View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by NASA Technical Reports Server

D8

N79-15596

THE INTERNAL MODEL: A STUDY OF THE RELATIVE CONTRIBUTION OF PROPRIOCEPTION AND VISUAL INFORMATION TO FAILURE DETECTION IN DYNAMIC SYSTEMS*

By Colin Kessel and Christopher D. Wickens

Department of Psychology, University of Illinois

SUMMARY

The development of the internal model as it pertains to the detection of step changes in the order of control dynamics is investigated for two modes of participation: whether the subjects are actively controlling those dynamics or are monitoring an autopilot controlling them. A transfer of training design was used to evaluate the relative contribution of proprioception and visual information to the overall accuracy of the internal model. Sixteen subjects either tracked or monitored the system dynamics as a 2-dimensional pursuit display under single task conditions and concurrently with a "sub-critical" tracking task at two difficulty levels. Detection performance was faster and more accurate in the manual as opposed to the autopilot mode. The concurrent tracking task produced a decrement in detection performance for all conditions though this was more marked for the manual mode. The development of an internal model in the manual mode transferred positively to the automatic mode producing enhanced detection performance. There was no transfer from the internal model developed in the automatic mode to the manual mode.

INTRODUCTION

Over the past few years there has been a great deal of research directed at the problem of determining the differences between operators and monitors of dynamic systems (References 1-7). While the conclusions reached by these authors do not always coincide, there is a general consensur that a greater understanding of the different processes operating in the two modes of participation is necessary for the successful integration of automated systems in the workplace.

We have provided a detailed theoretical analysis of the processes involved in the two modes of participation (Reference 7). Briefly, this analysis has argued that one way in which the differences between modes of participation can be studied is by determining the relative sensitivity of operators versus monitors in a failure detection task.

---

*This research was funded by the Life Sciences Program, Air Force Office of Scientific Research, Contract Number F44620-76-C-0009. Dr. Alfred Fregly was the scientific monitor of the contract.

84 PAGE INTENTIONALLY

Three attributes were identified that would seemingly facilitate failure detection in the controlling mode: (i) a smaller variability of the internal model of the system; (ii) the options of testing hypotheses about the nature of the dynamics by introducing signals into the system; and (iii) a greater number of information channels available upon which to base failure detection decisions. It was recognised however, that this latter advantage may be mitigated to the extent that: a) adaptation takes place reducing the strength of visual error information and, b) proprioceptive sensitivity is less than visual.

In comparison the monitoring mode was also characterised by two attributes that could facilitate detections: a greater "strength" of the visual signal (if adaptation by the autopilot does not take place) and a lower level of operator workload.

The study conducted (Reference 7) to test the above theoretical analysis found that detection performance in the manual mode was faster and only slightly less accurate than the autopilot mode. Furthermore the observed manual superiority was attributed to the additional proprioceptive information resulting from operator control adaptation to the system change. It is possible that some contribution to manual mode superiority in our prior study resulted from the greater internal model consistency in that mode. However this hypothesis was assumed to be doubtful because a within subjects design was employed, so that the same subjects participated in both automatic and manual conditions. Thus the internal model developed in manual conditions would presumably be available to facilitate detection in the automatic conditions as well.

In order to generate a greater distinction between the internal model employed in the two modes, the present study employed a between subject design using a transfer of training technique. This procedure enables an examination of the development of internal models, in the two modes of participation, and subsequently measures their impact upon transfer to the other mode.

It was hypothesized that this technique would increase the differential performance in detection between the two modes of participation while at the same time demonstrating that the internal model developed in the manual mode can subsequently be utilized to facilitate automatic mode failure detection performance.

METHOD

Subjects:

The subjects were 18 right-handed male university students. Subjects were paid a base rate of $2.50 per hour but could increase their average pay by maintaining a high level of detection performance.

C-2

Apparatus:

The basic experimental equipment included a 7.5 x 10 cm Hewlett Packard Model 1300 CRT display, a spring-centered, dual-axis tracking hand control with an index-finger trigger operated with the right hand, and a spring loaded finger controller operated with the left. A Raytheon 704 16-bit digital computer with 24k memory and A/D, D/A interfacing was used both to generate inputs to the tracking display and to process responses of the subject. The subject was seated on a chair with two arm rests, one for the tracking hand controller and one for the side-task finger controller. The subject's eyes were approximately 112 centimeters from the CRT display so that the display subtended a visual angle of 1.5^o.

Tracking tasks. The primary pursuit-tracking task required the subject to match the position of a cursor with that of a target which followed a semi-predictable two-dimensional path across the display. The target's path was determined by the summation of two non-harmonically related sinusoids (.05 and .08 Hz) along each axis with a phase offset between the axes. The position of the following cursor was controlled jointly by the subject's control response and by a band-limited forcing function with a cutoff frequency of .32 Hz for both axes. Thus the two inputs to the system were well differentiated in terms of predictability, bandwidth, and locus of effect (target vs. cursor). The control dynamics of the tracking task were of the form $Y_c = \frac{1-\alpha}{s} + \frac{\alpha}{s^2}$ for each axis, where α was the variable parameter used to introduce changes in the system dynamics. These changes, or simulated failures, were introduced by step changes in the acceleration constant α from a normal value of .3, a mixed velocity and acceleration system, to $\alpha = .9$, a system that approximates pure second order dynamics that requires the operator to generate considerable lead in order to maintain stable performance.

As the loading task, the Critical Task (Reference 8) was employed. This was displayed horizontally in the center of the screen and required the subject to apply force to the finger control in a left-right direction to maintain the unstable error cursor centered on the display. The value of the instability constant λ in the dynamics $Y_c = \frac{k\lambda}{s - \lambda}$ was set at a constant subcritical value. Two values ($\lambda = 0.5$ and $\lambda = 1.0$) were employed on different dual task trials.

Experimental Design and Task:

Three groups were used in the transfer of training design (see Figure 1). Group one transferred from manual (MA_I) on session one to automatic (AU_{II}) on session two; group two transferred from automatic (AU_I) to manual (MA_{II}) while group three was the control group for the automatic condition and monitored in the automatic mode in both sessions ($AU_{I(C)}$ and $AU_{II(C)}$). The control group for the manual group (MA_{II}) was MA_I. The various group comparisons are represented in Figure 1 by arrows and will be referred to at greater length in the results section.

Each group participated in six consecutive days of data collection. These were divided into two sessions; 3 days in each session with each session comprising 1 training day and two experimental days. Subjects in group one for example participated in 3 manual (MA_I) sessions and then transferred to 3 automatic (AU_{II}) sessions.

In the manual (MA) condition the subject performed the tracking manually while in the autopilot (AU) condition, his role in the control loop was replaced by simulated autopilot control dynamics consisting of pure gain, effective time delay, and a small added remnant. Each trial, MA or AU, lasted 150 seconds.

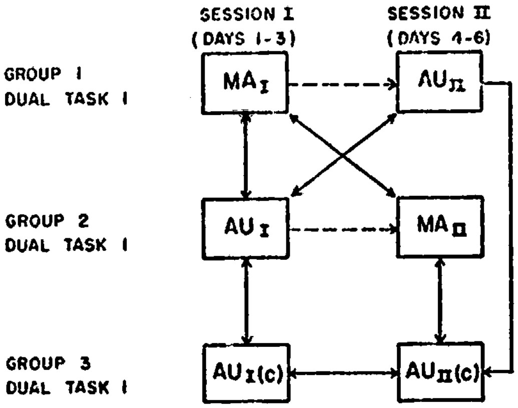


Figure 1: Experimental design and group comparisons

Training Day: The training day was designed to give the subject maximum experience and practice with the system. Subjects therefore received extensive practice tracking (or monitoring) with both prefailure and postfailure dynamics. Following this, they observed and then detected the step changes in dynamics. Practice with the critical side task was also included.

The presentation of the failure was generated by an algorithm that assured random intervals between presentations and allowed the subject sufficient time to establish baseline tracking performance before the onset of the next change. Task logic also ensured that changes would only be introduced when system error was below a criterion value. In the absence of this precaution, changes would sometimes introduce obvious "jumps" in cursor position.

During the detection trials, the detection decision was recorded by pressing the trigger on the control stick. This response presented a "T" on the screen and returned the system to normal operating conditions of the prefailure dynamics. If the subject failed to detect the change, the system returned to normal after six seconds via a 4 second ramp. On the basis of pretest data, it was assumed that six seconds was the interval within which overt responses would correspond to detected failures and not false alarms. The subjects were told to detect as many changes as possible as quickly as possible.

Experimental Days: The training day was followed by two consecutive experimental days. After four refresher trials in the AU or MA modes (depending upon the condition) with the side task, and a number of demonstrated failures, the subjects performed 15 experimental trials: 5 single task, tracking (or monitoring) only; 5 tracking with the easy critical task ($\lambda = 0.5$); and 5 tracking with the difficult critical task ($\lambda = 1.0$). The order of presentation was randomized. Each trial contained an average of 5 failures per trial with a range of 4 to 6.

The subject was instructed to "do the side task as efficiently and accurately as possible," and told to maintain that task at a standard level of performance. After each trial the subject received feedback about both his side task and detection performance. The instructions, feedback and payoff schedule, therefore, clearly defined the side task as the loading task while allowing the tracking and detection tasks to fluctuate in response to covert changes in available attentional resources (Reference 9).

ANALYSIS

Detection performance was assessed in terms of the accuracy and latency of responses. In computing the accuracy measure, signal detection theory analysis based upon the method of free response was employed (Reference 10). This technique accounts for the presence of hits and false alarms in the data; and the semi random occurance of failures within a trial. The area under the ROC curve (A[ROC]) was employed as the final accuracy measure (Reference 11). Further details of this analysis procedure may be found in Wickens and Kessel (Reference 7, 12).

The A(ROC) measure and the latency measure were then plotted in the form of a joint speed-accuracy measure depicted in Figure 2. "Good" performance is represented by points lying on the upper left, in the region of fast accurate response. Performance was quantified by projecting the point locus obtained onto the positive diagonal performance axis. The performance scale is computed as (10 times A[ROC] - LATENCY) and will be called the "derived performance score." This procedure produces a performance index that ranges from 0 for chance level of accuracy with a latency of 5" to 10.0 for perfect detection with 0 second reaction time. The units assigned to this performance index are somewhat arbitrary but are based on the observation that the overall variability (standard deviation) of the raw latency scores were found to be roughly 10 times the variability of the A(ROC) measure.

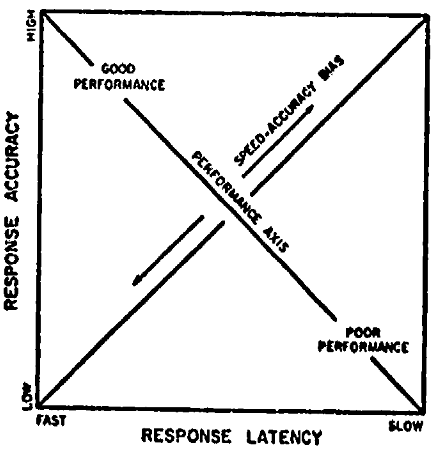


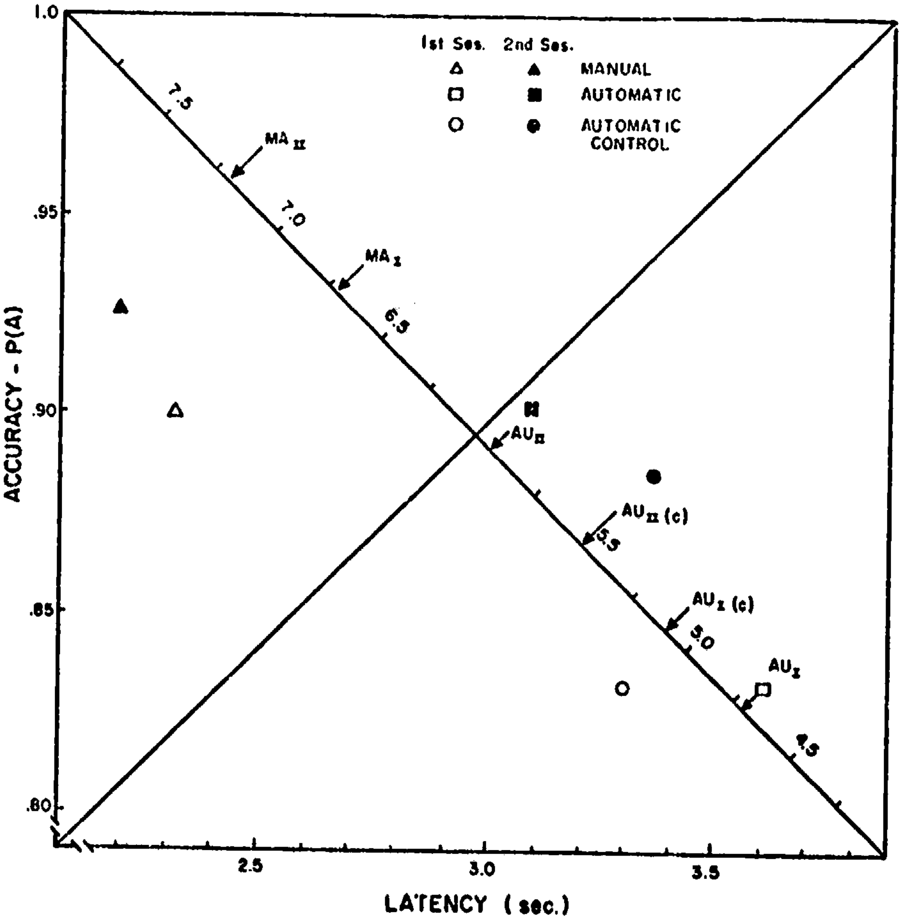
Figure 2: Speed-accuracy representation of detection performance

RESULTS

Averages and standard deviations were computed for the accuracy (A[ROC]), the latency and the derived performance measures following the rational and the procedures outlined in the preceding section.

The group averages for all three measures are presented graphically in Figures 3 and 4. Figure 3 represents the results for the single task condition while Figure 4 represents the dual task, workload condition collapsed over both levels of dual task difficulty (the rationale for this procedure is discussed below). The symbols in Figures 3 and 4 represent the group results in the speed-accuracy space, while the arrows and labels depict the derived performance scores for the various groups along the performance axis. In figures 5, 6, and 7 the experimental groups are plotted with the average derived performance score on the Y-axis.

The presentation of the results of the detection of failures will be divided into three sections. The first presents the results for each mode of participation, and represents a replication of the Wickens and Kessel (Reference 7) study with the between subjects design, the second examines the results of the loading task, while the third reports the results of the transfer of training experiment. Group differences were analyzed by means of a 3-way Analysis of Variance-ANOVA (groups x dual task x experimental days).



ORIGINAL PAGE IS
OF POOR QUALITY

Figure 3: Effect of participatory mode and experimental condition on detection performance-Single Task

(a) Mode of Participation

The most pronounced effect in the experimental data is the consistent superiority of MA over AU detection. This statistically reliable effect is clearly evident in the derived performance score shown in Figures 5 and 6 and was tested by contrasting group AU_I with MA_I ($F_{1,10} = 18.4$ $p < .001$). Examination of Figures 3 and 4 reveals that the MA superiority is reflected in detection latency ($F_{1,10} = 13,66$, $p < .01$), as well as accuracy ($F_{1,10} = 15.55$, $p < .01$).

While these findings essentially replicate the Wickens and Kessel (Reference 7) study, it is important to note that the extent of MA superiority observed in the present results is greatly enhanced. In fact the magnitude of the MA-AU difference in the desired performance score is roughly five times its value obtained in the previous within-subject design. Contrasting the two studies, one finds that AU performance is unchanged, but MA performance in the present results is reliably superior to its level in the previous study ($t_9 = 2.18$, $p < .05$). These findings add strength to the argument

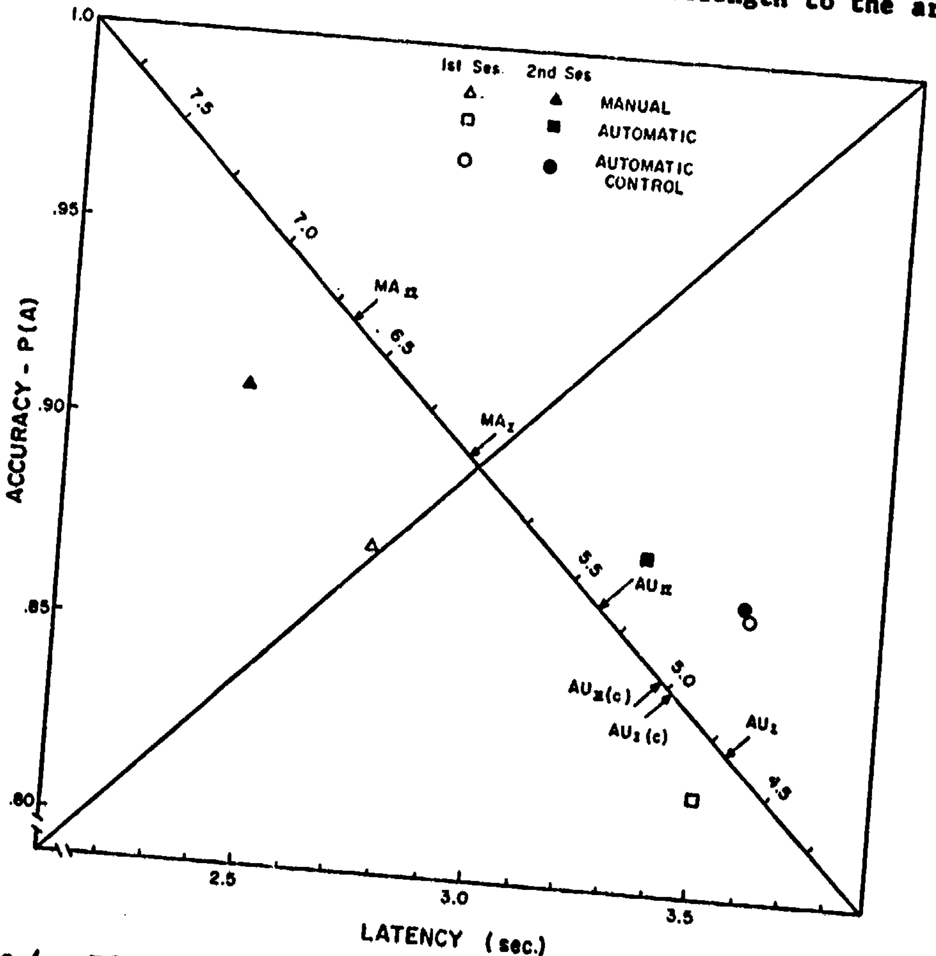


Figure 4: Effect of participatory mode and experimental condition on detection performance_Dual Task

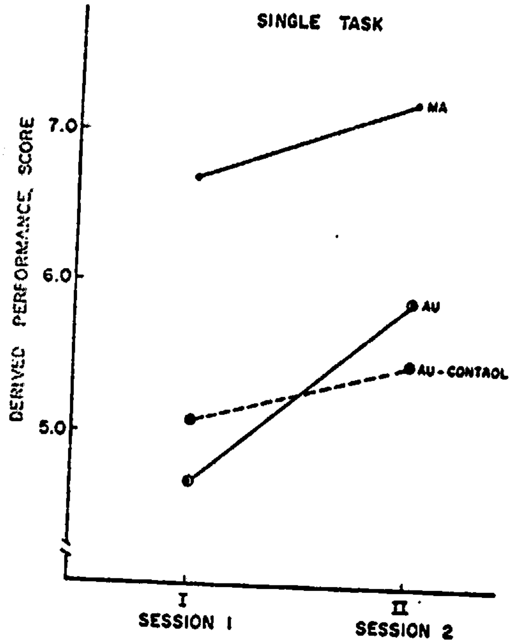


Figure 5

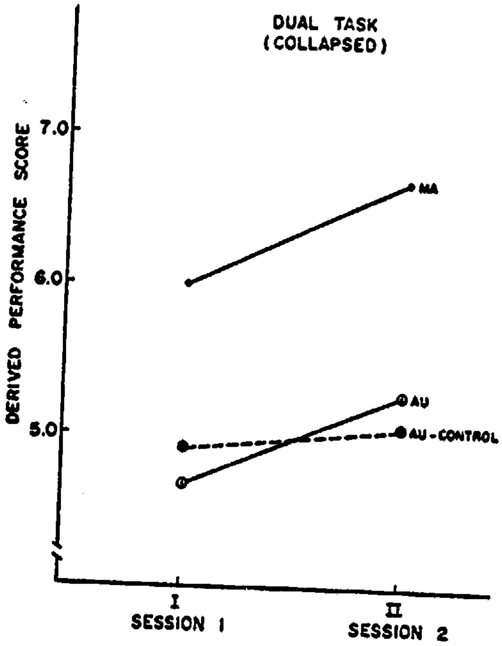


Figure 6

Detection performance as a function of experimental condition

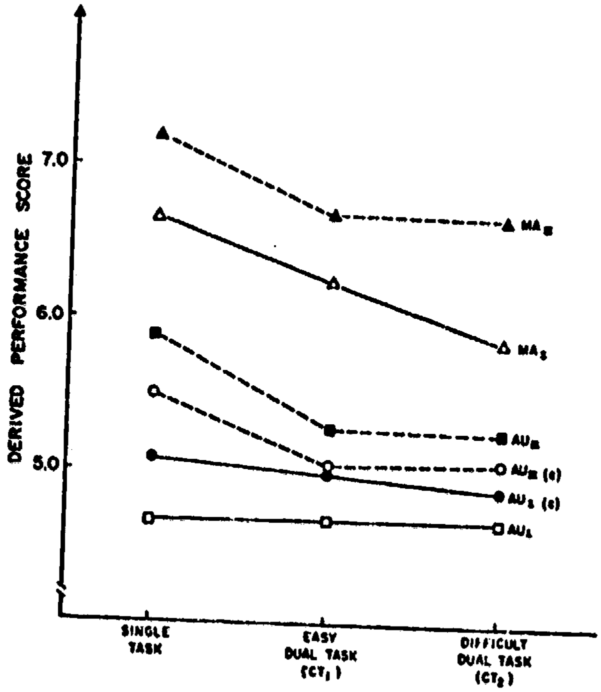
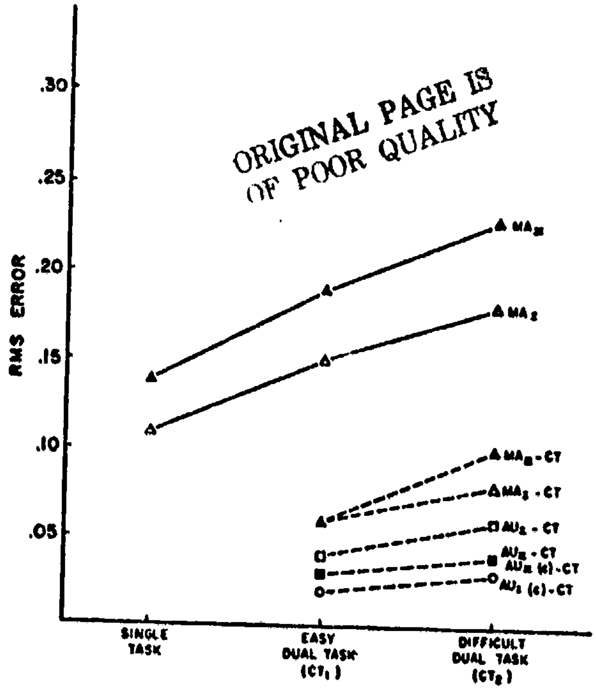


Figure 7: **Effect of dual task on detection performance**



ORIGINAL PAGE IS
OF POOR QUALITY

Figure 8: **Effects of dual task on manual tracking and critical task performance**

ORIGINAL PAGE IS
OF POOR QUALITY

that internal models developed separately tend to be more consistent, less variable and more sensitive to system changes.

By comparing the single task performance in MA_{II} with AU_{II} (see Figures 3 and 5) it is possible to determine whether MA superiority is maintained after prior training in the other mode of participation. From Figures 3 and 5 we can see that while this difference has been reduced somewhat, the overall MA superiority remains intact. This MA_{II}-AU_{II} group difference is also statistically reliable ($F_{1,10} = 6.76$, $p < .05$).

(b) Critical Task

The impact of the critical tracking task may be evaluated both as it affected detection performance (Figure 7) and, in the MA mode, as it affects performance of the primary tracking task (Figure 8). From Figure 7, it is evident that the introduction of the CT produced a decrement in detection. As might be expected, the decrement in the MA mode was somewhat more pronounced. While there was no decrement for the AU_I groups there is a substantial decrement for the AU_{II} groups, equivalent to the decrement of both the MA groups. For both the MA_I-AU_I and the MA_{II}-AU_{II} analyses, task loading showed a statistically reliable effect ($F_{2,20} = 3.60$, $p < .05$; $F_{2,20} = 5.45$, $p < .025$ respectively). It should be noted however that the primary impact of this effect is localized in the introduction of the critical task, and not with the increase in its difficulty level, a point born out by further statistical analysis. (The near equivalence of the two dual task conditions was the justification for collapsing detection performance over the two conditions in further analysis.)

Figure 8 reveals that the critical task had a clear influence on MA tracking performance, both with its introduction, and with the increasing difficulty. Analysis performed on the MA_I and MA_{II} data alone[1] indicated that the effect was statistically reliable ($F_{2,20} = 45.97$, $p < .001$).

Finally, Figure 8 reveals slight, but consistent, decreases in critical tracking performance that occur as a result of increasing λ. These increases were found to be statistically reliable for all the groups. Since the subjects were all treating the critical task as a loading task it can be concluded that the increase in λ fact did serve to divert attentional resources from the primary tracking/detection process.

(c) Transfer of Training

Manual Mode. In determining the relative amount of transfer to the manual mode resulting from prior automatic training, the MA_{II} group is compared with its control group MA_I (Figure 1) which essentially had no prior experience in the failure detection task.

---

[1] Naturally AU "tracking" performance remains unaffected by critical task difficulty level.

From Figures 3 through 7 it can be seen that in general there is an overall MA_{II} superiority over MA_{I} for both single and dual task conditions. However the ANOVA failed to reveal these differences to be statistically reliable. Examination of the data on a day by day basis reveals that the overall MA_{I}-MA_{II} difference is due to large differences that exist on day 1 which appear to dissipate completely when the two groups are compared on day 2 performance. This finding can be seen as support for the basic hypothesis that exposure to prior AU tracking and the development of an internal model based on visual cues only, produces only a small and transient facilitation of subsequent development of the internal model based on MA tracking.

Automatic Mode. The degree of transfer resulting from prior MA training to the AU mode is reflected in the performance of subjects in condition AU_{II}, and the comparison of this performance with that of the control group ($AU_{I(C)}$-$AU_{II(C)}$). In Figures 5 and 6, it is evident that the latter group failed to benefit at all from prior AU training, an observation supported by the lack of statistical reliability of the main effect when $AU_{I(C)}$ and $AU_{II(C)}$ are compared. In marked contrast, Figures 5 and 6 suggest that the AU_{II} group in fact showed considerable benefit from their prior MA training when their performance is contrasted with that of the AU_{I} group. In Figure 5, the magnitude of this effect is seen to be considerably larger than the effect for the control group or for the MA_{I}-MA_{II} contrast discussed in the preceding section.

The statistical reliability of this improvement on the single task data was assessed by a groups (AU_{I} vs. AU_{II})x days (Day 1 vs. Day 2) 2 x 2 ANOVA.

Both main effects were statistically reliable. This indicates that (a) both groups improved with practice (over two days) in their respective AU conditions ($F_{1,10} = 14.77$, $p < .001$). (b) More crucially, from the viewpoint of the hypothesis under investigation, the AU_{II} group performed reliably better than did the AU_{I} group ($F_{1,10} = 5.19$, $p < .05$). It is of course possible to argue that this effect resulted from greater exposure to and familiarity with the overall experimental environment experienced by the AU_{II} group and not to transfer of the internal model. However this interpretation appears unlikely because the control group failed to show any such "generalized" transfer.

We can conclude that there is a transfer from MA to AU. The AU_{I}-AU_{II} differences are very large and statistically reliable and as such support the basic hypothesis that while there are different sets of cues operating, the MA condition produces an internal model of the system that can be utilized to advantage in subsequent automatic monitoring.

SUMMARY AND CONCLUSIONS

The major results can be summarized as follows:

1) Detection of step increases in system order when the operator remains in the control loop (MA mode) is considerably faster and more accurate than

when he is removed (AU mode). This finding is consistent with both the findings of Young (Reference 2) and of Wickens and Kessel (Reference 7).

2) The manual mode superiority was found to be more pronounced in this between subject design than the previous within subject study (Reference 7). This difference can be attributed to the fact that the subjects were allowed to develop separate internal models for either the manual or the automatic mode, thereby producing models that were always appropriate for the mode of participation employed.

What is interesting in contrasting the two studies is the fact that AU performance is virtually identical. The effect of the between-subjects manipulation instead seems to have been to produce a large improvement in MA detection.

This result suggests that in the previous experiment the AU internal model was developed unhindered by the concurrent development of the MA internal model while the reverse situation did not hold. It would appear that the development of the MA internal model in the previous experiment was somehow subject to interference from the AU model development, suggesting that subjects were paying attention to non-relevant, visual cues. It has been argued (Reference 7) that the sensitivity to proprioceptive information is reduced relative to visual information particularly when the two sources are available at the same time and are conveying conflicting information (References 13, 14, 15). In the AU mode the subjects have only visual cues as information while in the MA mode both visual and proprioceptive information is available. Thus in the previous study, during the development of the MA internal models there were times when these cues might be in conflict and subjects tended to fall back on the visual cues learned in the AU mode. This produced an over-emphasis on the visual cues and a subsequent degrading of the crucial proprioceptive information. The introduction of the between subject design forced subjects to develop separate internal models based upon the relevant cues available within each condition--a situation that has enhanced the MA-AU differences found in the previous experiment.

3) The overall MA superiority is evident in both single and dual task conditions. The effect of adding the Critical Task was to reduce the overall detection performance via a reduction in the accuracy of detections and an increase in response latencies. The impact of the second task was more marked for the MA condition than the AU condition. This result is consistent with the fact that the critical tracking task, placing heavy demands upon the subject's response mechanism, produced an increase in interference at the structural, motor level of performance in the MA mode that was not present in the AU mode of operation. Increasing the difficulty of the subcritical loading task appeared to have little effect on detection performance in either mode, although it did serve to disrupt tracking performance.

4) An analysis of the transfer of training experiment shows that there is very little transfer from the automatic mode to the manual mode. This fact adds further weight to the argument that the development of the internal model for the manual mode cannot utilize to advantage the internal model developed

ORIGINAL PAGE IS
OF POOR QUALITY

for the automatic mode. The addition of the proprioceptive channels and the interactive describing function in the manual mode appears to require the development of a separate and unique internal model.

5) There does appear to be positive transfer from the manual mode to the automatic, a finding that supports the basic hypothesis outlined above that while there are different sets of cues operating, the MA mode produces an internal model of the system that can be utilized to advantage in subsequent automatic monitoring.

6) Finally, the successful transfer from manual to automatic and the lack of transfer from the automatic to the manual modes tends to add weight to the basic hypothesis outlined above. This hypothesis states that the internal models developed in different modes of participation are relatively independent and therefore care must be exercised in extrapolating expected results in one mode of participation from performance in the other.

REFERENCES

1. Vreuls, D., et al. Pilot Failure Detection Performance with Three Levels of Fault Warning Information. Bunker-Ramo Corp., Report No. SRDS-RD-68-9, 1968.
2. Young, L.R. On Adaptive Manual Control. IEEE Transactions on Man-Machine Systems, Vol. MMS-10, pp. 292-331, 1969.
3. Ephrath, A.R. Detection of System Failures in Multi-axes Tasks. Proceedings of the 11th Annual NASA-University Conference on Manual Control. NASA TMX, 1975, 62, 464.
4. Curry, R.E. and Ephrath, A.R. Monitoring and Control of Unreliable Systems. T.B. Sheridan and G. Johannsen, Eds., Monitoring Behavior and Supervisory Control, New York: Plenum Press, 1976.
5. Johannsen, G., Pfendler, C. and Stein, W. Human Performance and Workload in Simulated Landing-approaches with Autopilot-failures. In T.B. Sheridan and G. Johannsen, Eds., Monitoring Behavior and Supervisory Control, New York: Plenum Press, 1976.
6. Ephrath, A.R. and Curry, R.E. Detection by Pilots of System Failures During Instrument Landings. IEEE Transactions on Systems Man & Cybernetics. Vol. SMC-7, No. 12, pp. 841-848, 1977.
7. Wickens, C.D. and Kessel, C. The Effects of Participatory Mode and Task Workload on the Detection of Dynamic System Failures. 13th NASA Conference on Manual Control, NASA, TMX-73, 170, 1977.

ORIGINAL PAGE IS
OF POOR QUALITY

8. Jex, R., McDonnell, J.D. and Phatak, A.V. A 'Critical" Tracking Task for Manual Control Research. IEEE Transactions on Human Factors in Electronics, Vol. HFE-7, pp. 138-144, 1966.

9. Wickens, C.D. In N. Moray, Ed., Mental Workload: Theory and Measurement. New York, Plenum Press, 1978, (in press).

10. Watson, C.S. and Nichols, T.L. Detectability of Auditory Signals Presented without Defined Observation Intervals. Journal of Acoustic Society of America, Vol. 59, pp. 655-668, 1976.

11. Green, D.M., and Swets, J.A. Signal Detection Theory and Psychophysics. New York: John Wiley and Sons, Inc., 1966.

12. Wickens, C.W. and Kessel, C. The Effects of Participatory Mode and Task Workload on the Detection of Dynamic System Failures. Institute of Aviation Tech. Report ARL-77-8/AFOSR-1977.

13. Posner, M.I. et al. Visual Dominance: An Information-processing Account of its Origins and Significance. Psychological Review, 1976, 83, No. 2, 157-171.

14. Klein, R.M. and Posner, M.I. Attention to Visual and Kinesthetic Components of Skills. Brain Research, Vol. 71, pp. 401-411, 1974.

15. Jordan, T.C. Characteristics of Visual and Proprioceptive Response Times in the Learning of a Motor Skill. Quarterly Journal of Experimental Psychology, Vol. 24, pp. 536-543, 1972.