AIAA 80-1573R

Model-Based Handling Qualities Assessment Technique for Large Commercial Transports

William H. Levison*
Bolt, Beranek, and Newman, Inc., Cambridge, Mass.

A model-based technique for assessing longitudinal-axis transport handling qualities is described. Features of this procedure, which is based on the optimal control model for pilot/vehicle systems, include 1) the capability to treat "unconventional" aircraft dynamics, 2) a relatively free-form pilot model, 3) a scalar metric for attentional workload, and 4) a well-defined manner of proceeding from descriptions of the flight task environment and requirements to a prediction of pilot opinion rating. The method is able to provide a good match to a set of pilot opinion ratings obtained in a manned simulation study of large commercial aircraft in landing approach.

Introduction

MANUFACTURERS of commercial aircraft require more general and more reliable methods of predicting aircraft handling qualities than currently exist. Published criteria have been developed primarily for military aircraft and have been validated largely for high-performance aircraft such as fighters. Furthermore, these criteria are based largely on vehicle open-loop response and are only vaguely task-oriented.

This paper summarizes the results of a study performed by Bolt, Beranek, and Newman, Inc., with the aid of Douglas Aircraft Company, to develop and test a model-based technique for predicting the influence of aircraft response parameters and other relevant factors on pilot opinion ratings. While the procedure is intended to have general application, the focus in this paper is on large transports. Further documentation of this study is provided in Ref. 1.

Vehicle-Centered Handling Qualities Criteria

Handling qualities specifications are based almost exclusively on open-loop vehicle response characteristics,[2] and criteria are specified for both transient response and frequency response characteristics. Requirements of this sort allow the aircraft manufacturer to evaluate aircraft performance through a series of relatively straightforward in-flight tests.

Despite the relative convenience with regard to compliance testing, application of vehicle-centered handling qualities specifications to large commercial transports is limited in a number of ways. For example: 1) existing handling qualities criteria have been developed primarily for high-performance military aircraft; 2) most existing criteria are based on simple models of aircraft dynamics in which phugoid and short-period response characteristics can be distinguished—a condition not necessarily met by aircraft having relaxed static stability and substantial control augmentation; 3) turbulence effects are largely ignored; 4) the display environment, which may influence overall mission suitability, is not considered; and 5) present methods do not consider effects of dynamic aeroelasticity.

Model-Based Schemes for Predicting Handling Qualities

Pilot/vehicle analysis can allow considerably greater insight into the handling qualities of an aircraft control system than can be obtained by analysis of open-loop response, and the demands made on the pilot can be explored. The effects of external disturbances and control/display parameters, as well as inherent pilot limitations, can be considered. Furthermore, assessment schemes based on pilot/vehicle analysis are not constrained to deal with "conventional" dynamics and are thus potentially more general than techniques based solely on vehicle modal response characteristics.

Until recently, application of pilot/vehicle analysis to studies of vehicle handling qualities has been based primarily on servo (or "classical") control techniques. Perhaps the most comprehensive effort of this type has been conducted by R.O. Anderson and his associates in the development of the "Paper Pilot" analysis scheme.[3] This scheme relates pilot rating to metrics of both closed-loop system performance and pilot workload, and it introduces the concept that the pilot operates to minimize his rating score.

Pilot rating is assumed to be an explicit function of system performance and pilot lead requirements. (Lead compensation is the index of pilot workload in this scheme.) A frequency response pilot model is used; and pilot parameters are adjusted to minimize pilot rating. Good matches to experimental data have been obtained for a variety of control tasks through appropriate formulation of the rating expression and adjustment of the relative weighting coefficients associated with performance and workload (i.e., pilot lead).[3-6]

While the "Paper Pilot" scheme realizes many of the advantages of a model-based approach, its applicability is limited by lack of general rules for choosing the form of the rating expression and for quantifying the various weighting coefficients. Other factors limiting the generality of this and other procedures based on servo-theory models include 1) the use of a relatively constrained fixed-form pilot model; 2) the need to assume specific loop closures prior to analysis; 3) a cumbersome treatment of pilot workload, especially when multiple loops are closed; and 4) the inability to account directly for factors related to the perceptual environment (e.g., perceptual resolution limitations, whole-body motion cues).

Hess[7] has described a scheme for predicting pilot ratings based on optimal (or "modern") control theory. He suggests an index of performance that consists of a weighted sum of integral- (or mean-) squared error and control terms. "Error" is a vector quantity that consists of the system variables that the pilot wishes to maintain within acceptable limits. The pilot is assumed to adopt control and estimation strategies that minimize this performance index.

Presented as Paper 80-1573 at the AIAA 7th Atmospheric Flight Mechanics Conference, Danvers, Mass., Aug. 11-13, 1980; submitted Oct. 8, 1980; revision received July 1, 1981.

*Senior Scientist. Member AIAA.

Hess tested his scheme against 19 different configurations covering a range of pilot ratings. "Cost" coefficients of the quadratic performance index were chosen to match experimental scores, and pilot-related model parameters were chosen partly on the basis of previous results and partly to match observed performance. Pilot ratings could be matched to within ± 1 rating unit by a linear relationship between pilot rating and the logarithm of the performance index. More recently, Schmidt has used this prediction scheme as the basis for a model-based control design procedure.[8]

Although not validated as a reliable predictive tool, Hess' procedure lays the foundation for a scheme that seems to overcome some of the limitations inherent in techniques based on classical servo analysis. The basic form of the performance index is consistent across tasks; the form of the pilot model and nature of loop closures are determined by the optimal pilot model and need not be specified by the user; a scalar metric of workload is provided; and factors related to perceptual environment are considered.

Perhaps the most severe limitation of the optimal-model-based approach, as developed so far, is the requirement to specify numerous task- and pilot-related model parameters. Another limitation, in the opinion of this author, is the lack of a metric for information-processing workload that is distinct from the performance metric.

The methodology described in this paper builds upon the work of Hess and encompasses a pilot rating prediction scheme based on the optimal-control model for pilot/vehicle performance. Emphasis is placed on the predictive aspects of the procedure, and a rationale is offered for selecting model parameters on the basis of an adequate description of the task and in the absence of experimental data. In addition, a well-defined model parameter is suggested as a potential scalar workload metric for the purpose of predicting closed-loop performance as a function of attentional workload.

Methodology

Basic Approach

The handling qualities assessment scheme is based on the following assumptions: 1) pilot rating is a function of the flight task; 2) for a given flight task there exist one or more critical subtasks which serve as the primary determinants of pilot rating; 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; and 5) a reliable model exists for predicting performance/workload tradeoffs for relevant flight tasks.

The following steps are required for predicting an average pilot rating for a specific situation.

1) Task definition: Pilot opinion ratings are task dependent. For example, the rating associated with a specific vehicle, relative to other vehicles or other configurations of the same basic airframe, may not be the same in final approach as, say, in high-altitude cruise. Therefore separate assessments must be made for each flight task of interest.

2) Subtask definition: Use of the methodology requires a quantitative description of the specific task or subtask for which predictions are to be obtained. For example, if ratings are desired for landing approach, a critical aspect of that task (say, ILS tracking) must be quantified. Task specification requires a linearized description of vehicle dynamics plus a quantitative description of the external disturbances and command inputs.

3) Define performance criteria: Performance criteria must be defined in precise quantitative terms. In order to obtain performance/workload predictions with the pilot/vehicle model used in this procedure, a quadratic performance index containing error- and control-related terms must be specified. The user must specify both the terms to be included in the performance index as well as values for the cost weighting coefficients. Cost weighting coefficients based on assumed maximum allowable values are suggested. As illustrated below, these coefficients are determined partly from the physical constraints of the flight control system, partly from objective performance requirements of the closed-loop system, and partly from pilot preference. The performance criterion used in the rating expression should be a monotonic function of this quadratic performance index.

4) Predict performance/workload tradeoff: The "optimal-control" pilot/vehicle model is used to predict performance as a function of information-processing workload. "Workload"—considered synonymous with "attention" in the context of the model—is defined in terms of a model parameter relating to the human operator's signal-to-noise ratio characteristics.

5) Predict pilot rating: The results of the preceding step are used in a rating expression to predict the pilot rating. If experimental data are available for the task of interest, a regression analysis is performed to "calibrate" the independent parameters of the rating expression; in this case, absolute rating predictions are obtained. In the absence of such calibration data, rating parameters are adjusted on the basis of previous results, and rating predictions are interpreted on a relative basis with regard to predictions obtained for other vehicle configurations.

Pilot/Vehicle Model

The analytic technique described in this paper is built around the so-called "optimal control" model for pilot/vehicle systems. The theoretical foundation for this model has been described in the literature, and the model has been validated for both simple laboratory tracking tasks as well as for more complex control situations. As discussed above, this model has also been shown to yield good handling qualities predictions.[7]

Only key features of the model are summarized here. Readers unfamiliar with the optimal control model are directed to the references cited in Ref. 1.

The human operator is assumed to adopt strategies of state estimation and control that minimize a scalar quadratic performance index. For airplane piloting tasks, this performance index consists of "error" terms relating to path, attitude, speed, and control variables. Pilot-related limitations reflected in the model include information processing delay, response bandwidth limitations, and limitations associated with attention sharing and perceptual resolution.

Attentional and perceptual limitations are accounted for by a set of "observation noise" parameters. Each perceptual variable utilized by the pilot is assumed to be perturbed by a Gaussian white noise process linearly independent of other such noises and of external inputs to the system. In the case of an idealized single-variable tracking task, the variance of each observation appears to scale with the variance of the associated perceptual variable. In this case, response randomness is accounted for by a noise/signal ratio. A more complex definition of the observation noise variance has been derived for situations in which perceptual resolution limitations are important.[9]

The model is able to reproduce pilot response behavior in a number of simple laboratory tracking tasks with a nearly constant value of noise/signal ratio of about 0.01 (i.e., -20 dB). The consistency of this parameter across tasks and across subject populations suggests that it reflects a basic central-processing (rather than perceptual or motor) limitation, and these results have led to the following model for central attention sharing:

$$P_i = (P_0/f_i)(1/f_t) \tag{1}$$

where f_t is the fraction of attention devoted to the tracking task as a whole; f_i is the subfraction of such attention devoted to display variable y_i; P_i is the noise/signal ratio associated with y_i; and P_0 is the baseline noise/signal ratio associated with a high-workload single-variable tracking task (typically, −20 dB).

The attention-sharing model of Eq. (1) has a theoretical base[10] and has been validated in a study of multiaxis tracking by Levison, Elkind, and Ward,[11] who found that this model yielded accurate predictions of multiaxis system performance. Wewerinke[12] has also obtained generally good agreement between subjective workload assessments and a "workload index" based partly on this model.

The model parameter f_t, representing attention to the flight control task as a whole, serves as the metric for workload in the proposed handling qualities assessment scheme. Because it is a scalar quantity, it may be used in a straightforward manner to assess handling qualities for multivariable, multiaxis flight control tasks. Unlike workload metrics used in alternative model-based prediction schemes, the attention parameter defined here has a theoretical as well as empirical basis.[10,11]

Because the predicted "cost" for a given task increases monotonically with increasing noise/signal ratio, and because noise/signal ratio is related inversely to the attention parameter f_T, cost is a monotonically decreasing function of "workload" as we have defined it here. Thus, if other independent model parameters are kept fixed, tradeoff curves of performance vs workload can be predicted.

Prediction of Pilot Rating

In keeping with Anderson's philosophy,[3] pilot rating is predicted by means of a mathematical expression that includes both performance and workload effects. In general, "performance" is defined in terms of a scalar function of the signal deviations predicted by model analysis. As described above, "workload" is synonymous with the total attention to the task, f_t, which affects performance through the noise/signal ratio.

Best results in this study were obtained through use of a performance metric defined as the joint probability of one or more system variables being outside their respective "limits" (i.e., maximum desirable values). The following alternative philosophies were tested and found to yield good replications of experimentally obtained pilot ratings. 1) Pilot rating is determined by the performance achievable at some particular level of workload. 2) Pilot rating is determined by the workload required to achieve some criterion level of performance. 3) Pilot rating is a continuous function of both performance and workload, and the pilot operates at a workload that minimizes the numeric value of his rating (i.e., achieve the best rating).

These philosophies were implemented, respectively, by the following rating expressions:

$$R = 1 + 9\frac{S}{S - S_0}\Big|_{A=A_0} \tag{2}$$

$$R = 1 + 9\frac{A}{A - A_0}\Big|_{S=S_0} \tag{3}$$

$$R = 10\left[\frac{S}{S+S_0} + \frac{A}{A+A_0}\right] \tag{4}$$

$$1 \le R \le 10$$

where R is the predicted pilot rating on the Cooper-Harper scale[13]; S is predicted performance in terms of a probability as defined above; A is the attention model parameter [equivalent to f_t of Eq. (1)]; and S_0 and A_0 are constants of the rating expressions.† For convenience, we shall refer to these rating expressions as the "performance model," the "attention model," and the "minimum-rating model."

†Numerical values for A_0 and S_0 may vary from one expression to the next.

Data Base

The data base used for developing and testing the handling qualities prediction scheme was obtained from two sources: 1) an experimental study performed by Douglas Aircraft Company in 1975,[14] and 2) the results of a questionnaire submitted, during the course of this study, to the test pilots who participated in the Douglas study.

Description of Experiments

A manned simulation study was conducted by Douglas to explore the applicability of various handling qualities criteria to longitudinal flying qualities of large transport aircraft in the landing approach. Criteria that were evaluated included several vehicle-centered criteria from MIL-F-8785B,[2] vehicle-centered criteria from other sources,[14] and a pitch tracking criterion involving a closed-loop pilot model.[15] The simulation study is described in detail by Rickard[14]; a summary of the experiments is given below.

The Douglas study explored a total of 42 vehicle configurations. The first group of 26 configurations were obtained by selecting stability derivatives typical of wide-body aircraft and either varying the simulated c.g. location from far forward to far aft of the neutral point, or by varying a single stability derivative. Configurations of the second group were obtained by specifying vehicle frequency-response characteristics and then solving for the stability derivatives. All handling-qualities variations were confined to the longitudinal control axis; lateral-directional aircraft parameters were kept fixed throughout the experiment to provide response characteristics typical of a wide-body transport having Level 1 handling qualities.

Five Douglas test pilots performed evaluations of these configurations on a six-degree-of-freedom moving-base simulator. Each evaluation typically consisted of two ILS approaches: the first performed in the absence of simulated atmospheric disturbances, the second in the presence of simulated zero-mean turbulence. Approach was initiated at a range of 7.4 n.mi. from runway threshold at an altitude of 1500 ft on the extended runway center line. The 3-deg glide slope was intercepted at a range of about 4.7 n.mi.; the pilot flew down the glide slope relying on ILS instrumentation for path information to an altitude of about 700 ft, at which point the pilot transitioned to a visual display for flare and touchdown.

The test pilots were encouraged, in general, to perform maneuvers that would aid in their evaluations (e.g., intentionally impose and then eliminate a path or attitude error), but no specific set of maneuvers was required. A single Cooper-Harper rating was given by each pilot for the pair of still-air and turbulent-air approaches for each configuration. Some configurations were evaluated more than once by some of the test pilots. Evaluations were performed on the basis of approach performance only; flare and touchdown characteristics were not considered.

Configurations Explored in the Bolt, Beranek, and Newman Study

The rating expression described in Eqs. (2-4) were tested against eight configurations selected from the first group used in the simulation study. These configurations were chosen to span a range of pilot ratings as well as a range of handling qualities problems. The test pilots were assumed to utilize the ILS instrument, attitude indicator, and airspeed indicator as their primary displays during the instrument-flight portion of the simulated approach.

Zero-mean turbulence was simulated in the three linear and three rotational degrees of freedom in the Douglas study. Turbulence models (based on models suggested in the flying qualities specifications[2]) were used to provide disturbances to

Table 1 Pilot opinion ratings

Configuration[a]	Pilot rating Mean	SD	$d\gamma/dV$ level	ω_{sp} vs n/α level	Static stability
1	2.5	1.5	1	1	Yes
3	4.3	2.3	1	4	Yes
4	4.2	2.1	1	4	No
5	5.3	1.6	1	4	No
8	8.3	2.1	4	1	Yes
15	6.7	1.5	1	4	No
16	7.7	2.5	1	4	No
21	6.2	3.5	4	2	Yes

[a]Mean rating for five pilots, configurations 1,3,4,5; mean rating for three pilots, configurations 8,15,16,21.

longitudinal-axis variables. Rms u- and w-gust levels were fixed at 7.8 and 6.5 ft/s, respectively. Further details on gust models and simulated vehicle response characteristics are given in Ref. 1.

Performance Requirements

Application of the prediction scheme described above requires that one or more specific subtasks be selected for analysis and that performance requirements be specified for each subtask. To obtain this information, a questionnaire was administered to four of the five test pilots who had participated in the 1975 manned simulation study. Through this questionnaire the pilots were requested to 1) state whether or not pilot ratings were determined by longitudinal handling characteristics; 2) specify whether ratings were based mainly on the instrument-flight portions of the approach; 3) specify, in order of priority, the subtasks that were important determinants of pilot rating; and 4) specify in as quantitative a manner as possible the "desired" and "acceptable" levels of performance for each subtask. A sample of the questionnaire is provided in Ref. 1.

All four pilots agreed that lateral-directional handling qualities were quite satisfactory and that pilot ratings were influenced primarily by longitudinal handling characteristics. They all stated that the instrument-flight phase was more important in determining ratings.

On the average, highest priority was given to tasks involving transient maneuvering (glide-slope capture, correcting self-induced height error). Next in importance were tasks requiring continuous regulation of height error (altitude station-keeping prior to glide-slope acquisition, post-acquisition glide-slope tracking). Open-loop response and correction of pitch and airspeed mistrim were of substantially less importance overall in terms of influencing pilot opinion.

Obtaining quantitative comments related to performance requirements was considerably more difficult than anticipated. Only two of the four pilots provided quantitative responses, and only one of these differentiated between "desired" and "adequate" performance.‡ As described further on, the responses from these two subjects provided the basis for the performance criteria used in the model analysis.

Pilot Ratings

Mean and standard deviations of the pilot ratings obtained in the Douglas study are given in Table 1, along with handling qualities levels as determined from two of the vehicle-centered criteria considered by Rickard. Rating statistics were derived by first averaging multiple ratings (where such existed) for each pilot for each configuration, and then using these averages to compute a mean and standard deviation (SD) across subjects for each configuration. Table 1 shows both a wide spread of average pilot ratings as well as a variety of handling qualities problems. The short-period (ω_{sp}) response criterion predicts adverse handling qualities for five of the configurations—four of which exhibit static instability. Two of the remaining configurations, on the other hand, exhibit adverse flight path stability ($d\gamma/dV$).

‡To aid the pilot in making this distinction, "adequate" performance was defined in the questionnaire as corresponding to the boundary between a rating of 6 and 7, whereas "desired" performance was to be associated with a rating of 1.

Test of Methodology

The prediction scheme described above was applied to the data base obtained in the 1975 Douglas study. In order to apply this scheme, twenty independent model parameters had to be specified. As the following discussion demonstrates, eighteen of these parameters were defined largely on the basis of task analysis, tempered by some engineering judgement. Once selected, these parameter values were held fixed throughout the analysis; only the two parameters of the rating expression were adjusted to match experimental data.

Problem Definition

The methodology described in this paper was applied to the task of final approach, exclusive of landing. On the basis of the questionnaire submitted to the Douglas test pilots, two specific subtasks were initially selected for study: continuous glide-slope tracking in turbulence, and recovery from intentional glide-slope offset. Preliminary exploration of the latter (transient) task was performed, but resources permitted a complete analysis of only the continuous tracking task. Therefore discussion is confined to tests based on the continuous tracking task.

Although continuous in nature, glide-slope tracking following capture is not, strictly speaking, a steady-state task because of time variations in various task parameters. Nevertheless, because these time variations are relatively slow, piecewise-steady-state analysis can yield meaningful predictions of pilot/vehicle performance trends at various points along the glide path.

A "frozen-point" analysis was performed at a simulated altitude of 1000 ft. Parameters of the turbulence model appropriate to this altitude (see Ref. 1) were chosen for this analysis.

Weighting coefficients for the quadratic performance index were selected as the squared reciprocals of the maximum allowable deviations (or "limits") on important system variables—a procedure that has been followed with apparent success in previous applications of the optimal-control pilot model.[9,16] Limits of 117-ft height error (corresponding to 1-dot glide-slope deviation at an altitude of 1000 ft) and 10-knots (16.9 ft/s) airspeed error were chosen on the basis of pilot commentary. Limits of 40-lb stick force (10-deg elevator deflection), 60-lb/s force rate, and 21,500-lb thrust were chosen, in part, on the basis of physical constraints of the control system. A limit of 10,750-lb/s rate of change of thrust was chosen to induce a control-related lag time constant of about 2 s; this selection was based on the assumption that the pilot would not make continuous wide-band throttle movements during approach.

No limits (i.e., no terms in the quadratic performance index) were associated with either sinkrate error or attitude variables. Penalties on attitude variables were omitted because no limitations on such variables were specified by the test pilot; sinkrate error was omitted from the performance index to prevent overemphasis on height-related variables. Despite the lack of explicit performance penalties on attitude variables, the penalties on control-related variables constrained the model to predict a reasonable "mix" of height, attitude, and control deviations.

The pilots were assumed to make longitudinal-axis flight-control inputs primarily on the basis of perceptual information obtained from the ILS, attitude, and airspeed instruments. Rate information was also assumed to be obtained from the ILS and attitude indicators. Thus the set of perceptual variables assumed for model analysis consisted of height, sinkrate, pitch, pitch rate, and airspeed errors.

Table 2 Display- and performance-related model parameters

Variable[a]	Limit	Coefficient	Threshold	Residual noise	Relative attention
h	117	7.31 E-05	9.3	0	0.22
$\dot{h}$	...	...	37	0	
θ	...	...	0.43	0.5	0.22
q	...	...	1.72	0	
u_i	16.9	3.5 E-03	1.9	0	0.22
δ_e	40	6.25 E-04	...	...	...
$\dot{\delta}_e$	60	2.78 E-04	...	...	...
δ_t	21,500	2.16 E-09	...	...	...
$\dot{\delta}_t$	10,750	8.65 E-09	...	...	...

[a] h = altitude error, ft; θ = pitch change, deg; q = pitch rate, deg/s; u_i = airspeed relative to moving air mass, ft/s; δ_e = force on the control column; δ_t = thrust deviation from trim, lb.

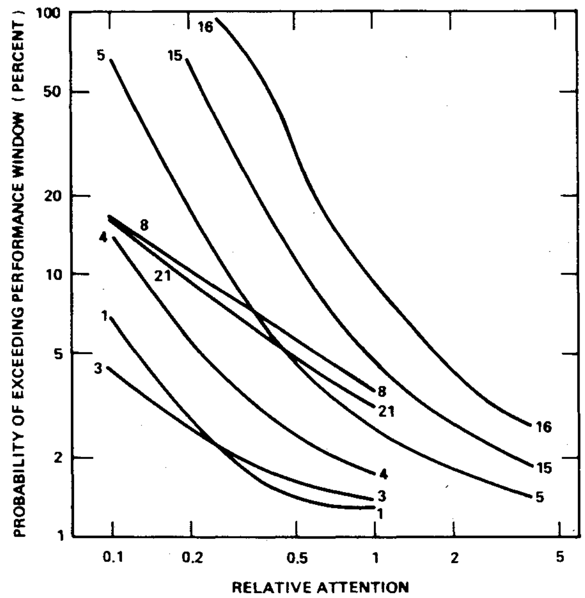

Fig. 1 Prediction of performance vs workload curve. Number corresponds to configuration number of Rickard.[14]

Attention was assumed to be divided equally between the ILS, attitude, and airspeed instruments; no attention-sharing penalties were considered between displacement and rate information from the same physical display. On the basis of analysis performed in a previous analytic study of landing approach,[14] 34% of the pilot's relative attention was assumed to be "lost" because of large eye movements required to scan the flight-control instruments. Thus fractional attentions of 0.22 were associated with the ILS, attitude, and airspeed displays.

Effective perceptual thresholds were computed from the display gains (i.e., inches of display deflection per unit change in problem variable), the eye-to-display distance, and assumed values of perceptual resolution limitations based on previous laboratory experiments as described by Levison.[1] A residual noise was also associated with perception of pitch attitude change. Display and performance-related model parameters are given in Table 2.§

Prediction of Performance/Workload Tradeoffs

Performance/workload tradeoffs were predicted for each of the eight configurations. For purposes of predicting handling qualities, "performance" was defined as the probability of one or more system variables exceeding maximum allowable values. To obtain an approximation to this joint probability, system variables were treated as independent Gaussian variables, and the probability was computed as

$$\mathrm{Pr} = 1 - \prod_i (1 - \mathrm{Pr}_i) \tag{5}$$

where Pr_i is the probability that the ith variables of interest will lie outside its prescribed boundary, and Pr is the probability that at least one such variable is out of bounds. The probability Pr_i was readily computed from the predicted variance of the ith system variable. (Since we considered steady-state conditions, all variables were assumed to be zero-mean Gaussian processes.) Workload was represented in the analysis by the attentional variable f_i; the f_i were adjusted to reflect attention-sharing as shown in Table 2.

A noise/signal ratio $P = 0.01$ was associated with a relative attention of unity $(3 \times 0.22 + 0.34)$. Thus variations in attentional workload were reflected by changes in the noise/signal ratios according to Eq. (1).

Predictions of performance vs attentional workload are shown in Fig. 1. Values of attention shown on the abscissa are relative to that inferred from data obtained in a standardized laboratory tracking task. That is, unity attention is intended as a benchmark level of workload and does not necessarily relate to maximum effort or capability. Thus, for configurations in which predicted performance is especially sensitive to attention, predictions are shown for relative attentions greater than unity.

The trends shown in Fig. 1 are consistent with the pilot ratings given in Table 1. Except for configuration 8, the ordering of the performance/workload curves is consistent with the ordering of the pilot ratings. For relative attentions of 0.5 and greater, predicted performance for the remaining seven configurations follows the trend of the ratings. Operation on these results to yield predicted pilot ratings is discussed below.

Table 3 Independent parameters for the rating expression

Expression	S_0,%	A_0
Performance model	5.3	0.50
Attention model	5.0	0.47
Minimum-rating model	10.0	2.0

Predicted Ratings

The three rating expressions presented in Eqs. (2-4) were applied to the performance/workload tradeoff curves to provide a test of the proposed methodology. Values were assigned to the independent parameters of each expression as shown in Table 3.

The value of A_0 of the performance model was chosen to represent a moderate-to-high workload level, and the corresponding value for S_0 was found through a regression procedure that minimized the mean-square difference between predicted and experimental pilot ratings, normalized with respect to the variance of each experimental rating. The value for S_0 of the attention model was selected to represent a moderate-to-stringent performance requirement, and the value for A_0 was found through a similar regression analysis.

Because of the lack of a tractable analytic expression relating performance to workload, the parameters S_0 and A_0 of the minimum-rating model were not found through a computerized regression analysis. Rather, pairs of integers were explored on a trial-and-error basis to provide a good match to experimental pilot ratings. The predicted (minimum) rating for a given configuration was obtained by superim-

§The rather large threshold of 37 ft/s associated with perception of sink rate stems from the assumption that the pilot uses the velocity of the glide-slope indicator as his primary source of sinkrate information.[1]

posing the predicted performance/workload tradeoff curve (Fig. 1) on the curves of constant rating, shown in Fig. 2.

Because of the difficulty in matching the predicted pilot ratings of configuration 8, ratings for this configuration were omitted from all three regression analyses.

Figure 3 provides a graphical comparison of predicted vs experimental pilot ratings for the three rating expressions. Dashed lines indicate boundaries of one rating unit. The three rating schemes performed about equally well on the average and were able to match six of the eight experimental ratings to within one rating unit. Configuration 8, which was matched least well, was omitted from the regression analysis.

Prediction errors may be compared against the variability of the experimental data in Fig. 4. Experimental ratings are indicated by filled circles, with brackets to indicate one standard deviation; open symbols indicate predictions obtained with the three rating expressions.

Except for configuration 8, predicted ratings are within one standard deviation of the experimental mean. Even for the worst case, the prediction error is well within two standard deviations of the mean. Thus the reliability of the predicted ratings is commensurate with the reliability of the experimental data.

Discussion of Results

The generally good match between "predicted" and experimental pilot opinion ratings demonstrates the validity of the model-based approach described in this paper. The technique is shown to replicate experimental results reasonably well across a set of conditions that spans a range of handling qualities levels and problems. Because the procedure is based on a pilot/vehicle model of considerable generality and demonstrated validity, this scheme ought to be valid for other aircraft configurations and, with appropriate definitions of performance requirements, other flight tasks as well. Further study is required to compare this technique against other model-based procedures and to further compare the usefulness of the three rating expressions tested in this study.

Resources did not permit a detailed study of the inability to obtain a good match to the experimental rating for configuration 8. The differences between the average ratings for configurations 8 and 21 (which differences are predicted to be negligible) were apparently not due to training effects; these two configurations were presented to the test pilots in a balanced order.

It should be noted that all tests of the proposed methodology have been based on steady-state analysis appropriate to conditions at a single altitude. Although steady-state-like tasks were important determinants of pilot opinion, transient-response behavior was also important. There may have been some aspects of glide-slope capture and other transient maneuvers that were especially adverse for configuration 8.

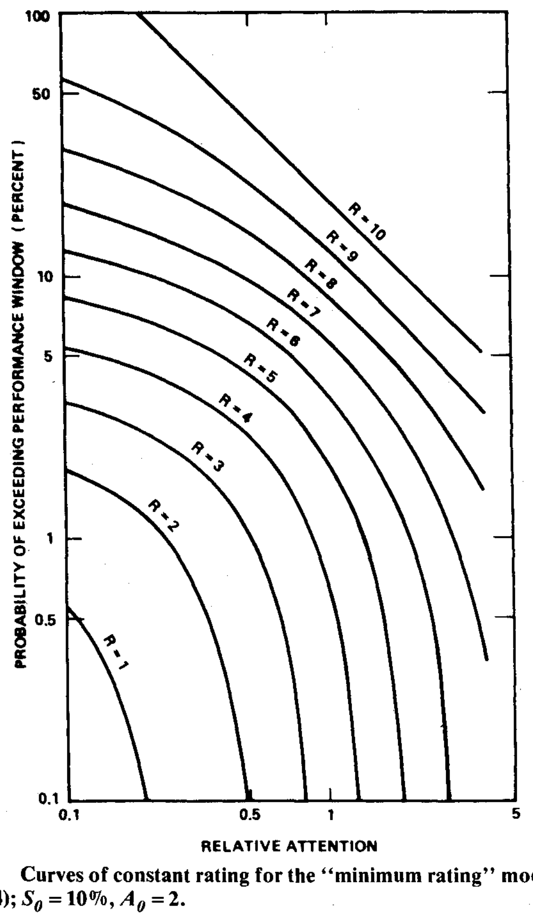

Fig. 2 Curves of constant rating for the "minimum rating" model of Eq. (4); $S_0 = 10\%$, $A_0 = 2$.

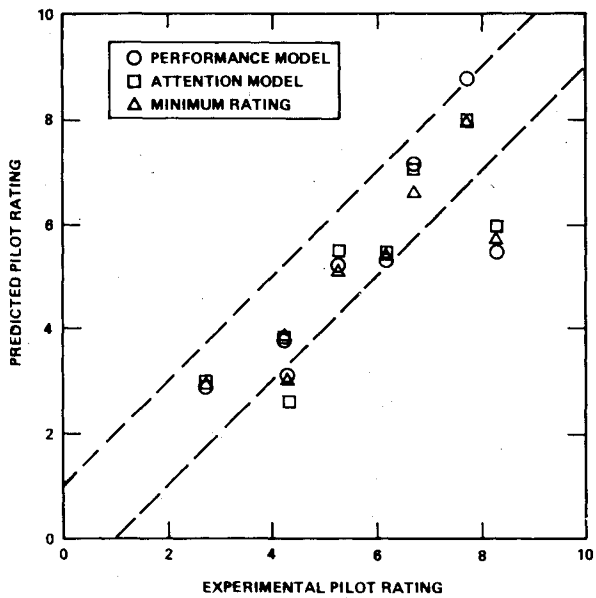

Fig. 3 Predicted vs average experimental pilot ratings.

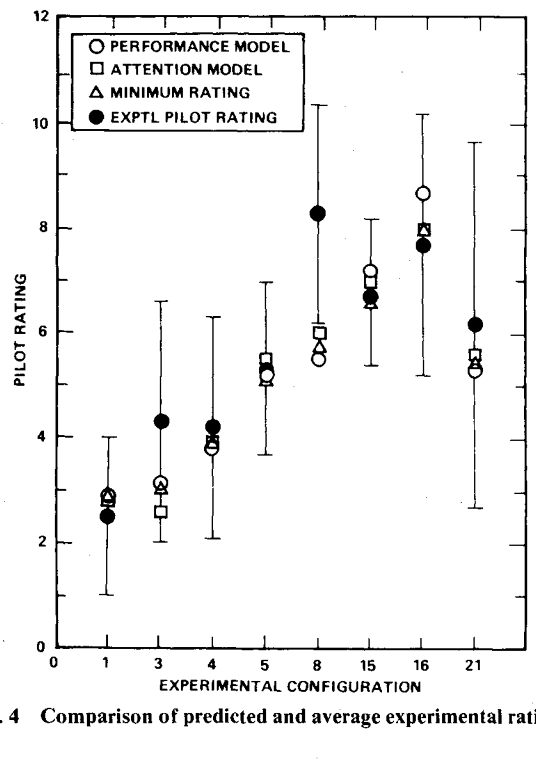

Fig. 4 Comparison of predicted and average experimental ratings.

The data base obtained in the 1975 Douglas simulation study—which was utilized because of its relevance to large transport aircraft—does not allow a thorough test of the model-based prediction scheme. Hindsight reveals the following methodological deficiencies in this regard: 1) lack of measures such as rms performance scores, pilot describing functions, etc., to provide a test of the model's ability to predict objective performance measures; 2) large pilot-to-pilot variability in opinion ratings; and 3) a relatively nonspecific evaluation procedure, leaving each individual pilot to determine the relative configuration of the various maneuvers and subtasks to the global rating score.

Conclusions and Recommendations

A technique based on the optimal-control model for pilot/vehicle systems has been developed for assessing aircraft handling qualities. Three variations of this technique provide a good match to opinion ratings obtained in a manned simulation study of large commercial transports in landing approach.

The model-based technique developed in this study has a number of features which should enhance its applicability to other aircraft configurations and other flight tasks and should allow wider application than alternative handling qualities prediction schemes:

1) One is able to proceed in a well-defined manner from a description of the task environment and of task requirements to a prediction of pilot opinion ratings. The general form of the rating expression and of the underlying pilot model is invariant across applications.

2) No constraints are placed on the nature of the vehicle response and the pilot model is relatively free form. Thus "unconventional" aircraft dynamics may be considered.

3) A scalar metric for attentional workload is expressed in terms of a model parameter related to the signal/noise properties of the pilot's response. Thus the analytic treatment of attentional workload is independent of the details of the flight task.

4) The effects of display parameters, turbulence, and other environmental factors on pilot opinion rating may be considered via parametric variations.

Encouraging results obtained with the model-based technique tested in this study warrant further research to provide a more rigorous test of the procedure and to determine its range of validity. Such a study should be designed to avoid the methodological deficiencies cited above. In particular, evaluation procedures should be standardized so that 1) all pilots perform the same maneuvers, 2) all test pilots weight the maneuvers in the same manner when assigning an overall rating to the aircraft, and 3) if practical, pilots should be encouraged to adopt a common set of performance expectancies and evaluation criteria.

Objective measures of system performance and pilot response behavior should be obtained, in addition to pilot opinion ratings, to provide a more rigorous test of the method. Aspects of the prediction scheme remaining to be explored include 1) the correlation between predicted and measured rms "error" scores, 2) the degree to which actual pilot loop closures conform to the optimal closures predicted by the model, and 3) the degree to which the coefficients of the analytic rating expression need to be recalibrated for different aircraft types or different mission requirements.

Acknowledgments

The research described in this paper was supported by NASA Langley Research Center under Contract NAS1-5529. This contract included a subcontract to Douglas Aircraft Company to provide experimental data and other consulting services. William W. Rickard was project engineer for the Douglas effort.

References

[1] Levison, W.H., "A Model-Based Technique for Predicting Pilot Opinion Ratings for Large Commercial Transports," NASA Contractor Rept. 3257, April 1980.

[2] "Flying Qualities of Piloted Airplanes," Military Specification, MIL-F-8785B (ASG), 1969.

[3] Anderson, R.O., "A New Approach to the Specification and Evaluation of Flying Qualities," Wright-Patterson AFB, AFFDL-TR-69-120, 1970.

[4] Arnold, J.D., Johnson, R.B. and Dillow, J.D., "Pitch Paper Pilot Revisited," *Proceedings of the Ninth Annual Conference on Manual Control,* MIT, Cambridge, Mass., 1973, pp. 75-81.

[5] Naylor, F.R., Dillow, J.D., and Hannen, R.A., "Roll Paper Pilot," *Proceedings of the Ninth Annual Conference on Manual Control*, MIT, Cambridge, Mass., 1973, pp. 83-92.

[6] Stone, J.R., "Paper Pilot Ponders Supersonic Transports," *Proceedings of the Ninth Annual Conference on Manual Control*, MIT, Cambridge, Mass., 1973, pp. 95-101.

[7] Hess, R.A., "Prediction of Pilot Opinion Ratings Using an Optimal Pilot Model," *Human Factors*, Vol. 19, No. 5, 1977, pp. 459-475.

[8] Schmidt, D.K., "Optimal Flight Control Synthesis via Pilot Modeling," *Journal of Guidance and Control*, Vol. 2, July-Aug. 1979, pp. 308-312.

[9] Baron, S. and Levison, W.H., "Display Analysis with the Optimal Control Model of the Human Operator," *Human Factors*, Vol. 19, No. 5, Oct. 1977, pp. 437-457.

[10] 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, pp. 189-218.

[11] 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.

[12] Wewerinke, P.H., "Human Operator Workload for Various Control Situations," *Proceedings of the Tenth Annual Conference on Manual Control*, Wright-Patterson AFB, Ohio, 1974, pp. 167-192.

[13] Cooper, G.E., "The Use of Pilot Rating in the Evaluation of Aircraft Handling Qualities," NASA TN D-5153, April 1969.

[14] Rickard, W.W., "Longitudinal Flying Qualities in the Landing Approach," *Proceedings of the Twelfth Annual Conference on Manual Control*, NASA TM X-73, 170, May 1976, pp. 559-584.

[15] Neal, T.P. and Smith, R.E. "An In-Flight Investigation to Develop Control System Design Criteria for Fighter Airplanes," AFFDL-TR-70-74, Vols. I and II, 1970.

[16] Levison, W.H., "Analysis and In-Simulator Evaluation of Display and Control Concepts for a Terminal Configured Vehicle in Final Approach in a Windshear Environment," NASA Contractor Rept. 3034, Aug. 1978.