N82-13687 D22

FURTHER TESTS OF A MODEL-BASED SCHEME FOR PREDICTING PILOT OPINION RATINGS FOR LARGE COMMERCIAL TRANSPORTS

by

W. W. Rickard
McDonnell Douglas Corporation
Douglas Aircraft Company
Long Beach, California

and

W. H. Levison
Bolt Beranek and Newman Inc.
Cambridge, Massachusetts

ABSTRACT

A methodology is demonstrated for assessing longitudinal-axis handling qualities of transport aircraft on the basis of closed-loop criteria. Six longitudinal-axis approach configurations were studied covering a range of handling quality problems that included the presence of flexible aircraft modes. Using closed-loop performance requirements derived from task analyses and pilot interviews, predictions of performance/workload tradeoffs were obtained using an analytical pilot/vehicle model. A subsequent manned simulation study yielded objective performance measures and Cooper-Harper pilot ratings that were largely consistent with each other and with analytic predictions.

INTRODUCTION

A certain dichotomy is associated with the assessment of flying qualities. From the pilot's point of view, the flying qualities of an airplane, in a given task, relate to the degree to which satisfactory performance can be achieved with reasonable workload levels. Nevertheless, flying quality specifications are written in terms of open-loop vehicle response characteristics to help the airplane manufacturer comply with the specifications. Accordingly, considerable effort has been expended to find the combination of aircraft response parameters that will reliably predict task performance and pilot workload.

In contrast to open-loop vehicle-centered criteria, pilot/vehicle model analysis allows one to explore issues related to closed-loop performance as well as to workload demands made on the pilot. The effects of external disturbances and control/display parameters, as well as inherent pilot limitations, can be considered. Perhaps most important, predictive schemes based on pilot/vehicle analysis are not constrained to "conventional" dynamics and can therefore be applied to flying quality studies of aircraft having high-order reponse characteristics.

Hess[1] and Levison[2, 3] have proposed similar schemes, based on the optimal control model (OCM) for pilot/vehicle systems, for predicting pilot-opinion ratings. Levison's scheme was recently tested against data obtained in a previous simulation study of commercial transport handling qualities.[4] Results of this test were sufficiently encouraging to warrant further exploration of the methodology.

This paper summarizes the results of a subsequent study by Bolt Beranek and Newman Inc. and Douglas Aircraft Company in which the analytic scheme for assessing longitudinal axis handling qualities of commercial transport aircraft in the landing approach was rigorously tested. Study goals included development of closed-loop performance criteria, a tightly constrained manned simulation to yield Cooper-Harper opinion ratings[5] with minimal interpilot variability, and compilation of a data base of objective performance measures suitable for methodological development. An additional goal was to explore the effects of simulating flexible modes of transport aircraft, and to determine whether or not the analytic scheme would predict these effects.

The prediction scheme is based on the following assumptions:

1. Pilot rating is a function of the flight task
2. For a given flight task, one or more critical subtasks exist which serve as the primary determinants of pilot ratings
3. Performance requirements are well defined for each critical subtask
4. Pilot opinion is based partly on the degree to which desired performance is achieved and partly on the information-processing workload associated with the task
5. A reliable model exists for predicting performance/workload tradeoffs for relevant flight tasks.

These assumptions lead to the procedure diagrammed in Figure 1. In effect, the analytic prediction scheme parallels the procedure that would be followed in performing a well-controlled handling quality simulation study, the major difference being use of the optimal control pilot/vehicle model rather than a human to obtain pilot ratings and other performance measures.

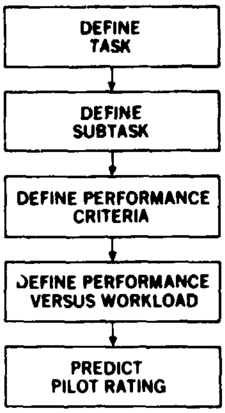


FIGURE 1. PROCEDURE FOR PREDICTING PILOT RATING

The following rating expression was developed in the preceding study and used in the current effort:

$$R = 10\left[\frac{\sigma}{\sigma + \sigma_0} + \frac{A}{A + A_0}\right] \qquad (1)$$

$$1 \leq R \leq 10$$

where R is the predicted Cooper-Harper pilot rating, σ is the probability that one or more important system variables will exceed its maximum acceptable value, and A is a measure of the relative attention (i.e., workload) associated with the task. The pilot is assumed to operate on the performance/workload tradeoff curve, predicted by the OCM, so as to minimize R.

A good fit with the experimental data of the preceding study was found using $\sigma_0 = 0.1$ and $A_0 = 2$. (Since "attention" is considered relative to that appropriate to a standardized laboratory tracking task[6, 7] rather than to some assumed capacity, values greater than unity are possible.)

The paper is organized into the following topic areas: (1) task definition, (2) pre-experiment analysis to select experimental configurations and predict aircraft handling qualities, (3) description of the experiment, (4) experimental results and analysis, and (5) conclusions.

TASK DEFINITION

The flight tasks to be performed by the test pilots were defined and closed-loop criteria were specified for model analysis. As in the preceding study,[2] the task was that of piloting a simulated large commercial transport aircraft in the final approach. Three subtasks were defined: (1) altitude stationkeeping prior to glideslope capture, (2) glideslope capture, and (3) post-capture tracking of the glideslope. Flare and landing were not considered, and were not performed in the simulation study. For purposes of pre-experimental model analysis, zero-mean turbulence as defined by the Dryden model[8] was assumed, with q and r components omitted.

To ascertain closed-loop requirements, interviews were held with five potential test pilots to determine what they considered to be maximum acceptable values, or "limits," for important system variables in moderate turbulence. (In general, the pilots interpreted a "maximum" value as an excursion indicative of poor approach performance.) Assessments were obtained for each of the flight subtasks, and for various altitudes with respect to the glideslope tracking subtask. Since model analysis was performed for frozen-point conditions appropriate to glideslope tracking at an altitude of 500 feet, the subjective acceptable excursions for that condition are reported here.

On the average, the following zero-to-peak allowable excursions from trim were specified:

Glideslope: 1/2 dot
Sink rate: 250 feet per minute
Airspeed: 7.6 knots
Pitch: 3.5 degrees
Stick: 28 percent of maximum excursion
Thrust: 4 percent of aircraft weight

For airspeed and sink rate excursions, which had asymmetric criteria, the above values reflect one-half the distance between upper and lower bounds. The limit on thrust represents a distillation of the pilot responses, which were expressed in different units (inches throttle movement, percent N_1, change in EPR) by different pilots. The pilots agreed that there was also a subjective limit to pitch rate but, as they could not assign a quantitative value to this parameter, it was excluded from the list of performance requirements.

Although the pilots did not provide subjective limitations to rate-of-change of stick and throttle, "limits" for these quantities were defined partly to satisfy certain mathematical requirements of the optimal control model and partly to satisfy physical constraints. A stick rate limit of 28 percent maximum slew rate was assumed, and the limit on rate-of-change of thrust was set equal to one-half the limit on thrust deviation to reflect the low bandwidth operation of this control.

To provide the scalar quadratic performance index needed to obtain model solutions, weighting coefficients were defined as the reciprocals of the squares of maximum allowable values. Thus, an rms deviation of a given system variable equal to its "limit" contributed one unit to the overall "cost."

PRE-EXPERIMENT ANALYSIS

This phase of the study consisted of two tasks: (1) preliminary selection of candidate aircraft configurations and (2) pre-experimental model analysis. The objectives of the latter task were to select six configurations for the simulation study and to obtain predictions of pilot ratings and closed-loop performance measures to allow rigorous testing of the analytic methodology.

PRELIMINARY SELECTION OF CANDIDATE CONFIGURATIONS

The pilot rating data from the test reported by Rickard[4] were used to demonstrate the feasibility of using the OCM to estimate pilot ratings. A thorough validation of this procedure, however, requires comparison not only of the pilot rating data but also of the workload and tracking performance data for human pilots with model predictions. These data had not been recorded in that test. Thus, another test was planned in which all the needed data would be recorded.

The first task undertaken in designing the test was the selection of a set of configurations to be evaluated. A primary goal was to produce data with high statistical reliability. This meant that many replications of the test matrix, or repeated evaluations of each configuration, would be needed. This meant that the test matrix would have to be small to keep costs reasonable.

One would prefer a large test matrix so one could evaluate the model against a wide range of airplane characteristics. Since the test matrix had to be kept small, it was decided that the configurations should be chosen to vary the most important properties. Among the issues considered critical in flying qualities today are relaxed static stability, control augmentation, and structural dynamics. A set of eight configurations was designed which varied these properties.

The flying qualities of the eight configurations, as predicted by existing criteria, are shown in Table 1. There are columns for five -8785 criteria: (1) short period frequency versus acceleration sensitivity ($\omega_{n_{sp}}$ versus n/α), (2) short period damping (ζ_{sp}), (3) phugoid damping (ζ_{ph} or $T_{2_{ph}}$), (4) static stability, and (5) flight path stability ($d\gamma/dV$). The pilot, of course, cannot be asked to rate these individual criteria; his rating of longitudinal flying qualities represents their sum. Since the -8785 provides no guidance on how to combine the pieces, one must use his own judgment. The judgment used here was to represent the "-8785 OVERALL" as the worst of the five preceding columns. The next column, labeled "BANDWIDTH," is a flying quality prediction using a frequency domain pilot-in-the-loop criterion.[4] This criterion has been demonstrated reliably to predict pilot opinion of longitudinal maneuvering dynamics. As such, it is not sensitive to $d\gamma/dV$, which is a measure of long-term flight path response. It was shown by Rickard[4] that the combination of the Bandwidth and flight path stability estimates, labeled BW + $d\gamma/dV$, yields an estimate of longitudinal flying qualities more accurate than the one labeled -8785 OVERALL. The criteria in Table 1 are the tools used to design a matrix of eight configurations for this test. Only six were simulated. Model analysis was used as described in the next section to eliminate two from the test matrix.

Configuration 1 is by time-honored tradition the baseline. According to the estimates, it has Level 1 longitudinal flying qualities. Configurations 8 and 3 explore a progression of increasing static instability, having times to double of 7.7 and 2.4 seconds, respectively. Configuration 2 was chosen to explore the issue of flight path stability, with $d\gamma/dV = 0.34$, where 0.24 is the Level 3 limit. Configurations 4 and 5 explore the issue of control augmentation. They are the same airplane, an advanced supersonic transport, without and with a full-state feedback flight control system which was designed using implicit model following. The unaugmented airplane has very poor flying qualities, while the augmented version has fair to good flying qualities, depending on the criteria used.

TABLE 1

FLYING QUALITY LEVELS OF TEST CONFIGURATIONS 1 THROUGH 8

CONFIG NO.	$\omega_{n_{sp}}$ vs n/α	ζ_{sp}	ζ_{ph} or $T_{2_{ph}}$	STATIC STABILITY	$d\gamma/dV$	-8785 OVERALL	BAND-WIDTH	BW + $d\gamma/dV$
1	1	1	1	STABLE	1	1	1	1
2	1-1/2	1	1	STABLE	WORSE THAN 3	WORSE THAN 3	1	2-1/2
3	WORSE THAN 3	2	WORSE THAN 3	UNSTABLE	1	WORSE THAN 3	3	3
4	WORSE THAN 3	1	WORSE THAN 3	UNSTABLE	3	WORSE THAN 3	3	3
5	2	1	1	STABLE	1	2	1	1
6	3-1/2	2	1	STABLE	1	3-1/2	1	1
7	3	1	1	STABLE	1	3	1	1
8	WORSE THAN 3	1	WORSE THAN 3	UNSTABLE	1	WORSE THAN 3	2	2

The last two configurations, 6 and 7, explore the effect of structural dynamics on flying qualities. Both have the same rigid-body equations, with Configuration 7 having two additional degrees of freedom representing structural dynamics. The criteria indicate that these configurations will be rated the same. The -8785 criteria, which cannot estimate the effect of structural modes, predict Level 3 flying qualities. The short period frequencies are too low and the damping ratio unacceptable or too high. The Bandwidth criterion, which should be able to predict this effect as it makes no assumption about model order, predicts Level 1 flying qualities.

MODEL ANALYSIS

Before obtaining model predictions, it was necessary to specify various independent model parameters relating to the pilot's information-processing limitations. The reader is referred to documentation from the preceding study[2] for methodological details.

Parameters reflecting limitations on the pilot's information-processing capabilities were defined. An effective perceptual threshold was associated with each perceptual variable assumed to be used by the pilot. On the basis of laboratory tracking experiments,[9] thresholds of 0.05 degree and 0.2 degree per second visual arc were assumed, respectively, for perception of the displacement and velocity of a given display indicator. Analysis of the cockpit displays yielded the following perceptual thresholds, in problem units, for an altitude of 500 feet: (a) 4.7 feet height error, (b) 19 feet per second sink rate error, (c) 4.3-degree pitch error, (d) 1.7 degrees per second pitch rate, and (e) 1.9 feet per second airspeed error. The rather large threshold associated with perception of sink rate error was a consequence of assuming that the pilot attempts to obtain this information from the velocity of the glideslope indicator. In addition, a "residual noise" of 0.5 degree was associated with perception of pitch error to account for the lack of an explicit zero-error reference.

To simplify the analysis, the pilot was assumed to pay equal attention to glideslope, pitch, and airspeed indicators (and was assumed to obtain both displacement and rate information from all but the airspeed indicator). In addition, 34 percent of the attention was assumed "lost" in scanning. Thus, a relative attention of unity corresponded to relative attentions of 22 percent each to glideslope, pitch, and airspeed variables. As described in the literature, attentional and perceptual factors determined the observation noise variance associated with each perceptual input.[10] The remaining independent model parameters were time delay (0.29 second to account for both pilot and control-actuator delays) and motor noise covariance (-50 dB, relative to predicted control-rate variance).

Curves of predicted performance versus relative attention, generated via the optimal control model, are shown in Figure 2a. The quadratic performance index was based on assumed maximum allowable excursions for important system variables, as described earlier in this paper. Variations in "attention" were reflected by appropriate manipulation of the baseline observation noise/signal ratio as described in Levison.[2] As recommended by the military flying quality specifications,[8] the analysis was performed using a Dryden gust model having parameters appropriate to an altitude of 500 feet and longitudinal and vertical rms gust amplitudes of 10 and 6.6 feet per second, respectively.

Figure 2a shows the following trends:

1. Best achievable performance (i.e., lowest cost) with the baseline aircraft (Configuration 1) and the augmented AST (Configuration 5).
2. Worst performance, and greatest sensitivity to attentional workload, with the unstable Configurations 3 and 4.
3. Intermediate performance with the configuration having a mild instability (Configuration 8) and the vehicle having adverse dγ/dV (Configuration 2).
4. Negligible effects due to simulation of flexible modes (Configurations 6 and 7).

As noted earlier, the scope of the manned simulation study was limited to six experimental configurations. On the basis of this analysis, Configurations 4 and 5 were dropped from further consideration as they appeared to be similar in terms of performance/workload tradeoffs to Configurations 3 and 1, respectively.

Application of the rating expression of Equation (1) to the performance/workload predictions shown in Figure 2a yielded unreasonably large Cooper-Harper ratings (e.g., a rating of 8 for the baseline configuration). Partly for this reason, and partly because the -8785B backup document[11] indicates that the initial choice of gust intensities represents a low probability (1 percent) of occurrence, gust intensities were halved for subsequent analysis and experimentation. The reduced levels represent a 50-percent probability of occurrence.

In addition to the reduction in gust levels, changes in other independent model parameters were made prior to reanalysis. The allowable performance "window" for glideslope error was increased to 1 dot to reflect published Category II specifications.[12] The performance window for sink rate was increased from around 4 feet per second to 7.5 feet per second to allow for the fact that, in actual flight, the flare maneuver would substantially reduce the sink rate prior to impact. We assumed that the pilot would obtain sink rate information from the vertical speed indicator, and we decreased the perceptual threshold to 0.8 feet per second to reflect assumed visual resolution capabilities with respect to this instrument. The maximum acceptable value for the rate of thrust change was reduced to one-fifth the corresponding limit on thrust deviation to more strongly reflect the pilot's aversion to frequent changes in throttle setting. Finally, the OCM was used to predict optimal allocation of attention.

The six configurations retained for the simulation study were reanalyzed as described above: the resulting performance/workload tradeoffs are shown in Figure 2b. As is the case with the initial analysis, the penalty for relatively low attention is greatest for Configuration 3, and inclusion of flexible modes

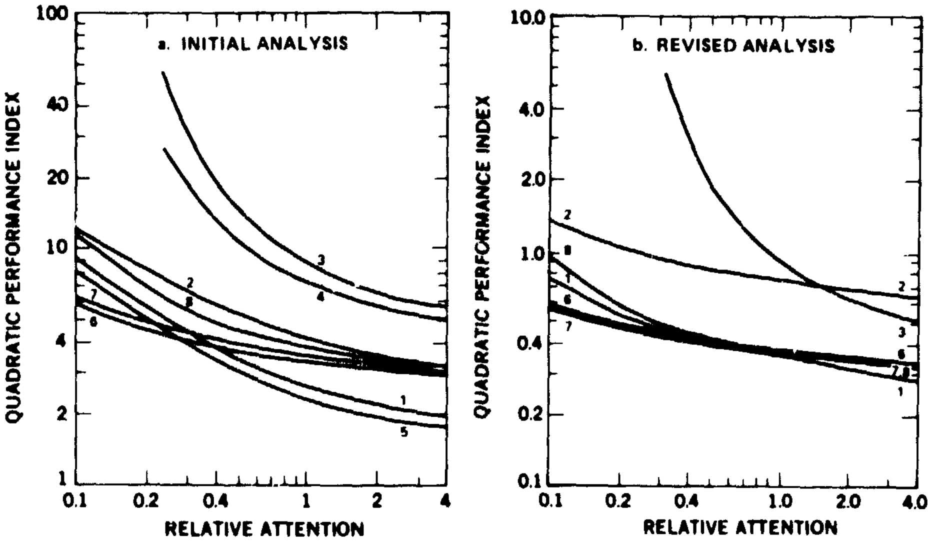


FIGURE 2. PERFORMANCE AND WORKLOAD TRADEOFFS

has little predicted influence. The predicted performance/workload tradeoff curves are compressed, however, with little separation among the curves for Configurations 1, 3, 7, and 8 at all but the lower attentional levels. Application of the rating expression of Equation (1) yields predicted ratings (shown later in this paper) that range from Level 1 to Level 3 and are consistent with those observed experimentally for Configurations 1 and 3 in the preceding study.[2]

DESCRIPTION OF EXPERIMENTS

SIMULATION CHARACTERISTICS

The simulation model used with all configurations was a complete airplane. Both longitudinal and lateral-directional degrees of freedom and controls were provided. The controls (column, wheel, and pedals) were DC-10 hardware. Control feel, force gradients, and motion limits were based on the DC-10. A full set of DC-10 instruments was provided. A flight director display was available but not used as this would affect workload and performance and, thus, pilot opinion of flying qualities. Actuator and engine dynamics typical of wide-body aircraft were simulated. Standard linearized equations of motion were used in the simulation. Euler integration of the differential equations was performed at 20 hertz. The actuators and other elements with fast dynamics were simulated using difference equations.

The simulator is Douglas' research and development motion base simulator. The motion platform is shown in Figure 3 and the cab interior in Figure 4. The cab is a DC-10 cockpit with stations for captain, first officer, flight engineer, and observer. Synthetic outside vision is available but was not used in this

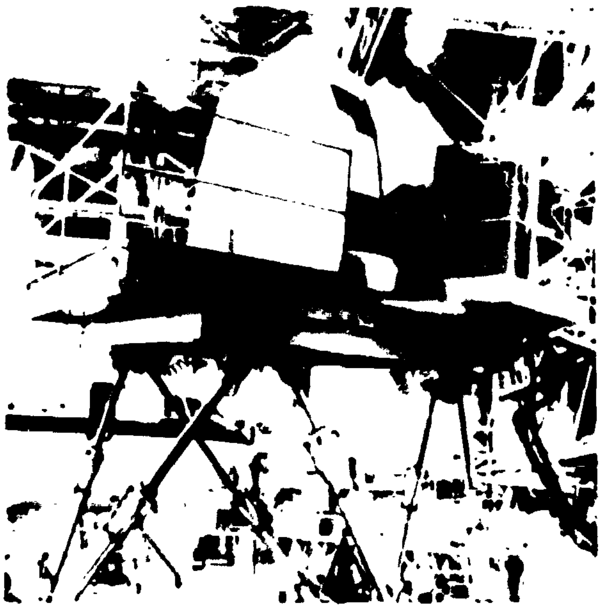


FIGURE 3. SIX-AXIS MOTION BASE SIMULATOR

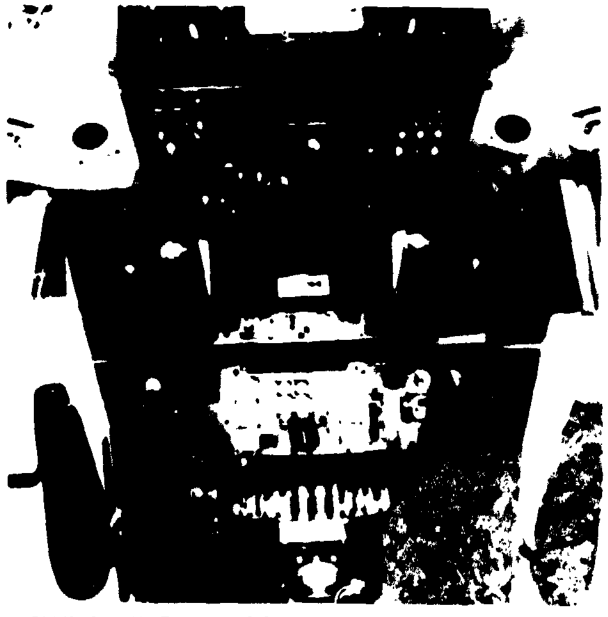


FIGURE 4. MOTION BASE SIMULATOR TRANSPORT COCKPIT

experiment. Motion limits for the platform are given in Table 2 for a moving mass of 22,000 pounds. The bandwidth for small inputs is 1 hertz, which can be boosted to at least 2 hertz by use of pre-emphasis filters on the motion drive signals. The evaluation pilot sat in the captain's seat and the test engineer in the first officer's seat. The engineer controlled all aspects of the experiment from this position, once the computers were started, using the control box shown in Figure 5. The box has six 3-position toggle switches, six momentary switches and sense lights, and five 16-position thumbwheels with LED readouts above. The box is on an umbilical so that it can be moved around the cab. The software "reads" these switches, performs the commanded functions, and displays the appropriate information on the displays.

In this experiment, only a few switches were used. One of the 3-position switches was used to turn turbulence off or on, one thumbwheel was used to select the configuration, and the pilot number was set using another thumbwheel. Three digits of the LED display showed the run number, another the pilot number, and another the configuration number currently being used by the computer. Another panel contained pushbutton switches to control start, stop, reset, and other operational functions.

The task flown was a manual instrument landing system (ILS) approach using raw rather than director, glideslope, and localizer data. Plan and side views of the approach geometry are shown in Figure 6. The Dryden turbulence model[8] was used, with 50th percentile intensities and scale lengths for an altitude of 500 feet. This model actually varies with airspeed and altitude, but was "frozen" in the experiment to match the stationary nature of the OCM. A sum-of-sines implementation was used, which concentrates the power at discrete frequencies. Twelve discrete frequencies from 0.0838 to 12.57 radians per second were used.

EVALUATIONS

The evaluations were made by four Douglas experimental test pilots, all of whom had prior experience in motion-base-simulator evaluations of flying qualities. Before the evaluations were begun, a checkout pilot flew the entire test matrix. No discrepancies were found except that, contrary to all predictions, Configurations 6 and 7 were unflyable. The coupling between airspeed and column movement was so tight that loss of control was inevitable if flight-path control was attempted. These were the elastic AST configurations. The problem was an unusually high value of X_q. This had been noticed in the pretest analysis, but was accepted when the flying-quality estimates turned out to be reasonable. Configurations 9 and 10 were developed to replace 6 and 7 by reducing X_q to a small value and increasing X_u to compensate. The flying quality parameters and estimates (see Table 3) for 9 and 10 were virtually identical to those for 6 and 7, yet 9 and 10 flew quite well.

Of the four pilots involved in the evaluations, two made four replications of the test matrix and two made five. Each session

TABLE 2
MOTION LIMITS

AXIS	EXCURSION	VELOCITY	ACCELERATION
HEAVE	±42 IN.	39 IN./SEC	1.65 g
SWAY	±67.5 IN.	67 IN./SEC	2.43 g
SURGE	±65 IN	71 IN./SEC	1.50 g
ROLL	±30.7 DEG	35.6 DEG/SEC	7.8 RAD/SEC²
PITCH	±33.3 DEG	33.6 DEG/SEC	7.8 RAD/SEC²
YAW	±38.7 DEG	36.3 DEG/SEC	7.9 RAD/SEC²

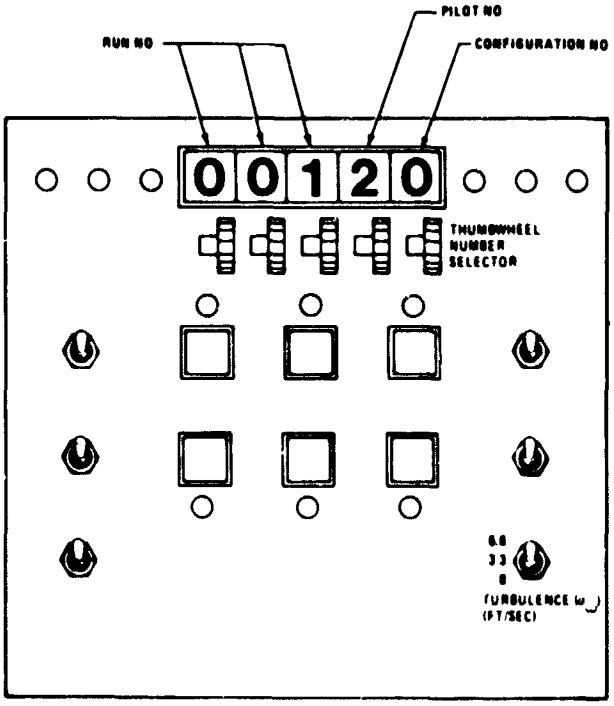


FIGURE 5. SIMULATOR CONTROL BOX

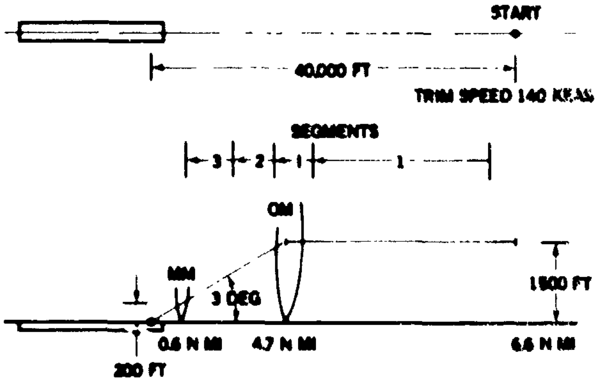


FIGURE 6. APPROACH GEOMETRY

began with a briefing in which the test procedure and performance standards were reviewed. The pilot then flew several approaches to warm up, or get used to, the equipment and procedure. He then flew two approaches with each configuration (turbulence off, then on). The configurations were presented in pseudo-random order, with the order balanced across replications. The "turbulence off" runs were flown to allow additional practice, to isolate the turbulence effects, and to gather data for the development of a glideslope capture strategy. The pilots could do whatever maneuvers and experiments they wanted in these runs after the intercept maneuver. In the "turbulence on" runs, however, they were told to track the ILS to the performance standards at all times. A replication of the test matrix took 1-1/2 to 2 hours. A total of 319 approaches were flown for data in the test; 25 more were flown by the checkout pilot. At the end of the test, the pilots were quizzed again about a number of items, including the performance standards they flew to in the test and how they allocated their attention.

A great deal of objective and subjective data was taken. Time histories of 50 variables were recorded on digital magnetic tape at 5 hertz for every approach. Stripchart records of 16 variables were made. The mean, root mean square, maximum and minimum values, and standard deviation of 15 variables were computed on-line and output on a line printer at the end of each run. Instantaneous values of 10 variables at 10 points along the approach were also printed out. The line printer was also used to record bookkeeping information, such as run start time and date, run number, configuration number, pilot number, etc., to reduce the test engineer's workload. The subjective data taken included Cooper-Harper pilot ratings, effort ratings for the three subtasks and three aspects of control, and pilot comments. The engineer made brief handwritten notes to supplement the complete record made by the cockpit voice recorder

EXPERIMENTAL RESULTS

Statistical analysis was performed on both Cooper-Harper ratings and closed-loop performance metrics. Ratings were first averaged across replications to obtain an average rating per condition per pilot. Population means and across-subject standard deviations were then computed from the individual subject means.

Statistics on system "errors" (deviations from trim) were computed for the three steady-state-like segments of the approach. Results for the final segment of the glideslope tracking task – corresponding to descent from approximately 700 to 200 feet altitude – are reported here. The mean and variability components of each error variable were analyzed separately. Only response variability is reported here, as only that error component can be compared with model predictions for the glideslope tracking task (remember that the external disturbances were zero-mean processes). Mean error is primarily reflective of piloting strategy (e.g., carry excess airspeed, "duck under" the glideslope) and therefore is not treated directly by the OCM. In general, the variability component was dominant. A variance score was first computed for each error variable of interest within a given replication. The square root of this measure was then treated as the basic error score. Note that this measure reflects within-trial variability, not run-to-run or pilot-to-pilot variability. Error scores computed in this manner were then subjected to the same statistical analysis as that described earlier for the pilot ratings.

TABLE 3

FLYING QUALITY LEVELS OF TEST CONFIGURATIONS 1 THROUGH 10

CONFIG NO.	$\omega_{n_{sp}}$ vs n/α	ζ_{sp}	ζ_{ph} or $T_{2_{ph}}$	STATIC STABILITY	$d\gamma/dV$	-8785 OVERALL	BAND-WIDTH	BW + $d\gamma/dV$
1	1	1	1	STABLE	1	1	1	1
2	1-1/2	1	1	STABLE	WORSE THAN 3	WORSE THAN 3	1	2-1/2
3	WORSE THAN 3	2	WORSE THAN 3	UNSTABLE	1	WORSE THAN 3	3	3
4	WORSE THAN 3	1	WORSE THAN 3	UNSTABLE	3	WORSE THAN 3	3	3
5	2	1	1	STABLE	1	2	1	1
6	3-1/2	2	1	STABLE	1	3-1/2	1	1
7	3	1	1	STABLE	1	3	1	1
8	WORSE THAN 3	1	WORSE THAN 3	UNSTABLE	1	WORSE THAN 3	2	2
9	3	2	1	STABLE	1	3	1	1
10	2	1	1	STABLE	1	2	1	1

Measured and predicted pilot opinion ratings are presented in Figure 7a. Across-subject standard deviations (designated by brackets) were generally less than one rating unit. Thus, the experimental technique yielded rating predictions that were reasonably consistent across pilots. The trend of the experimental ratings agreed well with pre-experimental model predictions: Configurations 1, 8, 9, and 10 were rated similarly whereas Configurations 2 and 3 received ratings that were appreciably more adverse. The major discrepancy between prediction and experiment was the relative compression of the simulation results, with the "better" configurations receiving Level 2 rather than the predicted Level 1 ratings. In addition, Configurations 2 and 3 were rated nearly the same on the average, whereas the analytic technique predicted a 2-unit spread.

In a previous study in which lateral-directional characteristics were considered to reflect Level 1 handling qualities, the "baseline" Configuration 1 received an average pilot rating in the Level 1 range.[4] Lateral characteristics were less favorable for the study reported here, receiving rating scores in the Level 2 range. Thus, we suspect that the greater-than-expected rating scores for Configurations 1, 8, 9, and 10 reflected, in part, an interaction with the lateral-axis tasks. (Model predictions were based on the assumption that the lateral-axis task would present no appreciable handling-quality problems.)

Configuration 2 was also explored in the previous study. In that study, as well as in the current one, the rating score obtained in the simulation study was higher than predicted analytically. As discussed shortly, this model/experiment difference may be due in part to a failure of the analytic scheme, as described so far, to consider the adverse effects of requiring loop closures that are not part of the pilot's standard repertoire.

Predicted and experimental measures of the quadratic performance index are compared in Figure 7b. Two sets of model predictions are shown: scores obtained with relative attentions corresponding to minimum ratings as determined from the expression of Equation (1), and scores corresponding to a relative attention of unity. Although measured scores were considerably greater than predictions, predicted trends were confirmed. As with the rating scores, performance scores for Configurations 1, 8, 9, and 10 were similar, whereas substantially greater (less favorable) scores were observed for Configurations 3 and 4.

A comparison of predicted and measured "error" variability scores for selected response variables is given in Figure 8. Again, measured scores were greater than analytic predictions, but trends related to the effects of vehicle characteristics were generally in agreement. In particular, the analytic procedure correctly predicted that relatively large elevator deflections would be required for Configuration 3, whereas large thrust changes would be required for Configuration 2. Overall, the two methods of predicting objective performance scores replicated experimental trends with similar fidelity.

Additional model analysis was conducted to determine methods for obtaining a more accurate assessment of the adverse handling qualities associated with Configuration 2, and for predicting the severe controllability problems found experimentally with Configurations 6 and 7. Compared to the baseline configuration, these three required a strategy that relied more heavily on throttle for height control and elevator for speed control: Configuration 2 because of adverse $d\gamma/dV$ characteristics, and Configurations 6 and 7 because of a high pitch/speed coupling. This observation suggested a simple technique for analytically detecting handling quality problems associated with undesirable throttle activity: model analysis was performed with and without the throttle control active. To test the discrimination of the procedure, model analysis was

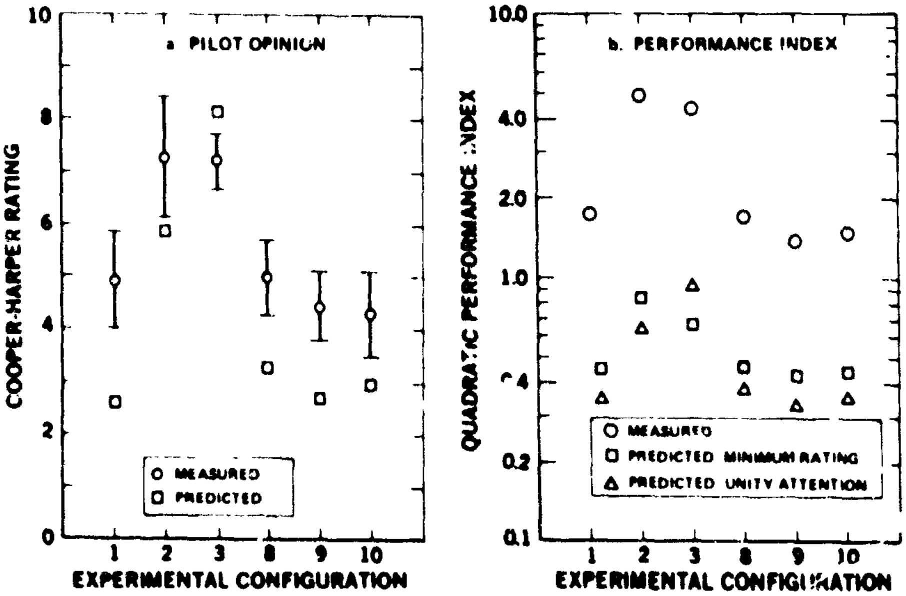
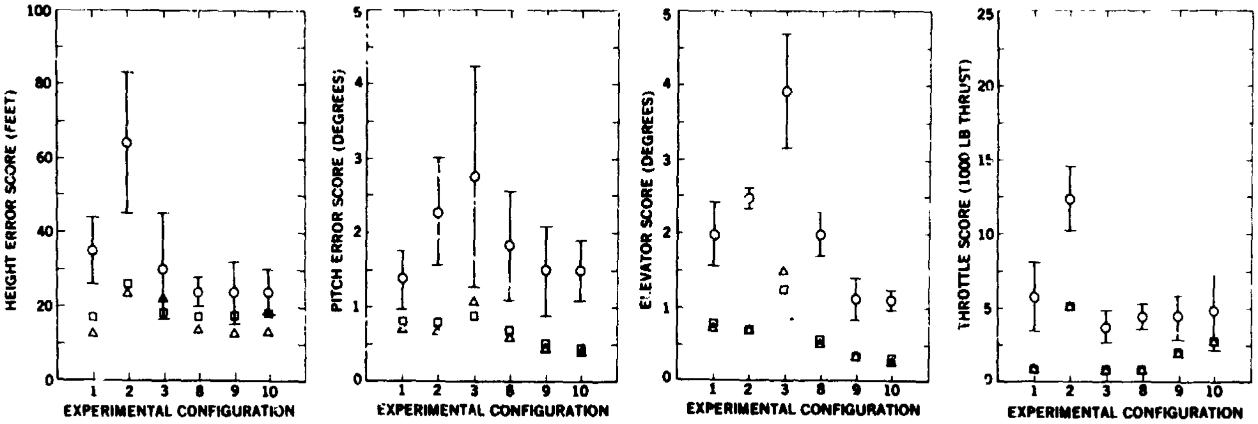


FIGURE 7 COMPARISON OF CRITERIA PREDICTIONS AND RESULTS



a. HEIGHT ERROR

b. PITCH ERROR

c. ELEVATOR

d. THROTTLE

FIGURE 8. PREDICTED AND MEASURED PERFORMANCE AND WORKLOAD SCORES

performed again for Configuration 1 (baseline), Configuration 3 (greatest instability), and Configuration 9 (configuration 6 modified to reduce the pitch-speed coupling). This analysis was performed with the baseline observation noise/signal ratio adjusted to reflect unity relative attention.

Figure 9 shows that this method readily identified handling quality problems related to throttle activity. The predicted quadratic performance indices for Configurations 1, 3, and 9, while different from each other, were relatively unaffected by the exclusion of throttle control. On the other hand, omission of throttle control caused the performance metric to more than double for Configuration 2 and to increase nearly sevenfold for Configuration 6. Thus, a model comparison of this sort appears to be a simple device for predicting handling quality difficulties caused by requirements for significant throttle activity.

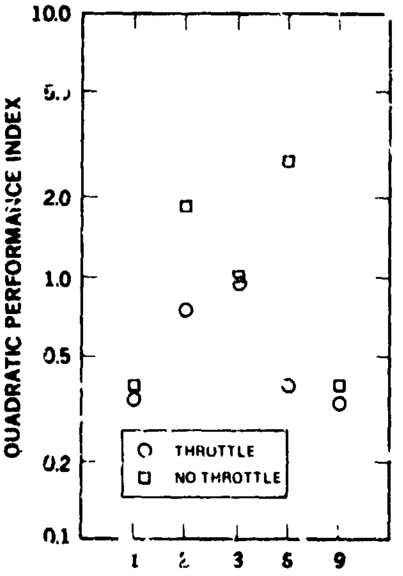


FIGURE 9. EFFECT OF THROTTLE CONTROL ON PREDICTED PERFORMANCE INDEX

CONCLUSIONS

A number of conclusions can be drawn from the analytical and experimental results of the subject study program.

1. A closed-loop criterion, or estimator of flying qualities, has been developed and validated against experimental data. The development of the model comprising the criterion was based on the characteristics of the task being modeled, not on the data. That is, it is not simply a model fit to a set of data, it is a simulation model. The characteristics used to develop the model include pilot preferences (determined from interviews), human capabilities (determined from laboratory experiments), the physics of the situations, and engineering judgment.
2. In the experiment, repeatable Cooper-Harper pilot opinion ratings were obtained by strict experiment protocol. Important aspects of the experiment design were:
   a. Design of a task with only zero-mean, stationary, random disturbances
   b. Well-defined subtasks and associated performance standards
   c. Adequate pilot familiarization.
3. A data base has been developed and recorded which includes pilot ratings and objective performance measures. The data have been used in validating the optimal control model as described herein, and will be used for further development in the future.
4. The model correctly predicted an absence of an effect of the structural modes simulated on handling qualities.
5. The model correctly predicted the experimental trends in workload, performance, and pilot ratings. This was an actual prediction: it was done before the experiment.
6. The ability of the model to predict low-frequency flight path control problems was demonstrated. These problems are detected using a two-step process. First, model predictions are made assuming throttle as a control. Then model predictions are made without throttle as a control

and compared with the previous analysis. If there is a major degradation, flight path control problems can be expected.

7. The experimental error scores were about twice those predicted analytically. The potential causes of this discrepancy include:
   a. Perceptual and indifference thresholds were perhaps greater in a realistic flight task than in a laboratory tracking task
   b. Possible discontinuous control behavior by test pilots
   c. Interference between longitudinal and lateral axes
   d. Less training in this experiment than in many laboratory tracking experiments.

   The discrepancies are not readily attributable to differences between predicted and actual tradeoffs of error and control: path, altitude, speed, and control variations were all greater than predicted.
8. Given the state of the art, the analytic scheme developed here is recommended for use in predicting important trends rather than absolute performance scores.
9. It is not necessary to select attention according to the rating expression, Equation (1). The quadratic performance index predicted for constant attentional workload mimics trends of pilot ratings. These results confirm an earlier study of Hess.[1]

REFERENCES

1. Hess, R.A., "Prediction of Pilot Opinion Ratings Using an Optimal Pilot Model", Human Factors, 1977, 19(5).
2. Levison, W.H., "A Model-Based Technique for Predicting Pilot Ratings for Large Commercial Transports", NASA CR-3257, April 1980.
3. Levison, W.H., "A Model-Based Technique for Predicting Pilot Opinion Ratings for Large Commercial Transports", Proc. of the 16th Annual Conference on Manual Control, Cambridge, MA, May 1980.
4. Rickard, W.W., "Longitudinal Flying Qualities in the Landing Approach", Proc. of the 12th Annual Conference on Manual Control, NASA TM X-73, 170, May 1976.
5. Cooper, G.E., and Harper, R.P., Jr., 'The Use of Pilot Rating in the Evaluation of Aircraft Handling Qualitites", NASA TN D-5153, April 1969.
6. Levison, W.H., Elkind, J.I., and Ward, J.L., "Studies of Multi-Variable Manual Control Systems: A Model for Task Interference", NASA CR-1746, May 1971.
7. Levison, W.H., "A Model for Mental Workload in Tasks Requiring Continuous Information Processing", Mental Workload Its Theory and Measurement, Plenum Press, New York and London, 1979.
8. "Flying Qualities of Piloted Airplanes", Military Specification, MIL-F-8785B (ASG), 1969.
9. Levison, W.H., "The Effects of Display Gain and Signal Sandwidth on Human Controller Remnant", AMRL-TR-70-93, March 1971.
10. Baron, S., and Levison, W.H., "The Optimal Control Model: Status and Future Directions", Proc. of the 1980 International Conference on Cybernetics and Society, Cambridge, MA, October 1980.
11. Chalk, C.R., Neal, T.P., Harris, T.M., Pritchard, F.E., "Background Information and User Guide for MIL-F-8785B (ASG), 'Military Specification-Flying Qualtities of Piloted Airplanes' ", AFFDL-TR-69-72, August 1969.
12. "DC-10 Flight Study Guide", Douglas Aircraft Company, June 1975.