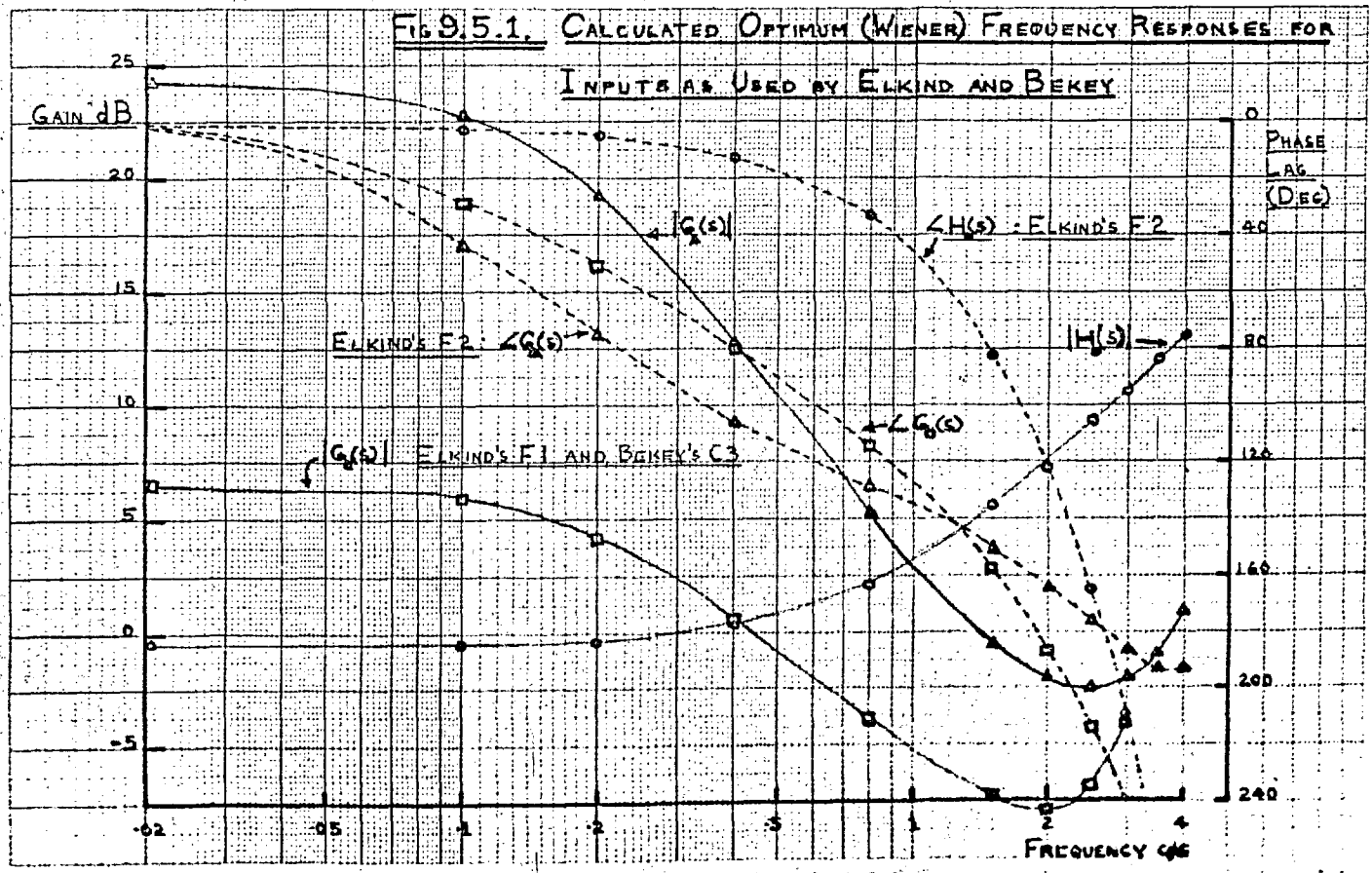


Fig 9.5.1. CALCULATED OPTIMUM (WIENER) FREQUENCY RESPONSES FOR INPUTS AS USED BY ELKIND AND BEKEY



response

(2) According to the model, the operator could not directly sense acceleration,but had to resort to a differencing of velocity. The optimum continuous filter utilised direct sensing of acceleration.

(3) Calculation shows that the optimum continuous filter has an error variance equal to 5% of the input variance. Operators displayed an error variance 2 to 3 times as large. The greater part of this error varaince can be accounted for by the operators' inability to achieve the optimum linear overall loop transfer. One would not expect the operator's overall loop transfer to match the optimum continuous filter exactly. Due to inevitable estimation noise,the operator cannot measure his state exactly ; also,he cannot operate on his measured input without statistical errors of response. These facts indicate that the actual optimum transfer would differ from that of the corresponding Wiener filter. One of the weknesses of impulse response optimisation,as used to derive the form of the optimum continuous filter,is that it does not suggest a suitable structure for a corresponding feedback control system in which the effects of errors arising within the system are minimised. The method gives an essentially open loop formulation,and this often leads to an overall loop transfer with unsatisfactory control loop characteristics.

(4) The effective pure delays in operators' responses differ among themselves,and from the average delay assumed in formulating the optimum linear filter. These differences would be expected to lead to differing relative weightings of position,velocity,and acceleration information.

(5)No cost has been associated with control action in deriving the optimum transfer.

As a result of the above arguments it was concluded that the operator does approach quite closely to the optimum transfer relevant to his own estimation and execution noise characteristics. This conclusion was consistent with the fact that differences were observed between operators ,although they exhibited a similar basic pattern of behaviour. These differences would be accentuated by differences in operators' noise characteristics. A larger amount of operator noise would not only contribute directly to the error variance,but would also cause the transfer adopted by the operator to diverge further from the absolute optimum transfer of which he might be capable in the absence of such noise....

9.5. Relation of Experimental Results to the Data of Other Workers in the Light of Impulse Response Optimisation

As mentioned in Section 7.6.,the calculated frequency domain characteristics of the best-fit models correspond well,as regards general form,with similar characteristics measured by Elkind (10) and by Bekey (2). It is possible to explain such divergences as may be observed in terms of differing statistical structures of the input functions used by these workers. It was concluded,as described in the previous Section,that the operator adopts a quasi-optimum transfer. Now the optimum transfers calculated for Elkind's and Bekey's inputs are shown in Fig.9.5.1. Comparison of these transfers with the measured transfers shown in Figs.7.6.12 - 14.,lends further weight to this conclusion. It also shows that the differences between characteristics can be explained very satisfactorily as being largely due to differing input statistics. For reasons given in Section 9.4.,one would expect differences between the measured transfers and the optimum transfer of the corresponding Wiener filter.

One point worthy of note is that both calculated and measured open loop transfers show an upturn in gain in the region of 2 cps. This indicates that the single lag plus delay model would be considerably improved by the addition of a high frequency lead term. The results described in Chapter 8 provide numerical substantiation for this statement. This is also in accord with the addition of further lead and lag terms as proposed by other workers (11,12). There is little to be gained by considering models of equivalent open loop transfer possessing a large number of lags and leads, because little further improvement could be made to the fit of experimental data,and in any case the linear continuous model is very much an idealised representation of operator characteristics.

According to Elkind,the quality of operators' prediction showed considerable deterioration with increasing input bandwidth, even under pursuit tracking conditions. For inputs with an infinite rate high frequency cut-off,prediction appeared to be very poor for bandwidths of 1.6 cps. or above. Although Elkind may have used too low a figure for overall delay in considering his results,it is significant that they do tend to confirm that the operator cannot sense higher derivatives directly,since these would assume greater importance as input bandwidth increased.

The above results also tend to support the sampling hypothesis,as advanced in Chapter 4,because one would expect a rapid deterioration in performance as the input cut-off frequency approached the average half sampling frequency exhibited by the operator. Further evidence is provided by the fact that,in the compensatory configuration,operators were prepared to track an input with a 2.4 cps bandwidth,but found it impossibleqto track one with a bandwidth of 4 cps. This lends support to the half sampling frequencies around 3½ cps. employed in the model. This value of sampling frequency would also explain why operators were prepared to attempt pursuit tracking of an input of 4 cps. bandwidth : the aliasing

of the input function would not be so troublesome as that of the error function, because the latter possesses higher relative power density at the high frequency end of the spectrum under all conditions except those where tracking is completely ineffectual.

The sampled model, as described in Chapter 5, embodied direct sensing of velocity and position ; acceleration could only be sensed through a differencing action. The fact that the model so proposed gave a good fit to the operators' transfers tended to confirm that they were not capable of directly sensing higher derivatives. Further evidence for this is provided by the conclusion reached by Elkind (10), that, for a given break frequency, it made little difference to the operators' transfers whether an input with a triple break or one with an infinite rate of cut-off was used. The optimum Wiener transfer for a triple break would involve only terms up to the second derivative, whereas that for an infinite rate of cut-off would theoretically require the measurement of an infinite number of derivatives, and this input would be perfectly predictable. The operator is apparently not able to use these higher derivative terms.

9.6. Consideration of the Operator's Optimum Seeking Behaviour

The conclusions of Sections 9.4. and 9.5. were that, for the type of task studied experimentally, the operators were capable of achieving characteristics near to those of the corresponding linear 'least squares' filter which might be designed from a knowledge of relevant input and noise statistics. The purpose of the present section is to consider further the operators' criteria for specifying, and methods of achieving, the best possible performance.

The definition of the overall goal of a tracking task is inherent in the instructions given to the operator. In order to achieve this goal as nearly as possible, it is necessary for the operator to interpret his instructions so as to arrive at a subjective criterion of performance. It is not generally possible to postulate 'a priori' the exact form of this performance criterion. However, inferences can be made as to the effective form of criterion adopted for a particular tracking task, at a particular stage of training, by examining data and noting operators' subjective impressions. As implied above, one would expect the form of the performance criterion to change during training, and that such change would be most rapid during initial task familiarisation.

For compensatory tasks, the chief requirement is usually the minimisation of some function of error. In particular tasks there may be further requirements ; e.g. in an aircraft it is usually required to constrain acceleration to lie within certain limits - in space vehicle control it may be required to minimise fuel expenditure. In an exacting task the operator may also need to organise response so

as to minimise some function of muscular control effort,in order to avoid undue fatigue. The relative weighting attached to this latter factor may be expected to depend to a considerable extent on the 'feel' of the control stick or manual output sensor.

As shown in Section 9.4.,for the simple compensatory task studied experimentally,the operator's error criterion was effectively equivalent to the mean square error criterion. Elkind (10) also found that such a criterion could be related to compensatory and pursuit data for a range of Gaussian inputs. Both results relate to tracking with a simple controlled element (unity transfer) and free-moving manual control,under instructions to keep the error as small as possible at all times. It should be remembered that the above results are for trained operators; there was subjective evidence suggesting a change of criterion during training. Most operators reported that,on first attempting the task,they tended to concentrate on the largest errors,and practically ignored errors below about 1½ cms.

It is significant that the above results were obtained in connection with free-moving manual controls,so that weighting of the control effort would probably be relatively small. However,the nature of this weighting can be inferred from the form of the operator's response,as outlined in Section 4.10. A velocity-triangle form of response would result from minimising the instantaneous absolute magnitude of output acceleration during a sampling interval,while simultaneously satisfying velocity and position constraints at sampling instants. The response could also arise as a result of some other,nonlinear weighting of acceleration.

There is physiological evidence in support of a weighting of acceleration similar to that described above. Minimising a function of acceleration corresponds to minimising a function of muscular force. Now muscular force is developed by recruiting a suitable number of muscle fibres (Section 2.3.) ; the greater the force,the larger the number of fibres required and the greater the expenditure of energy. Thus,the total energy dissipated in muscular tissue whilst achieving a given movement is quite strongly dependent on the maximum force required. It is not so dependent on the mechanical work done by the muscle,and in the limit,there is no such dependence,as when fatigue occurs through applying a constant force to an immovable object.

The differences observed in the forms of individual operator's output motion support the hypothesis that they attempted to minimise a function of both manual control effort (acceleration) and error. Operators who weighted error most heavily,and therefore returned the highest values of performance index,also displayed velocity triangle phenomena most clearly. The poorest performance was returned by operator D,who also displayed a manual filtering action most prominently. This filtering action corresponds to attaching such a high weighting to output acceleration that the response at sampling instants is not effectively dominated by requirements of error minimisation. The above hypothesis also accounts for the subjective impression that,when 'trying really hard',there was a significant increase in muscular effort. Some operators also reported that their

arm was physically tired after completing 2 or 3 runs during which they were highly motivated to improve their performance.

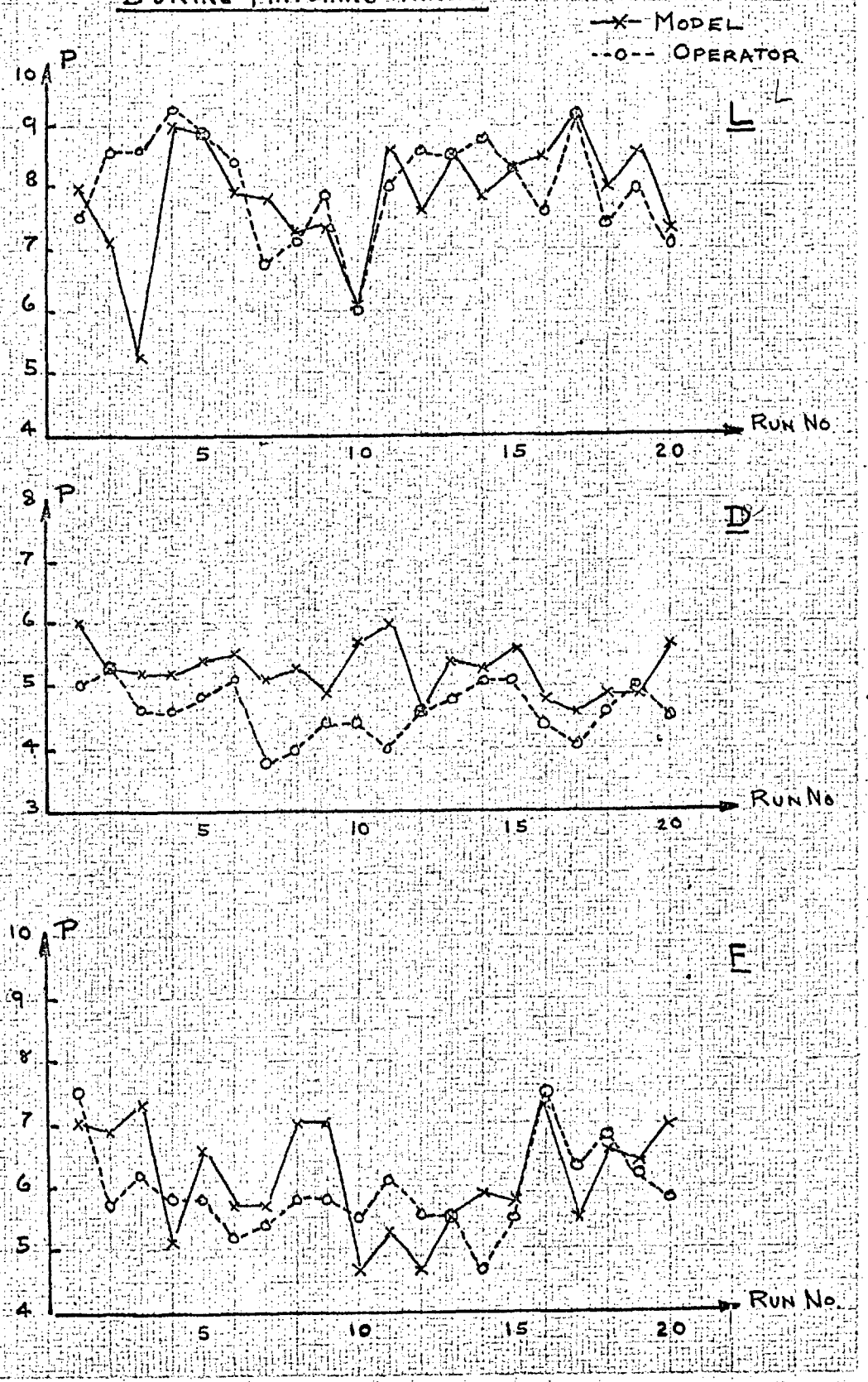
It has been suggested (2) that the operator's adaptation to input or error statistics might take the form of a direct variation in average sampling interval. However,it seems most probable that the operator's sampling action results from the inherent structure of the cerebrum and neuromuscular system,and on the basis of this there seems little justification for the above hypothesis,as put forward. One may postulate a modified form of the hypothesis - that under certain conditions the operator may not perceive,compute,or act upon a sample. Thus,average sampling interval would be lengthened by an occasional 'missed' sample. This would imply a modification not only of average sampling rate,but also of the form of the probability density function of sampling interval length,so that the effects of such action would be very complex. Experimental results obtained during the present study (Section 4.7.) indicated that there was,at most,a marginal effect due to dependence of sampling interval on input velocity.

The effects of variations in sampling interval have been considered (2) in terms of a variable sampler feeding a fixed plant of simple form. It seems more likely,however,that the operator's action is to programme a required hand motion over the succeeding sampling interval,whatever its length. In this case the plant parameters must be considered to vary in sympathy with the sampling interval. The results of such analysis must,in any case,be viewed with suspicion,since such simple models represent gross approximations to a complex structure.

One may also consider the operator's optimising action from the point of view of varying the parameters of a sampled data model, while average sampling frequency is kept constant. If no further assumption is made,then theoretical analysis presents extreme difficulty,because parameters would be related to input and error functions in a complex,and unknown,way. However,one could instrument the problem on an analogue computer.

The effect of continuous variations in parameters may be approximated by considering the effect of parameter variations at discrete intervals of time. Such variations were effectively made during the course of model matching trials. These parameter variations cannot be considered as being strictly representative of the operator's attempts to reach an optimum,since their purpose was to achieve a good operator-model match. However,they do represent an optimising action in the sense of minimising the variance of the algebraic difference between operator and model errors. It was found that the run-to-run time sequence of model performance was remarkably similar to that observed in the case of operators (see Fig.9.6.1.). Only a small part of these fluctuations could be ascribed to factors common to model and operator ; i.e. fluctuations in input statistics and the tendency to track operator performance. Also,it proved quite possible to follow the tracking performance of operator D,after initial task familiarisation (see Section 5.7.) without making gross changes in model parameters. Therefore it was inferred that the observed fluctuations in operator performance largely arose from small

FIG. 9.6.1. RUN-TO-RUN PERFORMANCE OF MODEL DURING MATCHING TRIALS



perturbations in response parameters associated with optimum seeking action, where the criterion of 'optimum' was itself subject to stochastic variation due to factors such as motivation, fatigue, etc.

The considerations outlined in the above paragraphs are consistent with the hypothesis that the part of the basic structure of the operator's response which is not innate, becomes established quite early on in training, except, perhaps, for very difficult tasks. Subsequent improvements in performance are then obtained by changing the parameters associated with this structure, or occasionally by its elaboration. Thus, the process of formulation of the model structure, described in Chapter 4, may be considered to be a good analogy of the operator's learning process.

The form of the operators' learning curves (Section 3.7.) agree very satisfactorily with the above hypothesis. The initial period of task familiarisation corresponds to the establishment of the basic response structure. Subsequent parameter variations then lead to a fluctuating, but generally improving performance. The phenomenon of learning plateaux corresponds to the discrete nature of those points in time at which elaboration of response structure occurs. Thus, the operator's method of achieving the best possible tracking response (according to his own error and control effort criteria) may be thought of as analogous to a multi-dimensional hill-climbing, where both parameters and structure of the tracking system are subject to change. The 'hill' itself must be considered to be subject to stochastic changes due to fluctuations in input statistics, controlled element dynamics, and the operator's own criterion of performance.

The characteristics described above are just those associated with a self-organising system. This might have been expected, since the human being represents the most sophisticated self-organising system known. However, the nature of his self-organising capabilities have usually been investigated in relation to quite complex tasks. It is interesting that they were also in evidence even when performing the relatively simple task forming the subject of the present experimental study. The implications of this are that, when considering the human operator's adaptive behaviour, full account should be taken of his capacity for self-organisation. Simple models of adaptation are unlikely to be representative of more than a small range of tasks.

9.7. Summary

Data concerning the operator's random sampling action was gathered on the basis of the correctness of the hypothesis of action embodied in the structure of the model. The form of histograms of the observed sampling intervals implied a probability density function of delays to step inputs consistent not only with data observed in experiments described previously, but also with independent experimental results obtained by Haynes (20). It was therefore concluded that the

form of the random sampling hypothesis advanced during the course of experimental investigation was basically correct. However, it was recognised that instantaneous sampling was an idealisation ; the hypothesis of a sampling time occupying 20 to 45 ms. was shown to be consistent with physiological data.

Published data concerning operator remnant were shown to be consistent with the form of the model's response, assuming reasonably independent sequences of disturbances at both input and output. At least part of the output disturbance could be related to randomness of sampling.

The implications of the structure of the model were in substantial accord with the 'Successive Organisation of Perception' model of learning and the relation between compensatory, precognitive, and pursuit tracking, as proposed by Krendel & McRuer (13).

Comparison of the model's transfer with that of a corresponding optimum Wiener predictor plus pure delay, indicated that when allowance was made for effects of remnant recirculation, the operator did approach quite closely to an optimum performance in terms of the 'least square error' criterion. Differences between experimental and published data were satisfactorily accounted for on this basis. The weight of evidence supported the hypothesis that the operator could not directly sense acceleration or higher derivatives in a purely visual error display configuration. Also, published data concerning the effects of input bandwidth supported the sampling frequency proposed in the model.

Preliminary considerations indicated that the operator's criterion of performance was generally quite involved, and represented an interpretation of instructions. It was concluded that, in the task investigated experimentally, the criterion was composite, with both error and muscular control effort contributing to 'cost', and was subject to change during training. The most likely hypothesis of the relation between average sampling interval and external task requirements, was that the operator 'missed' samples according to his own criteria.

Consideration of experimental data showed that the operator's optimising behaviour was manifestly that of a self-organising system, even for the relatively simple task investigated. This implied that the generality of simple mathematical models of adaptation was likely to be somewhat restricted.

Chapter 10 : CONCLUSIONS AND RECOMMENDATIONS

10.1. Summary of Results and Conclusions Drawn from Investigations

A review of literature indicated that much work had been devoted to describing human tracking characteristics in terms of linear continuous models. These were capable of generating outputs showing cross-correlations with operator outputs as high as .9 ,but generally exhibited poorer error cross-correlations,of the order of .5 or less. Also,the forms of the outputs of such models were much smoother than those of operators' outputs. Published data showed that for inputs of less than 1 cps. bandwidth,as much as 99½% of the operators output power was linearly coherent with the input signal. Sampled data models had been formulated to represent operator characteristics both in the time and frequency domains. These models had succeeded in giving a better simulation of the fine structure of the operator's output,but had not achieved cross-correlations significantly better than those shown by linear continuous models.

Preliminary investigation led to the conclusion that the most fruitful area of investigation was that of a compensatory task. The task studied involved tracking in one dimension with visual error display and manual output motion. From 'a priori' reasoning it was concluded that the most suitable input was a random signal generated by passing white noise through a filter consisting of three cascaded exponential lags; the most suitable break frequency was determined experimentally. This input was a zero-mean Gaussian signal,and results showed that other loop variates were also approximately zero-mean Gaussian signals. Hence a performance criterion of the form (input variance)/(error variance) was selected. According to this criterion,the effect of change of hand motion-to-display gain,and the introduction of mild nonlinearity,had little effect on operator performance. It was concluded that the study of nonlinearity in the control loop was beyond the scope of the present study,and that the appropriate feedback was provided by a simple unity transfer.

Results showed that operators exhibited learning phenomena, and it was concluded that at least 30 runs were required to obtain reasonably steady run-to-run performance. Preliminary investigation of operators' responses to discontinuous inputs showed that,to avoid the effects of precognitive response,randomness of both time and amplitude structure was required.

In order to investigate the validity of the continuous and of the various sampling hypotheses,a series of experiments were conducted utilising a sampled display. The results were that operators showed deterioration of performance even at sampling rates as high as 20/sec.,and could 'synchronise' to the display for sampling intervals in the range 100 to 200 ms.,when it was noted that their output motion consisted of approximately 'triangular' segments. It was concluded that the operator sampled both position and velocity at sampling instants, with an average sampling interval of approximately 150 ms.,and effected

output motion by programmes of 'velocity-triangle' form, occupying a similar time.

The quantitative results of tests conducted with step and ramp inputs showed no significant relations between any of the variables measured (reaction time,magnitude of input,overshoot,time from last input discontinuity). There was a tentative relation between magnitude of input and reaction time in the case of ramps,which,it was concluded, might possibly be due to slightly faster sampling of steeply sloping inputs,plus threshold effects. It was concluded that reaction time data were consistent with a random sampling hypothesis,with an average interval of about 150 ms.

Study of qualitative features of step responses showed that they could be largely accounted for by the mechanism proposed by Wilde (18),but this mechanism could not account for correction of velocity errors. Study of ramp responses showed that the operator could follow a constant input velocity by a relatively smooth hand motion. It was concluded that the operator sampled both position and velocity and attempted to match his hand motion to these by a combination of a velocity-triangle for position correction,and a velocity ramp for velocity correction,each running over the second complete sampling interval following the input discontinuity giving rise to the response. A linear analogue model was constructed to simulate the proposed mechanism of ramp response,and was found to give a good match to operators' responses,despite the idealisations of instantaneous sampling and sharp velocity triangle responses embodied in the model structure. However,when this model was compared directly with operators performing a continuous tracking task,rather disappointing results were obtained,and the model showed too much lag. It was concluded that the basic mechanism of ramp response must be extended to allow for prediction.

The model structure was extended according to the hypothesis that the operator employed linear prediction,but extensions were formulated so as to minimise the number of extra model parameters. Two subsidiary feedback loops were added to the structure,and this modified model was then found to be capable of very accurate simulation of operators' tracking,achieving cross-correlation between errors of the order of .9 . The analysis of this extended model was facilitated by the fact that it was both idealised and linear. The appropriate form of analysis was that of the matrix Z-transformation. The analysis showed that the form of the model's transfer was such that care was needed in defining the model's forward transfer from position error to position output: equivalent (closed loop) and actual open loop transfers were not the same. Analysis also indicated that the model's frequency response could be defined in terms of sampled functions,with good accuracy below 3 cps.

Direct comparison of the model with operators showed that, with suitable parameter settings,the model gave a very good simulation of both their ramp and continuous tracking responses. Consideration of Z-plane pole-zero plots of model transfers showed that they represented very stable transfers,in accord with observed operator characteristics.

Comparison of the shapes of covariance functions emphasised the excellent degree of fit attained by the model. Fit could be improved slightly by small changes in model parameters, but no improvement was effected by adding extra feedbacks for velocity prediction. It was concluded that extension of the model structure would probably not improve fit for the type of task being studied. Calculated spectral characteristics of best fit models showed features similar to those obtained by other workers ; observed discrepancies were due to differing input functions used by these workers. It was concluded that the above results offered a large measure of substantiation to the hypotheses embodied in the model structure, especially when it was recognised that these were represented in an idealised form.

The above hypotheses were further tested by conducting a series of tests involving temporary interruption of the operator's feedback, and substitution of a synthetic error signal, derived from the model, adjusted so as to fit the particular operator. The chief features of operator responses were oscillatory hand motion and drift. Subjectively, the operator perceived unusually large velocity errors. These factors eventually caused disturbance to the operator, so that the duration of tests had to be limited to 6 secs. It was concluded that these features were in accord with predictions made by considering the structure of the model and Z-plane pole-zero plots of its actual open loop transfer, but that the hypothesis of instantaneous sampling represented an idealisation of a sampling action in which the 'read-in' of a sample might occupy as long as 50 ms.

Comparison of error-output covariance functions showed similarity between operator and model, but also noticeable differences. It was concluded that these differences could be explained in terms of the difference between equivalent and actual open loop transfers of the model. Simple linear continuous models, of a form similar to that proposed in the literature, were fitted to averaged error autocovariance and error-output covariance functions by a regression technique. The best fit was obtained in respect of a model possessing a lag-lead combined with a pure delay and a simple gain. Fitted parameter values were of the same order of magnitude as those described by other workers. In the frequency domain, these simple fitted models showed characteristics which were a combination of equivalent and actual open loop model transfers. It was concluded that the form of the data processing gave a bias toward equivalent open loop transfer, and that the results were consistent with the form of the model, bearing in mind the idealisations made in formulating its structure.

A simple sampled model, with sampling interval of 100 ms., and a first order hold for data reconstruction, was also fitted to the covariance data. This gave a good fit to the data, but it was concluded that features of this model's response to discontinuous inputs were not consistent with experimental data.

It was concluded that the experimental results described above were consistent with the structure of the model as proposed, with the proviso that sampling should occupy a finite time. The

operator's sampling action was investigated further by obtaining data concerning the lengths of sampling intervals from output records,on the assumption that the hypothesis of action embodied in the model was correct. Subsequent calculations based on reasonable simplifying assumptions,including independence of the length of successive sampling intervals,then showed that observed histograms of sampling intervals led to probability density function of delays to step inputs consistent not only with data observed in the present study, but also with that obtained by Haynes (20). The random sampling hypothesis could be related to physiological data ,on the basis of a brain processing time of 100 ms.,and a time to 'read-in' a sample of between 20 and 45 ms. This conclusion agreed with conclusions reached concerning open loop test responses. The above points,plus the form of the histograms of sampling intervals,led to the conclusion that the random sampling hypothesis was basically correct,as previously advanced,but could be extended by postulating a finite sampling time, and a probability density function of sampling intervals running from about 90 to 220 ms.,with a peak in the vicinity of 130 ms.,and with lengths of successive sampling intervals practically independent. It was shown that the model response was consistent with the form of the remnant spectra obtained by Elkind (10),and that some of the remnant might arise from randomness of sampling action.

The implications of the model structure were found to be in complete accord with those of the 'Successive Organisation of Perception' model of learning proposed by Krendel&McRuer (13). It was concluded that the elaboration of model structure during the course of experimental investigations could provide a good analogy of the operator's process of learning. The observed features of precognitive tracking could be logically accounted for in terms of the structure of the model,as extended to allow for a very predictable input,and the random sampling hypothesis as described above.

Consideration of the transfer of an impulse response optimised filter indicated that,in the task studied,the operator did approach quite closely to an optimum transfer,in terms of the 'least square error' criterion. This fact largely accounted for observed differences between experimental and published results. These also led to the conclusion that the operator could not directly sense acceleration or higher derivatives with a purely visual display. More careful consideration of the operator's optimising action led to the conclusion that both error and muscular control effort contributed to 'cost',and that the exact form of the operator's subjective criterion of optimum varied during training,and in response to instructions.

From a consideration of both experimental results and physiology it was concluded that a relation between average sampling interval and external task requirements was likely to arise as a result of the operator's 'missing' occasional samples,according to his own error criterion. Thus,the overall effect was likely to be quite complex.

Experimental data concerning learning,the evolution of the

model structure,and model performance during matching trials,together with subjective data,led to the conclusion that the operator's action was manifestly that of a self-organising system,even for the simple form of task investigated experimentally. It was further concluded that simple mathematical models of the operator's learning and adaptation were likely to be somewhat restricted in their generality.

10.2. General Conclusions

The investigations described in this thesis led to the formulation of a linear sampled data model,capable of accurate simulation of the characteristics of human operators performing a continuous one-dimensional compensatory tracking task,with visual error display,manual output motion combined with simple unity feedback, and a Gaussian random input signal. This model was based on the idealised hypothesis that the operator sampled both instantaneous position and velocity,in a regular fashion and with a sampling interval in the region of 150 ms.,and then attempted to match his output position and velocity to those of the input by means of a hand motion consisting of a combination of velocity-triangle and velocity ramp functions running over individual sampling intervals. The original form of the model was based on study of the operator's responses to random step and ramp inputs,and,with suitable parameter settings,the extended form of the model could provide a good visual match to these responses.

The accuracy of simulation was such that cross-correlations between model and operator errors of the order of .8 to .9 could be obtained,corresponding to output cross-correlations of .98 to .99 . It was therefore concluded that the model was capable of matching substantially all that part of the operator's output which was linearly correlated with the input. Also,the ripple in the model's output gave a realistic simulation of the form of the operator's output,which was emphasised by considering output velocity traces. The structure of the model,combined with the random sampling hypothesis,could also account for the form of the remnant spectra reported in the literature.The close match of the model was also illustrated by comparison of covariance functions and consideration of spectral characteristics in relation to other workers' data. It was concluded that the sampled data model,as described above,was capable of giving a substantially better representation of operators' characteristics than other models so far described,especially in regard to response to discontinuous inputs.

A further test of model quality was provided by considering the results of tests in which the operator's feedback was interrupted while a synthetic error signal was provided by the model. It was concluded that the results of these tests were in accord with the hypotheses embodied in the model structure,but showed that the sampling

hypothesis needed extension to allow for a finite 'read-in' time. It was found that simple models could give a good fit to covariance data obtained during open loop operation, and that these models possessed a form, and values of parameters, similar to those obtained by other workers from measurement of operators' overall loop transfers.

From further consideration of operators' sampling action, through consideration of both experimental and physiological data, it was concluded that this action could be represented as random, with a probability density function of sampling interval length ranging from approximately 90 to 220 ms., possessing a peak in the vicinity of 130 ms. (although exact values would vary among individual operators). It was further concluded that the lengths of successive sampling intervals were not highly correlated, and that the 'read-in' time for each sample ranged from about 20 to 45 ms.

Experimental data indicated that the operator approached quite closely to an optimum transfer according to the 'least square error' criterion. From this, and through consideration of published data, it was concluded that the operator could not directly sense acceleration and higher derivatives of input under conditions of purely visual display.

From a careful study of experimental data, both objective and subjective, it was concluded that the operators manifested aspects of a self-organising system, even in the relatively simple task studied experimentally. The operator's subjective performance criterion involved both error and muscular control effort, and represented an interpretation of instructions; it was subject to change during training and due to other factors, such as motivation, emotion, fatigue, and level of physical fitness.

The reasoning and results presented above led to the postulate that a general representation of the operator's response in one dimensional compensatory tracking tasks with visual error display, could be made in the form of a model composed of a sampling of position and velocity information, according to the random sampling hypothesis already described, combined with a digital computation leading to output motion of an appropriate form, executed by means of individual programmes running during each sampling interval. The digital computation would represent an optimum digital processing of error and output data, with the criterion of optimum being a subjective funtion of instructions, error, and control effort. Thus this processing would generally be nonlinear. Also, its structure would alter during training, as more information was accumulated concerning the tracking task. The form of the output motion during sampling intervals would be such as to approximately minimise the muscular effort needed to effect a given movement, and might therefore be expected to be fairly strongly dependent on the feel of output transducer.

The above model is quite consistent with previously presented experimental data, since with a Gaussian input, fairly well attenuated at frequencies above 2 cps., and a simple controlled element, consisting of a unity transfer, one might reasonably expect a very nearly linear processing of error and output information as a result of the instruction

to minimise error at all times. Further, the fact that the control lever was light and practically frictionless would lead, through minimisation of a function of muscular force, to a velocity-triangle type of response.

The model can also explain the occurrence of increasingly nonlinear response as controlled element dynamics increase in complexity and overall lag. If the operator were restricted to an approximately linear form of response, then, for small error, the required output motion would represent a multiple differentiation, or quasi-differentiation, of the input signal ; it would therefore be likely to contain excessive power, much of which would be concentrated at high frequencies. Thus, the muscular effort required to match the output of the controlled element to the input signal would tend to be such as to cause rapid onset of fatigue. Also, it would not be possible to generate power at the higher frequencies by programmes of output motion of a discrete nature, each of which occupies upwards of 100 ms. The overall effect of increasing the controlled element complexity and lag would therefore be to increase the relative importance of the form and costing of output motion, thereby leading to a more nonlinear form of data processing.

The results of varying stick 'feel' obtained by Johnson (19) are also well explained on the basis of the above postulates. Johnson found that somewhat different forms of equivalent linear transfer were associated with different types of stick 'feel' force. Operators found it easy to adapt to inertial forces, but adaptation to viscous forces proved quite difficult, although it was eventually achieved. Now inertial forces are commonly encountered, and, according to reasoning presented above, the appropriate manual response programmes would still be in the form of velocity-triangles, though relative weighting of control effort would be increased. However, the appropriate manual response to viscous forces would be in the form of truncated velocity-triangles - practically velocity trapezoids - and it might reasonably be expected that operators could only achieve this after some training.

It was concluded, as a result of the above considerations, that the model of sampling and digital processing according to some subjective criterion of 'optimum', combined with an adaptive form of output response, as described above, was capable of accounting for experimental results, both of the present study, and as published in the literature. It was therefore concluded that this model constituted a valid representation of the operator's response in one dimensional compensatory tracking tasks, capable of direct logical extension to describe pursuit tracking. It was further concluded that the model was capable of a more satisfactory and complete representation of the characteristics of human tracking in these tasks, than the linear continuous or sampled data models so far proposed in the literature.

10.3. Recommendations for Further Investigations

The results and conclusions of the present study suggest several ways in which further investigations would probably be fruitful:-

(a) The study of the operator's response under conditions where the controlled element exhibits nonlinearity. There is evidence to suggest that,under certain circumstances,a nonlinear controlled element might result in better performance or greater convenience ; e.g. a power steering system has been developed for motor cars,in which steering sensitivity increases with steering wheel angle.

(b) The study of the effect of non-Gaussian input statistics,and of different instructions,on operator response.

(c) Allied to (a) and (b) above,the study of learning phenomena in relation to the evolution of a suitable digital processing of information. This would involve studying both short and long term variations in operators' characteristics.

(d) The study of the effects of 'feel' on output response,both in regard to overall transfer and to fine structure of output motion, with a view to discovering the form of the criterion of manual control effort,and its relative weighting under given instructions. The effects of 'active' feel forces (e.g. a control lever which actually accelerated in the direction in which it was already moving) might also prove extremely interesting in this connection...

(e) Study of tasks involving two dimensional tracking,with a view to extending the model described in Section 10.2. Greater dimensionality could logically be studied as a sequel.

(f) Study of tracking tasks involving simultaneous use of more than one sensory modality. For example,the use of a moving mock-up may allow direct sensing of acceleration via the semicircular canals. The chief point would be to evaluate how extra information was used in formulating response.

It is hoped that the above suggestions may be of value as proposals for research to advance the status of human operator studies.

REFERENCES

1. Goodyear Aircraft Corporation ; Final Report , Human Dynamic Study : Report No.GER-4750 , April 8 ,1952 (summarised in 11)

2. Bekey G.A.; Sampled Data Models for the Human Operator in a Control System : Ph.D. Thesis , UCLA ,1962 ; and Space Technology Laboratories Report No. 9990-6013-RU-000 .

3. Ward ; Ph.D. Thesis, reviewed in (2).

4. Lemay L.P. ; A Model to Represent the Human Operator in a Tracking Task : M.Sc(Eng) Thesis , University of London , 1961 .
see also,
Lemay L.P. & Westcott J.H. ; The Simulation of Human Operator Tracking Using an Intermittent Model : I.R.E. Professional Group on Human Factors in Electronics, International Congress on Human Factors in Electronics,1962.

5. Tustin A. ; The Nature of the Operator's Response in Manual Control and its Implications for Controller Design ; J.Inst.Elec.Eng. Part IIA , May 1947 .

6. Craik J.K. ; The Theory of the Human Operator in Control Systems : Brit.Jour.Psych.,Dec. 1947 and Mar. 1948 .

7. Vince M.A. ; The Intermittency of Control Movements and the Psychological Refractory Period , Brit.Jour.Psych.,1948,Vol.38.

8. Vince M.A. ; Rapid Responses and the Psychological Refractory Period : Brit.Jour.Psych.,1949 ,Vol. 40 .

9. Welford A.T. ; The Psychological Refractory Period and the Timing of High Speed Performance : Brit. Jour. Psych., Feb. 1952 .

10. Elkind J.I. ; Characteristics of Simple Manual Control Systems : Technical Report No.III ,MIT Lincoln Laboratory , 1956 .

11. McRuer D.T. & Krendel E.S. ; Dynamic Response of Human Operators : Wright Air Development Center , Technical Report 56-524 , October 1957 ; ASTIA Doc. No. NR AD-110693 .

12. McRuer D.T. & Krendel E.S. ; The Human Operator as a Servo System Element : Jour. Franklin Inst.,Vol.267,Nos. 5(May) & 6(June) 1959 .

13. Krendel E.S. & McRuer D.T. ; A Servomechanisms Approach to Skill Development : Jour. Franklin Inst.,Vol.269,No.1 ,Jan,1960 .

14. North J.D. ; The Rational Behaviour of Mechanically Extended Man : Boulton Paul Aircraft Ltd.,Wolverhampton,Eng., 1955 .

15. Henderson J.G. ; The Estimation of the Transfer Function of a Human Operator by a Correlation Method of Analysis : Ergonomics, Vol.2 ,No.3 ,May 1959 .

16. Elkind J.I. & Green D.M. ; Measurement of Time Varying and Nonlinear Dynamic Characteristics of Human Pilots : Bolt Beranek and Newman Inc., ASD Technical Report No. 61-225 . Dec. 1961 .

17. Rogers M.A. ; Interim Report on the Human Operator Investigation, Imperial College , 1958 .

18. Wilde R.W. ; A Study of the Human Operator : DIC Thesis , Imperial College , 1958 .
see also,
Wilde R.W. & Westcott J.H. ; The Characteristics of the Human Operator Engaged in a Tracking Task : Automatica , Vol. 1 , No.1 , Jan-Mar. 1963 .

19. Johnson D.W. ; The Dynamic Relationship of Force on a Control Handle to its Displacement,as Affecting Accuracy of Control : M.Sc(Eng) Thesis, University of London , 1960 .

20. Haynes D. ; DIC Thesis , Imperial College , 1963 .

21. Young L.R. ; A Sampled Data Model for Eye Tracking Movements : Sc.D. Dissertation , Dept. of Aeronautics and Astronautics, M.I.T., Cambridge , Mass. , 1962 . (also submitted to second IFAC conference, Basle, Switzerland , Sep. 1963).

22. Walsh E.G. ; Physiology of the Nervous System : (book) , 2 nd. ed. 1964 , Longmans ,London .

23. Wiener N. ; Cybernetics : (book) 2 nd. ed. 1961 , the MIT press and J.Wiley & Sons,Inc., New York .

24. Hammond P.H. ; An experimental Study of Servo Action in Human Muscular Control : Proc. Third International Conference on Medical Electronics , London 1960 .

25. Truxal J.G. ; Automatic Feedback Control System Synthesis : (book) 1955 , McGraw-Hill Book Co. Inc., New York .

26. Raoult J.C. ; Thesis for Doctor ès Sciences Degree - Etude de l'opérateur humain en tant qu'élément d'un système asservi : University of Toulouse , 1962 .
see also,
Naslin P. and Raoult J.C. ; Modèles Continus et

Echantillonnés de l'Opérateur Humain Placé dans une Boucle de Commande : Second IFAC conference ,Basle,Switzerland,1963.

27. Ezekiel M. : Methods of Correlation Analysis : (book) 2nd. ed. 1945 John Wiley & Sons Inc.,New York .

28. Blackman R.B. & Tukey J.W. ; The Measurement of Power Spectra from the point of view of communications engineering : (book) 1959, Dover Publications Inc., New York .

29. Bartlett M.S. ; An Introduction to Stochastic Processes : (book) 1960 , Cambridge University Press , Cambridge , England .

30. Jury E.I. ; Theory and Applications of the Z-transform Method :- (book) 1964 , John Wiley & Sons Inc., New York .

31. Merriam III C.W. ; Optimisation Theory and the Design of Feedback Control Systems : (book) 1964 , McGraw-Hill ,Inc., New York .

32. Matthews P.B.C. ; pp.219-288 in Physiological Reviews Vol.44 , 1964 .

33. Bamford R.S. ; A Simple Correlator for Gaussian Signals and a Low Frequency Gaussian Noise Source : DIC Thesis , Imperial College , 1958 .

34. Iakovlev V.P. ; Some Asymptotic Properties of Gaussian Random Processes : Radio Engineering & Electronics , Vol. 5 , 1960 , No.10 , p.265 et seq.

35. Feller W. ; An Introduction to Probability Theory and its Applications : (book) Volume 1 , 1957 , 2nd. ed. John Wiley & Sons Inc., New York .

APPENDIX I : Apparatus for Investigation of Compensatory Tracking

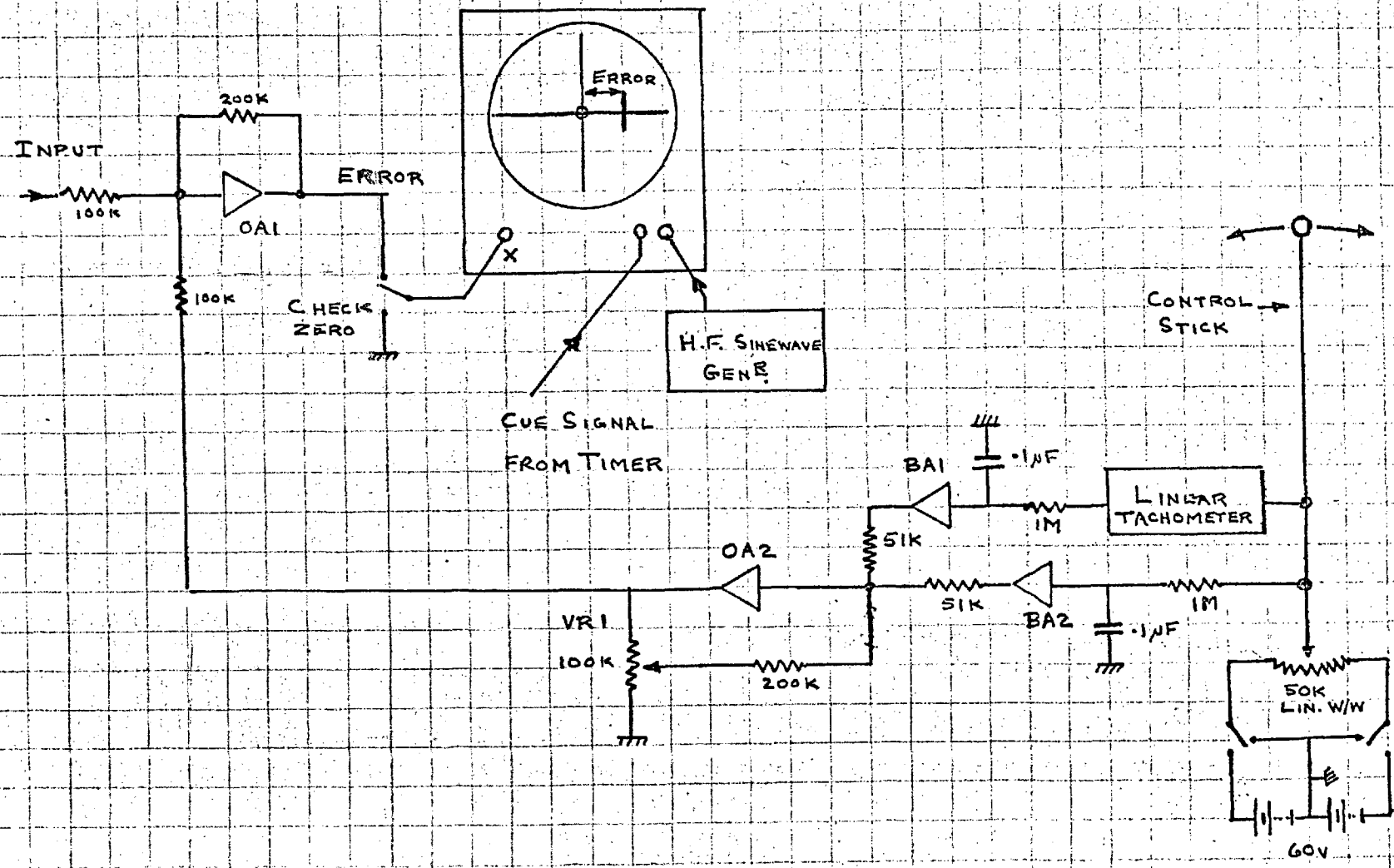
I.1. The Basic Operator Feedback Loop

The experimental investigation was concerned with compensatory tracking without any dynamics in the feedback path ; i.e. the controlled element was a simple gain term. The apparatus which constituted the operator feedback loop was therefore relatively simple, and is shown diagrammatically in Fig.AI.1.1. The display oscilloscope was provided with two independent inputs (positive and negative gain) on each of the X and Y axes. The display consisted of a vertical line whose horizontal distance from the centre line of the display graticule was a direct representation of the error between the operator's actual hand position and that required to exactly balance out the input signal. The vertical line was formed by applying a high frequency sinewave to one of the Y input terminals. During the periods immediately preceeding and following a run an additional 50 cps. cue signal,derived from a dekatron timer specially built for the purpose,was applied to the other Y input. This signal slightly modified the appearance of the line by adding a small spike to the top of it. Thus,the disappearance and reappearance of this spike informed the operator of the commencement and completion of the run. The function of the check-zero switch on the display oscilloscope was to remove the signal to the X deflection amplifier and to short the input terminal to ground. The X shift was then adjusted as necessary to bring the display line into coincidence with the centre line of the graticule.

The operator was seated facing the display,which was just below eye level,while his right hand grasped a control stick. Definite hand location was provided by a knob fixed to the top of the stick. The control stick was arranged to pivot about a short shaft near its base,which was free to rotate in bearings,so that the stick moved in a plane parallel to that of the display ; i.e. a plane normal to the operator's line of sight. The control stick was long in comparison to the distance through which the operator was required to move his hand,so that his horizontal hand motion could be considered to be linearly related to the angular motion of the control stick. The shaft of a precision 50 kilohm linear wire wound potentiometer was connected directly to that on which the control stick was pivoted. The ends of the potentiometer winding were maintained at + and - 60 volts,so that the voltage on the potentiometer slider was directly proportional to the angular position of the stick with respect to its central position. The central position itself could be arranged as required by rotating the potentiometer shaft with respect to the stick pivot shaft.

The potentiometer slider voltage suffered inevitably from quantisation effects (cogging) due to the finite thickness of the wire forming the winding. Although the quantisation was quite fine - less than ½ mm. on the display - it could be perceived by the operator. Any annoyance which this effect might have caused was prevented by passing

FIG. AI.1.1. DIAGRAMMATIC REPRESENTATION OF BASIC APPARATUS OF TRACKING LOOP



the signal through an RC filter with a time constant of 100 ms. This had the effect of causing a lag between hand movement and the movement of the display line. In order to compensate for this lag, a signal derived from a linear tachometer, the shaft of which was connected to the control stick about 3" from the pivot, was passed through a similar filter, and then via a variable gain amplifier (BA1) to be added to the filtered potentiometer signal at the input of an operational amplifier (OA2). The reason for using a linear tachometer was that it gave a smooth signal representing output velocity, unaffected by quantisation noise. The signal at the output of operational amplifier OA2 was made strictly proportional to stick position by adjustment of the gain of amplifier BA1, which effectively generated a zero in the overall response cancelling the pole due to the RC filter. The overall position gain from control stick knob to display line was adjusted by means of a potentiometer (VR1) in the feedback path of operational amplifier OA2. The normal setting was such that 1 cm. of movement at the control stick knob caused the display line to move 1 cm. in the same direction.

A general view of the tracking apparatus is illustrated at the end of the thesis.

I.2. Arrangements for Conducting a Tracking Run

Before commencing the experimental run, the operator checked the zero setting of the display oscilloscope. The stick position transducer potentiometer was then energised, and the dekatron timer was switched to operate. This timer was constructed so as to give a 50 secs. countdown period, followed by a 200 secs. run. During the countdown period the operator was instructed to commence tracking, and the circuits for computing standard deviations of error and output functions (see Appendix IV) were switched to 'operate' in readiness for the run. The operator was given a countdown at 1 sec. intervals over the last 10 secs.; at zero, the timer automatically activated the computing circuits and removed a small cue signal from the operator's display. After a run of 200 secs. the timer switched computing circuits to 'hold' and switched back the cue signal, thereby informing the operator of the end of the run.

The above procedure ensured accurate switching of computing circuits, and was most convenient for the investigator, who could concentrate on monitoring the operator's performance during the course of the run.

I.3. Circuit for Sampled Error Display

As described in Section 4.2.,experimental investigations were conducted in which the operator was presented with a sampled error function,derived from his continuous error function,which he did not see. The presentation was such that each sampled value of error was held until the next sampling instant,thereby minimising errors of perception on the part of the operator. The sampling frequency was derived from a square wave generator,and its exact value was under the control of the investigator.

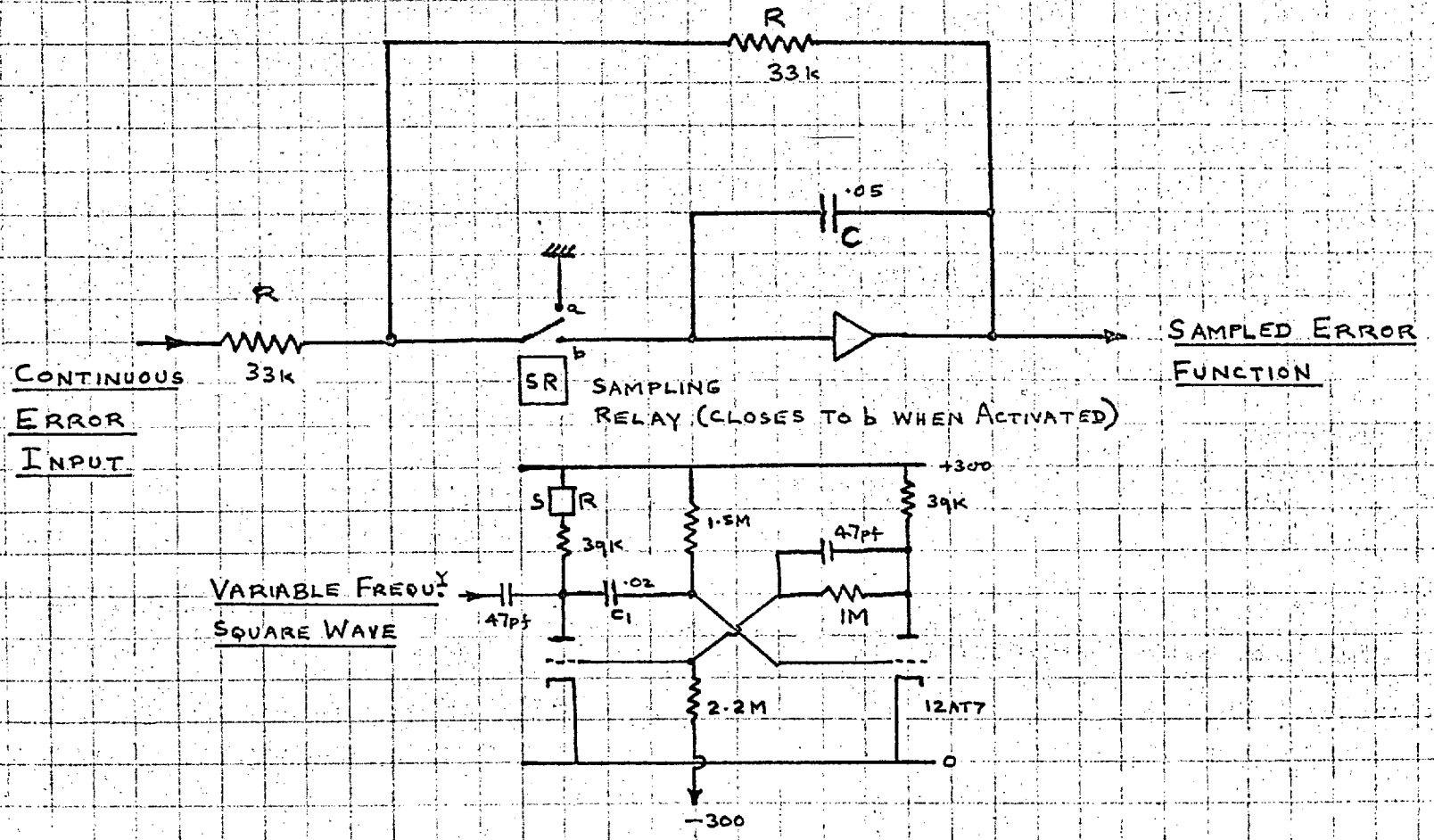
The circuit which was used to effect the above action is shown in Fig.AI.3.1.; its action was as follows :-

The sampling relay was activated by a monostable fed by the square wave generator. This monostable was designed to close the relay for about 7 ms. per cycle of the square wave. When the sampling relay closed,the capacitor C was discharged by R,and the output voltage went toward the negative of the input voltage on a time constant of CR (1.5 ms.). Thus,during the 7 ms. closure time the output reached the negative of the input with very small error. (The operator's continuous error signal changed by only a very small amount in 7 ms.). When the sampling relay opened, C was left effectively open circuited,and the output voltage remained steady. The value of the error at one sampling instant was thus held until the next sampling instant. There was some very small drift of output voltage between sampling instants,due to the finite internal resistance of the capacitor,the finite input impedance of the amplifier,and the inevitable grid current of the amplifier input stage.

I.4. Error Switching Circuit for Open Loop Tests

As described in Chapter 8 ,a series of investigations were conducted in which it was desired to present the operator with a 'synthetic' error signal for periods of about 6 secs. This synthetic signal was derived from the error signal present in the analogue model tracking loop ; both model and operator were fed with a common loop input signal. Presentation of this signal was accompanied by interruption of feedback from the operator's hand motion,so that his responses were essentially open loop. It was therefore necessary to effect a smooth transition from operator error to model error,in order to avoid giving the operator any visual cues as to what was happening. A jump in the error signal on switching back the operator's feedback was of little consequence,since by then the required recording of the operator's open loop response would have been made. It was therefore necessary to effect the changeover at a time when the model and operator errors were equal,and had a similar rate of change. It was clear that this could be effected most conveniently by means of an

Fig. AI.3.1. Circuit for Forming Sampled Error Function



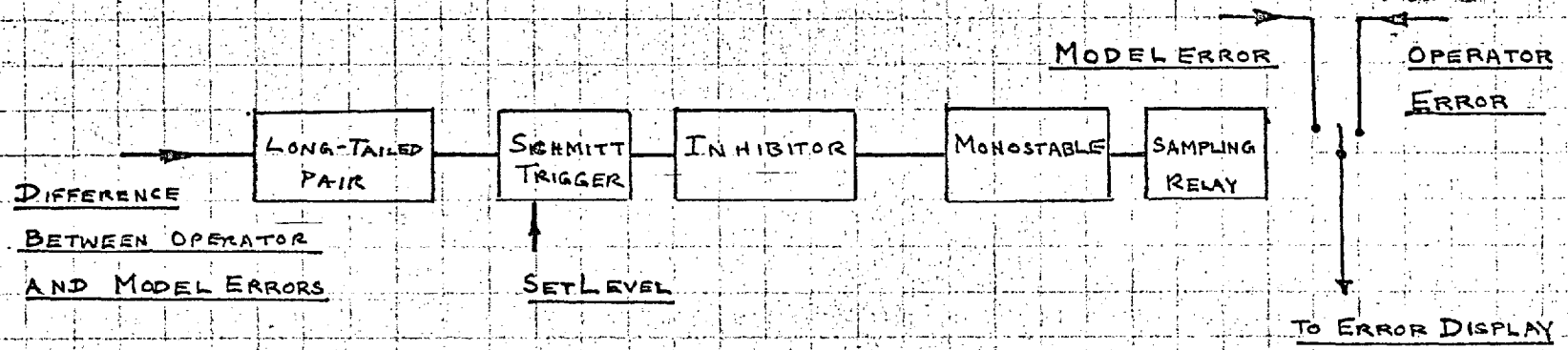
electronic circuit.

The scheme used to effect the error switchover according to the above requirements is shown diagrammatically in Fig.AI.4.1a. Referring to this figure, the action was as follows. The Schmitt trigger was set so as to trigger when the error difference signal passed through zero. Negative pulses from the Schmitt circuit were then passed through an 'inhibitor' circuit, and, under suitable conditions, triggered the monostable. This monostable operated a high speed relay so that the model error signal was switched to the operator's display for a period of 6 secs.

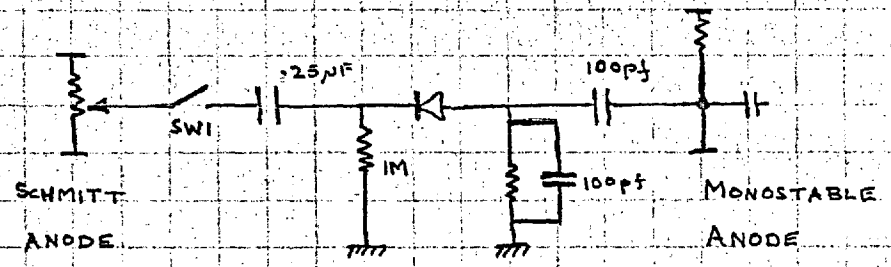
The function of the inhibitor circuit was to ensure that switching occurred only when the operator and model errors were reasonably matched in regard to velocity. It was based on the observation that model and operator position errors were always fairly closely matched, so that when their velocities were also well matched there would be relatively fewer crossovers of the error difference signal. Therefore it was required that the inhibitor circuit should prevent triggering of the monostable during periods of frequent crossovers. The circuit designed to accomplish this is shown in Fig.AI.4.1b. Referring to this, it may be seen that point D exhibits a positive potential which is greater during periods of frequent crossovers. Thus, negative pulses at the input to the monostable are reduced in amplitude. Suitable adjustment of the potentiometer in the Schmitt output anode thus ensured that the monostable was only triggered when there had been few crossovers in the previous second or so.

When it was desired to conduct an open loop test, switch SW1 was closed. This gave an initial large positive potential at D, which decayed with a time constant of ¼ sec. At the first suitable moment after this potential had decayed sufficiently, the monostable was triggered, and the operator was 'open looped' for 6 secs. The monostable was designed to be insensitive to further triggering pulses for a period of about 30 secs. or so, which gave ample time to open SW1, thereby preventing further spurious triggering.

Fig. AI. 4.1. Error Switching Circuit's Configuration



(a) Scheme of Operation



(b) Inhibitor Circuit

APPENDIX II : Low Frequency Gaussian Random Signal Generator

II.1. Requirements for Statistics of the Random Signal

As described in Section 3.2.,the experimental study of the human operator involved the use of a low frequency Gaussian random signal generator,to provide a suitable loop input signal. The requirement was for a source whose upper cut-off frequency was in the region of .5 to 1 cps. For bandwidths above 1 cps. the operator finds the tracking task extremely difficult,and is not able to follow the input with any accuracy. On the other hand,if the upper cut off frequency is too low there are undesirable psychological effects,in the nature of 'boredom',and statistical reliability of results also suffers.

To avoid undesirable fluctuation of estimates based on a run length of about 200 secs.,it was necessary to exclude ultra low frequency components.It was considered appropriate to attenuate components at frequencies less than about .01 cps.

II.2. Methods of Generation of a Low Frequency Random Signal

As explained in Section 3.2. ,it was considered desirable to employ a truly random input signal in the investigation of continuous tracking characteristics. The method most frequently used for generating such a signal is to utilise a 'noise' diode,thyratron,or similar device to generate a relatively 'white' noise,and then to pass this through a suitable filter to give the desired output spectrum. The application of this method becomes extremely difficult where,as in the present example, the bandwidth is very low. A bandwidth of 1 cps. would involve amplification of the order of a million times to yield a reasonable output signal amplitude. The difficulties of excluding fully the effects of extraneous disturbances,such as fluctuating mains noise,are obvious. Another point of difficulty concerns the difficulty of obtaining a reasonably flat spectrum. This can be overcome by means of a modulation and demodulation operation,but a very stable sinewave generator is required.

An alternative method is to derive a signal of reasonable amplitude with a spectrum flat over the region of interest,by means of a sampling or a nonlinear filtering operation. This signal is then filtered to provide the required random signal. For the random signal required in the present case,it was necessary that the input signal should have a flat spectrum from near D.C. to a point appreciably above 1 cps.,and that the rate of cut-off should not be too sharp. Various schemes may be adopted to derive such a signal. For example,Bamford (33) used a scheme involving sampling a high frequency noise source and holding the sampled value for 1/10 th. of a second. The aliasing thereby produced was such as to yield a reasonably flat spectrum.

Another scheme is to generate a pulse each time the output of a wide bandwidth noise source exceeds a given voltage level. If the level is high compared with the standard deviation of the noise, then successive pulses are approximately independent, so that the distribution of intervals between pulses is very nearly Poissonian. A random telegraph wave may then be produced by utilising these pulses to trigger a bistable circuit. The random telegraph wave has a very useful spectrum, but also possesses a violently non-Gaussian probability density function. It is therefore necessary to arrange that the spectrum extends well above the filter high frequency cut-off ; i.e. the time between transitions is small compared with the filter time constants, so that the process of multiple convolution which takes place during filtering produces a reasonably Gaussian output, as indicated by the central limit theorem. The above scheme was adopted in the present case, because it was convenient to implement. It is described more fully in the following Sections.

II.3. Probability Density Function of Intervals of a Pulse Sequence Generated by Level Sampling

The process of random pulse generation by level sampling a wide bandwidth noise signal is illustrated in Fig.AII.3.1a. When the amplitude of the wide bandwidth noise exceeds the reference level a pulse is produced. The output of the level sampler is then a series of pulses randomly spaced in time, with a certain distribution of intervals between pulses. It has been shown by Iakovlev (34) that if the statistics of the wide bandwidth noise satisfy certain conditions, the output pulse sequence tends to be Poisson distributed in time as the reference level tends to infinity. These conditions are that:-

(a) Both the wide bandwidth noise and its derivative have Gaussian probability density functions.

(b) The correlation function is doubly differentiable.

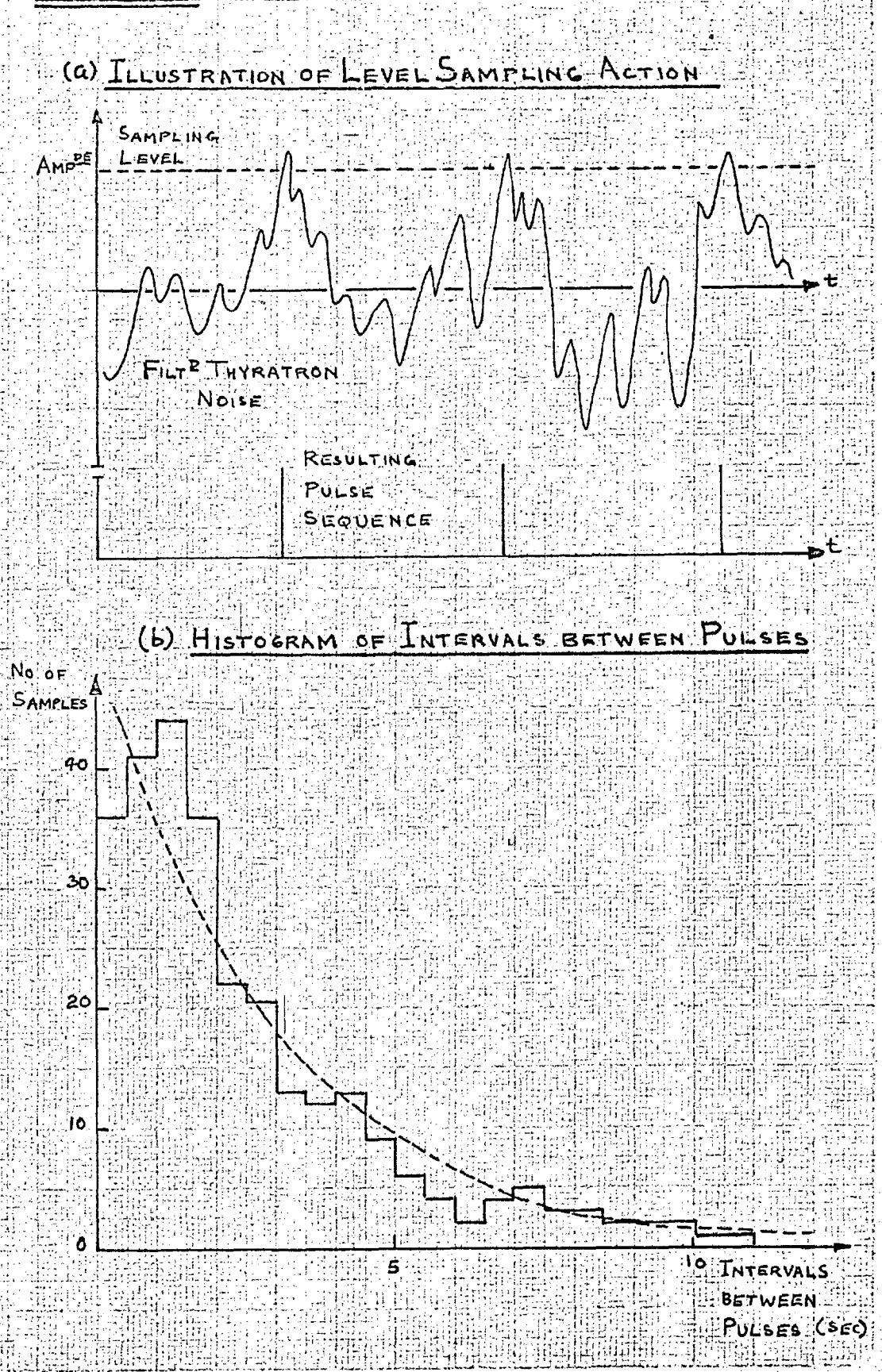
(c) If the correlation function is $R(\sigma)$, then

$$\lim_{\sigma \to \infty} \log|\sigma| . R(\sigma) = 0 \qquad \text{and} \qquad \lim_{\sigma \to \infty} \sqrt{\log|\sigma|} . R'(\sigma) = 0$$

where $R'(\sigma) = \frac{\partial}{\partial \sigma} . R(\sigma)$

A thyratron was chosen as a convenient practical source of wide bandwidth noise. Its output was passed through a low pass filter in cascade with a high pass filter with a relatively low cut-off frequency. Thus, to a first approximation , the correlation functions of

FIG. A II. 3.1. PULSE SEQUENCE FROM LEVEL SAMPLING



this filtered thyratron noise were :-

$$R(\sigma) = e^{-a\sigma} \quad \text{and} \quad R'(\sigma) = a.e^{-a\sigma}$$

Conditions (b) and (c) were therefore well satisfied.

It was not clear that condition (a) was satisfied. The filtered thyratron output seemed a good visual approximation to a Gaussian signal, but it was considered desirable to make an experimental investigation by recording output pulses and thence deriving the distribution of intervals between pulses. Results are shown in Fig.AII.3.1b., in the form of a histogram of interval lengths. This histogram was found to fit a Poisson probability density function at the 10% level of significance, according to the χ^2 test. The reason for the first two class intervals departing from the Poisson shape was probably due to the fact that, at the level setting used, the probability of obtaining two pulses close together was somewhat less than that forecast by a Poisson function, due to the finite bandwidth of the noise. However, random fluctuations could easily be responsible for this amount of deviation.

It was concluded that filtered thyratron noise was suitable as the basis of a level sampling scheme for the generation of a random square wave with an approximately Poisson distribution of intervals between transitions. The spectrum of this random square wave is considered in the following Section.

II.4. Spectral Density of a Random Square Wave with Approximately Independent Transitions

The purpose of this section is to consider the form of the power spectral density function of a random square wave having a Poisson probability density function of intervals between transitions, and the effect on this spectrum of a relative paucity of short intervals.

We first consider the ideal case of a random square wave having transitions between + and -1, independent of each other and of the time origin, and occurring at an average rate of λ/sec. Any particular sample of this square wave may then be considered as being drawn from a stationary ergodic ensemble of such random square waves. Let the value of the wave at time t be given by X(t). We may conveniently evaluate the spectral density function , $\Phi_{xx}(w)$, by evaluating and transforming the corresponding correlation function , $\rho_{xx}(\tau)$. Now we have :-

$$\rho_{xx}(\tau) = E\left[X(t).X(t+\tau)\right] \qquad \text{II.1.}$$

and we further note that :-

$$X(t).X(t+\tau) = + \text{ or } -1 \qquad \text{II.2.}$$

where the positive sign applies for an even number of transitions in the interval τ ,and the negative sign applies for an odd number of transitions. It may therefore be seen that :-

$$\rho_{xx}(\tau) = P(\text{even}) - P(\text{odd}) \qquad \text{II.3.}$$

where P(even) = probability of an even number of transitions in τ
and P(odd) = " " " odd " " " " τ

But the transition process is Poisson,therefore the probability of n transitions in the interval $\tau \geq 0$, P(n), is given by :-

$$P(n) = \frac{(\lambda\tau)^n}{n!}.e^{-\lambda\tau} \qquad \text{II.4}$$

We note that P(even) is given by the sum of P(n) for all even values of n ; a similar relation holds for P(odd). Therefore,for $\tau \geq 0$,we have :-

$$\rho_{xx}(\tau) = e^{-\lambda\tau}.\left[1 - \lambda\tau + \frac{(\lambda\tau)^2}{2!} - \frac{(\lambda\tau)^3}{3!} + \text{etc.}\right] \qquad \text{II.5.}$$

$$= e^{-2\lambda\tau}$$

Finally we have :-

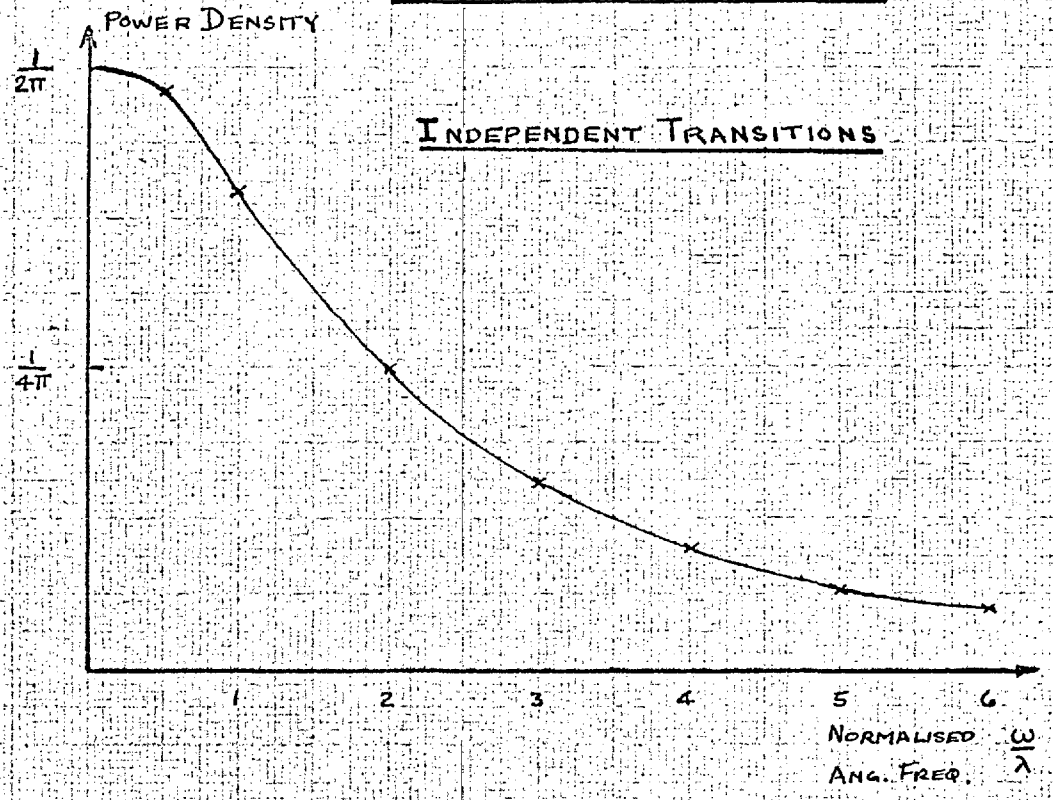
$$\Phi_{xx}(w) = \frac{1}{\pi}\int_0^{\infty} e^{-2\lambda\tau}.\cos(w\tau).d\tau \qquad \text{II.6.}$$

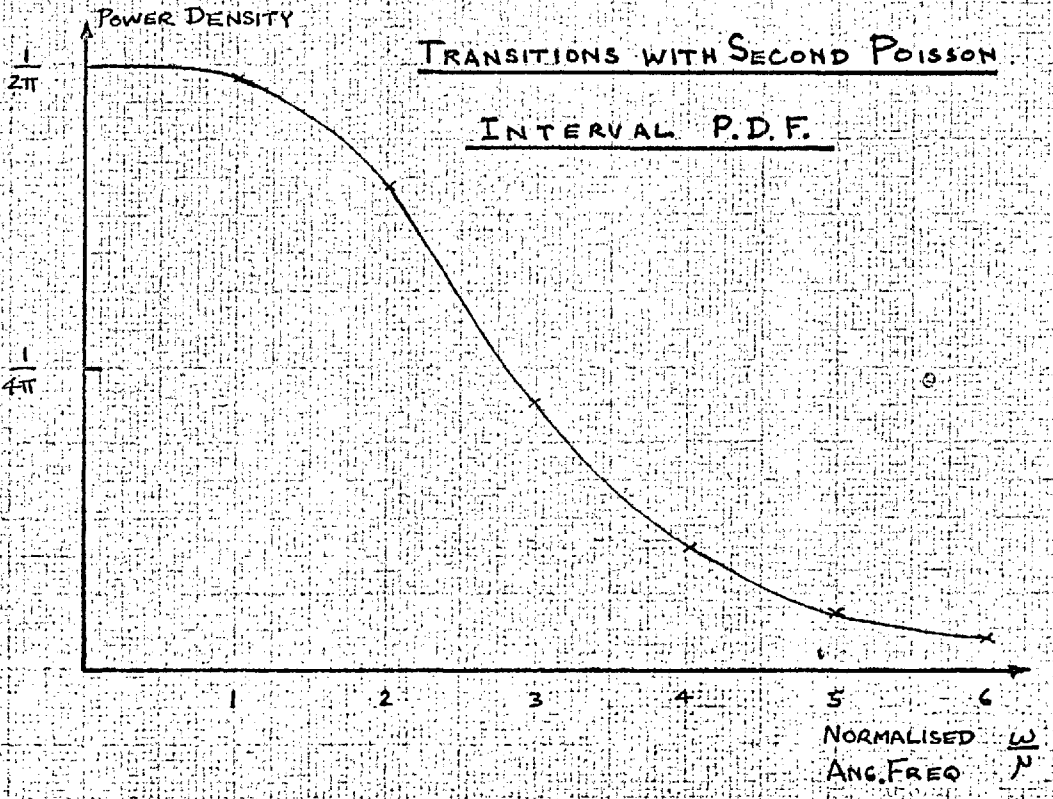
$$= \frac{2\lambda}{\pi}\,\frac{1}{w^2 + 4\lambda^2}$$

This function is sketched in Fig.AII.4.1. It may be seen that its slope is zero at w=0 and that the function is fairly flat out to $w = \lambda/2$, combined with a smooth and gentle cut-off for increasing w.

To evaluate the effect of having relatively fewer very short intervals,we consider the case of a square wave whose probability density function of intervals approximates a second Poisson. This may be considered as reasonably representative of the practical situation, since it exhibits relatively few very short intervals,with a probability density of 0 at τ=0. This wave may be related to the wave analysed above by noting that it may be generated by the same random process (random Poisson pulse sequence),but with transitions occurring at every second pulse. The analysis may also be carried out in a similar fashion. However,we now note that the correlation function is given by :-

FIG. A.II.4.1. POWER SPECTRAL DENSITY FUNCTIONS OF RANDOM SQUAREWAVES





$$\rho_{xx}(\tau) = e^{-\lambda\tau}.\left[1 + \lambda\tau - \frac{(\lambda\tau)^2}{2!} - \frac{(\lambda\tau)^3}{3!} + \frac{(\lambda\tau)^4}{4!} + \frac{(\lambda\tau)^5}{5!} - \frac{(\lambda\tau)^6}{6!} - \text{etc.}\right]$$

whence it may be seen that, for $\tau \geq 0$:-

$$\rho_{xx}(\tau) = e^{-\lambda\tau}.(\cos\lambda\tau + \sin\lambda\tau) \qquad \text{II.7.}$$

Thus, we may finally derive :-

$$\Phi_{xx}(w) = \frac{1}{\pi}\int_0^{\infty} e^{-\lambda\tau}.(\cos\lambda\tau + \sin\lambda\tau).\cos(w\tau).d\tau \qquad \text{II.8.}$$

$$= \frac{4\lambda^3}{\pi}.\frac{1}{w^4 + 4\lambda^4}$$

In terms of μ, the average number of transitions of the square wave per unit time ($\lambda = 2\mu$):-

$$\Phi_{xx}(w) = \frac{32\mu^3}{\pi}.\frac{1}{w^4 + 64\mu^4} \qquad \text{II.9.}$$

This function is illustrated in Fig.AII.4.1., in direct comparison with that of expression II.6. It may be seen that the spectrum is considerably flatter in the region $0 \leq w/\mu < 1$, but exhibits a rather sharper eventual rate of cut-off. The form of the spectrum is, if anything, more suitable than that of the ideal case, with perfectly independent transitions. It was therefore concluded that the scheme for generating a random square wave by level sampling a suitably filtered thyratron noise was appropriate for generating a noise function with a smooth spectrum, flat over the region of interest, and containing reasonable power density for very narrow bandwidth filtering.

If the linear filter fed by the above random square wave is to give the desired form of output, the random square wave must have a flat spectrum up to the filter break point, f_b, and must not exhibit a sharp decline within the range of frequencies up to at least $10.f_b$; this is also necessary to ensure that the filter output is a good approximation to a Gaussian signal. For the practical circuit the filter break point was around .5 cps. A mean transition rate of 10/sec or more was therefore quite satisfactory. The value finally chosen was of the order of 50/sec, so that the filter output was a very good approximation to a Gaussian signal.

II.5. Practical Scheme for Generation of Random Square Wave

The practical scheme adopted for generation of a random square wave suitable for subsequent filtering, is shown in Fig.II.5.1. The circuits used were all of standard form, and it was therefore considered unnecessary to elaborate on their detailed operation in the present Section. The main features of the circuits' operation are described in the following paragraphs.

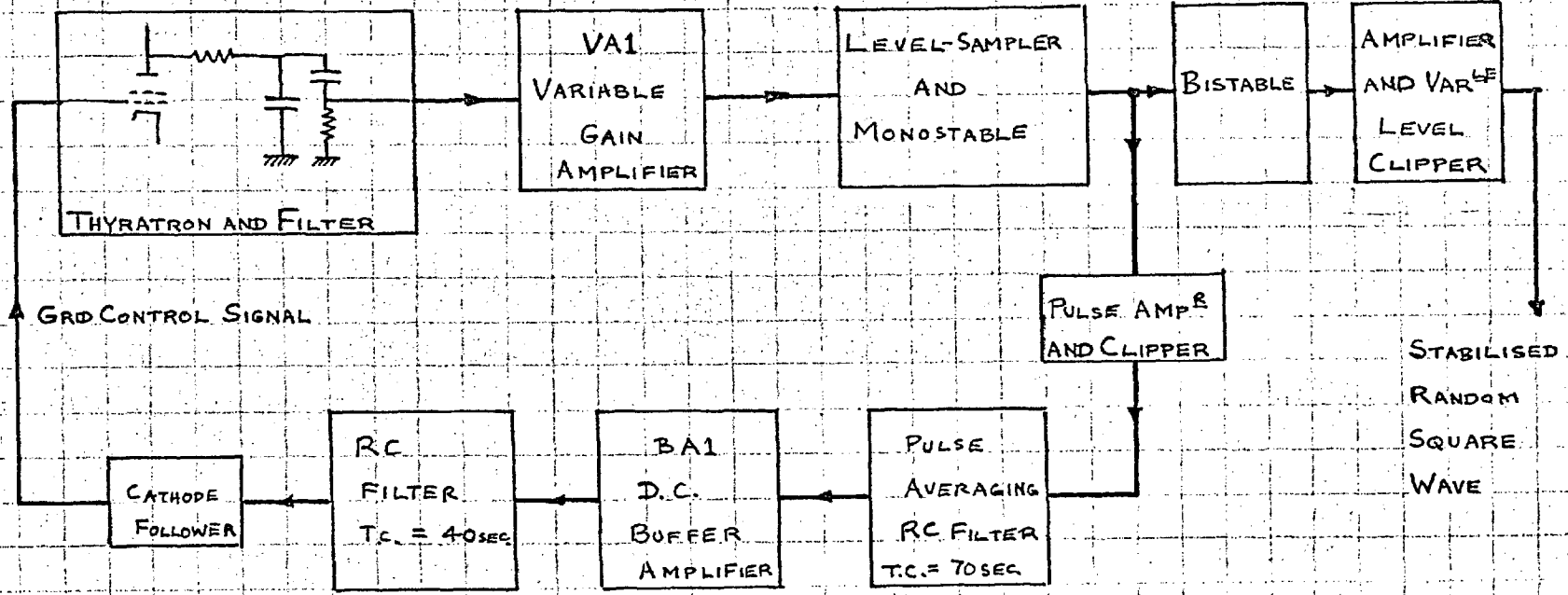
A suitable grid current was chosen for operation of the thyratron, so as to obtain a reasonable level and probability density function of noise from the output of the filter in its anode circuit. This filter possessed upper and lower cut-off points of 1.5 kcps. and 100 cps. respectively, which were chosen to give a reasonably wide bandwidth, consistent with complete interchangeability of thyratrons, and relative freedom from mains disturbances. The restriction of bandwidth resulted in some dependence between the times of occurrence of subsequently generated pulses, but, as pointed out in Section II.4, this resulted in an even better form of random square wave spectrum.

A sampling level of 50 volts was chosen as a reasonable compromise between freedom from drift effects, and requirements of amplification and standing voltage. The variable gain amplifier, VA1, was then adjusted to give about 50 pulses/sec. from the level sampling and monostable circuit. These pulses were passed to a bistable circuit, thereby generating a random squarewave, with transitions determined by the pulse sequence. This random square wave was then amplified and clipped by a variable level diode clipper, whence it was passed to a cathode follower for final output.

The output square wave possessed a well defined amplitude, but the average number of transitions per second were subject to drift, due to drift in D.C. levels, amplification (minimised by feedback), and thyratron noise level. Since stabilised power supply circuits were used this drift was not very serious, but it was considered desirable to apply a measure of overall stabilisation. This was accomplished by means of feedback from the mean rate of level sampler pulses to the thyratron grid current, which controlled the thyratron output noise level. The requirements for this feedback were that it should not operate too quickly, thereby degrading the statistical properties of the random pulse sequence, but that it should provide a reasonable long term measure of correction. This was achieved by means of the circuit described below.

Pulses from the output of the monostable were fed to a pulse amplifier and clipper, which was required in order to increase and standardise the areas of individual pulses. This amplified pulse sequence was then passed to a long time constant (70 sec.) RC circuit. Calculation shows that the ratio of the standard deviation to the mean level of the output of this circuit (assuming a mean level of zero in the absence of input pulses) is given by $1/\sqrt{2\mu T}$. Inserting the values $\mu = 50$, $T = 70$, it may be seen that this figure is little more than 1%. The output of this circuit therefore represented a good estimate of the mean number of pulses/sec., over the previous minute or so. However, it tended to give rather too much weight to pulses occurring in

Fig. AII 5.1 Scheme for Generating Stabilised Random Square Wave



the very recent past. This was corrected by passing the signal through a buffer amplifier,BA1,feeding another low pass RC filter,with a time constant of 40 secs. This gave a much more satisfactory overall filter weighting function ; the filter impulse response was initially zero, relatively small over the first 10 secs. or so,and exhibited a very smooth peak. The final filter output was passed to a cathode follower, which supplied grid current to the thyratron.

The sign of the overall feedback was determined by the requirement that the thyratron grid current should be reduced if the average level-sampler pulse rate increased. The appropriate magnitude of the gain could not be determined 'a priori' because of the complicated nonlinearities present in the loop ; e.g. an analytic expression for the relation between pulse rate and thyratron noise level was not available. An experimental approach was therefore adopted. At various gain settings of BA1 the transient and long terms response of the control loop was evaluated,by monitoring the signal at the output of BA1. The procedure was to interrupt the random pulse train feeding the first RC filter,and allow the circuit to settle. The pulse train was then re-applied,and readings of the monitored voltage were taken at one minute intervals. Typical responses are shown in Fig.II.5.2.

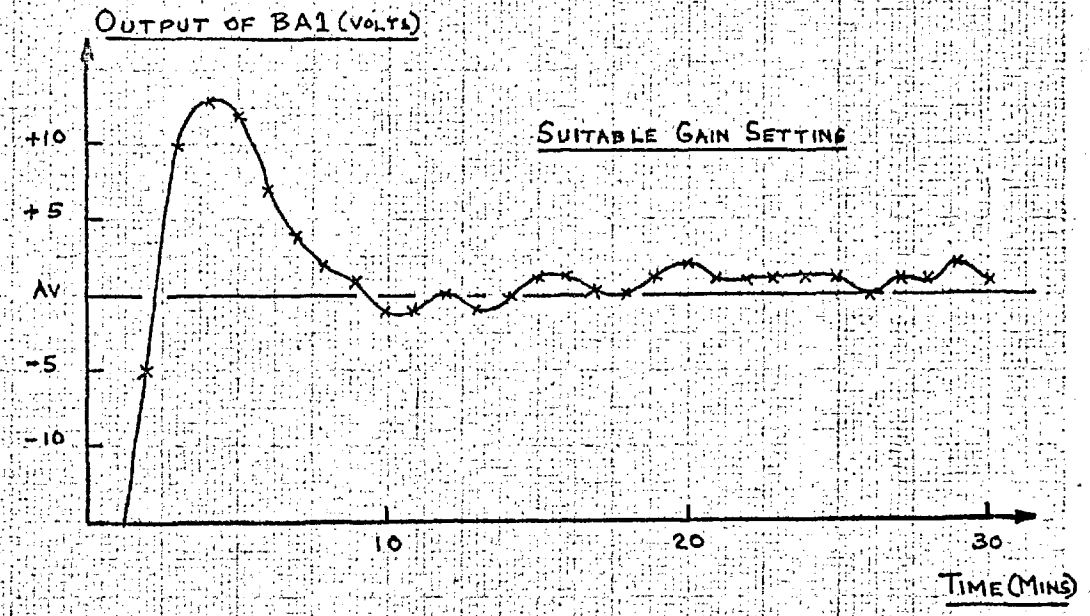
At a gain setting giving reasonable overall loop transient response (small overshoot),the long term stability of pulse rate was also very acceptable ; the fluctuations of monitored voltage could be accounted for as being entirely due to statistical fluctuations of the random pulse sequence,which might be expected to occur even under conditions of perfectly stable mean rate. Too high a gain setting resulted in a 'limit-cycling' action,while increasing the gain still further resulted in absolute instability.

When properly adjusted,the circuit described above generated a random square wave exhibiting very satisfactory stability both of amplitude and mean rate of transitions,combined with suitable spectral characteristics. The filtering of this wave is described in the following Section.

II.6. Filter for Deriving Low Frequency Random Signal from the Random Square Wave

The requirements for the spectral characteristics of the continuous random signal,used in tracking studies,are set out in Section 3.2. To obtain the required cut-off at high frequencies,three identical exponential lags were required ; the low frequency cut-off was accomplished by two CR lead circuits,which it was convenient to use for coupling filter stages. The circuit diagram of the filter is shown in Fig,II.6.1..To obtain a suitable output signal level (about 10 volts R.M.S.) it was necessary to have a stable overall gain of about 6. This was accomplished evenly over the first two stages,by employing triode buffer amplifiers with heavy cathode feedback.

FIG. AII.5.2. RESPONSES OF RANDOM SQUARE WAVE STABILISATION LOOP



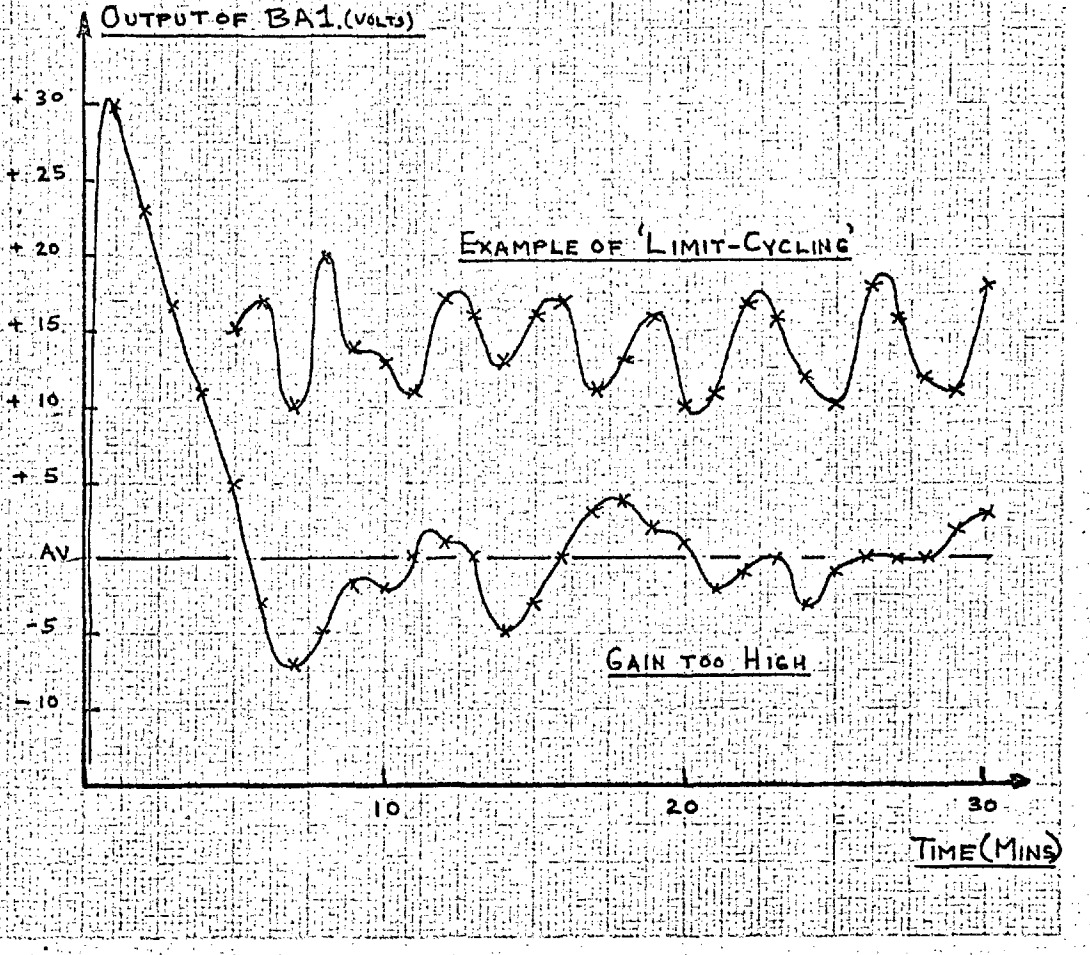
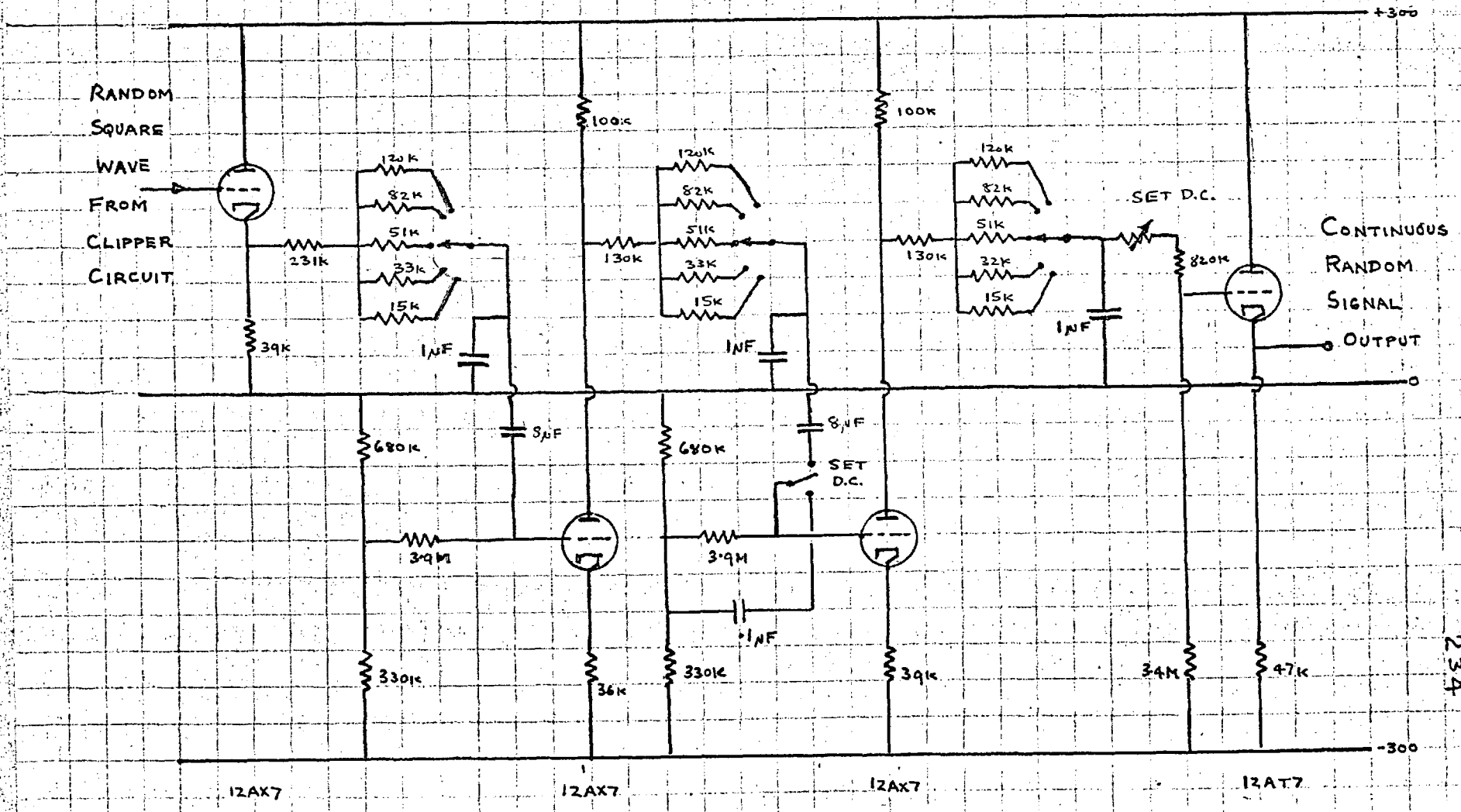


Fig. AII. 6.1. Circuit of Filter for Generating Continuous Random Signal



This arrangement served to more than balance out the reduction of the variance due to filtering, with the idea of keeping the standard deviation of the signal as small as possible compared with the allowable swings of the amplifier stages. The amplitude of the random square wave input was more than 50 volts, whereas the output of the first filter stage was of the order of 1 volt r.m.s., so that truncation effects were completely negligible. The largest standard deviation occurred at the last stage of filtering, but even here the ratio of the standard deviation to allowable swing was less than .1 . This amount of truncation of the Gaussian distribution was considered completely acceptable ; practically, the signal would never exceed the allowable swing.

Alteration of the upper cut-off frequency of the filter was provided for by means of switching suitable resistors, as shown in Fig.II.6.1. The change of standard deviation of the output caused by this change of filter transfer could be corrected, if desired, by alteration of the clipping level of the input square wave.

A segment of the continuous random signal generated by the filter is shown in Fig.II.6.2. Analysis of such recordings showed that the signal possessed a satisfactorily Gaussian distribution, according to the χ^2 test. The theoretical properties of the signal in the spectral and correlation domains are illustrated in Figs.II.6.3. and II.6.4. respectively. It may be noted that the cross-correlation between position and velocity and the autocorrelations of velocity and of acceleration are hardly affected by the presence of the lead terms in the filter transfer, as might be expected from consideration of the derivation of the corresponding spectral functions.
All the above illustrations refer to a signal possessing an upper cut-off frequency of .61 cps., as used in experiments.

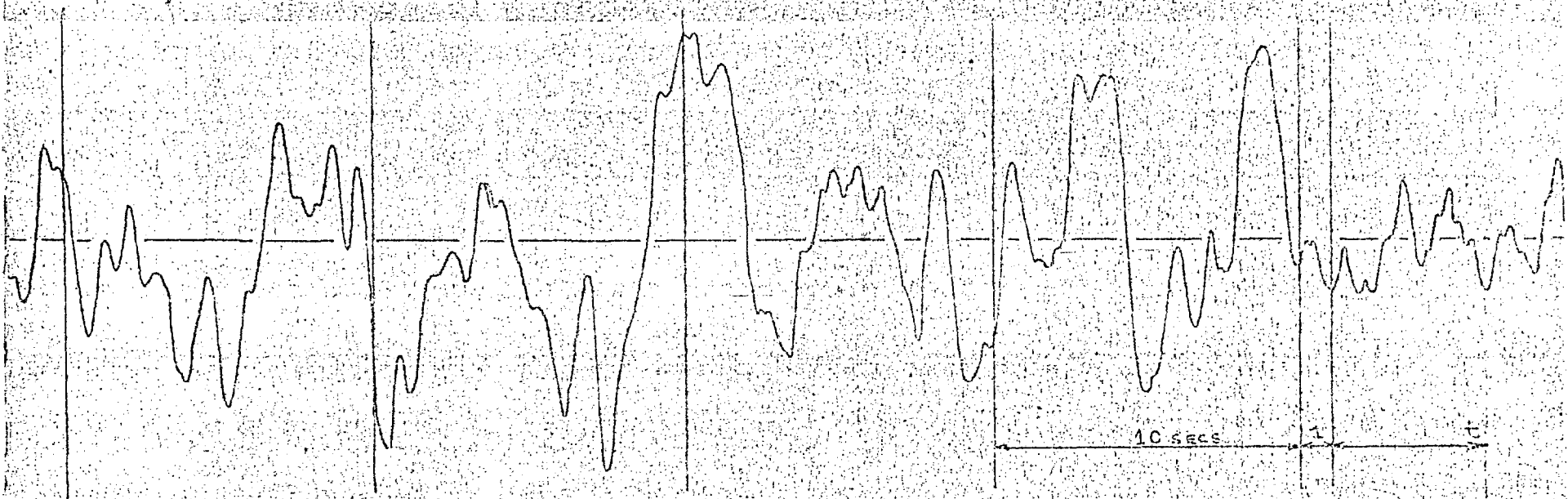


FIG. AII.6.2. SAMPLE OF OUTPUT OF CONTINUOUS RANDOM SIGNAL GENERATOR. (FILTER C/O AT 61 CPS)

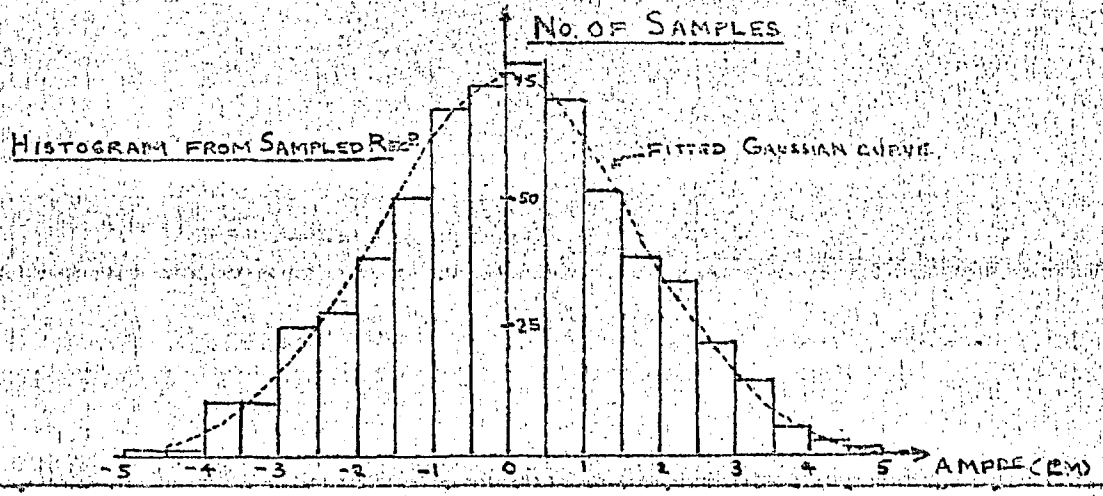


FIG. AII.6.3.

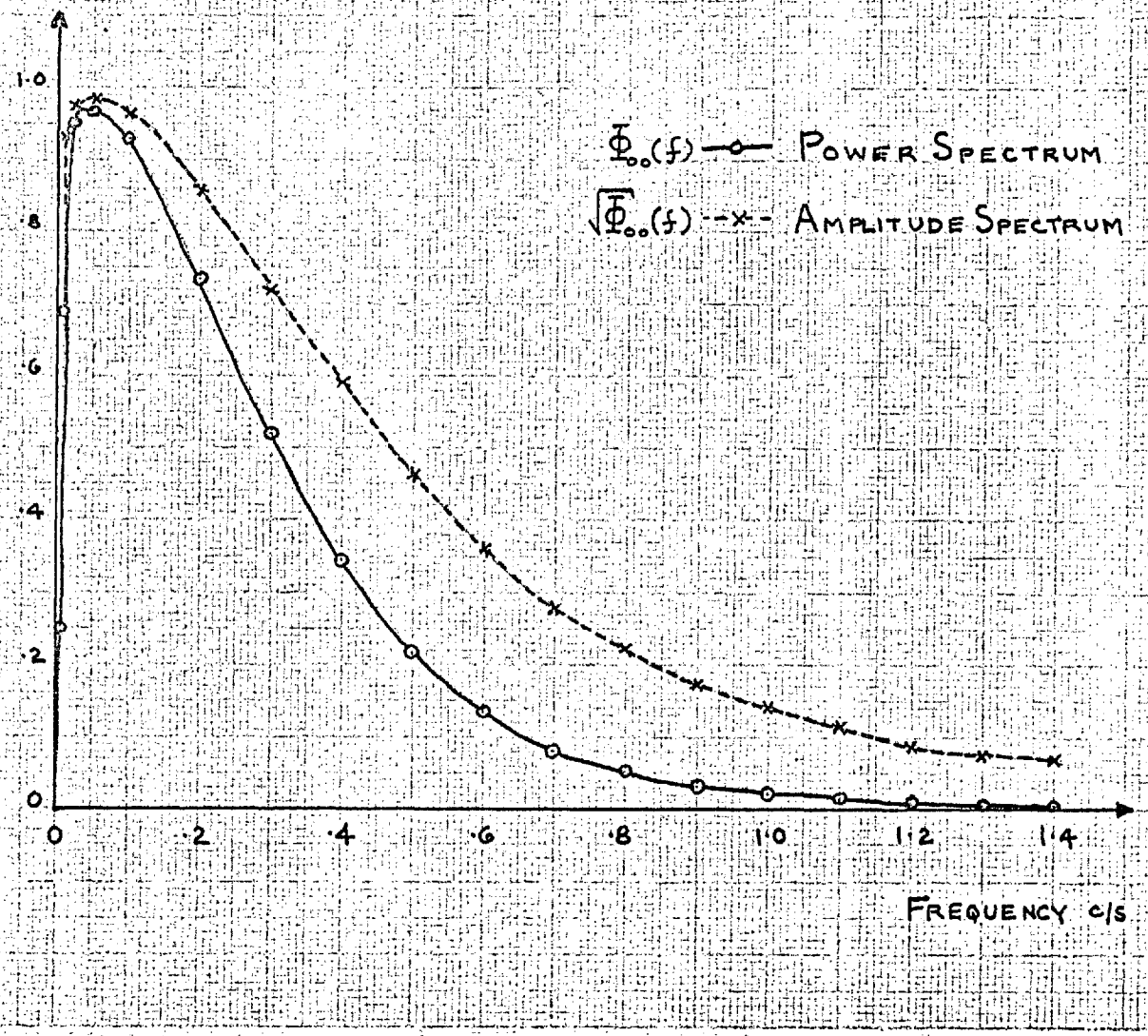
OUTPUT SPECTRUM OF NOISE GENERATOR

CALCULATED THEORETICALLY FROM :-

$$\Phi_{oo}(f) = \frac{k}{\left(1+\frac{f^2}{f_1^2}\right)^3\left(1+\frac{f_2^2}{f^2}\right)^2}$$

$f_1 = \cdot 61$ c/s $\qquad f_2 = \cdot 0045$ c/s

NORMALISATION $k=1$



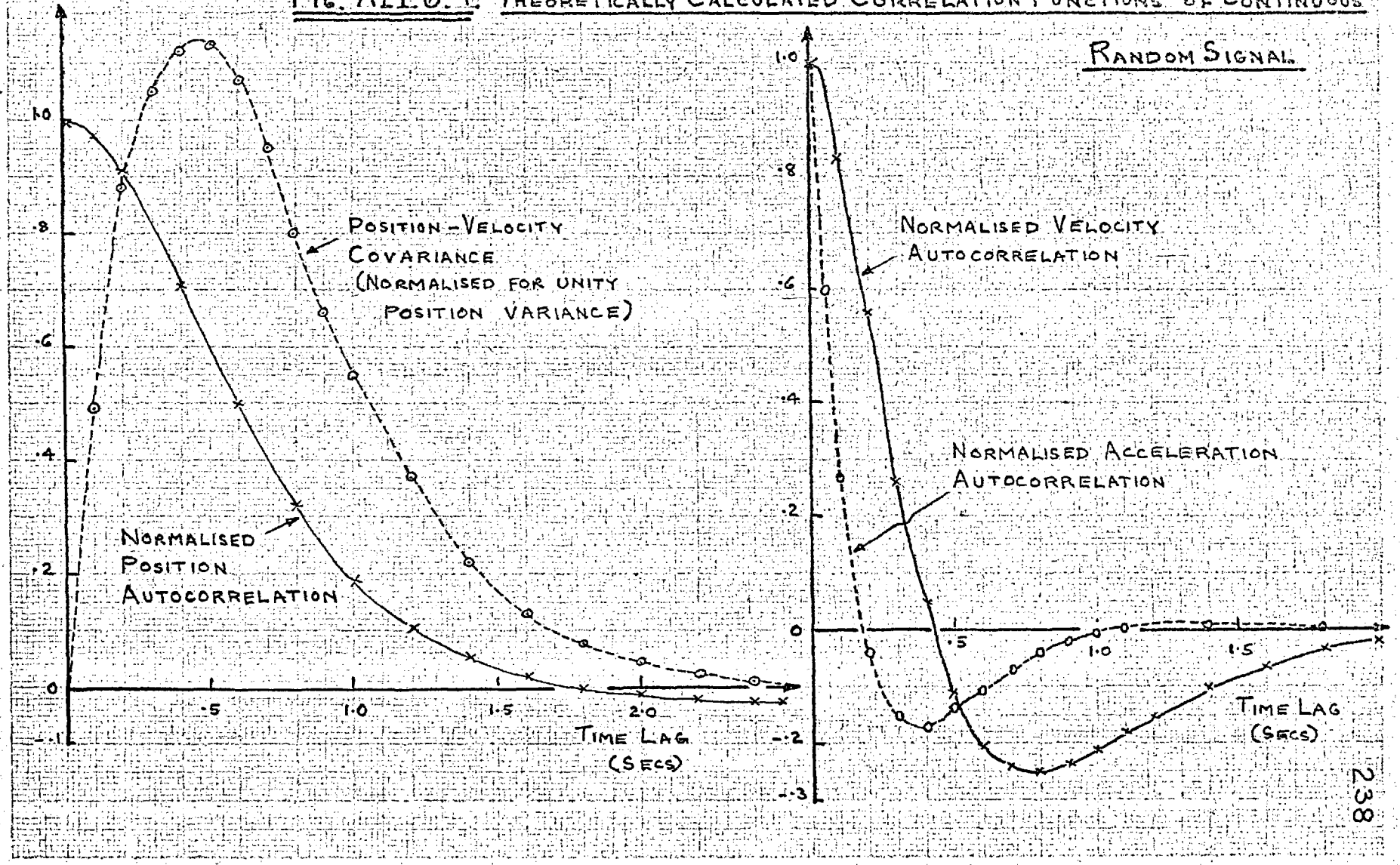


FIG. AII.6.4. THEORETICALLY CALCULATED CORRELATION FUNCTIONS OF CONTINUOUS RANDOM SIGNAL
1·0
·8
·6
·4
·2
0
-·1
·5
1·0
1·5
2·0
TIME LAG (SECS)
POSITION-VELOCITY COVARIANCE (NORMALISED FOR UNITY POSITION VARIANCE)
NORMALISED POSITION AUTOCORRELATION
1·0
·8
·6
·4
·2
0
-·2
-·3
·5
1·0
1·5
TIME LAG (SECS)
NORMALISED VELOCITY AUTOCORRELATION
NORMALISED ACCELERATION AUTOCORRELATION

APPENDIX III : Generation of Random Step and Ramp Functions

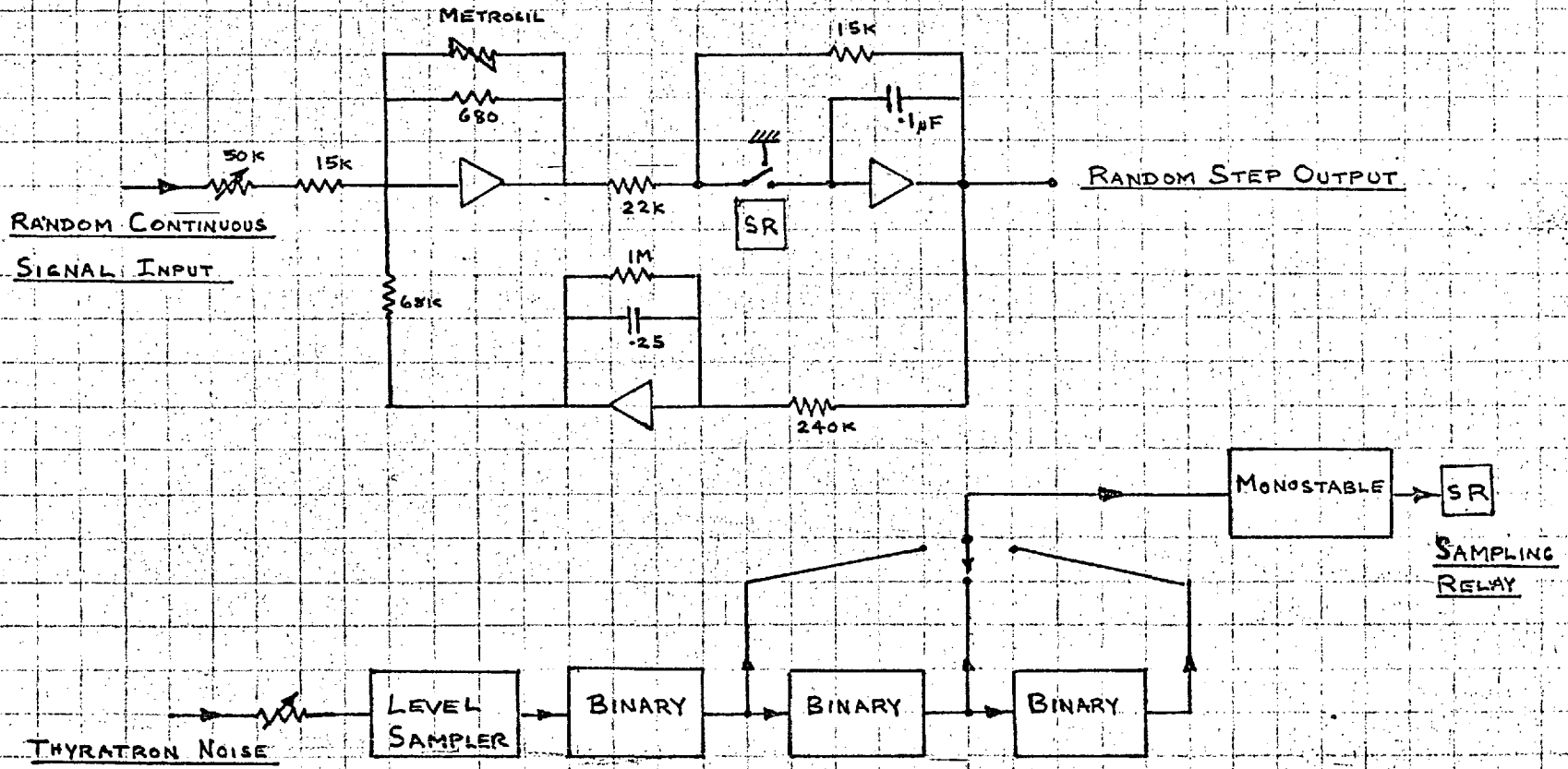
III.1. Generation of Random Step Function

The requirements for the random step function were set out in Section 3.4.,and it is noted here that they were for a function possessing randomness in both time and amplitude structure,with relatively few small steps. The circuit which was evolved for generating steps according to this requirement is shown in Fig.III.1.1. The operation of this circuit is described in the following paragraphs ; since component circuits were of standard form,their operation is not discussed in detail.

Randomness in the times of occurrence of steps was obtained by causing steps to occur according to a random pulse sequence obtained from the random signal generator. This pulse sequence was derived by level-sampling the stabilised thyratron noise signal (Section II.5.). The level-sampling circuit was exactly as used previously,with the same setting of sampling level. To obtain the slower rate of pulses required,the noise was attenuated by a potentiometer. This also allowed convenient and rapid setting of the exact rate of pulses needed ; the rate usually employed was of the order of 1/sec.,which resulted in practical independence between successive pulses. Steps occurring with approximately second,fourth,or eighth order Poisson interval probability density functions were then obtained by passing the random pulses into a chain of bistables,the outputs of which could be switched as required to trigger a monostable circuit driving a high speed sampling relay. This relay formed part of the sample and hold circuit generating the output step function.

The required non-Gaussian form of amplitude probability density function was obtained by feeding the sample and hold circuit from an operational amplifier with a metrosil in the feedback path. This gave an output approximately equal to the cube root of the input. The input to the nonlinear amplifier was effectively provided by a combination of a continuous random input derived from the random signal generator,and a feedback signal proportional to the magnitude of the last output step. This latter was provided by an operational amplifier exponential lag circuit ; the value of the lag was a compromise sufficient to minimise change of output while sampling took place,but nevertheless allow the output to settle before the next sampling time. The feedback gain was chosen so that,in the absence of external input,the loop exhibited a limit cycle of a suitable magnitude (about 10 volts),under which condition the output was a random square wave. The necessary degree of randomness was then obtained by feeding the loop with a continuous random signal,of the required variance. The feedback arrangement was necessary to obtain a non-Gaussian distribution of step heights,with relatively few small steps,yet it still allowed a satisfactory degree of independence between successive steps. The suitability of the amplitude structure of the step function was checked by experimental measurement,as described in the following Section.

FIG. A III.1.1 ARRANGEMENT FOR GENERATING RANDOM STEP FUNCTION



III.2. Statistics of Step Generator Output

The nonlinear nature of the step generator circuit precluded the direct derivation of analytic expressions for the amplitude probability density functions of the output step sequence or of step height. Numerical approximations to these could have been obtained by digital computation, but in the present case this was considered unwarrantable. It was therefore judged that the most appropriate method of estimating the statistics of the step generator output was that of direct measurement from recordings of step functions obtained under specified conditions.

Fig.III.2.1. shows histograms of step height obtained for three different levels of variance of the continuous random signal input, where the average rate of steps was sufficiently small for successive samples of the random input to be effectively independent. The forms of the histograms change as expected for increasing input variance, becoming decidedly less 'peaky'. They all clearly illustrate the relatively low probability of very small steps. These histograms therefore served to confirm that the step function possessed the required form of probability density function of step height.

The magnitude of the correlation between successive steps would be expected to vary from unity, for no input signal, down to a very low figure for large input variance. This was verified experimentally ; for the largest variance used, correlation between successive steps was of the order of -.2 . Steps used for studies of operators' responses were generated under conditions of moderate to large input variance, so that the magnitude of correlation of successive steps was less than .4 .

The time structure of the signal was measured for the case where one step was generated for every fourth pulse from the level-sampler. It was found that the experimental histogram of intervals fitted a fourth Poisson probability density function at the 10% level of significance, according to the χ^2 test. This was considered quite satisfactory for the purposes of step response investigations.

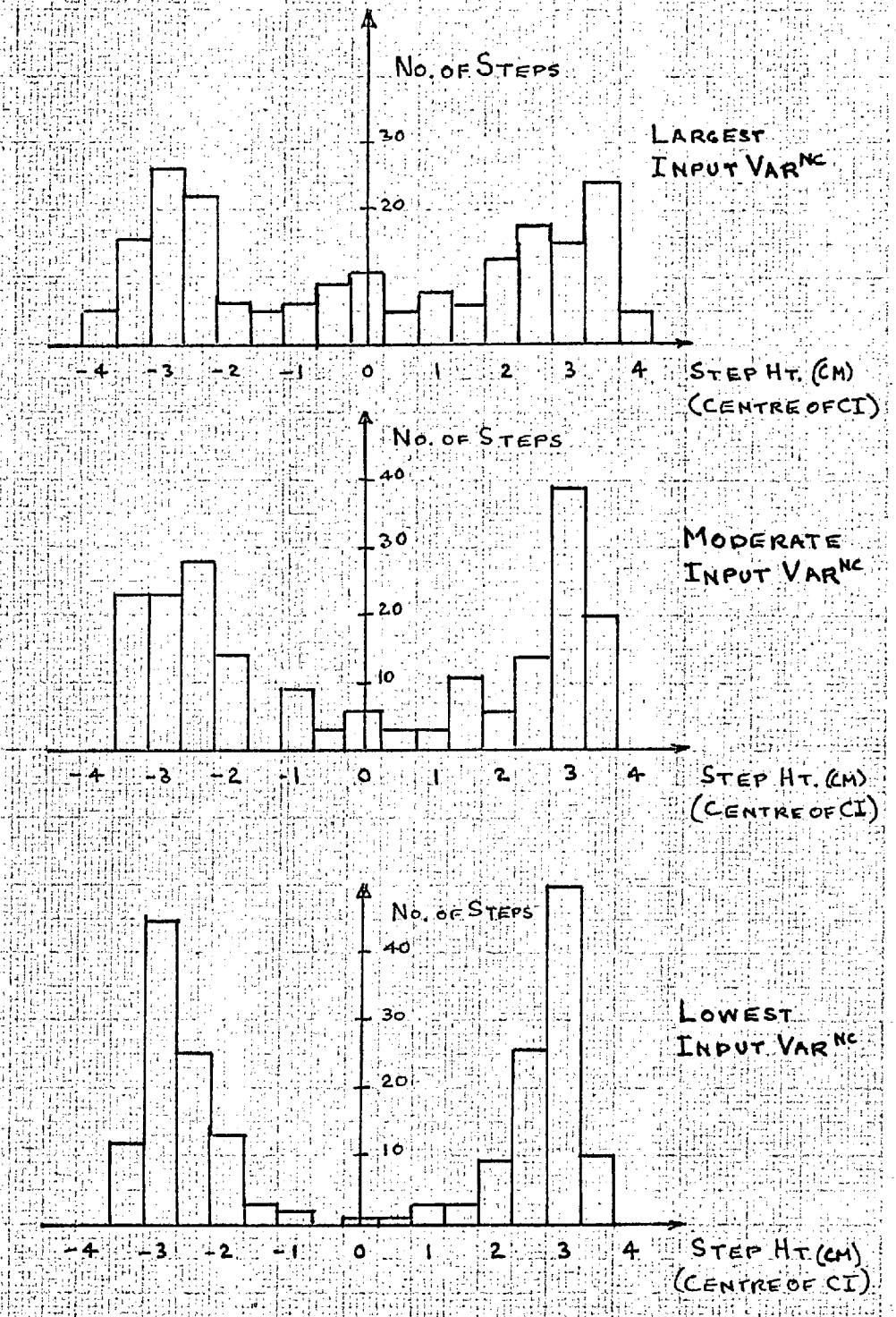
III.3. Random Ramp Function

The random ramp function was formed by passing the random step function described above through an operational amplifier quasi-integrator circuit. The ramp function therefore consisted of a series of gently curving exponentials, separated by 'corners' where there were discontinuities of velocity. The probability density function of the discontinuities of velocity was determined entirely by the properties of the step function, as was that of the length of the ramps. The statistics of these functions could therefore be controlled as required by suitable modification of the step function.

The reason for not using a pure integration to generate ramps

was that the output function which resulted exhibited nonstationarity. In terms of ensemble properties, the variance of the mean level increased with time ; the form of the output was similar to that obtained from 'coin-tossing' experiments (35). The specification of a quasi-integration implied a stationary ramp function, but the variance was directly proportional to both integrator gain and time constant. The choice of suitable values was a compromise between the desire for the greatest possible time constant (gentlest curvature of ramps), and experimentally useful discontinuities of velocity (reasonable integrator gain), consistent with reasonable variance. Suitable values of gain and time constant were determined experimentally; suitable integrator gain was the primary requisite, and, having set this, the time constant was set to give a reasonable variance...

FIG. A III. 2.1. HISTOGRAMS OF STEP HEIGHT FOR THREE LEVELS OF VARIANCE OF CONT^{S} RANDOM INPUT



APPENDIX IV : Estimation of Variance and Cross-Correlation from Integrated Moduli

IV.1. Estimation of Variance

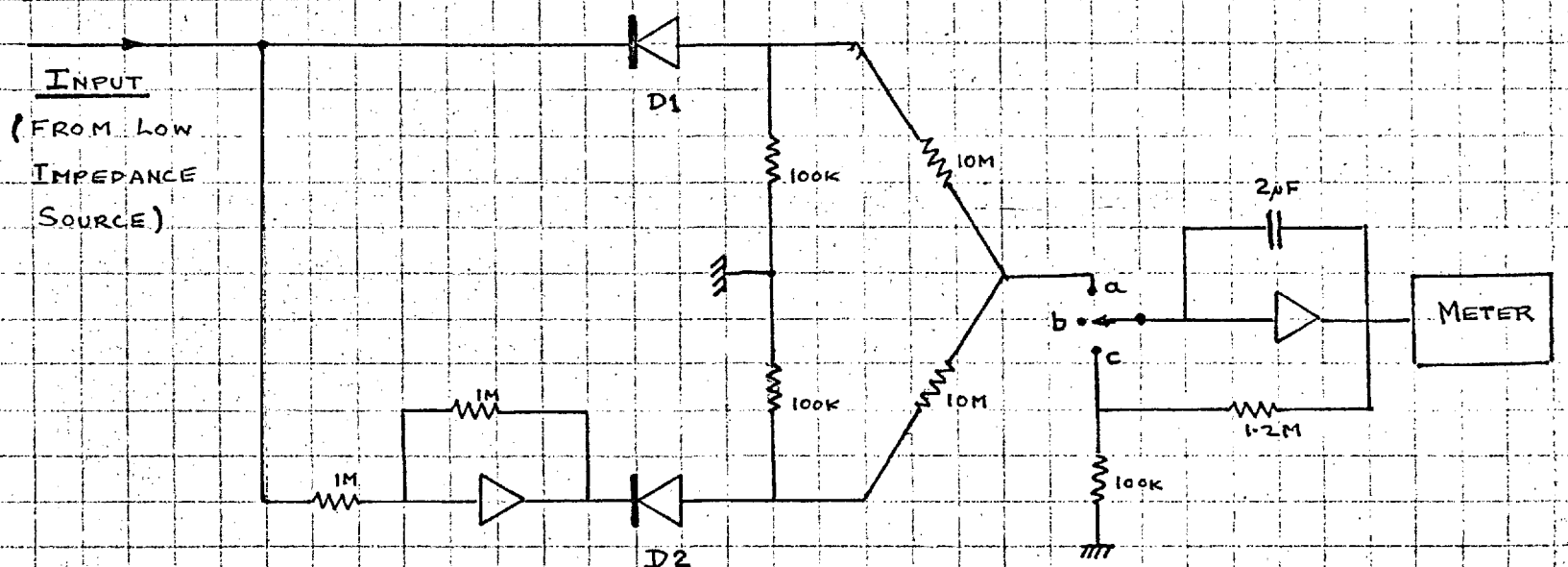
As discussed in Section 3.6.,a suitable performance index could be defined in terms of variances of error and input functions. Also, the calculation of the cross-correlation between model and operator errors (Section 5.2.) could be conveniently made from estimates of the variances of the errors and of their difference. It was therefore necessary to design an analogue computing circuit from the output of which an estimate could be made of the variance of the input function.

The most obvious method of computation was to square the instantaneous voltage whose variance was to be measured,and integrate the resulting signal during the course of an experimental run. The integrated value at the end of the run,combined with knowledge of the mean value,would then yield an estimate of the variance. The drawback to such an arrangement was that it required an accurate squaring circuit, or alternatively,an analogue multiplier. Since four simultaneous computations were required in model matching trials (input,operator error,model error,and difference of errors),this implied considerable complication and/or expense of circuitry.

The above difficulty was circumvented by taking into account the fact that the signals to be measured possessed zero-mean Gaussian probability density functions. It was therefore possible to estimate the variances from the integrated moduli of the signals. This fact enabled the use of a relatively simple computing circuit,involving the use of two operational amplifiers and two high quality diodes ; a diagram of this circuit is shown in Fig.IV.1.1. The operation of the circuit is quite straightforward ; if the input is negative,D1 conducts - if positive,D2 conducts a negative voltage with a magnitude equal to that of the input. The 100 kΩ resistors ensure that the diodes turn on and off sharply,thereby preventing either of the diode outputs from rising significantly above zero ; at the same time they are very much greater than the forward resistance of the diodes,except at voltages of the order of a few millivolts. Since signals to be measured possessed r.m.s. levels of at least 2 volts,the error incurred by this small effect was negligible. The time constant of the integrator could be any suitable value ; for the purposes of the present study it was chosen so as to give an output voltage between 10 and 100 volts at the conclusion of the tracking run,in order to obtain the best possible accuracy. The estimating properties of the integrated modulus were investigated by theoretical analysis,as described in the following paragraphs.

Let X(t) be the instantaneous value,at time t,of the signal whose variance is to be estimated,considered to be drawn from an ensemble of stationary ergodic Gaussian signals. Then :-

Fig. IV. 1.1. Circuit for Measuring Integrated Modulus



$$\lim_{T_N \to \infty} \frac{1}{T_N} \int_0^{T_N} |X(t)| \, .dt = E\left[|X(t)|\right] \qquad \text{IV.1.}$$

Now, in the case of a zero mean signal with variance σ^2, it may easily be shown that:-

$$E\left[|X(t)|\right] = \sqrt{\frac{2}{\pi}} \, .\sigma \qquad \text{IV.2.}$$

Equations IV.1. and IV.2. show that, in the limit, the integrated modulus divided by the integration time yields an unbiassed estimate of the standard deviation of the signal.

For the case where the signal has a mean value of μ, the expression corresponding to IV.2. is :-

$$E\left[|X(t)|\right] = \sqrt{\frac{2}{\pi}} \, .\sigma . e^{-\mu^2/2\sigma^2} + \mu . \mathrm{Erf}\left(\frac{\mu}{\sigma\sqrt{2}}\right) \qquad \text{IV.3.}$$

Comparison of expressions IV.2. and IV.3. shows that the presence of a mean value in the signal introduces an upward bias into the estimate of standard deviation. For small mean values this bias is very small ; e.g. the bias is 3% for a value of μ/σ as high as .14, corresponding to about 1½ volts for the input signal used in continuous tracking tests.

In the practical case the integration time was limited to 200 secs. This introduced a variance into the estimate of standard deviation, but did not introduce bias, because no subtraction was made to allow for the mean value of the sample. Since population mean values of the signals could be considered to be very small, a practically unbiassed estimate of variance was provided by taking a value proportional to the square of the integrated modulus of the sample.

The relative standard deviation of the estimate of standard deviation, obtained as described above, was evaluated by considering the integration over T_N to be equivalent to adding n independent samples of $|X|$, where $n = T_N/\tau$, and τ is the time for which successive samples of the signal being measured may be considered to be independent. Now the probability density function for $|X|$ is given by :-

$$p(|X|) = \sqrt{\frac{2}{\pi}} \, . \frac{e^{-X^2/2\sigma^2}}{\sigma} \quad ; \quad X > 0 \qquad \text{IV.4.}$$

whence it may be seen that the mean square value of $|X|$ is just σ^2 Since the mean is $\sqrt{(2/\pi)}.\sigma$, the variance is given by:-

$$\mathrm{var}(|X|) = \left(1 - \frac{2}{\pi}\right).\sigma^2 \qquad \text{IV.5.}$$

Now the variance of a sum of n independent samples of $|X| = n.\mathrm{var}(|X|)$ and therefore the standard deviation is $= \sqrt{n}.\sqrt{\left(1 - \frac{2}{\pi}\right)}.\sigma$

The mean of a sum of n independent samples of $|X| = n.\sqrt{\frac{2}{\pi}}.\sigma$

Therefore the relative standard deviation of the estimate of standard deviation, or its relative variation, v_r, is given by :-

$$v_r = \sqrt{\frac{1}{n}\left(\frac{\pi}{2} - 1\right)} = \sqrt{\frac{.57}{n}} \qquad \text{IV.6.}$$

For the computation of the standard deviation of the continuous random signal input, the effective value of n was approximately 200, whence :-

$$v_r = \sqrt{\frac{.57}{200}} = .05 \qquad \text{IV.7.}$$

Since n was large, the estimated standard deviation was approximately normally distributed, so that probable error was an appropriate measure of estimates' reliability. The relative probable error of estimate was approximately 3%.

For the estimation of the standard deviation of the error functions the effective value of n was approximately 500, whence:-

$$v_r = \sqrt{\frac{.57}{500}} = .03 \qquad \text{IV.8.}$$

Therefore the probable error was approximately 2%.

From IV.7. and IV.8. it may be seen that the probable error of the estimate of variance obtained by squaring the estimate of standard deviation was approximately 6% and 4% relative to the mean variance, for input and error functions respectively. However, it should be remembered that the operation of squaring caused the distribution of errors of estimate of variance to be non-Gaussian.

To investigate the effect of squaring the estimates of standard deviation, let us represent this estimate in the normalised form :-

$$s_n = (1 + x) \qquad \text{IV.9.}$$

where x is considered to be a zero mean Gaussian variate (approximately true for large n), with standard deviation = v_r.
Then we have :-

$$E\left[s_n^2\right] = E\left[(1 + x)^2\right] = 1 + v_r^2 \qquad \text{IV.10}$$

Expression IV.10. shows that bias is introduced by the squaring operation, but insertion of the values of v_r from IV.7. and IV.8. shows that this was extremely small in practice. The variance of the estimate of s^2 is given by :-

$$E\left[(s_n^2)^2\right] - E\left[s_n^2\right]^2 = E\left[(1 + x)^4\right] - (1 + v_r^2)^2 \qquad \text{IV.11.}$$
$$= 4.v_r^2 + 2.v_r^4$$

Insertion of values of v_r from IV.7. and IV.8. shows thet the above

variance was approximately $4.(.05)^2$ and $4.(.03)^2$ for input and error functions respectively ; the corresponding standard deviations were therefore .10 and .06 respectively.

The above analysis shows that the relative variation of the estimates of variance computed from integrated moduli were 10% and 7% for input and error functions respectively, and were subject to very small bias. These figures were considered quite satisfactory ; especially so because ratios were computed, and to the extent that variances were correlated, these ratios would show less relative variation.

IV.2. Estimation of Cross-Correlation from Integrated Moduli

As described in Section 5.2., a measure of correlation between operator and model errors was required for the purpose of model matching, which was biassed so as to take into account the fact that, for good matching, the variance of these errors should be comparable. A convenient measure was provided by estimating the variance of the difference signal d(t), formed by mutual subtraction of the model error, X(t) , and the operator error, Y(t) . Since the model was linear, the variance of this difference signal would be a minimum when the model matched all that part of the operator's error signal which was linearly coherent with the loop input. The procedure adopted was to define a regression line between Y and X having unit slope and zero intercept. The apparent coefficient of non-determination was then given by :-

$$\delta_{YX} = \delta_d^2/\delta_Y^2 \qquad ; \delta_d < \delta_Y \qquad \text{IV.12.}$$

Hence the (biassed) coefficient of correlation was given by :-

$$r_{YX} = \sqrt{(1 - \delta_d^2/\delta_Y^2)} \qquad \text{IV.13.}$$

It may be seen that r_{YX} could only be unity when $\delta_d = 0$, for which $\delta_X = \delta_Y$. It may be noted that the bias introduced is small for a good match between operator and model, as defined above.

The obvious way to estimate the unbiassed coefficient of correlation between errors would have been to use an analogue multiplier. A suitable multiplier was not available, however, but fortunately the true coefficient of correlation , r, could also be conveniently estimated by using the estimated variances of d, X, and Y, and noting that :-

$$\delta_d^2 = E\left[(Y - X)^2\right] = \delta_X^2 + \delta_Y^2 - 2.\delta_{YX} \qquad \text{IV.14.}$$

whence :-

$$r = \frac{1}{2}.\left\{\frac{\delta_X}{\delta_Y} + \frac{\delta_Y}{\delta_X} - 2.\frac{\delta_d^2}{\delta_Y \delta_X}\right\} \qquad \text{IV.15.}$$

The accuracy of estimation of r from IV.15. is poor for large difference between σ_Y and σ_X. However, for relatively small σ_d, and σ_X and σ_Y nearly equal, the accuracy is very good. The operation of subtracting X from Y then corresponds quite closely to estimating correlation by evaluating the variance of residuals from a regression line.

Expression IV.15. was used to estimate unbiassed correlation for the really good operator-model matches, where the above conditions were amply fulfilled. Moreover, an effectively large number (500) of samples were used to estimate σ_X, σ_Y, and σ_d. If anything, the estimated correlation was biassed slightly downward, because small differences in mean values of X and Y would be relatively larger compared to σ_d, thereby tending to give an overestimate of σ_d. A check was made by estimating cross-correlations from recordings of the error functions made simultaneously with analogue estimation. It was found that both methods gave practically identical results (within 2%) ; correlation estimated from recordings tended to be slightly higher.

APPENDIX V : Practical Realisation of the Elements of the Model

The theoretical diagram of pulse transfers of the sampled data model is shown in Fig.5.7.1. The purpose of the present Appendix is to enlarge on how these transfers were realised by practical circuit elements.

V.1. Generation of Velocity Signal

It may be seen that the theoretical circuit involved direct sampling of both position and velocity. To effect the latter it was necessary to differentiate the continuous error signal. A pure differentiation was not feasible, of course, and a compromise was required in the selection of a quasi-differentiator with a suitable time constant. It was found that a time constant of 1 ms. gave a satisfactory freedom from noise in the output signal, while retaining excellent accuracy for the purpose of obtaining samples of the velocity signal. The smallest possible noise was obtained by subtracting output velocity from input velocity directly, by using two identical capacitors (see Fig.AV.4.1.).

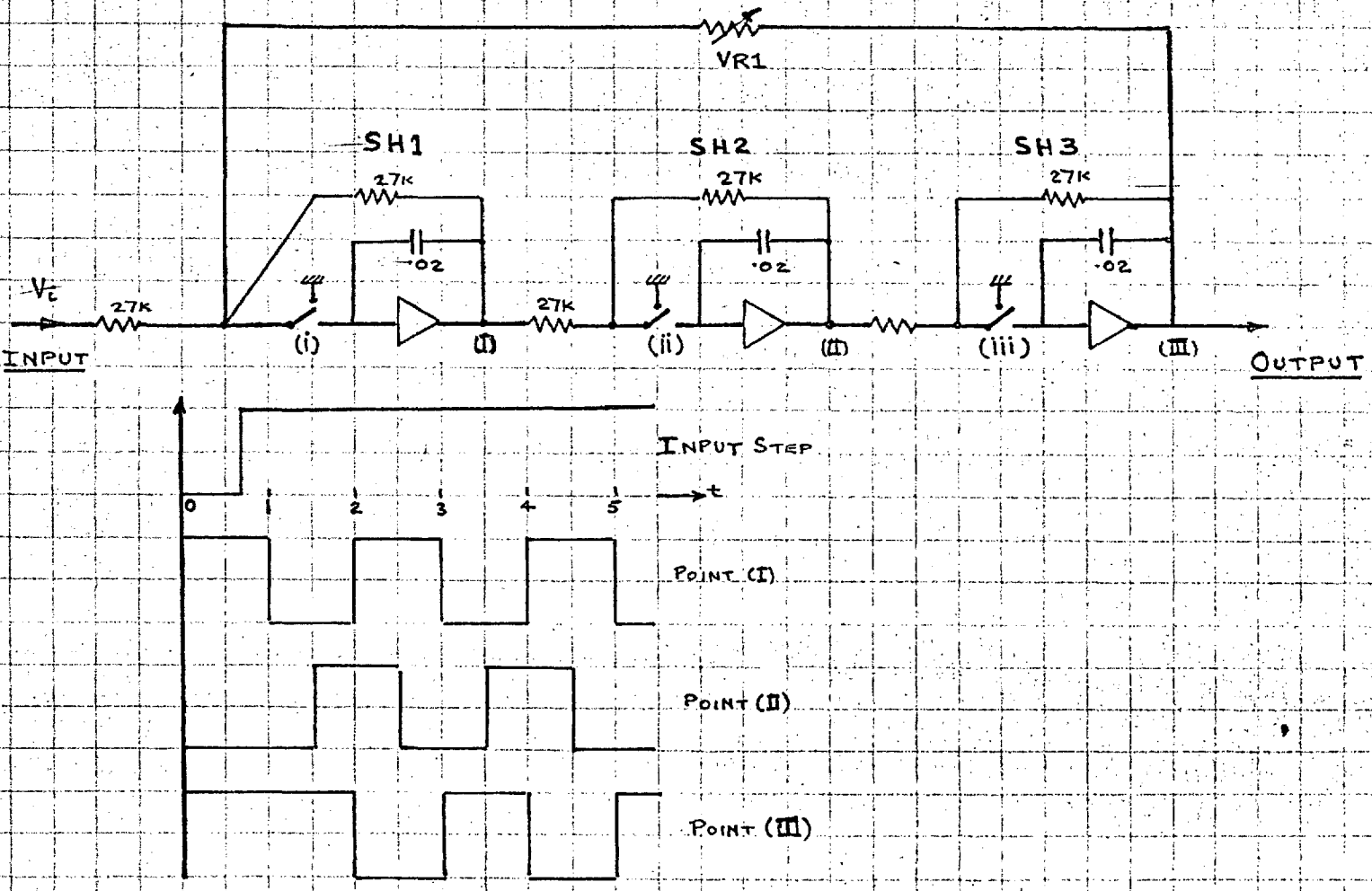
V.2. Velocity and Position Computation Loops

The theoretical circuit involved operations performed on impulses, the area of which was proportional to sampled values. Such operations could not be performed in practice, so that it was necessary either to approximate the impulses by pulses of similar areas but finite width and height, or work directly with sample and hold circuits. The form of the succeeding model transfers made it particularly convenient to work according to the latter method, so that it was necessary to design the computation loops to work in this manner.

The form of the computation loop circuit is shown in Fig.AV.2.1. The basic operation of the sample and hold circuits (SH1, SH2, SH3) has already been described in Section I.3. The operation of the relays was arranged so that relays (i) and (iii) closed at sampling instants, while relay (ii) closed at the midpoints of sampling intervals, so that an overall delay of one sample interval was obtained. The chief factor determining the minimum allowable closure time was the requirement that SH1 should charge satisfactorily after SH3 had completed its charging. Accordingly, relay closure time was set at 5 ms., which allowed plenty of time for the hold circuits to charge fully, since their time constant during sampling was only ½ ms. This closure time was such that functions being sampled changed by only a negligible amount during the sampling operation. Also, it was a small fraction (3%) of the sampling interval, and since each hold circuit was effectively followed by a double integration, it had a negligibly small effect on the model's output response.

The circuit's operation was such that at the end of each sampling

Fig. AV.2.1. Circuit of Velocity Computation Loop for Analogue Model



operation the output of SH1 was equal to the inverse of the input voltage, $-V_i$, minus a proportion of the output voltage determined by the potentiometer VR1, which controlled the overall loop gain. At the midpoint of the succeeding interval the output of SH2 became equal to the inverse of SH1 output. Then at the next sampling 'instant' SH3 output became equal to the inverse of SH2 output, thereby becoming equal to SH1 output, one sampling interval before.

The action of the circuit may be made clearer by considering its response, with loop gain set at unity, to an input step of height h occurring between sampling 'instants' 0 and 1 (see Fig,AV.2.1.). Assuming zero initial conditions, at time 1+ the output of SH1 is -h . At time 1½+ the output of SH2 is h , and therefore at time 2 the output of SH3 goes exponentially to -h , with time constant ½ ms. The output of SH3 is added directly to h at the input , so that the input to SH1 is an exponential decay toward zero on a time constant of ½ ms., running from the start of the sampling 'instant'. The output SH1 therefore goes toward zero at time 2+ rather more sluggishly than if presented with a step. However, the length of the sampling instant is sufficient for it to achieve a final value satisfactorily close to zero. With unity loop gain, the circuit goes on to generate a square wave of amplitude h at the output, as illustrated.

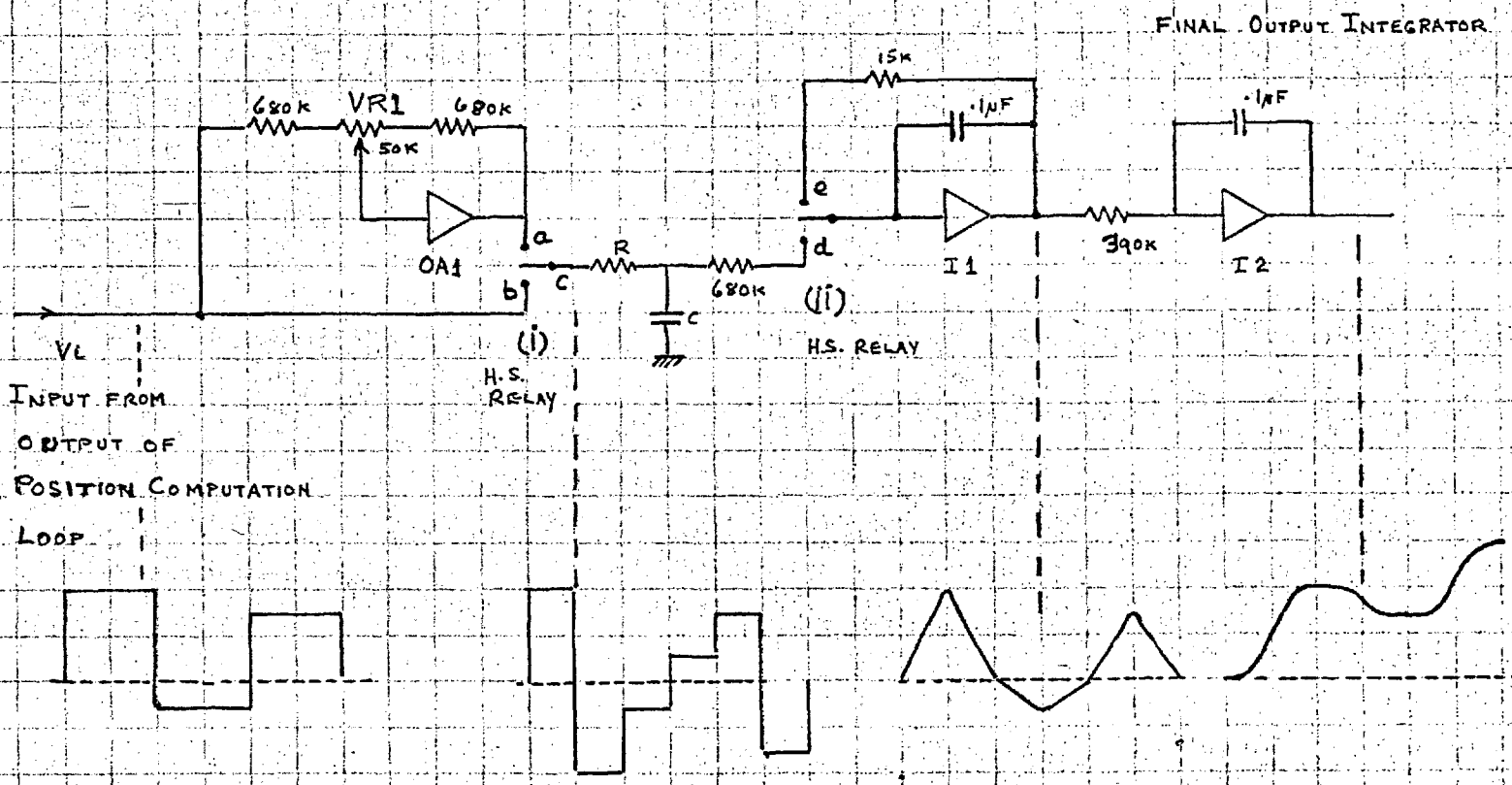
V.3. Generation of Velocity Triangle Response

The specifications for the velocity-triangle response generating circuit required that it should generate a velocity-triangle beginning and ending at sampling instants, and exhibiting a discontinuity of slope at the midpoint of the sampling interval, such that the area of the triangle would be proportional to the input voltage. The circuit designed to effect this action is shown in Fig.AV.3.1. Its operation was as follows.

The input voltage, V_i , was derived from the position computation loop, and therefore consisted of a step waveform, constant during sampling intervals. The voltages at points a and b were therefore approximately equal in magnitude, but opposite in sign, when potentiometer VR1 was correctly adjusted. Relay (i) was arranged to switch from a to b at sampling instants, and from b to a at the midpoints of sampling intervals. The voltage at point c was therefore in the form of a 'double-square' between sampling instants, as illustrated. The areas of each of the squares were proportional to V_i. The RC filter possessed a short time constant, and fulfilled the function of defining the input to the integrator I1 during switching. It also effectively delayed the changeover of the squarewave slightly, and rendered the resultant velocity triangle slightly more symmetrical - i.e. it corrected for the slight asymmetry caused by the finite length of sampling 'instants'.

Relay (ii) was arranged so as to be switched to d during sampling intervals, thereby causing the integrator to generate a

Fig. AV.3.1 Circuit For Generating Velocity Triangles



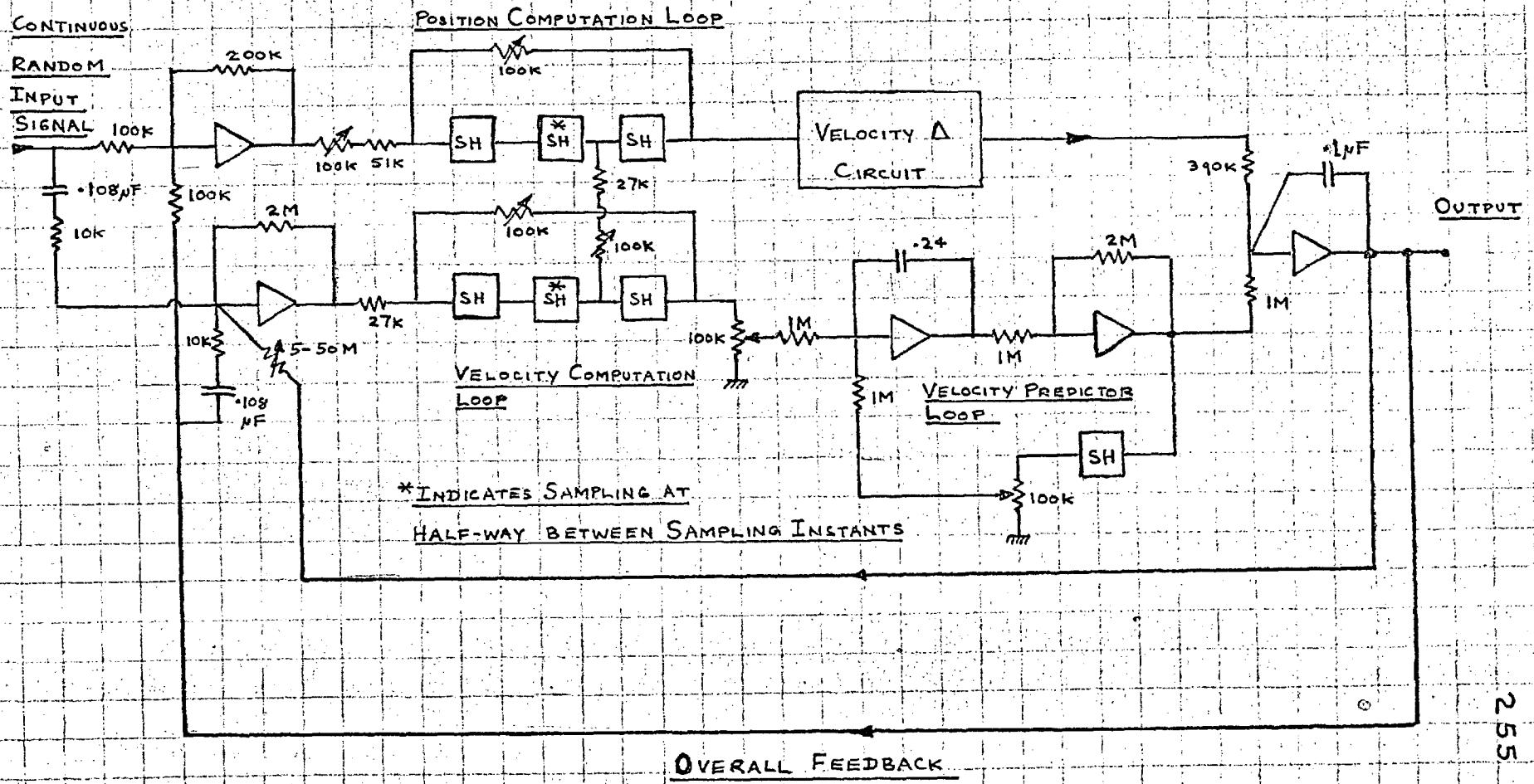
velocity-triangle according to the 'double square wave' input. During sampling 'instants' relay (ii) was switched to e ,which reset the integrator I1 ; i.e. it corrected any small velocity error at the end of the sampling interval. This resetting action was very necessary in practice,because errors of initial conditions and small errors of circuit operation would be cumulative,and lead to a standing velocity at sampling instants. The areas between velocity triangles and the zero axis would then be incorrect,and give a false response to V_i. Also,the spurious velocity output would lead to the sensing of a false velocity error. No matter how carefully VR1 was set to cause velocity triangles to end at exactly zero velocity,the above effects could not be avoided without a resetting action.

The action of the circuit was to produce velocity-triangles over each sampling interval,with areas proportional to the input voltage, V_i ,during the same sampling interval. These triangles exhibited a small section of zero velocity at sampling 'instants',corresponding to the duration of the sampling operation,so that they did not affect the sampling of velocity error ; this action was just that required to avoid spurious interaction between velocity and position corrections. The velocity-triangles generated were slightly shorter (3%) than the overall sampling interval,but this effect was easily allowed for when calculating the effective gain of the circuit. Final position output was obtained by integrating the velocity triangles,which led to a 'double-parabola' with zero velocity at sampling instants. The action of the circuit is analysed in Section VI.1.,in terms of modified Z-transforms.

V.4. Overall Analogue Model Circuit

The overall circuit of the analogue model is shown diagrammatically in Fig.AV.4.1. The features of chief interest have been described in the above Sections,and will not be further enlarged upon. The diagram is presented to illustrate the arrangement of analogue components to simulate the theoretical model structure.

FIG. AV. 4.1. CIRCUIT DIAGRAM OF ANALOGUE MODEL



APPENDIX VI : Forms and Stability of Model Z-transfers

VI.1. Transfer of the Velocity-Triangle Generating Circuit

The action of the velocity-triangle generating circuit is shown, in idealised form, in Fig.AVI.1.1. This action may be conveniently analysed in terms of the modified Z-transformation, as described in the following paragraphs.

It may be noted that the representation of the action of velocity-triangle generation requires the specification of two ranges of the parameter m , specifying times between sampling instants, because there is a discontinuity of slope at $m = \frac{1}{2}$. Representing the input at time nT as $V_i(n)$ and the output at time (n+m)T as $V_o(n+m)$, we may write the following equations :-

$$V_o(n+m) = V_o(n) + \tfrac{1}{2}(mT)^2.V_i(n) \qquad ;0 \leq m < \tfrac{1}{2} \qquad \text{VI.1.}$$

and,

$$\lim_{m \to 1} V_o(n+m) = V_o(n) + \tfrac{1}{2}T^2.V_i(n) \qquad \text{VI.2.}$$

whence:-

$$V_o(n+m) = V_o(n+1) - \tfrac{1}{2}(1-m)^2.T^2.V_i(n) \qquad ;\tfrac{1}{2} \leq m < 1 \qquad \text{VI.3.}$$

where $\frac{1}{2}(1-m)^2.T^2$ represents the portion of the response remaining uncompleted. Combining equations VI.2. and VI.3. we finally get :-

$$V_o(n+m) = V_o(n) + \tfrac{1}{2}T^2.\left\{1 - 2(1-m)^2\right\}.V_i(n) \qquad ;\tfrac{1}{2} \leq m < 1 \qquad \text{VI.4.}$$

Taking modified Z-transforms on both sides of equations VI.1. and VI.4., and assuming zero initial conditions, we get :-

$$z.V_o(z,m) = z.V_o(z,0) + \tfrac{1}{2}(mT)^2.V_i(z) \qquad ;0 \leq m < \tfrac{1}{2} \qquad \text{VI.5.}$$

and:-

$$z.V_o(z,m) = z.V_o(z,0) + \tfrac{1}{2}T^2.\left\{1 - 2(1-m)^2\right\}.V_i(z) \qquad ;\tfrac{1}{2} \leq m < 1 \qquad \text{VI.6.}$$

where we have used :- $z.V_i(z,0) = V_i(z)$

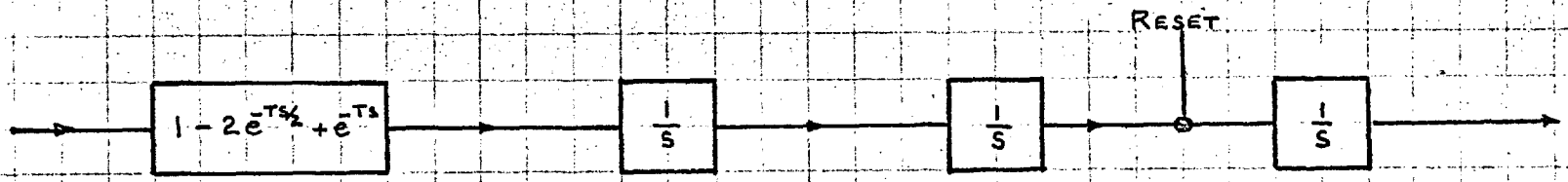
From equation VI.6. we obtain :-

$$\lim_{m \to 1} z.V_o(z,m) = z.V_o(z,0) + \tfrac{1}{2}T^2.V_i(z) = z^2.V_o(z,0) \qquad \text{VI.7.}$$

Whence we obtain :-

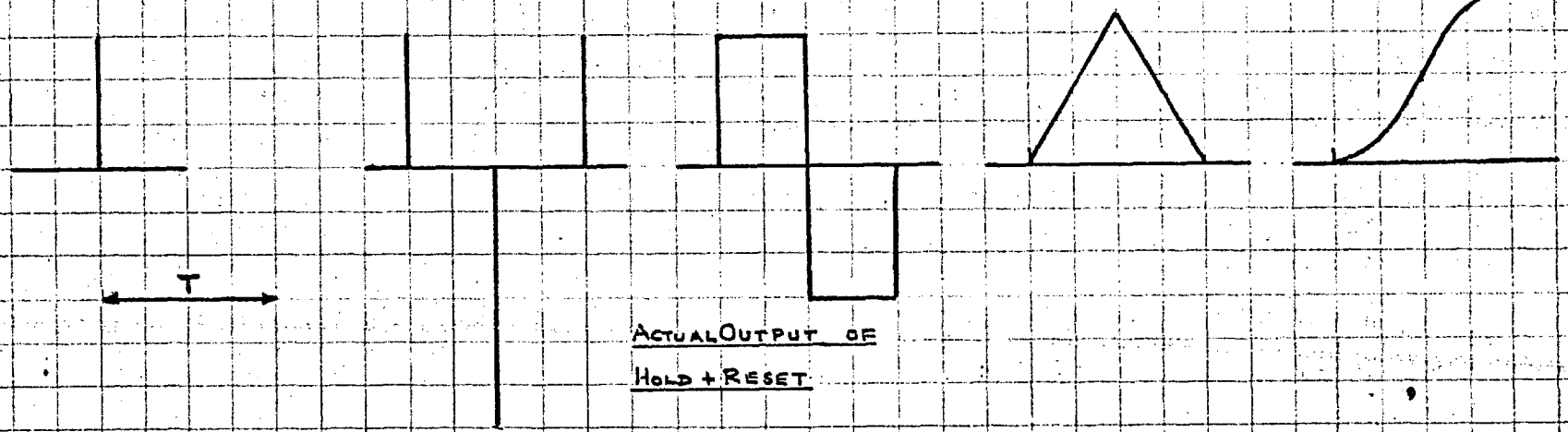
$$V_o(z,0) = \tfrac{1}{2}T^2.\frac{1}{z(z-1)}.V_i(z) \qquad \text{VI.8.}$$

FIG. A.VI.1.1. REPRESENTATION OF ACTION OF VELOCITY TRIANGLE GENERATOR



TRAIN OF IMPULSES REPRESENTING PROGRAMMED CHANGES OF POSITION OUTPUT

RESET ENSURES THAT TRIANGLE BEGINS AND ENDS AT ZERO VELOCITY



ACTUAL OUTPUT OF HOLD + RESET

INPUT IMPULSE (AREA SHOWN BY HEIGHT)

IMPULSES TO INTEGRATORS

ACCELERATION RECTANGLES

VELOCITY TRIANGLE HEIGHT $T/2$

OUTPUT PARABOLA HEIGHT $T^2/4$

Then substituting for $V_o(z,0)$ in equations VI.5. and VI.6. we get:-

$$V_o(z,m) = \tfrac{1}{4}T^2 . \frac{1}{z(z-1)} . \left\{1 + 2m^2(z-1)\right\} . V_i(z) \quad ; 0 \leq m < \tfrac{1}{2} \qquad \text{VI.9.}$$

and,

$$V_o(z,m) = \tfrac{1}{4}T^2 . \frac{1}{z(z-1)} . \left\{1 + (z-1)\left[1 - 2(1-m)^2\right]\right\} . V_i(z) \quad ; \tfrac{1}{2} \leq m < 1 \qquad \text{VI.10.}$$

The overall transfer may be represented as :-

$$P(z,m) = \tfrac{1}{4}T^2 . \frac{1}{z(z-1)} . \Lambda(z,m) \qquad \text{VI.11.}$$

where the function $\Lambda(z,m)$ assumes the appropriate form according to whether $0 \leq m < \tfrac{1}{2}$ or $\tfrac{1}{2} \leq m < 1$,as indicated in equations VI.9. and VI.10.

VI.2. Explicit Forms of the Model Transfers

Sections 6.1. and 6.2. have described the methods by which the Z-transfer characteristics of the model were evaluated. This Section presents explicit forms of the transfers in terms of model parameters. During the course of evaluation of the explicit model transfers,it was found convenient to write the products of the model parameters in the following shorthand form :-

$P_1 = g_s k_2$ $\qquad$ $P_4 = \tfrac{1}{4} g_d g_1 k_m k_1$ $\qquad$ $P_7 = P_1 P_4$

$P_2 = g_v k_d k_m k_2 k_3$ $\qquad$ $P_5 = \tfrac{1}{2} g_d g_v k_m k_2 k_3$ $\qquad$ $P_8 = P_2 P_3$

$P_3 = \tfrac{1}{4} g_p k_m k_1$ $\qquad$ $P_6 = P_1 P_3$ $\qquad$ $P_9 = \tfrac{1}{4} g_1 k_1 k_d k_m k_1$

where the g and k parameters are shown in Fig.6.1.1. It may be noted that $P_2 P_4 = P_5 P_9$.

With the above substitutions the elements of matrix $[A(z,m)]$ become :-

$$a_{pp}(z,m) = P_3 T^2 . \frac{\Lambda(z,m)}{(z+1)(z-1)}$$

$$a_{pv}(z,m) = P_9 T^2 . \frac{\Lambda(z,m)}{(z-1)(z+1)^2} + \tfrac{1}{2} P_2 T^2 . \left\{ \frac{m^2(z-1)^2 + (2m+1)(z-1) + 2}{z(z+1)(z-1)^2} \right\}$$

$$a_{vp}(z,m) = 0$$

$$a_{vv}(z,m) = P_2 T . \frac{1}{z(z+1)} . \left\{ m + \frac{1}{z-1} \right\}$$

The elements of matrix $[B(z,m)]$ become :-

$$b_{pp}(z,m) = -p_4T^2.\frac{\Delta(z,m)}{(z-1)(z+1)^2} - p_5T^2.\left\{\frac{m^2(z-1)^2+(2m+1)(z-1)+2}{z(z+1)(z-1)^2}\right\}$$

$$b_{pv}(z,m) = -\tfrac{1}{2}p_1T^2.\left\{\frac{m^2(z-1)^2+(2m+1)(z-1)+2}{z(z-1)^2}\right\}$$

$$b_{vp}(z,m) = -2p_5T.\left\{\frac{m(z-1)+1}{z(z+1)(z-1)}\right\}$$

$$b_{vv}(z,m) = -p_1T.\left\{\frac{m(z-1)+1}{z(z-1)}\right\}$$

Considering the limits as $m \to 1$ in the above elements, the following relations result :-

$$[A(z,1)] = \begin{bmatrix} p_3T^2.\frac{1}{(z+1)(z-1)} & ; & p_9T^2.\frac{z}{(z-1)(z+1)^2}\cdot\frac{z}{} + \tfrac{1}{2}p_2T^2.\frac{1}{(z-1)^2} \\ 0 & ; & p_2T.\frac{1}{(z-1)(z+1)} \end{bmatrix}$$

$$[B(z,1)] = \begin{bmatrix} -p_4T^2.\frac{z}{(z-1)(z+1)^2} - p_5T^2.\frac{1}{(z-1)^2} & ; & -\tfrac{1}{2}p_1T^2\frac{(z+1)}{(z-1)^2} \\ -2p_5T.\frac{1}{(z+1)(z-1)} & ; & -p_1T.\frac{1}{(z-1)} \end{bmatrix}$$

Execution of the required manipulations of $A(z,1)$ and $B(z,1)$ shows that the actual open loop Z-transfer of the model is given explicitly by :-

$$h(z) = z^4 + h_3.z^3 + h_2.z^2 + h_1.z + h_0$$

where $h_3 = T.p_1$; $h_2 = \{-2 + T^2.p_1 + T^2(p_4 + p_5)\}$

$h_1 = \{-T.p_1 + T^2(-p_4 + 2p_5) + p_1p_4.T^3\}$; $h_0 = \{1 - T.p_1 + T^2.p_5\}$

$$H(z,1) = \begin{bmatrix} h_{pp}(z,1) & ; & h_{pv}(z,1) \\ h_{vp}(z,1) & ; & h_{vv}(z,1) \end{bmatrix}$$

where :-

$$h_{pp}(z,1) = p_3T^2(z - 1 + T.p_1)(z + 1)$$

$$h_{pv}(z,1) = z^2.\left\{T^2(\tfrac{1}{2}p_2 + p_9)\right\} + z.\left\{T^2(p_2 - p_9) + T^3(p_1p_9)\right\} + \tfrac{1}{2}p_2T^2$$

$$h_{vp}(z,1) = - 2p_3p_5T^3$$

$$h_{vv}(z,1) = T.p_2(z - 1)(z + 1)$$

and the closed loop (input-output) Z-transfer is given explicitly by :-

$$q(z) = z^4 + q_3.z^3 + q_2.z^2 + q_1.z + q_0$$

where $q_3 = T.p_1$; $q_2 = \left\{-2 + T(p_1 + p_2) + T^2(p_3 + p_4 + p_5)\right\}$

$$q_1 = \left\{- T.p_1 + T^2(-p_4 + 2p_5) + T^3(p_6 + p_7)\right\}$$

$$q_0 = \left\{1 + T.(-p_1 - p_2) + T^2(-p_3 + p_5) + T^3(p_6 + p_8)\right\}$$

and :-

$$\left[Q(z,1)\right] = \begin{bmatrix} q_{pp}(z,1) & ; & q_{pv}(z,1) \\ q_{vp}(z,1) & ; & q_{vv}(z,1) \end{bmatrix}$$

where :-

$$q_{pp}(z,1) = z^2.\left\{T^2.p_3\right\} + z.\left\{T^3.p_1p_3\right\} + \left\{T^2.p_3\left[-1 + T(p_1 + p_2)\right]\right\}$$

$$q_{pv}(z,1) = z^2.\left\{T^2.(p_9 + \tfrac{1}{2}p_2)\right\} + z.\left\{T^2(p_2 - p_9) + T^3.p_1p_9\right\} + \tfrac{1}{2}T^2.p_2$$

$$q_{vp}(z,1) = - 2p_3p_5T^3$$

$$q_{vv}(z,1) = T.p_2(z^2 - 1 + T^2.p_3)$$

The closed loop input-error transfer is given by q(z) as above, combined with the matrix :-

$$\left[q(z).I - Q(z,1)\right]$$

as shown in equation 6.2.11.

VI.3. Qualitative Examination of Closed-Loop Stability of Model

As shown in Section 6.1., when examining the stability of the closed loop model transfer, it is sufficient to consider the output vector at sampling instants only. Therefore the closed loop stability of the model may be investigated by considering the polynomial $q(z)$, which contains zeros corresponding to all the poles of the closed loop model response at sampling intervals. If a qualitative estimate of stability is all that is required, then the problem reduces to that of investigating the magnitude of the dominant zero (or zeros, if a complex conjugate pair) of $q(z)$. Several methods are available, as described in the following paragraphs.

The Companion Matrix Method involves the formation of a companion matrix, $[M]$, to the polynomial $q(z)$; $[M]$ is the matrix satisfying :-

$$q([M]) = [0]$$

(by the Cayley-Hamilton theorem every matrix satisfies its own characteristic polynomial)

In general there are an infinite number of matrices which share $q(z)$ as their characteristic polynomial, since if $[M]$ is such a matrix, then so is $[PMP^{-1}]$, where P is any nonsingular matrix. It is convenient to write $[M]$ as :-

$$[M] = \begin{bmatrix} -q_3 & -q_2 & -q_1 & -q_0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

Now the companion matrix has eigenvalues which are equal to the roots of the characteristic polynomial, $q(z)$. If the zeros of $q(z)$ are contained within the circle $|z| = 1$, then $|\lambda(\max)| < 1$, where $|\lambda(\max)|$ is the modulus of the largest eigenvalue of $[M]$. Moreover, it is known that if $|\lambda(\max)| < 1$, then :-

$$\lim_{n \to \infty} [M]^n = [0]$$

Therefore, to test whether $q(z)$ has zeros within the unit circle we may form,

$[M]^2$, $[M]^4$, $[M]^8$, etc., and check that the elements of the resultant matrices tend to zero as $n \to \infty$. The rapidity with which this process occurs gives an idea of the size of $|\lambda(\max)|$.

The Bilinear Transformation,

$$z = \frac{w + 1}{w - 1}$$

, maps the interior of the unit circle $|z| = 1$ into the left half of the w-plane ; i.e. the point set $|z| \leq 1$ maps into the point set $\mathrm{Re}(w) \leq 0$, and $q(z) \to p(w)$.

The Routh-Hurwitz criterion may now be applied directly in the w-plane. This will directly show any singularities of the polynomial p(w) in the right-half w-plane ,and hence of q(z) outside |z|= 1.

An alternative method for determining the number of distinct zeros in the w-plane to the right of the imaginary axis (not on the axis) is provided by constructing a Sturm sequence as follows. Putting x = jw (so that as w traverses the imaginary axis,x traverses the real axis), the polynomial r(x) = j.p(jw) is formed. A Sturm sequence for r(x) is then constructed,and it is noted that the first two members of this sequence are R(x) and I(x),where ,

$$r(x) = R(x) + j.I(x)$$

The remaining terms of the sequence may be obtained by a process of successive division. If $f_k(x)$ is the k.th member of the sequence and,

$q_k(x) = f_k(x)/f_{k+1}(x)$, then $f_{k+1}(x)$ may be found from :-

$$f_{k+1}(x) = q_{k-1}f_k(x) - f_{k-1}(x)$$

Since $f_1(x)$ and $f_2(x)$ are as specified above,the entire sequence may be derived . There are generally n+1 members of the sequence for a polynomial of degree n ; if there are less,then the terms calculated as above contain a common factor which must be divided out.
Then the function $V(x_o)$ is defined as equal to the number of sign changes passing from left to right along the vector,

$$\{sgn.f_1(x_o);sgn.f_2(x_o);sgn.f_3(x_o);.....;sgn.f_{n+1}(x_o)\}$$

Finally,a theorem due to Routh is used ; this states that P ,the number of zeros in the upper half x-plane (right half w-plane) is given by

$$P = \tfrac{1}{2}\{n + V(\infty) - V(-\infty)\}$$

where n is the degree of the polynomial r(x),or p(w).
The above method is,in fact,the basis of the derivation of the Routh-Hurwitz determinants, and possesses advantages for large problems requiring the use of computing machinery.

VI.4. Evaluation of the Roots of h(z) and q(z)

The roots of the polynomials q(z) and h(z) were generally obtained by use of the Newton-Raphson iteration. An initial guess at a root z_o was made ,and then refined by the well known formula,

$$z_1 = z_o - \frac{q(z_o)}{q'(z_o)}$$

The method possessed the advantage of rapid convergence, provided initial convergence can be obtained ; no practical difficulty was experienced in this respect.

An alternative method of estimating the roots of a polynomial is that of Bernouilli ; this method leads to an estimate of the dominant root (or roots if there is a dominant complex conjugate pair). A real dominant root z_1 is given by :-

$$z_1 = \lim_{k\to\infty} \frac{S_{k+1}}{S_k}$$

and dominant complex conjugate roots of modulus $|z_1|$ and $\arg \pm\phi$ are given by :-

$$2|z_1|.\cos\phi = \lim_{k\to\infty} \frac{S_{k-1}S_{k+2} - S_kS_{k+1}}{S_{k+1}S_{k-1} - S_k^2}$$

$$|z_1|^2 = \lim_{k\to\infty} \frac{S_kS_{k+2} - S_{k+1}^2}{S_{k+1}S_{k-1} - S_k^2}$$

where S_k in the above expressions represents the k.th Newton sum, and is defined as :-

$$S_k = \sum_{v=1}^{n} z_v^k \quad , \text{ where the } z_v \text{ are the roots of the polynomial of order n.}$$

Newton first showed that the value of S_k could be obtained from the coefficients of the polynomial, without directly evaluating its roots. For example, in the case of the polynomial $q(z)$, which was of the fourth order, S_k was given by :-

$$S_o = 4 \; ; \; S_1 = -q_3 \; ; \; S_2 = q_3^2 - 2q_2 \; ; \; S_3 = 3(q_3q_2 - q_1) - q_3^3$$

and, for $k \geq 4$, S_k was given by the relation :-

$$S_k + q_3S_{k-1} + q_2S_{k-2} + q_1S_{k-3} + q_0S_{k-4} = 0$$

Bernouilli's method was used to evaluate the roots of $h(z)$ for operator D, which were all complex. It was found to be very lengthy and subject to trouble from rounding error. It was concluded that an initial guess made from examination of the behaviour of the polynomial along real and imaginary axes, followed by a search procedure in the complex plane, represented an easier method for hand calculation.

APPENDIX VII : Optimum Linear Filters Including Pure Delay

As discussed in Sections 9.4. and 3.2.,the operator's overall loop transfer could be represented in the form of a pure delay ,T_d, cascaded with a linear continuous transfer having the form of a rational polynomial,or ratio of such polynomials,in the Laplace operator s. It was required to derive the form of this latter transfer which would give least mean square error between loop input and output. The input function could be considered as being drawn from a zero-mean,stationary ergodic random process,so that the optimum transfer was given by the optimum Wiener filter for prediction time T_d. It was convenient to derive this optimum transfer according to the method given by Truxal (25),since this facilitated simultaneous calculation of the relative mean square error of prediction.

We first consider the case of an input whose spectral density is given by:-

$$^{2}\Phi_{rr}(s) = \frac{1}{(s + a)^2.(a - s)^2} \qquad \text{VII.1.}$$

The required continuous transfer is considered to be composed of two parts,so that:-

$$G(s) = G_a(s).G_b(s) \qquad \text{VII.2.}$$

$G_a(s)$ is selected so as to cancel out the left half plane poles and zeros of $^{2}\Phi_{rr}(s)$,so that:-

$$G_a(s) = (s + a)^2 \qquad \text{VII.3.}$$

This initial part of the transfer G(s) thus converts the input spectral density to the form of white noise. For convenience,this white noise is considered to be composed of a sequence of perfectly random pulses.

The best form of the transfer $G_b(s)$,without regard to physical realisability,is given by:-

$$npr G_b(s) = \frac{e^{sT_d}}{(s + a)^2} \qquad \text{VII.4.}$$

so that the overall transfer would represent a perfect prediction. The non physically realisable impulse response corresponding to this $npr G_b(s)$ is given by inverse transforming equation VII.4.:-.

$$npr g_b(t) = \frac{1}{2\pi j} \int_{-j\infty}^{j\infty} \frac{e^{s(t + T_d)}}{(s + a)^2} ds \qquad \text{VII.5.}$$

By carrying out the complex integration, the non physical impulse response may be finally derived as :-

$$_{npr}g_b(t) = (t + T_d).e^{-a(t+T_d)} \qquad ; t \geqslant -T \qquad \text{VII.6.}$$

But physical realisability requires that the optimum physical transfer must have no response before $t = 0$. Since the output of $G_a(s)$ is considered as composed of a number of uncorrelated pulses, the response to each pulse may be considered separately. Then it may be seen that the optimum physically realisable impulse response, $g_b(t)$, is that which approaches most closely to the physically unrealisable response for $t \geqslant 0$. Therefore :-

$$g_b(t) = (t + T_d).e^{-a(t+T_d)} \qquad ; t \geqslant 0 \qquad \text{VII.7.}$$

$g_b(t)$ may be transformed to obtain the corresponding Laplace transfer:-

$$G_b(s) = \frac{\{(1 + aT_d) + sT_d\}}{(s + a)^2}.e^{-aT_d} \qquad \text{VII.8.}$$

Whence the overall transfer is given :-

$$G(s) = \{(1 + aT_d) + sT_d\}.e^{-aT_d} \qquad \text{VII.9.}$$

To evaluate the mean square error as a proportion of the input variance, it is most convenient to consider impulse responses. Now the output may be considered as composed of a sum of independent impulse responses, all of the same shape. Therefore the mean output power, for the physically unrealisable filter, is given by :-

$$_{npr}W = k.\int_{-T_d}^{\infty} {}_{npr}g_b^2(t).dt \qquad \text{VII.10.}$$

and for the physically realisable filter, by :-

$$W = k.\int_{o}^{\infty} g_b^2(t).dt \qquad \text{VII.11.}$$

Now the physically unrealisable filter has a mean square error of zero, i.e. output = input. The difference between the output powers of the realisable and unrealisable filters represents the error power of the optimum filter. Therefore the relative mean square error is given by :-

$$rmse = ({}_{npr}W - W)/{}_{npr}W \qquad \text{VII.12.}$$

By considering equations VII.6.,7,10,& 11.,it may be seen that the relative mean square error is given by :-

$$\text{rmse} = \frac{\int_0^{T_d} \text{npr} g_b^2(\delta).d\delta}{\int_0^{\infty} \text{npr} g_b^2(\delta).d\delta} \quad ; \delta = t + T_d \qquad \text{VII.13.}$$

The following expression results from carrying out the above integrations :-

$$\text{rmse} = 1 - \tfrac{1}{2}(2 + 4aT_d + 4a^2T_d^2).e^{-2aT_d} \qquad \text{VII.14.}$$

Carrying through the above analysis for the case of an input spectrum with third order cut-off gives the following relations :-

$$^3\Phi_{rr}(s) = \frac{1}{(s + a)^3.(a - s)^3} \qquad \text{VII.15.}$$

$$G(s) = \left\{1 + (s+a)T_d + \frac{(s+a)^2T_d^2}{2!}\right\}. e^{-aT_d} \qquad \text{VII.16.}$$

$$\text{rmse} = 1 - \tfrac{1}{3}(3 + 6aT_d + 6a^2T_d^2 + 4a^3T_d^3 + 2a^4T_d^4).e^{-2aT_d} \qquad \text{VII.17.}$$

Substitution of practical values for T_d and a (Section 9.4.) shows that the relative mean square error is reduced from .32 to .05 in going from a second to a third order cut-off spectrum.

It is interesting to note that the form of G(s) for an n th. order cut-off spectrum of the type shown in equation VII.15. is given by the following equation :-

$$G(s) = e^{-aT_d}.\left\{e^{(s+a)T_d}\right\}_{n \text{ terms}} \qquad \text{VII.18.}$$

where the last terms on the R.H.S. represent an expansion truncated to n terms. Thus, as the order of spectral cut-off increases, the G(s) function approaches to a perfect prediction. This could not be realised in practice, because the absence of noise could not be assumed when considering high order differentiations.

APPENDIX VIII : Distortion of Error-Output Statistical Functions Due to Recirculation of Remnant

To investigate the qualitative effect of the recirculation of the remnant term on the statistical functions relating operator error and output, we may represent the operator loop as in Fig.VIII.1a., where loop variates are also defined. For the purposes of this analysis, these loop variates are assumed to be drawn from stationary, ergodic random processes.

Working in the frequency domain, we may derive the following equations:-

$$D(s) = G(s).E(s) \qquad \text{VIII.1.}$$

and

$$C(s) = D(s) + N(s) = G(s).E(s) + N(s) \qquad \text{VIII.2.}$$

Multiplying both sides of equation VIII.2. by $E^*(s)$, where * represents the operation of taking the complex conjugate, we get:-

$$E^*(s).C(s) = E^*(s).G(s).E(s) + E^*(s).N(s) \qquad \text{VIII.3.}$$

Whence it may be seen that:-

$$\Phi_{ec}(s) = G(s).\Phi_{ee}(s) + \Phi_{en}(s) \qquad \text{VIII.4.}$$

By considering the relations between loop variates, it may be seen that:-

$$E(s) = \frac{1}{1 + G(s)}.\left[R(s) - N(s)\right] \qquad \text{VIII.5.}$$

By definition, the remnant, $N(s)$, is not linearly coherent with the input, $R(s)$. Therefore it may be seen that:-

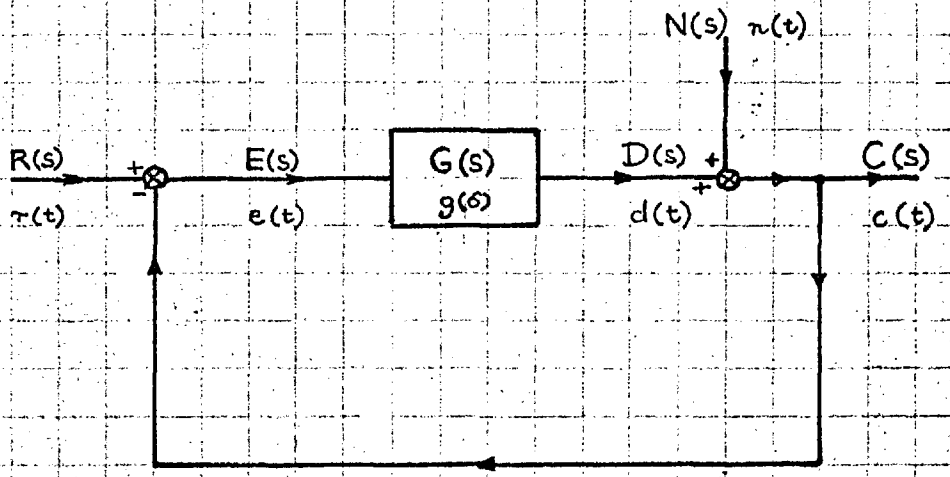
$$\Phi_{en}(s) = \frac{-1}{1 + G^*(s)}.\Phi_{nn}(s) \qquad \text{VIII.6.}$$

Combining equations VIII.4. and VIII.6.:-

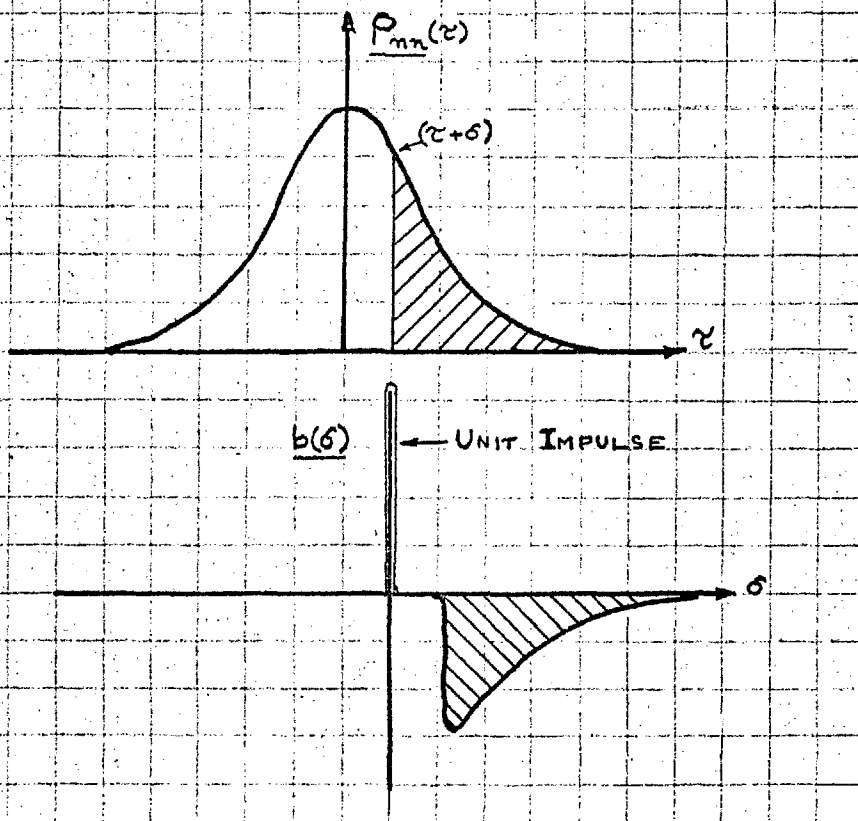
$$\Phi_{ec}(s) = G(s).\Phi_{ee}(s) + \frac{-1}{1 + G^*(s)}.\Phi_{nn}(s) \qquad \text{VIII.7.}$$

The above equation shows how the presence of the remnant term distorts the error-output cross power spectral density function. The effect of the remnant term on time domain functions may be evaluated by taking inverse transforms on both sides of equation VIII.7., and specifying $g(\sigma)$ and $b(\sigma)$ as the weighting functions corresponding to the transfers $G(s)$ and $\frac{1}{1 + G(s)}$, respectively.

FIG. VIII.1. CONTINUOUS SYSTEM WITH ADDED NOISE



(a) REPRESENTATION OF OPERATOR IN TRACKING LOOP



(b) REPRESENTATION OF CONVOLUTION OF $P_{nn}(\tau)$ AND $b(\sigma)$

Thus, the following equation holds in the time domain:-

$$\rho_{ec}(\tau) = \int_{-\infty}^{\infty} g(\sigma).\rho_{ee}(\tau-\sigma).d\sigma - \int_{-\infty}^{\infty} b(-\sigma).\rho_{nn}(\tau-\sigma).d\sigma \quad \text{VIII.8.}$$

By manipulation of the dummy time variable, σ, in the last term of the R.H.S. of the above equation we may derive:-

$$\rho_{ec}(\tau) = \int_{-\infty}^{\infty} g(\sigma).\rho_{ee}(\tau-\sigma).d\sigma - \int_{-\infty}^{\infty} b(\sigma).\rho_{nn}(\tau+\sigma).d\sigma \quad \text{VIII.9.}$$

The last term on the R.H.S. of this equation represents a distortion of the error-output cross-correlation due to the presence of remnant. The effect of this term may be qualitatively appreciated by considering the following reasoning.

In the present case the transfer G(s) is considered to be physically realisable, so that $b(\sigma) = 0$ for all $\sigma < 0$. Thus the convolution integral representing distortion may be considered as:-

$$\int_{0}^{\infty} b(\sigma).\rho_{nn}(\tau+\sigma).d\sigma$$

This convolution is represented graphically by the shaded areas of the functions shown graphically in Fig.VIII.1b., the shapes of which have been selected as being fairly representative of functions encountered in practice. It may be seen that the degree of distortion caused by the remnant is greatest at small negative values of τ and becomes rapidly smaller as τ increases.

APPENDIX IX : Spectral Analysis and Model Fitting to Time Domain Data

IX.1. Causation of Difficulties of Model Fitting in the Spectral Domain

The difficulties of model fitting in the spectral domain have been enumerated in Section 8.5. The manner in which these difficulties arise may be seen by reference to Fig.IX.1.1. Here, the process of spectral analysis is represented in terms of determining the weights to be attached to the outputs of a finite number of filters, whose frequency characteristics correspond to the particular lag or spectral window used in the analysis. The covariance matrix of the filter outputs is used to determine the weighting factors (b's) as indicated in the figure. If the outputs of the filters can be considered as orthogonal to each other (i.e. if the record length is long enough), then all matrix elements will be approximately zero, except those on the diagonal. These conditions are approximated in spectral analysis as usually performed, using very long lengths of record. Therefore only small error is incurred by ignoring the non-diagonal elements, as implicit in spectral analysis. However, with relatively short record lengths, as possessed by open loop test data, the non-diagonal elements are of comparable magnitude to those on the diagonal, and they cannot be ignored without incurring serious error.

IX.2. Model Fitting by Direct Use of Time Domain Data

Direct fitting of models to time domain data, in the form of covariance functions, was based on the use of equation 8.5.2. (q.v.) The first step in the fitting procedure was to represent the effective model Z-transfer, G(z), in terms of numerator and denominator polynomials :-

$$G(z) = \frac{N(z)}{D(z)} = \frac{n_0 + n_1 z + n_2 z^2 + n_3 z^3 + \text{etc.}}{1 + d_1 z + d_2 z^2 + d_3 z^3 + \text{etc.}} \qquad \text{IX.1.}$$

whence the following equation was derived :-

$$D(z) \cdot \underline{\Phi}_{eo}(z) = N(z) \cdot \underline{\Phi}_{ee}(z) \qquad \text{IX.2.}$$

Now, the nature of the Z-transformation is such that :-

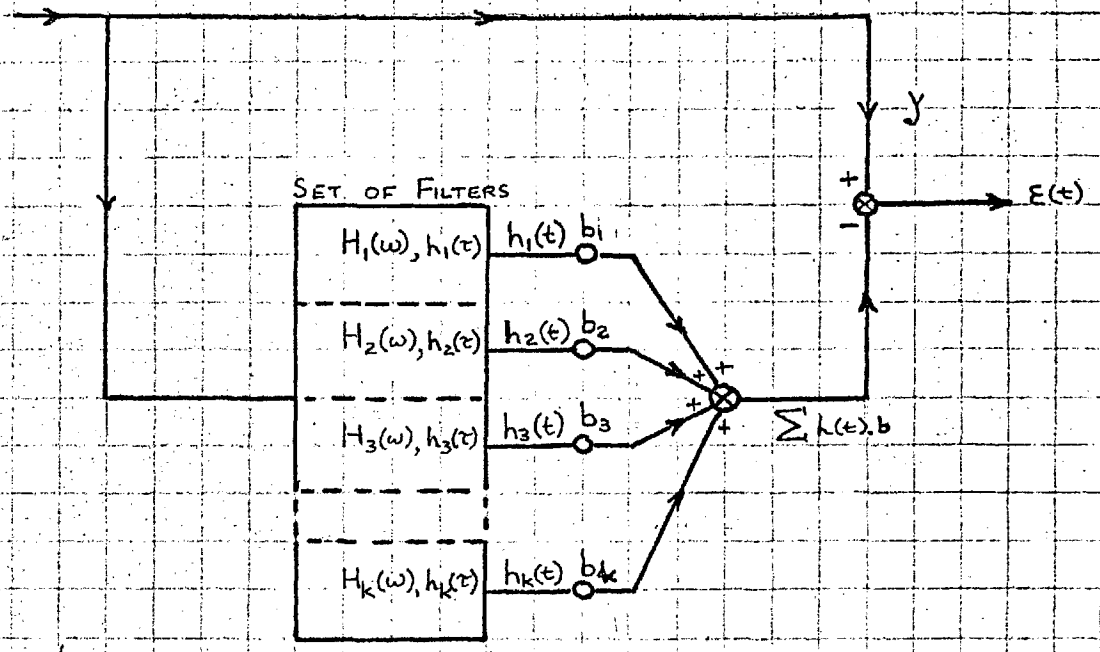
$$\underline{\Phi}_{eo}(z) = \sum_{n=-\infty}^{\infty} \delta_{eo}(n\Upsilon) \cdot z^{-n} \qquad \text{IX.3.}$$

where $\delta_{eo}(n\Upsilon)$ is the value of the error output covariance function

FIG. IX.1.1.

REPRESENTATION OF SPECTRAL ANALYSIS IN TERMS OF WEIGHTING A SET OF FILTER OUTPUTS

SIGNAL TO BE ANALYSED



EQUATION TO DETERMINE WEIGHTING FACTORS, b_1 TO b_k FROM $\min_b \overline{\varepsilon(t)^2}$

$$\begin{bmatrix} \overline{h_1^2} & \overline{h_1 h_2} & \overline{h_1 h_3} & \cdots & \overline{h_1 h_k} \\ \overline{h_2 h_1} & \overline{h_2^2} & \overline{h_2 h_3} & & \\ \overline{h_3 h_1} & \overline{h_3 h_2} & \overline{h_3^2} & & \\ \vdots & & & \ddots & \\ \overline{h_k h_1} & \cdots & & & \overline{h_k^2} \end{bmatrix} \cdot \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_k \end{bmatrix} = \begin{bmatrix} \overline{h_1 y} \\ \overline{h_2 y} \\ \overline{h_3 y} \\ \vdots \\ \overline{h_k y} \end{bmatrix}$$

— INDICATES SAMPLE AVERAGE

at lag $n\Upsilon$. A similar relation holds for $\Phi_{ee}(z)$. Therefore, equating the coefficients of like powers of z on each side of IX.2. results in the following equation :-

$$\sigma_{eo}(n\Upsilon) + d_1.\sigma_{eo}(\overline{n+1}.\Upsilon) + d_2.\sigma_{eo}(\overline{n+2}.\Upsilon) + \text{etc.}$$
$$= n_o.\sigma_{ee}(n\Upsilon) + n_1.\sigma_{ee}(\overline{n+1}.\Upsilon) + n_2.\sigma_{ee}(\overline{n+2}.\Upsilon) + \text{etc.} \qquad \text{IX.4.}$$

Substitution of the available covariance data in equation IX.4. leads to an overdetermined set of equations for the coefficients of n and d . One may thus regard the problem of estimating these coefficients as one of estimating partial regression coefficients from a linear multiple regression analysis on the covariance function data. Knowledge of the values of the coefficients may then be used to calculate the parameters of the equivalent continuous transfer. The analysis may be extended to cover the case where G(z) contains a delay not equal to an integral multiple of the data sampling interval, Υ, by writing the delay in the form $z^{-(k+m)}$, where m is fractional. It is then necessary to substitute for n in the left hand side of equation IX.4., and some interpolation must be used to obtain values of the error covariance function between data sampling points.

A modification of the above reasoning allows the direct fitting of a continuous model. If one postulates a model transfer function of the form :-

$$G(s) = \left\{\frac{a_o + a_1 s + a_2 s^2 + a_3 s^3 + \text{etc.}}{1 + b_1 s + b_2 s^2 + b_3 s^3 + \text{etc.}}\right\} e^{-(k+m)\Upsilon . s} \qquad \text{IX.5.}$$

where k is an integer, and m fractional, an equation involving the covariance functions and their derivatives may be obtained as follows :-

$$\left\{(1 + b_1.\frac{\partial}{\partial\tau} + b_2.\frac{\partial^2}{\partial\tau^2} + b_3.\frac{\partial^3}{\partial\tau^3} + \ldots\text{etc.}).\sigma_{eo}(\tau)\right\}_{\tau = n\Upsilon}$$
$$= \left\{a_o + a_1.\frac{\partial}{\partial\tau} + a_2.\frac{\partial^2}{\partial\tau^2} + a_3.\frac{\partial^3}{\partial\tau^3} + \ldots\text{etc}).\sigma_{ee}(\tau)\right\}_{\tau = (n+m)\Upsilon}$$

Substitution of sets of values for the covariances and their derivatives then gives regression equations to determine the values of the a's and b's. It may be appreciated that, without direct knowledge of the derivatives of covariance functions, there are severe limitations on the possible complexity of G(s) ; even the estimation of first derivatives of the covariance functions will generally introduce error into the regression data.

APPENDIX X : Calculation of Probability Density Functions Associated with Random Sampling

Section 9.1. describes the basis on which histograms of the operators' sampling intervals were obtained. From these observed histograms, it was required to calculate the corresponding probability density functions of the pure delay of operators' responses to step inputs, assuming the hypothesis of action described in Section 4.10. to be correct.

The first step was to approximate the observed histograms of sampling interval, I, by a continuous probability density function, p(I), defined to exist over a range of intervals, $a \leqslant I \leqslant b$, and to be zero outside this range. According to the aforementioned hypothesis, the total pure delay, L, observed in operators' step responses, has two components:-

(1) A delay of D, due to the step input not occurring at a sampling instant.

(2) A delay of I_s, equal to the length of the succeeding sampling interval.

Thus the total pure delay is given by:-

$$L = D + I_s \qquad \text{X.1.}$$

To calculate the probability density function of L, r(L), we theoretically need to calculate the joint probability density function of D and I_s. However, the problem is simplified if it may be assumed that there is no correlation between the lengths of successive sampling intervals, as specified in Section 9.1. Then we need only calculate q(D), the probability density function of delay D, since D is independent of I_s, and $p(I_s) = p(I)$.

In calculating q(D) it is reasonable to assume that the input steps are widely spaced, relative to the operator's sampling intervals, and therefore have an equal probability of landing anywhere in any sampling interval. It may be noted that each step must land somewhere, so that the probability of a step occurring within an interval must be such that the total probability summed over all intervals is unity. (It is here assumed that sampling instants are of infinitesimal length, so that the probability of a step occurring at a sampling instant is zero). Then we proceed as follows:-

Let the probability of a step occurring in an interval of time δt be $\mu.\delta t$.

Then the total probability of a step occurring anywhere in an interval of length I is given by,

$$\mu.I$$

Therefore the total probability of a step occurring anywhere in an interval of any length is given by,

$$\int_t \mu.I$$

$$P_a = \int_a^b \mu.I.p(I).dI = \mu.E[I] \qquad X.2.$$

But it has already been shown that P_a must be unity, therefore:-

$$\mu = 1/E[I] \qquad X.3.$$

Now, the probability of a delay in the range D to D + δD arising from a step occurring in an interval of length in the range I to I + δI, is given by,

$$q(D).\delta D.\delta I = p(I).\delta I.\mu.I.\frac{\delta D}{I} \qquad 0 < D \leq I \qquad X.4.$$

$$= 0 \qquad D < I$$

The above equation may now be integrated to obtain:-

$$q(D) = \mu \int_a^b p(I).dI \qquad 0 < D \leq a \qquad X.5.$$

$$= \mu \int_D^b p(I).dI \qquad a < D \leq b$$

$$= 0 \qquad b < D$$

There are thus three distinct ranges for the function q(D).

From equations X.1. and X.5. we may obtain a convolution equation for r(L):-

$$r(L) = \int_x^y p(I).q(L-I).dI \qquad X.6.$$

where the limits of integration and the form of q(D) are dependent on the magnitude of L. There are thus five distinct ranges for r(L) :-

$$r(L) = \mu \int_a^L p(I).dI \qquad a \leq L \leq b \qquad X.7.$$

$$= \mu \int_{L-a}^b p(I).dI + \mu \int_a^{L-a} p(I).F(D).dI \qquad b < L \leq a+b$$

$$= \mu \int_a^b p(I).F(D).dI \qquad a+b < L \leq 2b$$

$$= 0 \qquad 0 \leq L < a \quad \text{and} \quad 2b < L$$

Substitution of the appropriate form of q(D) and the limits a and b in the above equations thus enables the direct calculation of r(L).

As an example, consider the case where I has a square density function over the range .1 $\leq$ I $\leq$ .2 sec. It may be seen that :-

$$\mu = 20/3 = 6.66^* \qquad \text{X.8.}$$

and
$$q(D) = 6.66^* \qquad 0 < D \leqslant .1 \qquad \text{X.9.}$$

$$= 6.66^* \int_D^{.2} 10.dI = 6.66^*(.2 - D) \qquad .1 < D \leqslant .2$$

$$= 0 \qquad .2 < D$$

Thus q(D) is trapezoidal in form.

Making the appropriate substitutions and integrations in equations X.7.,we obtain the following expressions for r(L):-

$$r(L) = 6.66^*(10L - 1) \qquad .1 \leqslant L \leqslant .2$$

$$= 3.33^*(-2 + 4L - 10L^2) \qquad .2 < L \leqslant .3$$

$$= 3.33^*(4 - 10L)(.4 - L) \qquad .3 < L \leqslant .4$$

$$= 0 \qquad 0 < L < .1$$

and

$$.4 < L$$

The form of the above r(L) is illustrated in Fig.9.1.2.

APPENDIX XI : Frequency Response of Sampled Data Systems

The analysis of the frequency response of the sampled data model, presented in Section 6.3., was based on the results presented in the following paragraphs.

For the case of a simple sampled data system, with stable transfer G(z) and input R(z), the output, C(z) , is given by :-

$$C(z) = G(z).R(z) \qquad \text{XI.1.}$$

Consider a complex sinusoidal input , $R(nT) = e^{jwnT}$, so that :-

$$R(z) = \frac{z}{z - e^{jwT}} \qquad \text{XI.2.}$$

Then the z-transform of the sampled output is given by :-

$$C(z) = G(z).\frac{z}{z - e^{jwT}} \qquad \text{XI.3.}$$

The output sample sequence may be derived by means of the inversion theorem :-

$$C(nT) = \frac{1}{2\pi j}.\int_{\Gamma} G(z).\frac{z^n}{z - e^{jwT}}.dz \qquad \text{XI.4.}$$

where Γ is the contour formed by describing the unit circle in the z-plane in a counter clockwise direction.

Now, for a stable system, transient terms die away with the passage of time. Because the steady state frequency response is required, it is necessary to consider only the residue of the integrand at the pole due to the forcing function, $z = e^{jwT}$.

Accordingly, the output at sampling instants is given by :-

$$C(nT) = G(e^{jwT}).e^{jnwT} \qquad \text{XI.5.}$$

which may be seen to be equivalent to a sampled sinusoid. The simplest sinusoidal function which fits C(nT) at sampling instants is given by :-

$$C(t) = G(e^{jwT}).e^{jwt} \qquad \text{XI.6.}$$

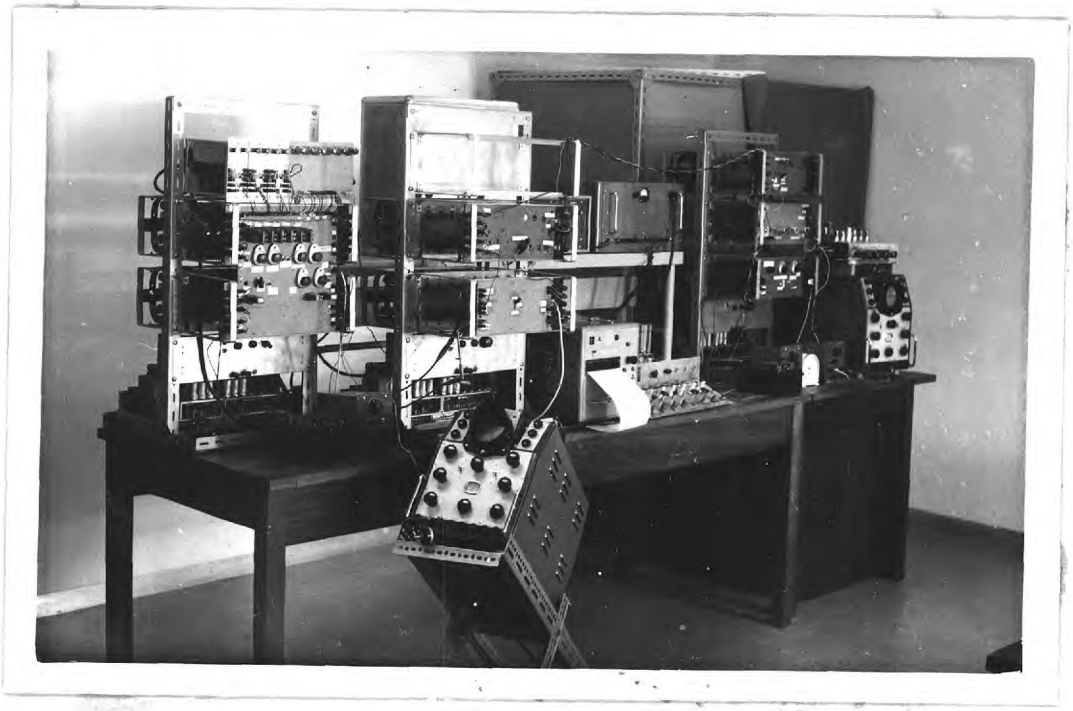
although, of course, there is an infinity of other sinusoids of different frequencies which will also fit C(nT). By analogy with linear systems operating continuously in time, the frequency response function may be

defined in terms of the relative magnitude and phase of that sinusoid of the same frequency as the input, which exactly matches the output at sampling instants. This definition dismisses departure of the output from the 'fitted' sinusoid, between sampling instants, as 'ripple'. Neglect of this ripple is usually of little practical consequence, because most control systems are of a low-pass nature.

On the basis of the above definition, the system frequency response function, $G(jw)$, is given by :-

$$G(jw) = \left[G(z)\right]\Big|_{z = e^{jwT}} \qquad \text{XI.7.}$$

More accuracy can be obtained by taking into account the output response between sampling instants, as explained in Section 6.3.



General View of Apparatus

Operator Tracking Configuration and Associated Apparatus