N82-13714 D49

A MODEL OF THE HUMAN OBSERVER AND DECISION MAKER

by P.H. Wewerinke

National Aerospace Laboratory NLR
The Netherlands

SUMMARY

In this paper a model is described of the human observer and decision maker monitoring a dynamic process. The decision process is described in terms of classical sequential decision theory by considering the hypothesis that an abnormal condition has occurred by means of a generalized likelihood ratio test. For this, a sufficient statistic is provided by the innovation sequence which is the result of the perception and information processing submodel of the human observer. On the basis of only two model parameters the model predicts the decision speed/accuracy trade-off and various attentional characteristics.

A preliminary test of the model for single variable failure detection tasks resulted in a very good fit of the experimental data.
In a formal validation program a variety of multivariable failure detection tasks was investigated. The task variables were the number, the bandwidth and the mutual correlation of display variables and various failure characteristics.
A very good overall agreement between the model and experimental results showed the predictive capability of the model. In addition, the specific effect of almost all task variables was accurately predicted by the model.

INTRODUCTION

With increasing complexity and automation of man-machine systems the human operator's role shifts from controller to supervisor. In the context of transport aircraft operation this is very much the case after the introduction of automatic approach and landing systems and the future microwave landing system (MLS).
The last two decades considerable research effort has been devoted to the study of human control behavior. One result is a number of mathematical tools, of which the state-space, time-domain optimal control model has been shown to provide a general framework adequately describing the human processing of information provided by a dynamic system (Refs 1-4). This can be extended to other cognitive functions involved in monitoring an automatic system, detecting system failures, making decisions, etc. The insight in this higher mental functioning is still rather incomplete although some attempts have been made to investigate and to model failure detection and simple decision making behavior (Refs 5-9).

This paper summarizes the results of a theoretical and experimental analysis of the human observer and decision maker in multivariable failure detection tasks. In the next section a model of the human decision maker is formulated in terms of multivariable classical sequential decision theory, accounting for the important effect of correlated information. In the subsequent section the model is tested against the results of a single variable task experiment reported in reference 8. Next, a formal model validation experiment is discussed. The latter results are extensively presented in reference 10.

MODEL OF THE HUMAN OBSERVER AND DECISION MAKER

It is assumed that the human perceives information of a linear dynamic system which is described by a Gauss-Markov random sequence. Based on the known dynamics of this system and the perceived information (i.e. noisy observations), the human makes the best estimate of the system state. This is described in standard linear estimation theoretical terms (Kalman filter, Refs 11-12) and is part of the well documented optimal control model (Refs 1-2) but included in figure 1 for the ease of reference.

Now, in the normal mode of operation, the discrepancy between perceived and expected information (the so-called innovation sequence n_k) is a zero mean Gaussian purely random sequence (Ref. 12) with covariance N_k. It is assumed that abnormal system operation, as caused by errors in display instruments, malfunctioning of the system and excessive system disturbance levels (e.g., large windshears in aircraft operation) can be represented by a deterministic process, as such unknown to the human observer but detected on the basis of a non-zero mean innovation sequence whose statistic is sufficient to make decisions (test hypotheses) when the system is completely observable.

In terms of classical sequential decision theory (Refs 13-14) a so-called generalized likelihood ratio test can be formulated. The test amounts to the comparison of the probability of a non-zero mean with the probability of a zero mean innovation sequence assuming that the human operator makes a short-term estimate of the mean of the innovation sequence on the basis of the sample mean of m past observations ($\hat{n}_k$ in figure 1).
It can be derived (Ref. 15) that the effect of each observation at stage k on the (log of the)likelihood ratio is given by

$$\Delta L_k = \tfrac{1}{2}\, \hat{n}_k'\, N_k^{-1} \hat{n}_k \qquad (1)$$

under the assumption that the sample mean $\hat{n}_k$ is constant during m observations. The accumulating effect of each observation on the total (log of the) likelihood ratio is given by the recursive expression

$$L_k = L_{k-1} + \Delta L_k \qquad (2)$$

assuming that the innovation sequence is a white noise sequence (independent samples) which is exact in the normal mode of operation. The number of observations, based on which the decision is made, is chosen such that ΔL_k is, on

the average, not decreasing; in other words, using only confirming evidence to make the decision. Otherwise, the positive semi-definite elements ΔL_k would have an accumulating effect on the likelihood ratio.
When the likelihood ratio L_k (representing the total evidence of abnormal system operation) is equal to, or larger than, a decision threshold T, the decision is made that an abnormal condition has occurred. This decision threshold can conveniently be related to the accepted (or assumed) risk according to (Ref. 14) $T=(1-P_M)/P_F$, with P_M the miss probability (i.e. of no response to an abnormal condition) and P_F the false alarm probability.

Following reference 14 an expression can be derived (Ref. 15) for the average number of samples used to make the decision that the system is operating abnormally, which is, for a given sample rate, uniquely (linearly) related to the average detection time.
This average number of observations K, given the abnormal condition, is given by

$$\bar{K}= \frac{2P_F T \ell n T}{\overline{E\{\bar{n}'N^{-1}\bar{n}\}}} \tag{3}$$

where $\overline{(.)}$ indicates the average over the ensemble and E{(.)} is the average over the sequence.
Equation (3) gives a relationship between the average number of observations and the decision error probabilities (P_F and P_M), for a given innovation covariance N and the non-zero mean failure state sequence which is, however, a given task variable. Thus the only human decision model parameters are the short-term average sample size m and the innovation covariances which depend exclusively on the human observation noise covariances (Ref. 15). The model output is the average failure detection time corresponding with given (or assumed) error probabilities.

Previous studies (Ref. 2) support the hypothesis that the observation noise covariance scales with the mean-squared value of the corresponding signal. Thus for display variable j $V_j = P_0 E\{y_j^2\}$, $j= 1,\dots.\ell$. P_0 represents a nominal, "full attention" noise ratio (typically 0.01π). In case a display variable is not looked at (foveally), it is assumed that the corresponding observation noise is infinite, thus neglecting peripheral viewing. Inserting this expression in equation (3) results, after some matrix manipulation (Ref. 15) in a very simple expression for the average "detection time"

$$\bar{K}= \frac{2P_F T \ell n T}{E_t\{\bar{n}_e^2\}} \tag{4a}$$

with

$$\bar{n}^2_{e_j} = \bar{n}^2_{r_j}(1-p_{r_j}) \tag{4b}$$

is the effective, relative, estimated sample mean-square; the subscript r indicates the normalization by the (constant) observation noise covariance V_j, p_j is the estimation error covariance of display variable j, and E_t indicates the average over the ensemble, and the sequence, and the display variables.

The expected value of $\tilde{n}_e^2$ over the display variables involves the probability distribution of the human attention to these variables. When an optimal distribution (i.e., yielding a minimal detection time) of the human attention is assumed, the only remaining human decision model parameters are the short-term average sample size m and the overall level of attention P_0.

The general model structure (the only assumptions are that the dynamic sy em is linear and that abnormal conditions can be represented by a deterministic process) accounts for the effect of a variety of task variables such as the number and bandwidth of display variables, the correlation among them and the perceived failure characteristics. These variables are included in the selected task configurations investigated in the validation program discussed in the next section.

MODEL VALIDATION

In order to test the validity of the model the results of two experimental programs were considered and compared with the model predictions. The first experiment is reported in reference 8 in which observers were required to detect a change in the mean of a stationary stochastic process. The experimental results were used for a preliminary validation of the human decision model and to "calibrate" the model with respect to the short-term average sample size m. Next, a formal model validation program is described which is reported in reference 15.

Preliminary model validation

In the first experiment the (two) subjects were instructed to detect, as soon as possible, the occurrence of a non-zero mean component while observing a second order, zero mean process. Both step and ramp failures with four possible amplitudes were introduced yielding 8 experimental conditions. Model results of the average failure detection time were obtained assuming the decision error probabilities of 0.05 which were also obtained in the experiment and on the basis of a typical overall level of attention P_0 of 0.01π. The remaining model parameter m was selected so as to obtain the best overall match with the experimental results. As shown in figure 2 the resulting value of m of four seconds yields an excellent fit to the experimental data. This can be expressed in the linear correlation coefficient between model and experimental results of 0.99 and in the ratio of model detection times and corresponding experimental values (t_m/t_e). This ratio was on the average, 0.98, with a standard deviation of 0.09. Thus the constant value of m of 4 seconds yielding a good fit for all the experimental conditions seems a human operator-related parameter. This value of four seconds, which ties in very well with the short-term memory span typically ranging up to five seconds for visual stimuli (Ref. 16), will be assumed and kept constant in following validation experiment.

Formal model validation

A variety of tasks was specified so as to present the most crucial combination of the task variables considered: number, bandwidth and mutual correlation of the displays and failure characteristics. Two failure types were investigated either appearing on one display (display failure, F_i) or on two correlated displays (system failure, F_s). Also the effect of prior knowledge about the failure type was investigated.
Up to four-display tasks were considered consisting of two separate (independent) identical processes. Each process could be observed via two displays: a relatively high bandwidth variable y1 (second order process with a break-frequency of 1.2 rad/s) was additionally filtered (first order filter with a time constant of four seconds). The output of this filter which is correlated with y1 ($r \approx 0.5$) was displayed as y2. This process was duplicated resulting in a four-display process (y1 to y4), of which the displays were two by two correlated.
The resulting 8 configurations and display situation are summarized in table 1. The configurations were investigated for two failure rates (a ramp with a slope of 0.1 standard deviation of the display position per second and a slope of 0.2 σ_{y_i} /s) resulting in 16 experimental conditions. A system failure appeared as a ramp on y1 which was additionally filtered and subsequently displayed. For a detailed presentation of the foregoing tasks and the model and experimental results (of three subjects, being general aviation pilots; twelve replications per condition), the reader is referred to reference 15. In this paper only the principal results are summarized aimed at a test of the human decision model.

Failure detection times

Model predictions of the (ensemble) mean failure detection times for all the 16 failure detection tasks were obtained on the basis of the two constant model parameters: the overall level of attention $P_o = 0.01\pi$ and the short-term average sample size m = 4 seconds. Based on the previous results a value of the false alarm probability of 0.05 was assumed. Additional model assumptions and procedural details are discussed in reference 15.
The results of two subjects achieving an overall false alarm probability P_F of 0.05 could be compared directly with the model predictions. The result is summarized in figure 3. The linear correlation coefficient between the model predictions and the experimental mean failure detection times is 0.86 which reflects a very good overall predictive capability of the human decision model. The ratio of the model and experimental failure detection time t_m/t_e is an other measure for the agreement between the model and experimental results. On the average (over all 16 tasks) this ratio is 0.98. The standard deviation is 0.12 which is comparable with the reliability of the experimental estimated means.
This reliability of the data is included in figure 4 showing the model and experimental failure detection times per configuration. Apart from the mean value, also the standard deviation of the mean value estimate ($\sigma/\sqrt{N}$, with σ the standard deviation of the raw data) is given. For almost all configurations

the model predictions agree rather well with the experimental failure detection times.
For one configuration (Conf. 4) the model predictions clearly disagree with the experimental results. A very plausible explanation of this discrepancy (which was confirmed during the debriefing of the subjects) was that the subjects did not realize (use) the system failure dynamics but assumed that the system failure appeared simultaneously on both (correlated) displays. The model results based on this assumption are also shown in figure 3 and indicated with model refinement. In that case, the linear correlation coefficient is 0.95; the mean ratio t_m/t_e is 1.01 with a standard deviation of 0.09.

The agreement between the model predictions and experimental results with respect to the specific task variables is summarized in table 2. The configurations involved in the pair-wise comparison between the configurations which differ with respect to the specific task variable (only) are indicated.
Comparing the measured and model results shows that the effect of display bandwidth, of additional (correlated) displays and of the failure rate is excellently predicted by the model. The predicted interference between uncorrelated displays (because of the human attention sharing involved) is larger than obtained experimentally. The model predictions are based on a constant level of attention. However, the physiological (heart rate) measures obtained during the experiment suggest a small, but statistically significant, increase in attention with an increase in displays which can easily explain the small difference in interference. The effect of the failure type is discussed before. It can be seen that the model refinement and the experimental results agree closely.
The experimental results of the third subject reflect a distinctly different decision strategy. He made no false alarms and his failure detection times were, on the average, 40 % higher. Yet his results correlated well with the model predictions (r = 0.80). Model results based on a very low value of P_F yield an arbitrary good overall correspondance with the measured failure detection times (detection times increase monotonically with decreasing P_F). However, virtually the same linear correlation coefficient is obtained. As discussed in reference 15 this signifies that the predicted effect of the various configurations on the failure detection time match the results of this subject with an accuracy of about 12 % (the standard deviation of t_m/t_e is 0.12).

Scanning behavior

Various attentional characteristics can be derived from the foregoing model of the human observer and compared with eye scanning data obtained in the experimental program.
Combining eqs (1), (3) and (4), it can easily be seen that the (ensemble) average effect on the likelihood ratio of one observation of display variable j is given by

$$\overline{\Delta L_j} \doteq \tfrac{1}{2}\, \overline{\hat{n}^2_{e_j}} \cdot \overline{\delta_j} \qquad (5a)$$

where $\bar{\delta}_j$ indicates the probability of attending to j. In the normal mode of operation ($\bar{n}$= 0) an alternative expression can be obtained (Ref. 15)

$$\overline{\Delta L}_j \overset{\text{normal}}{\underset{\text{mode}}{=}} \tfrac{1}{2}\, \Delta p_{r_j} \cdot \bar{\delta}_j \qquad (5b)$$

where Δp_{r_j} is the reduction in the estimation error covariance due to observing variable j.

Equations (4) and (5) show that the optimal allocation of attention among the displayed variables (i.e. yielding the maximum ΔL and thus, the minimal detection time) is obtained for the maximum expected value of $\tilde{n}_e^2$. The model predicts (Ref. 15) that the optimal fraction of attention to the high bandwidth display(s) is varying between 0.4 and 0.7 (the total attention to the high and low bandwidth display(s) is 1.0) somewhat depending on task specifics and failure characteristics. As the failure detection time is relatively insensitive to the division of attention in this region, a relatively constant fraction of attention to the high bandwidth display(s) for all configurations is predicted by the model, say between 0.5 and 0.6 (enhanced by the randomized block design and the corresponding transfer of training). This agrees very well with the experimental dwell fractions. The average dwell fraction on the upper display was not depending on the configurations and varied between 0.5 and 0.55.

The optimal allocation of attention in the time domain - thus the optimal scanning strategy for a given task - is described by eq. (5). Various model predictions can be derived from the attention allocation model and compared with the corresponding eye scanning measures. This is discussed in reference 15. For illustrative purposes, consider the normal mode of system operation, for which situation the effect of observing is described by eq. (5b). The model predicts that there is a scanning preference for high bandwidth display variables as the estimation error covariance Δp increases with display frequency, for a given amount of time. Furthermore, it follows from eq. (5b) that there is a scanning preference for correlated display variables, because observing variable j yields an additional reduction in Δp_i if variable j is correlated with i which results in a maximum total ΔL and a minimum (average) failure detection time.

These model predictions can be compared with the experimental eye scanning data in terms of display link values. The experimental results agreed well with these qualitative predictions. The fraction of links between the correlated displays was about two times the fraction of links between the uncorrelated displays (0.64 versus 0.36; in the failure mode this division is 0.70 versus 0.30 which increase is predicted on the basis of eq. (5a) as discussed in reference 15). Furthermore, the fraction of scans (number of observations) towards the high bandwidth displays was two times the fraction of scans towards the low bandwidth displays.

The foregoing analysis illustrates the predictive capability of the attention allocation model which may be a powerful tool in the study of human information processing tasks and display design problems.

CONCLUDING REMARKS

In this paper a model of the human observer and decision maker is summarized. The model consists of two parts. A submodel of the human observer is formulated in linear estimation theoretical terms including the perception of the displayed information of a linear(ized) process and the information processing stage which is described by a Kalman filter. The resulting innovation sequence provides a sufficient statistic for the decision process. In terms of classical sequential decision theory the hypothesis is considered that an abnormal condition has occurred by means of a likelihood ratio test. An abnormal condition is represented by a deterministic process which has to be detected on the basis of noisy observations of a normally zero-mean stochastic process. On the basis of only two model parameters (short-term average sample size m and the overall level of attention P_0) the model predicts the (ensemble) mean failure detection time and various attentional characteristics.

A preliminary test of the model for single variable failure detection tasks resulted in a very good fit of the experimental results for a constant value of m of four seconds.

This constant value for m and a typical value for the overall level of attention were used to predict the mean failure detection times and scanning characteristics of a variety of multivariable failure detection tasks which were investigated in the formal validation program. The task variables were the number, the bandwidth and the mutual correlation of display variables and various failure characteristics.

A very good overall agreement between the model and experimental results both in terms of failure detection times and eye scanning measures showed the predictive capability of the model. In addition, the specific effect of almost all task variables was accurately predicted by the model.

REFERENCES

1. Kleinman, D.L. and Baron, S.: Manned Vehicle System Analysis by means of Modern Control Theory. NASA CR-1753, June 1971.

2. Baron, S. and Levison, W.H.: Display Analysis with the Optimal Control Model of the Human Operator. Human Factors, 1977, 19(5).

3. Wewerinke, P.H.: Performance and Workload Analysis of In-Flight Helicopter Missions. Paper presented at the 13th Annual Conference on Manual Control, MIT, June 1977 (also NLR MP 77013 U).

4. Wewerinke, P.H.: The Effect of Visual Information on the Manual Approach and Landing. Paper presented at the 16th Annual Conference on Manual Control, MIT, May 1980 (also NLR MP 80019 U).

5. Phatak, A.V. and Kleinman, D.L.: Current Status of Models for the Human Operator as a Controller and Decision Maker in Manned Aerospace Systems. AGARD symp. on Automation in Manned Aerospace Systems. WPAFB, Dayton, USA, 1972.

6. Levison, W.H. and Tanner, R.B.: A Control-Theory Model for Human Decision Making. NASA CR-1953, December 1971.

7. Wewerinke, P.H.: Human Monitoring and Control Behavior - Models and Experiments. NLR TR 77010 U, October 1976.

8. Gai, E.G. and Curry, R.E.: A Model of the human Observer in Failure Detection Tasks. IEEE Trans. on Systems, Man and Cybernetics. Vol. SMC-6 no. 2, February 1976.

9. Gai, E.G. and Curry, R.E.: Failure Detection by Pilots during Automatic Landing: Models and Experiments. J. Aircraft, Vol. 14, no. 2, February 1977.

10. v.d. Graaff, R.C.: An Experimental Analysis of Human Monitoring Behavior in Multivariable Failure Detection Tasks. NLR TR 81063 U, June 1981.

11. Bryson, A.E. and Ho, Y.C.: Applied Optimal Control. Blaisdell Publishing Company, 1969.

12. Mehra, R.K. and Peshon, J.: An Innovation Approach to Fault Detection and Diagnosis in Dynamic systems. Automatica, Vol. 7, 1971.

13. Wald, A.: Sequential Analysis, J. Wiley, New York, 1947.

14. Sage, A.P. and Melsa, J.L.: Estimation Theory with Applications to Communications and Control. McGraw-Hill, 1971.

15. Wewerinke, P.H.: A Model of the Human Decision Maker Observing a Dynamic System. NLR TR 81062 U, March 1981.

16. Sheridan, T.B. and Ferrell, W.R.: Man-Machine Systems: Information, Control and Decision Models of Human Performance. MIT Press, Cambridge, Ma. 1974.

Table 1a
Task configurations

CONF.	DISPLAY	FAILURE
1	y1	F_1
2	y2	F_2
3	y1, y2	F_1
4		F_s
5		F_1, or F_2, or F_s
6	y1, y3	F_1 or F_3
7	y1, y2, y3, y4	
8		F_i, or F_{s_j}

Table 1b
Display situation

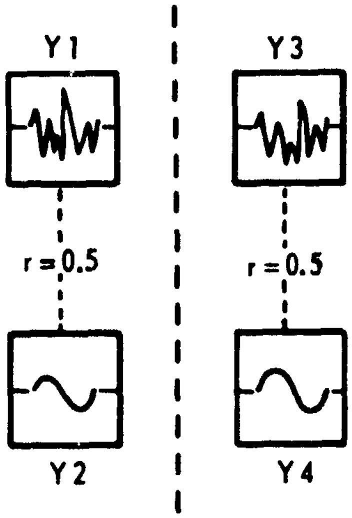

Table 2
Specific effect of task variables on model and experimental failure detection times

EFFECT OF	CONFIGURATIONS	RATIO t_i/t_j measured m	measured s	model m	model s
bandwidth (low/high)	1, 2, 5, 8	1.01	0.12	1.00	0.08
additional display info.	1, 3, 6, 7	0.76	0.06	0.77	0.07
interference	1, 3, 5, 6, 7, 8	1.11	0.05	1.24	0.05
failure type (system/display)	3, 4, 5, 8	1.43	0.33	1.24 (1.37)	0.06 (0.21)
failure rate (high/low)	all	0.67	0.05	0.69	0.07

(.) model refinement

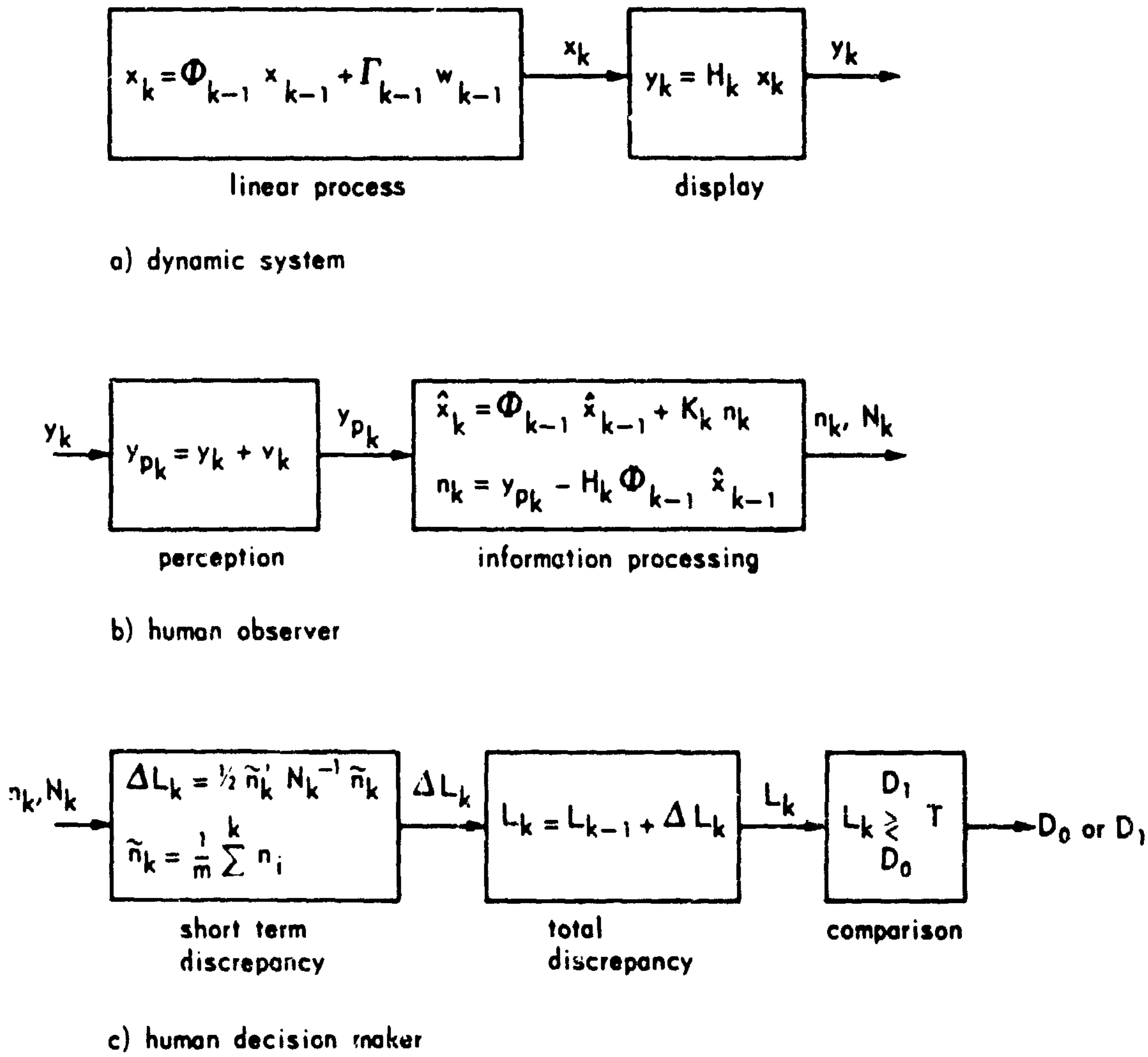

Fig. 1 Model of the human decision maker observing a dynamic process

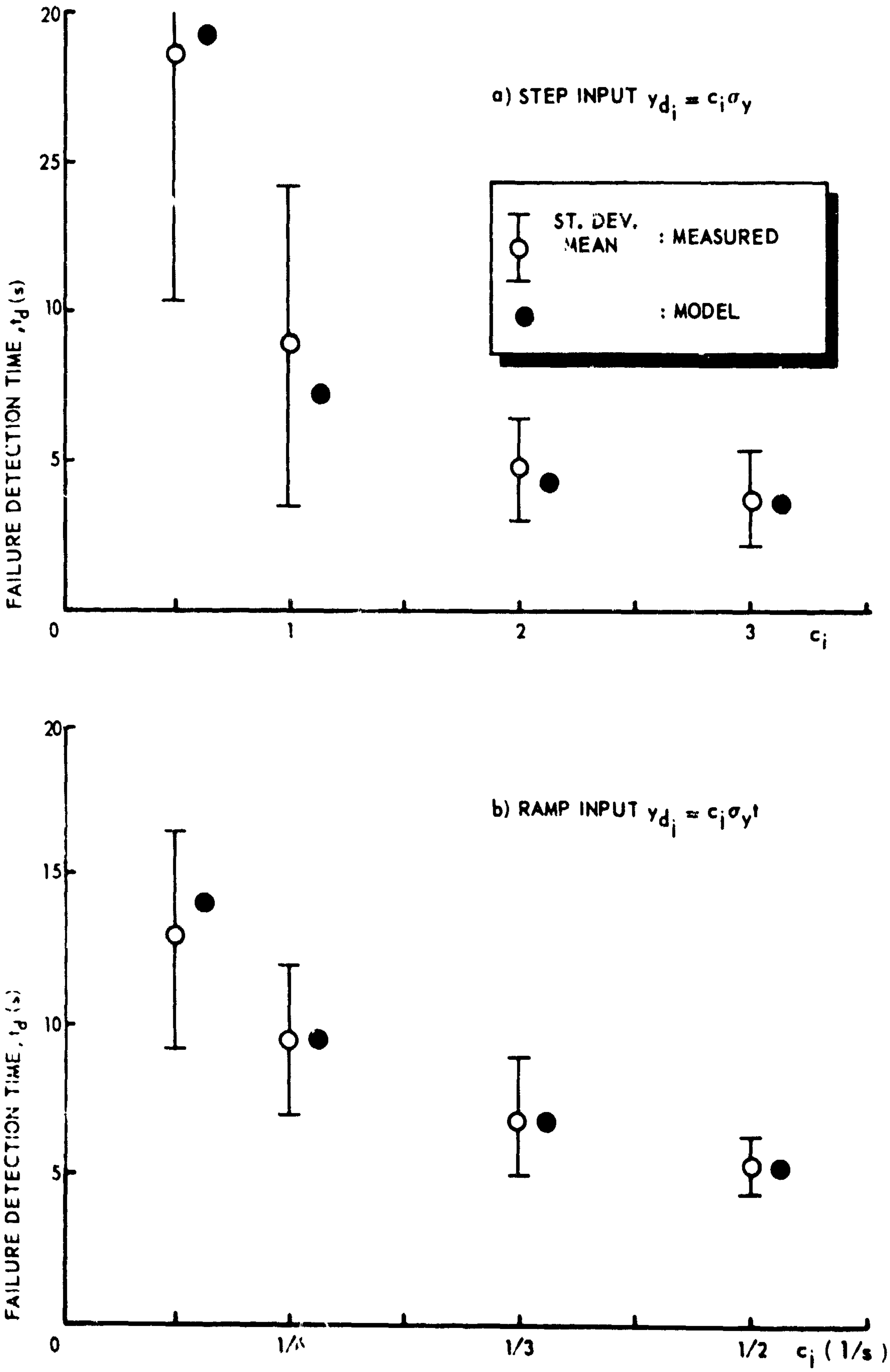

Fig. 2 Experimental and model failure detection times for various step and ramp inputs

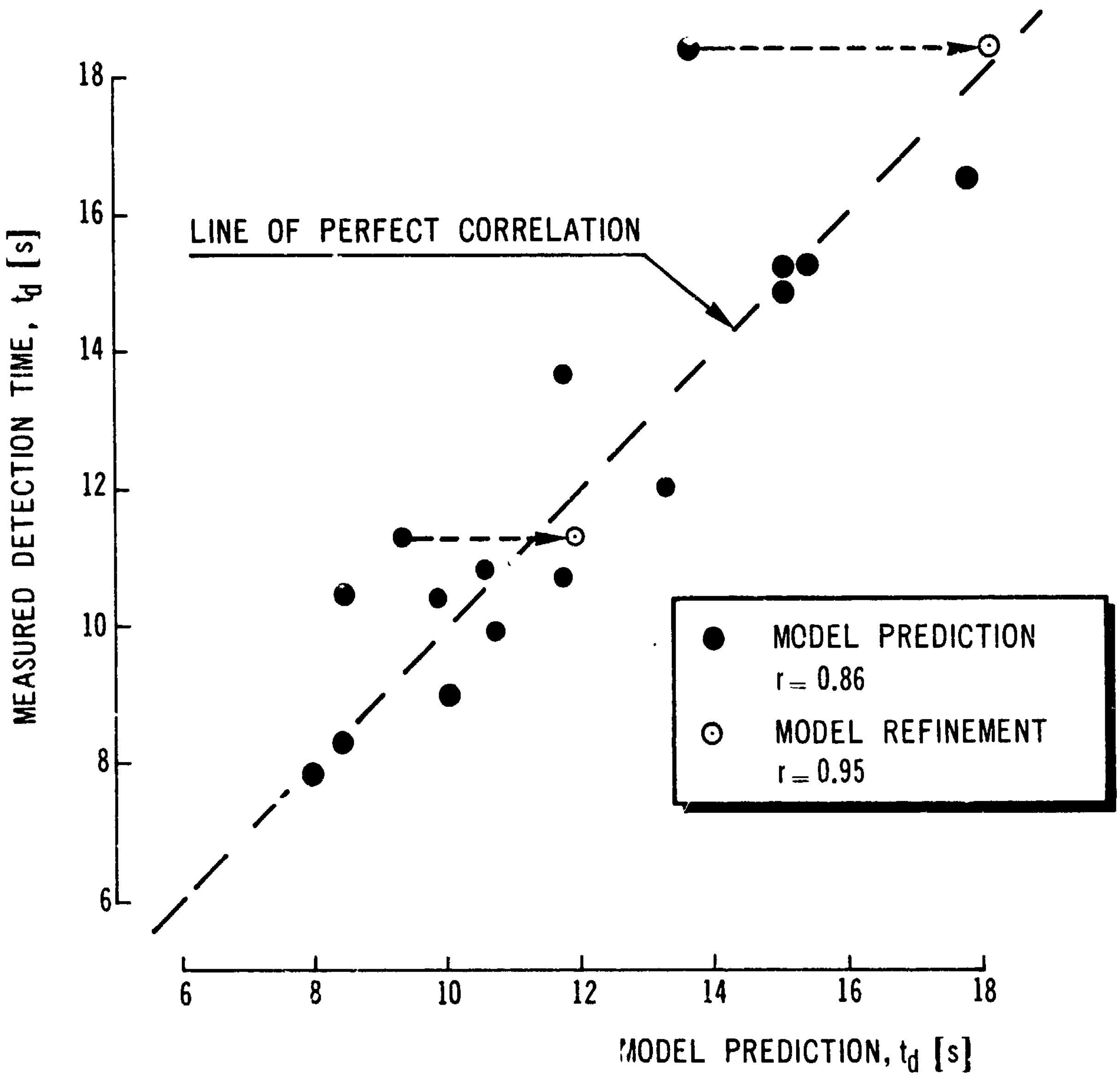

Fig. 3 Overall comparison of model predictions and experimental results of two subjects

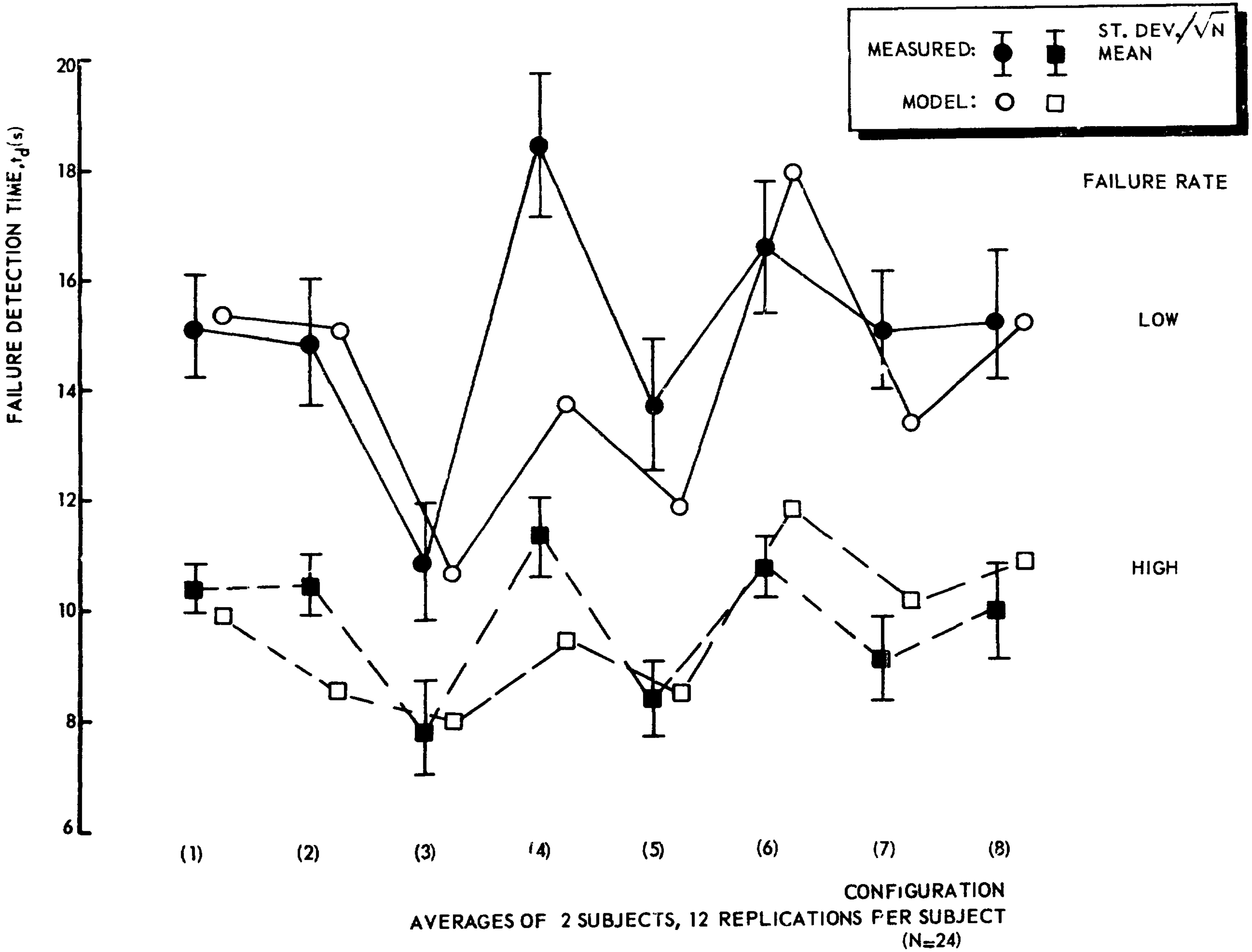

Fig. 4 Model and experimental failure detection times for all configurations