PILOT ORIENTED PERFORMANCE MEASUREMENT

Joe De Maio, Herbert H. Bell and John Brunderman
Air Force Human Resources Laboratory/Operations Training Division

ABSTRACT

Flight simulators provide a complete quantitative record of a pilot's flying performance. Evaluating this record is complicated by the volume of data and by its fine detail, dozens of flight parameters, sampled many times per second. Automated performance measurement systems (APMS) reduce the volume of data to an amount which is manageable and understandable. The usual APMS is aircraft state oriented. The APMS keys on aircraft state (e.g., X-Y position, bank angle) to define intervals over which performance data are integrated. This APMS is relatively insensitive to pilots' intentions and so may average performances which had differing objectives, based only on their having occurred at the same point during the task sequence. An alternative APMS has been developed which is pilot oriented. This APMS defines measurement intervals based on control inputs. Control inputs are identified by discrete changes in flight path. These intervals are psychologically relevant in that they begin with a goal-directed control input and end with a countervailing input. By relating performance in the pilot defined intervals to state defined intervals, it is possible to quantify performance on given flight segments (e.g., a level turn), and to specify factors which lead to a given level of performance.

INTRODUCTION

As man-machine systems have become more complex and costly, the need for effective measurement of operator performance has increased dramatically. Performance measurement systems (PMS) are needed which will permit assessment not only of total man-machine system performance but also of performance factors contributing to total performance. To accomplish this sort of assessment measures are needed which permit the decomposition of performance into perceptual, information processing and physical control performance components.

The need to decompose overall performance into its components is particularly apparent when there is an interactive effect of task difficulty factors on overall performance. For example Rinalducci (1) examined the performance of pilots in maintaining level flight in an F-16 simulator. Rinalducci used two measures of performance, mean altitude above ground level (AGL) and RMS deviation from 200 ft AGL. Both measures were sensitive to the effects of visual cue and airspeed manipulations as well as to difference between straight and turning flight. In addition both measures were sensitive to interactions between these variables. One variable, visual cues, is clearly a perceptual/informational factor. Neither the other two variables nor the interactions are amenable to intuitive labeling. Since the performance measures do not permit analysis of component processes, it can be shown that these factors affect performance, but not how they do so.

Attempts to study component control processes have followed two lines. One approach is to use a discrete stimulus such as a cross-wind gust (2) to elicit a control input. Because the input was elicited, it is possible to obtain timing information, which shows the contribution of perceptual, subject and control task factors to input latency and effectiveness. This approach provides information not only about how well an operator controls the vehicle but also about the effectiveness of the operator's response. The limitation of the approach is that it can only be applied when inputs which are made in response to a discrete environmental change. The ad lib control inputs which operators make frequently in unperturbed, steady-state operation are not amenable to this approach.

An approach which has been used to study ad lib control inputs is simply to measure the rate of control inputs. An input rate measure which has been used in the study of driving is steering reversal rate (SRR), the rate at which the steering wheel is reversed through a small finite arc. Greenshields (3) has found SRR to increase with traffic density. McLean and Hoffman (4) have found SRR to be affected by lane width, speed and preview. Hicks and Wierwille (5) found control task difficulty to affect SRR.

Although SRR is sensitive to the effect of task variables on ad lib control performance, it has drawbacks. MacDonald and Hoffman (in 6) found addition of a secondary task to affect SRR differently in simulated and actual driving. More importantly, SRR is often uncorrelated with overall steering performance (6).

Control reversal rate has also been employed in flight control research. As in driving, control reversal rate is sensitive to changes in the difficulty of the flying task, but the sources of this sensitivity are obscure. Blomberg, Pepler and Speyer (7) used elevator position reversal rate (EPRR) to measure control performance in the A-300 aircraft. Blomberg et al found introduction of an electronic flight information system (EFIS) to increase EPRR. Other measures of flying performance showed the EFIS to improve pilot performance. Introduction of an autopilot, which reduced the difficulty of the control task by controlling horizontal position, caused EPRR to decrease. As in driving control reversal rate is sensitive to flying task difficulty, but the factors underlying this sensitivity are not clear.

For measures of ad lib control inputs to be useful, indices of input effectiveness, such as are available for elicited control inputs, are needed. A PMS is presented here, which employs measures of effectiveness of ad lib inputs. Two assumptions underlie this PMS: (1) the conditions which prevail when the input is made serve to elicit the input and (2) the qualitative effect of the input reflects the operator's intention in making it. That is, if an input causes the vehicle to change direction of travel, then the operator's intention was to change direction. While this assumption may not apply to the totally naive operator, it is reasonable for one having even minimal skill.

Measures based on the above assumptions are used to decompose control performance into a perceptual task component and a physical control task component. Following description of the PMS, data are presented to show effects of perceptual and control task difficulty variables on the performance measures. These data were gathered in a flight simulator visual system evaluation, the results of which are presented elsewhere (8). The intent here is simply to describe the functioning of the PMS.

THE MEASUREMENT SYSTEM

Because the intent of the PMS is to provide measures of performance which are sensitive to the pilot's intentions moment by moment, both overall and component performance measures need be defined specifically with reference to the flying task considered. The PMS presented below is used to evaluate performance in maintaining level flight at a specified altitude.

Four performance measures are presented (see Table 1). Two measures relate to overall control performance: Target Altitude (TA) is the mean of the local altitude minima and maxima; Altitude Range (AR) is the mean difference between local maxima and minima. These measures give the altitude the pilot is attempting to maintain and the degree to which the aircraft varies about that altitude. Target Altitude and AR are analogous to conventional measures of mean altitude and standard deviation. Two measures, Smoothness (S) and Critical Error Rate (CER) are used to decompose performance into its components. These measures are based on attributes of individual control inputs.

Control inputs of interest are those made through the aircraft stick. Unlike previous approaches, which have defined inputs in terms of control manipulandum displacement, the present PMS defines inputs in terms of their effect on the velocity vector. This approach has the advantage that it employs a functional criterion for defining an input, rather than an arbitrary one.

Table 1
Performance measures for level flight

OVERALL PERFORMANCE MEASURES
Target Altitude (TA):
Mean of local altitude minima and maxima.
Altitude Range (AR):
Mean difference between local altitude maxima and minima.
COMPONENT PERFORMANCE MEASURES
Smoothness (S):
Proportion of critical control inputs.
Critical Error Rate (CER):
Ratio of lag distance to lag time for critical control inputs.

A control input is designated by a change in sign of the vertical acceleration. If the vertical acceleration is positive (increasing rate of climb), a control input is said to have been made when the vertical acceleration becomes negative (decreasing rate of climb). This definition is analogous to that used for SRR. This functional criterion makes the PMS highly sensitive, since control inputs are identified according to a criterion which adapts to all task relevant factors.

Once we have determined that a control input has been made, it is necessary to determine the degree of effectiveness of the input. Since control inputs are made to affect velocity, they can only fall into one of two categories: (1) those after which velocity changes sign (direction of travel changes) and (2) those after which velocity does not change sign (direction of travel remains the same). With regard to error control, these two classes of control input effectiveness have psychological and task relevance because in the former case the input causes a decrease in error, while in the latter error continues to increase.

In the PMS critical inputs are those which reverse the direction of travel and so decrease error, and non-critical inputs are those which do not alter the direction of travel. Efficient control might be expected to involve a relatively large proportion of critical inputs. A greater proportion of non-critical inputs might result in less efficient control since many inputs do not result in error reduction.

In the PMS Smoothness (S) is defined as the proportion of control inputs which are critical inputs. Smoothness has a value of 1.0 when all inputs are critical inputs made for the direct purpose of velocity control. If, on the other hand, no inputs were made for the purpose of altering the direction of travel, that is, none were critical, then S would have a value of 0.0. A higher value of S represents more efficient control.

As the distinction between critical and non-critical inputs provides a finer grain analysis of control behavior than simple input rate, so a still finer grain analysis can be obtained by examining the effectiveness of critical inputs. Since critical inputs are made to reverse the direction of travel, their effectiveness can be determined by measuring the rate at which error accumulates following the input. An effective input is one which results in a low rate of error accumulation. In the PMS this measure is given by Critial Error Rate (CER), the ratio of the lag distance to the lag time from the critical input to altitude minimum/maximum.

The two measures, S and CER, permit the decomposition of control performance into behavioral components. For this decomposition to be useful, two things are necessary: (1) the component measures need be tied to psychologically relevant processes and (2) the contribution of the performance components to overall performance must be determined. The following analysis of control performance in a flight simulator addresses these issues through examination of flying performance in straight and turning flight under varying conditions of environmental visual cue quality.

FLYING PERFORMANCE EVALUATION

In the flying performance evaluation we shall examine the effect of two task difficulty factors on our measures of performance. Different types of task difficulty factors will be shown to affect performance components differently. The interaction of the components in overall performance will also be shown. Finally we shall show how overall performance reflects the strategy adopted by the pilot to cope with decrements in component performance.

One of the task difficulty factors addressed is the quality of out-of-the-cockpit visual cues provided the pilot. De Maio and Brooks (9) and De Maio et al (8) have used the slope (b) of an altitude estimation function to evaluate the altitude cueing effectiveness of simulator visual environments. Flying performance is examined in five environments, whose cueing effectiveness ranges from b=.2 to b=.8.

The second task difficulty factor is determined by the physics of flight. When an airplane is in wings level flight, the force of gravity is counterbalanced directly by the lift vector. When the aircraft is banked, a cosine component enters the lift equation. This cosine component increases the difficulty of the control task in proportion to the size of the bank angle up to 90°.

We will begin the performance analysis by looking at overall performance as measured by TA (see Fig 1). In both straight and turning flight TA increases substantially in those environments providing poor altitude cueing (b ≤ .5). Turning also causes an increase in TA in all visual environments, but this effect is not as great as that of visual cue quality.

Target Altitude measures the altitude the pilot attempts to hold. In order to see why pilots raise TA with increased task difficulty, we examine another overall performance measure, AR (see Fig 2). Since AR measure how precisely the pilot controls altitude, it drives TA in that the TA must be, at the least, high enough to preclude collision with the ground on minimum altitude excursions.

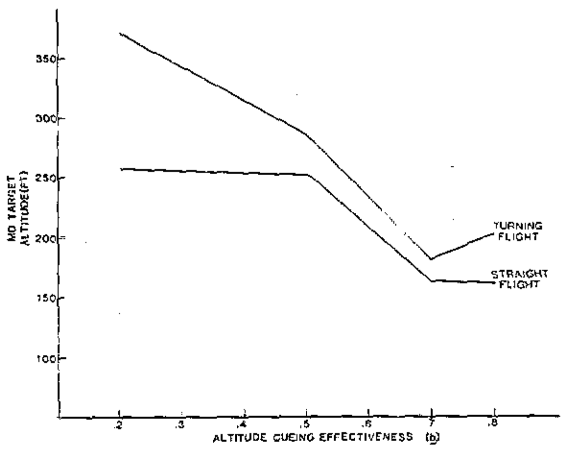

Figure 1. Target altitude for straight and turning flight

As was true with TA, AR is sensitive to both perceptual and motor control task difficulty factors. The pattern of the effects, however, differ substantially for the two variables. Altitude Range is greater in turns than in straight flight at all levels of altitude cueing. At low levels of altitude cueing this difference is very large, roughly 200 ft.

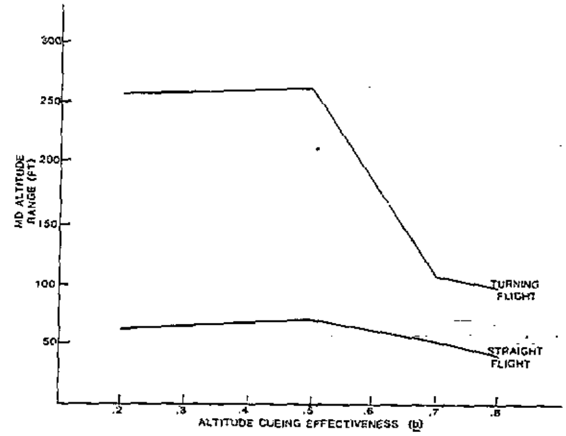

Figure 2. Altitude range for straight and turning flight

In straight flight AR increases only slightly (@ 30 ft) due to cue quality variation, and yet TA increases substantially. At the same time a much larger difference in AR due to turning under good cue conditions leads to little change in TA. In fact the function relating TA to visual cue quality in both straight and turning flight is much more like the AR function for turns than it is like that for straight flight. This similarity leads to the conclusion that control precision performance (i.e., AR) in turns is the determinant of TA in both turning and straight flight. Since turns must be executed, the pilot chooses a straight flight TA which permits maneuvering comfortably.

COMPONENT PERFORMANCE ANALYSIS

The above analysis of AR permits us to determine that task difficulty affects control precision. The analysis of TA shows how pilots might determine an appropriate altitude based on their ability to control altitude. What remains to be shown is how the perceptual and control task factors act individually and in concert to affect control precision. This analysis is accomplished by examining the performance components, S and CER.

Smoothness is a measure of control input efficiency, the proportion of inputs that are critical inputs (that is, effective in the sense of altering the aircraft's direction of travel). Examination of Fig 3 shows that S is highly sensitive to altitude cue quality but insensitive to control task difficulty (bank). Changing the quality of visual information available to the pilot affects the proportion of critical and non-critical inputs of the control inputs made. When cue quality is high, most inputs are made to change the direction of travel. On the other hand, when cue quality is low, a relatively small proportion of inputs is made for this reason.

It is reasonable to suppose that non-critical inputs, also, have a purpose. Since the proportion of these inputs increases when altitude cues are poor, these inputs may serve to give the pilot information needed for aircraft control. When altitude cues are good, only a small number of non-critical inputs is needed to provide flight control information, and the majority of inputs is made to effect flight control. When visual cues are poor, more non-critical inputs are needed, and so S declines.

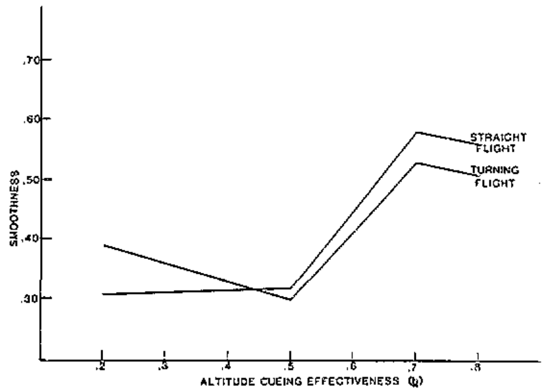

Figure 3. MD smoothness for straight and turning flight

The second component performance variable, CER, measures the effectiveness of individual critical control inputs; that is, how quickly error accumulates following an error reducing input. Since CER measures the responsiveness of the man-machine system, we might expect it to be differentially sensitive to control task difficulty factors. Examination of Figure 4 shows this sensitivity. Critical Error Rate increases from about 15 ft/sec in straight flight to about 34 ft/sec in turning flight. Yet CER does not vary systematically with altitude cue quality.

We have now identified two components of control performance: S, or input efficiency, and CER, or input effectiveness. These performance components show the differential *sensitivity to task difficulty factors that* permits us to determine how increases in difficulty affect the control process. Increased perceptual task difficulty leads to a decrease in S as more inputs are made to provide the pilot information and fewer for the express purpose of flight control. Increased control task difficulty leads to an increase in CER. When the control task is more difficult, inputs are less effective.

Conceptually the effects of variation in control efficiency (S) or input effectiveness (CER) can readily be related to overall control performance (AR). When inputs are less effective due to increased control task difficulty, the aircraft responds more slowly and AR increases. When input efficiency decreases, due to increased perceptual task difficulty, directional changes occur less frequently and again Altitude Range increases.

In order to demonstrate a quantitative relation between control efficiency and effectiveness and overall performance, a third variable must be introduced, that is, Effective Input Duration. Effective Input Duration (EID) is the time between inputs for non-critical inputs and the time between input and local maximum (minimum) for critical inputs. The difference in definition for critical and non-critical EID arises because the directional change following a critical input acts psychologically as an input. Effective Input Duration is about the same for critical and non-critical inputs, although the inter-input interval is about twice as long for critical inputs as for non-critical inputs.

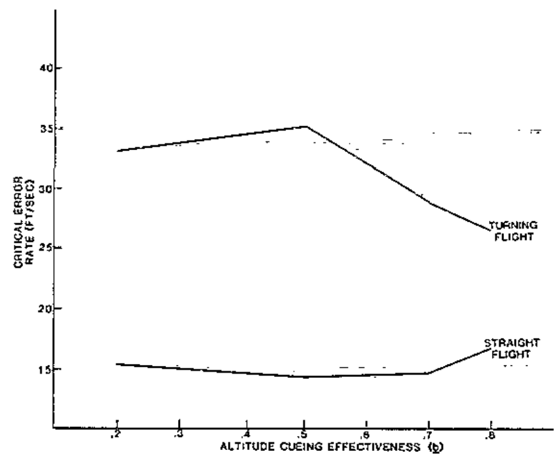

Figure 4. MD critical error rate for straight and turning flight

Effective Input Duration is not a particularly useful performance measure itself because, like the input rate measures discussed above (SRR, EPRR), it exhibits sensitivity to a variety of task factors affecting performance of both operator and aircraft. Yet the difference between EID and inter-input interval may be useful in explaining some of the conflicting input rate data in the literature. Since input rate is half as great for critical inputs as for non-critical, factors which affect S will affect *input rate even if EID is not changed.* Effective Input Duration too affects input rate when S may be unchanged. Depending on the magnitude and direction of these two effects, input rate may increase or decrease in response to the ensemble of task difficulty factors.

Altitude Range may be predicted from S, CER and EID using the function:

$$AR = 2*D*((S*C)+(1-S)*2C) \quad (1)$$

Where:

AR = Altitude Range, S = Smoothness
C = Critical Error Rate and
D = Effective Input Duration.

Equation 1 says the average excursion above or below the Target Altitude is determined by rate at which error accumulates for a particular type of input multiplied by the proportion of such inputs and by the effective input duration. The error accumulation rate for non-critical inputs is twice that for critical inputs since deceleration to zero does not occur. Using the median values for S, CER, and EID (see Table 2), Equation 1 predicts an AR of about 267 ft in turning flight with poor altitude cues ($b \leq .5$). The obtained AR was 260 ft. In wings level flight with good altitude cueing, predicted AR is 55 ft while actual AR is 49 ft.

DISCUSSION

Two performance measures have been developed which permit the decomposition of flight control

Table 2
Effective input duration (Sec)

Cueing Effectiveness	Straight	Turning
Low (b .5)	1.7	2.0
High (b .7)	1.3	1.4

performance into component processes. These component processes are differentially sensitive to factors affecting the difficulty of the flying task. Smoothness, the proportion of critical inputs, reflects the efficienty of control. Increasing the difficulty of the *perceptual component of the flight control task* leads to a decrease in S. Critical Error Rate measures the effectiveness of critical control inputs. This measure is sensitive to the difficulty of the physical control task itself.

In order to relate performance on these two difficulty specific components to overall control performance, a third performance component, sensitive to general task difficulty, is included. EID measures the amount of time the pilot "holds" an input. At higher levels of the task difficulty, EID increases. The precision of control, AR, is a multiplicative function of S, CER and EID.

The precision of control is a function of component control processes beyond the conscious, voluntary control of the pilot. The only voluntary control option open to the pilot is to select a TA which is a compromise based on both task requirements and limitations to control precision. The pilot adjusts this TA to permit accomplishment of the flying task within the constraints of perceptual and psychomotor limitations.

REFERENCES

1. Rinalducci, E. J. Visual Cues in the Simulation of Low Level Flight. Air Force Office of Scientific Research Report, Bolling AFB, 1981.

2. Wierwille, W. W., Casali, J. G., and Repa, B. S. *Driver Steering Reaction Time to* Abrupt-Onset Crosswinds, as Measured in a Moving-base Driving Simulator. Human Factors, 1983, 25, 103-116.

3. Greensheilds, B. D. Driving Behavior and Related Problems. Highway Research Record, 1963, 25, 14-32.

4. McLean, J. R. and Hoffman, E. R. The Effects of Restricted Preview on Driver Steering Control and Performance. Human Factors, 1973, 15, 421-440.

5. Hicks, T. G. and Wierwille, W. W. Comparison of Five Mental Workload Assessment Procedures in a Moving-base Driving Simulator. Human Factors, 1979, 21, 129-143.

6. MacDonald, E. A., and Hoffman, E. R. Review of Relationships Between Steering Wheel Reversal Rate and Driving Task Demand. Human Factors, 1980, 22, 733-739.

7. Blomberg, R. D., Pepler, RD and Speyer J. *Performance Evaluation of Electronic Flight* Instruments. Proceedings of the Second Symposium on Aviation Psychology, Columbus, OH, 1983.

8. De Maio J., Rinalducci, E. J., Brooks, R., and Brunderman, J. Visual Cueing Effectiveness: Comparision of Perception and Flying Performance. Proceedings of the Fifth Interservice/Indusry Training Equipment Conference, Washington, D.C., 1983.

9. De Maio, J., and Brooks, R. Assessment of *Simulator Visual Cueing Effectiveness by* Psychophysical Techniques. Proceedings of the Fourth Interservice/Industry Training Equipment Conference, Orlando, FL, 1982.

ABOUT THE AUTHORS

Dr De Maio, Dr Bell and Ms Brooks are Research Psychologists at the Air Force Human Resources Laboratory/Operations Training Division, Williams AFB, AZ. First Lieutenant Brunderman is a T-38 Instructor Pilot with the 82 FTW, Williams AFB, AZ.