Twenty-First Annual Conference on Manual Control

May 1986

NASA

National Aeronautics and
Space Administration

Twenty-First Annual Conference on Manual Control

Compiled by:
Richard A. Miller,
Richard J. Jagacinski, Ohio State University
for:
Ames Research Center
Moffett Field, California

May 1986

NASA

National Aeronautics and
Space Administration

Ames Research Center
Moffett Field, California 94035

TABLE OF CONTENTS

iii

FOREWORD

The Twenty-first Annual Conference on Manual Control was held at Ohio State University, Columbus, Ohio, June 17-19, 1985. Sponsorship of the conference by the NASA Ames Research Center was arranged by E. James Hartzell. The conference was co-hosted by the Department of Industrial and Systems Engineering and the Department of Psychology of Ohio State University.

This was the twenty-first in a series of conferences dating back to December 1964. These earlier meetings and their proceedings are listed below:

First Annual NASA-University Conference on Manual Control, The University of Michigan, December 1964. (Proceedings not printed)

Second Annual NASA-University Conference on Manual Control, University of Southern California, February 28 to March 3, 1967. (NASA-SP-128)

Third Annual NASA-University Conference on Manual Control, University of Southern California, March 1-3, 1968. (NASA-SP-144)

Fourth Annual NASA-University Conference on Manual Control, University of Michigan, March 21-23, 1968. (NASA-SP-192)

Fifth Annual NASA-University Conference on Manual Control, Massachusetts Institute of Technology, March 27-29, 1969. (NASA-SP-215)

Sixth Annual Conference on Manual Control, Wright-Patterson AFB, Ohio, April 7-9, 1970. (AFIT/AFFDL Report, no number)

Seventh Annual Conference on Manual Control, University of Southern California, June 2-4, 1971. (NASA-SP-281)

Eighth Annual Conference on Manual Control, University of Michigan, May 17-19, 1972. (AFFDL-TR-72-92)

Ninth Annual Conference on Manual Control, Massachusetts Institute of Technology, May 23-25, 1973. (Proceedings published by MIT, no number)

Tenth Annual Conference on Manual Control, Wright-Patterson AFB, Ohio, April 9-11, 1974. (AFIT/AFFDL Report, no number)

Eleventh Annual Conference on Manual Control, NASA-Ames Research Center, May 21-23, 1975. (NASA TM X-62,464)

Twelfth Annual Conference on Manual Control, University of Illinois, May 25-27, 1976 (NASA TM X-73,170)

Thirteenth Annual Conference on Manual Control, Massachusetts Institute of Technology, June 15-17, 1977. (Proceedings published by MIT, no number)

Fourteenth Annual Conference on Manual Control, University of Southern California, April 25-27, 1978 (NASA CP-2060)

Fifteenth Annual Conference on Manual Control, Wright State University, Ohio, March 20-22, 1979. (AFFDL-TR-79,3134)

Sixteenth Annual Conference on Manual Control, Massachusetts Institute of Technology, May 5-7, 1980. (Proceedings published by MIT, no number)

Seventeenth Annual Conference on Manual Control, University of California at Los Angeles, June 16-18, 1981. (JPL Publications 81-95)

Eighteenth Annual Conference on Manual Control, Wright-Patterson AFB, Ohio, June 8-10, 1982. (AFWAL-TR-83-3021)

Nineteenth Annual Conference on Manual Control, Massachusetts Institute of Technology, May 23-25, 1983. (MIT publication, no number)

Twentieth Annual Conference on Manual Control, Ames Research Center, Moffett Field, California, June 12-14, 1984. (NASA CP-2341)

Richard A. Miller
    Department of Industrial
    and Systems Engineering

Richard J. Jagacinski
    Department of Psychology

Twenty-First Annual Conference on Manual Control

Ohio State University

Columbus, Ohio

June 17-19, 1985

Descriptive Linear Modeling of
Steady—State Visual Evoked Response

by

William H. Levison
BBN Laboratories Incorporated
10 Moulton St.
Cambridge, MA   02238

Andrew M. Junker
AFAMRL

Kevin Kenner
Synergy, Inc.

ABSTRACT

AFAMRL is currently conducting a study to explore use of the steady—state visual—evoked electrocortical response as an indicator of cognitive task loading. Application of linear descriptive modeling to steady—state visual evoked response (VER) data obtained in the AFAMRL study is summarized in this paper. Two aspects of linear modeling are reviewed: (1) "unwrapping" the phase—shift portion of the frequency response, and (2) parsimonious characterization of task—loading effects in terms of changes in model parameters. Model—based phase unwrapping appears to be most reliable in applications —— such as manual control —— where theoretical models are available. Linear descriptive modeling of the VER has not yet been shown to provide consistent and readily interpretable results.

INTRODUCTION

Considerable effort has been devoted in recent years to the development of reliable metrics for pilot workload. Such metrics could be of value in the areas of cockpit design, pilot training, and flight operations. A measurement technique suitable for in—flight application could potentially warn of impending performance degradation and thereby allow timely remedial action. Assessment of workload in both simulated and operational flight tasks would enhance the identification of workload "bottlenecks", provide additional data for the evaluation of the crew/system interface, and, in general, provide information necessary for maintaining task workload within desired limits throughout a given mission.

Various studies have been undertaken in recent years to develop reliable metrics of pilot workload, including subjective estimates, primary and secondary task measures, and physiologic measures. Exploration of physiologic measures has been motivated by the desire to obtain one or more measures that are non—interfering with the primary mission and are not likely to be biased by the subject's preference for a given man/machine interface or his unwillingness to admit that a particular task is difficult.

1.1

AFAMRL is currently conducting a study to explore use of the steady-state visual-evoked electrocortical response as an indicator of cognitive task loading [1]. This paper summarizes the results to date of an effort to characterize the visual evoked response (VER) via linear descriptive modeling. Two applications of linear modeling are reviewed. Part I describes methods for "unwrapping" the phase-shift portion of the frequency response, an issue of concern when analyzing behavioral as well as physiological response. The central issue of this paper -- characterization of task-loading effects in terms of changes in model parameters -- is addressed in part II.

As of the writing of this paper, characterization of task loading effects is still in progress. Part II of this paper is consequently written in the style of a progress report.

PART I: PHASE UNWRAPPING

Nature of the Problem

To obtain the plots of amplitude-ratio ("gain") and phase-shift that are commonly used to characterize the response of linear systems, one typically employs the following procedure:

1. Compute Fourier transforms of the "input" and "output" time histories.

2. Divide Fourier coefficients (or cross-power spectral quantities) at frequencies of interest to obtain estimates of the frequency response as complex numbers.

3. Perform an appropriate nonlinear transformation to express the frequency response in terms of gain and phase-shift.

Various averaging techniques may be performed to enhance the reliability of the results as discussed in [2].

Procedures of this sort necessarily yield somewhat ambiguous phase-shift estimates, because phase repeats every 360 degrees. For example, a negative real number can be considered to have a phase shift of +180 degrees, -180 degrees, -540 degrees, etc. Therefore, we can shift any phase estimate by an integral multiple of plus or minus 360 degrees (one "cycle") and not be at variance with the data. In general, the frequency analysis scheme described so far must be accompanied by a procedure for "unwrapping" the phase in a meaningful way. Otherwise, the frequency shaping of the phase response will have a sawtooth appearance, since Fourier analysis schemes can only identify phase shift within a single cycle (typically, -180 to 180 degrees).

Techniques for Unwrapping the Phase Shift

Certain assumptions must be made in order to derive a method for unwrapping the phase. In the case of manual control data, we usually assume that phase varies relatively smoothly with frequency. That is, we assume that the frequencies at which we obtain frequency-response estimates are sufficiently close together so that successive phase estimates are unlikely to differ by more than 180 degrees. We simply unwrap the phase by adjusting the phase at each measurement frequency by the number of cycles required so that it does not differ from the preceding (in frequency)

estimate by more than 180 degrees. We also assume a reference point for the phase obtained at the lowest measurement frequency -- usually 0 or -180 degrees.

The assumption of a smoothly-varying phase response is not always justified, however. For example, unless the frequency-response measurements are finely quantized in frequency space, a highly-resonant system (especially one that is accompanied by significant pure delay) may well exhibit sharp changes in phase-shift in the region of the resonance.

If we wish to avoid the constraint that successive phase measurements differ by less than 180 degrees, we must assume that the phase and gain curves are related to each other in an orderly manner, and we must have a quantitative understanding of the analytic constraints (typically, a linear model) on the gain and phase curves. In this case, the experimental phase-shift is unwrapped with respect to a model-generated baseline.

Although we do not generally recommend that one "adjust the data to fit the model", such adjustments are entirely legitimate provided they are integral multiples of 360 degrees.

In general, the use of a model to unwrap the phase curve implies a model-matching exercise: a single iterative procedure is employed to jointly select parameters to best characterize the data and to unwrap the phase. Ideally, the model used for this purpose is a "theoretical model"; i.e., one that is expected on theoretical grounds to provide a good match to the data. Otherwise, a "descriptive" model may be employed which, while having no theoretical justification, is of a form that generates the type of qualitative frequency dependencies exhibited by the data.

The following procedure is suggested for unwrapping the phase via model analysis:

1. Use a theoretical model if one is available. Otherwise, select the least complex descriptive analytic model that seems likely to provide an acceptable match to the data.

2. For theoretical modeling, select an initial set of model parameters based on theoretical considerations or on previous modeling results. For descriptive modeling, important features of the frequency response may be analyzed to provide a reasonable initial parameter selection.

3. Using the current model parameters, predict gain and phase at each measurement frequency.

4. Readjust the experimental phase shift at each frequency, where necessary, by an integral multiple of 360 degrees until the experimental phase estimate is within 180 degrees of the corresponding model prediction.

5. Using an appropriate adjustment scheme and matching criteria, readjust independent model parameters to improve the match to the data.

6. Iterate on steps 3-5 until the matching criteria are satisfied. The resulting adjusted experimental phase curve is substituted for the sawtooth curve originally yielded by the Fourier analysis scheme.

This procedure is based on the assumption that frequency response data are to be matched. Other techniques for parameter adjustment might be employed if modeling is to be applied instead to the relevant time histories.

The validity of this procedure can be judged in a particular application in terms of the resulting model match. If a good qualitative match is obtained to both the gain and phase curves (note: experimental gain is not adjusted), then the resulting adjustments to the phase curve can be accepted as valid; otherwise the phase curve should be unwrapped using another model form.

Application of Model—Based Phase Unwrapping

Application of the model—based technique described above is demonstrated for both manual control and physiological response data. A theoretical model is used for the manual control data, whereas a linear descriptive model is employed for the physiologic data.

Manual Control Example

Figure 1a shows frequency—response data obtained in a recent simulation of an F—14 performing a steady—state gunsight tracking task [3]. The data points related to phase shift show sharp positive jumps at around 1 and 11 rad/sec because of the −180 and +180 degree boundaries on the Fourier analyzer.

Because these data were obtained in a tracking task employing a known task environment using linearizable vehicle dynamics, the optimal control model (OCM) for piloted systems was used to unwrap the phase. No model—matching was employed; rather, a single prediction of pilot response behavior was generated using pilot—related model parameters typical of those found to match human operator behavior in previous studies. The phase—shift curve was then used as a point—by—point baseline for unwrapping the experimental phase data. As shown in Figure 1b, the initial selection of model parameters gave a qualitatively good match to the data; there was no need to improve the model—match, via parameter adjustment, in order to demonstrate the validity of the unwrapped phase curve.

For this particular data set, the same phase unwrapping is generated by simply assuming that consecutive data points do not differ by more than 180 degrees. Nevertheless, in general, the results are more compelling if they are shown to be consistent with reasonable analytical constraints.

Application to Visual Evoked Response

At present, theoretical models of the type available for manual control do not exist for the visual evoked electrocortical response (VER). Unlike the manual control task, where a specific response strategy can usually be derived for accomplishing well—defined control objectives (particularly in a laboratory setting), the VER is not known to have a similar teleological foundation. Unless one is using the VER for biofeedback in a control loop, it is not clear why the electrocortical potentials recorded from the scalp should bear any particular relationship to the visual stimulus. Thus, to the extent that we rely on model analysis to unwrap the steady—state VER phase data, we must currently use descriptive models.

Figure 2a shows the average gain and phase data obtained from a single subject in an ongoing AFAMRL study of steady—state VER. (The details of this experiment are
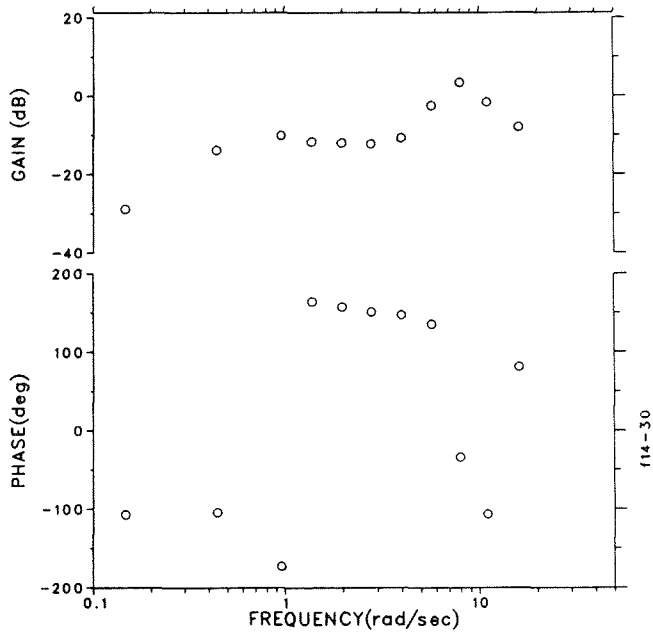
a)   Phase Unmodified   b)   Phase Unwrapped With OCM
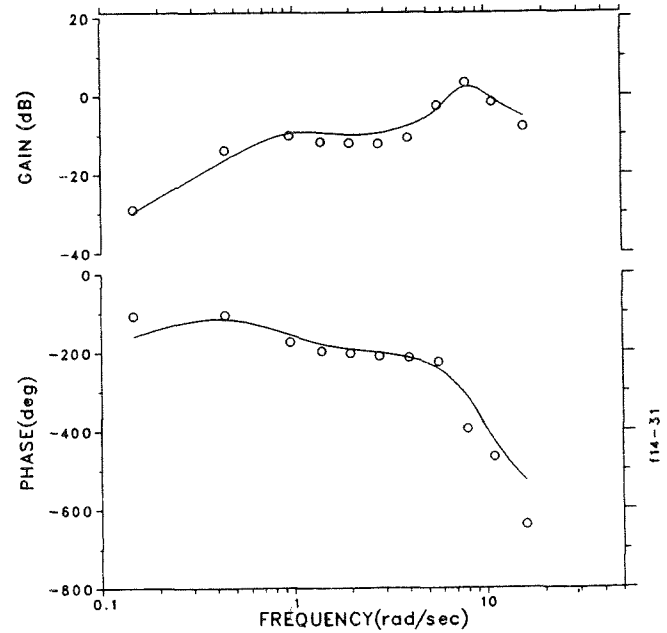
Figure 1.   Pilot Frequency Response, Simulated F-14
            Gunsight Tracking Task

briefly summarized later in this paper and in more detail by Junker et al [1]). The
unmodified phase curve shows upward-directed discontinuities at around 8, 15, and 20
Hz.

Because the gain curve has the general appearance of a second-order resonant
lowpass filter, a linear model of the following form was employed to unwrap the phase:

$$F(s) = \frac{K\,\omega_0\,e^{-sT}}{s^2 + 2\zeta\,\omega_0 s + \omega_0^2}$$

where the four independent model parameters are the asymptotic low-frequency gain
K, the natural frequency ω, the damping-ratio ζ, and the pure time delay T. (The
frequency variable "s" is not a model parameter.)

1.5

a) Phase Unmodified

b) Phase Unwrapped with
   Descriptive Linear Model
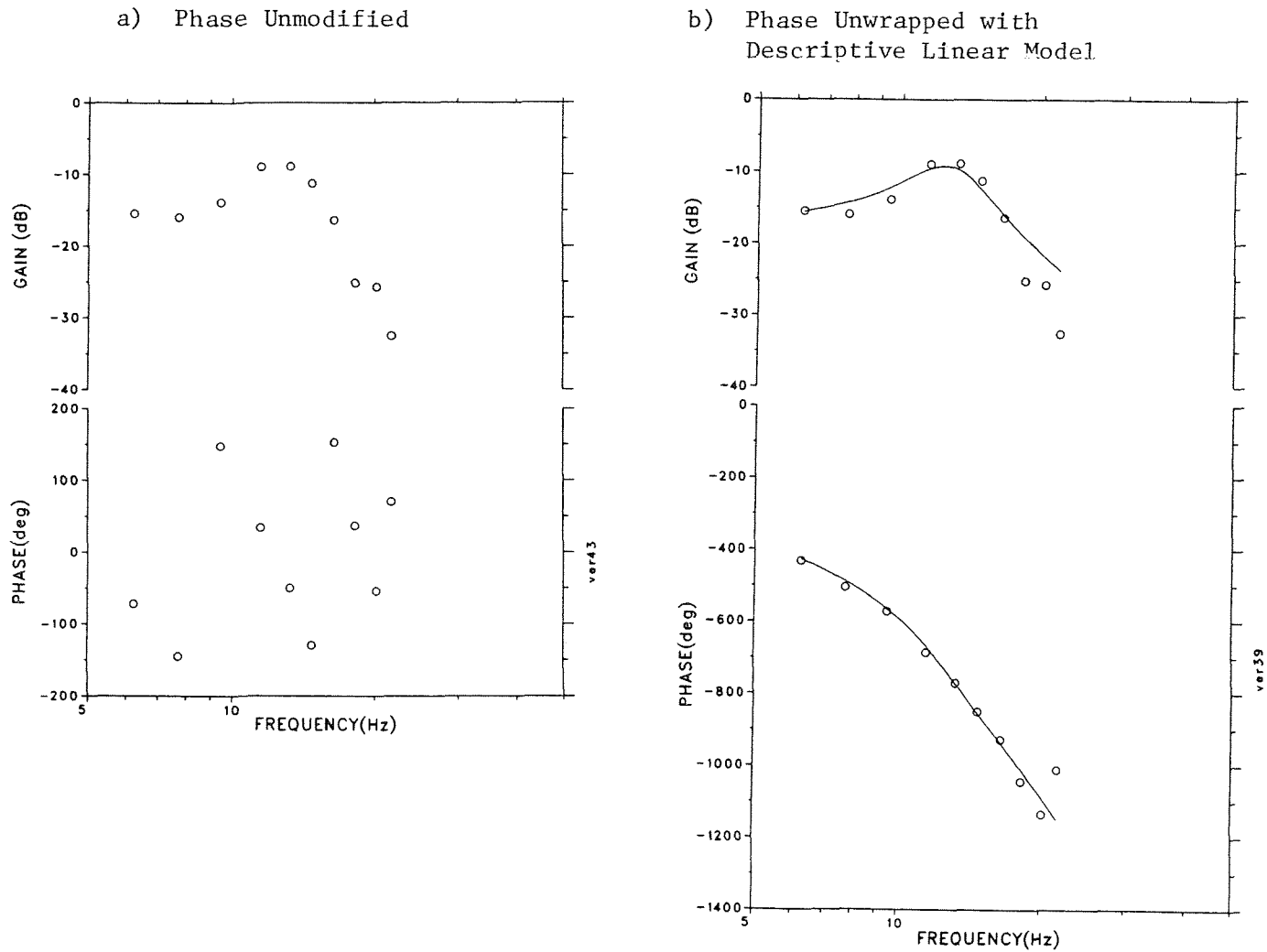
Figure 2.   Visual Evoked Response, Example 1

An initial selection of parameters was based on the apparent resonance frequency, the asymptotic low—frequency gain, and the difference between maximum and low—frequency gains. In addition, the monotonic and relatively sharp negative increase in phase shift with frequency suggested the presence of a pure delay term, which was also included in the model. The initial estimate of the delay was chosen on the basis of the slope of the phase curve after a preliminary unwrapping in which a 180—degree difference limitation was imposed.

A scalar model—matching error was defined as the rms difference between model predictions and experimental data, weighted inversely by the standard errors of the experimental data. (The unwrapped phase estimates were used for this computation.) Best—fitting model parameter values were identified using a quasi—Newton gradient search scheme similar to that employed previously in manual control studies [4,5].

1.6

Because the lowest measurement frequency was relatively large (5 Hz, compared to 0.15 rad/sec for the tracking data), we could not rely on the data of Figure 2a to determine the asymptotic zero—frequency phase shift. It was not obvious whether the asymptotic frequency would be referenced to 0 degrees (implying a positive low—frequency model gain), or −180 degrees (implying a negative gain). Accordingly, model analysis was performed with both positive and negative gains, and results were accepted from the model yielding the smallest matching error. ("Gain" here refers to the scale factor parameter K, specified as a real number, not the the amplitude—ratio portion of the frequency response, which is specified in logarithmic units.)

Analysis with the negative gain yielded a substantially lower matching error; the resulting phase curve is shown in Figure 2b. The relatively good qualitative match to the data suggests that the phase curve is likely to be valid, with the possible exception of the phase at the highest measurement frequency.

Application of the same model form to another VER data set is shown in Figure 3 for both positive model gain (Fig. 3a) and negative model gain (Fig. 3b). For this data set, the two model—matches yielded nearly identical matching errors, but the unwrapped phase curves differed by 360 degrees. Apparently, the −180 degree phase shift imposed on the model predictions by the negative gain shifted the predicted phase response sufficiently to require an extra 360 degrees of unwrapping in order to minimize model—data differences.

Because we have no theoretical basis for determining the asymptotic low—frequency phase shift (equivalently, the sign of the model gain parameter), and because the qualitative matches to the data sets are equally good (though different in detail), the two phase curves must be considered equally valid. Thus, the phase unwrapping remains to some extent ambiguous when a second—order resonant loss—pass filter is adopted as the model form. Other model forms might provide unambiguous results, but that would have to be determined from trial and error.

PART II: LINEAR MODELING OF STEADY—STATE VER

Background

Prior research has indicated that recorded scalp electrical potentials respond, to some extent, in a manner linearly related to the visual stimulus. There is, in addition, a strong nonlinear component of the response, plus a substantial amount of unrelated ongoing electrical activity that is present. Under proper stimulus conditions, the linear component of the response is large enough to allow its estimation with reasonable statistical confidence. Thus, this electrophysiological system lends itself to the analytical techniques employed in pilot/vehicle analysis —— i.e., to the measurement of describing function and remnant —— as has been demonstrated above in Part I. The focus of the ongoing research, to which this paper is addressed, is to determine whether such measures are sensitive to workload and other forms of stress.

As noted earlier, we cannot define a "purpose" for the visual evoked response, in the sense that we can for control response in a well—defined tracking task. Not only do we lack a theoretical model for what the evoked response ought to be, there is no obvious functional relation between the response (electrical potentials measured at the scalp) and the demands of the "task" (which may be no more specific than to attend to or fixate on the stimulus). Therefore, our basis for interpreting visually evoked response is not as solid as our basis for interpreting manual control response, and intra— and inter—subject variability tends to be substantially greater than with

1.7

a) Model with Positive Gain   b) Model with Negative Gain
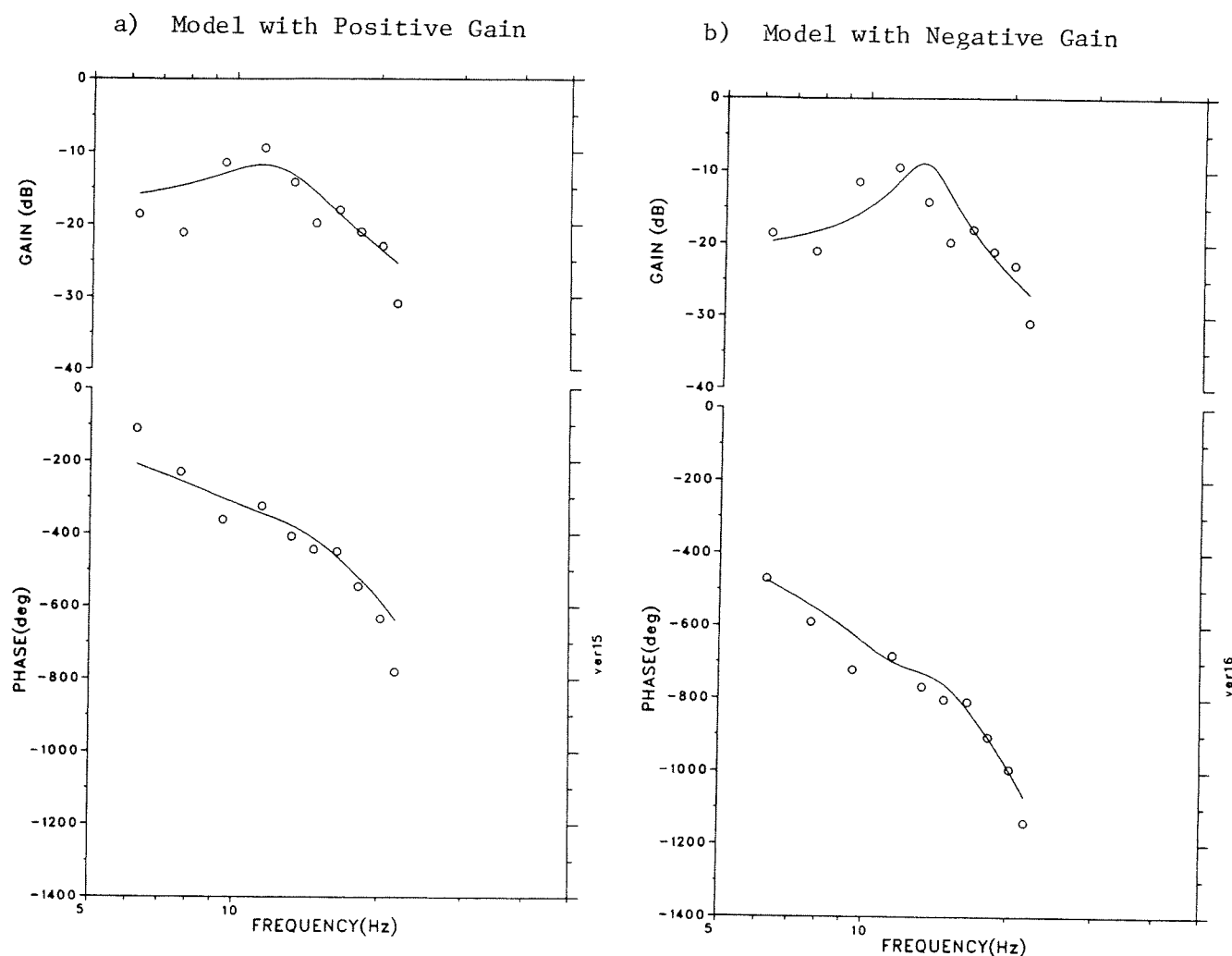
Figure 3.   Visual Evoked Response, Example 2
            Second-order Lowpass Resonant Filter

manual control response behavior. The averaging technique described elsewhere in these Proceedings by Levison [2] were developed largely to deal with this variability.

A number of research efforts have focused on obtaining a frequency—response description of the VER [6—9]. In what is perhaps the most comprehensive effort to date, Spekreijse [9] measured the VER using inputs consisting of single sinusoids (as opposed to a sum—of—sinusoids), or single sinusoids plus Gaussian noise. His work focused a great deal on characterizing the nonlinear aspects of the response. On the basis of numerous sub—experiments, Spekreijse concluded that nonlinear response components in the VER were due largely to memoryless rectification and saturation nonlinearities and that these nonlinearities were located prior to the "cortical selective process" If this model is correct, then nonlinear VER components are not influenced by the operator's cognitive state, and we are justified in characterizing task—related VER changes in terms of quasi—linear model parameters even though the VER may contain significant nonlinear response components.

1.8

More recently, Junker and Peio [10] obtained steady—state evoked responses to sum—of—sinusoids visual stimuli. They found that, although the nature of the frequency response varied from subject—to—subject, it appeared to be relatively stable for a given subject across replications, and to be influenced by the task environment. Preliminary analysis of their data revealed that, for at least some of the data sets, the frequency response could be reasonably well characterized by a second—order linear descriptive model.

Experiments

Details of the VER experiment are provided in a companion by Junker et. al. [1]. A brief overview is given here.

Electrocortical response was recorded from subjects exposed to spatially uniform light stimulus modulated by a complex sum of sinusoids. Ten sinusoidal components of uniform amplitude and random phasing were used, with component frequencies ranging from 6.25 to 21.75 Hz.

Three task loading conditions, provided in a balanced order, were explored: (a) no explicit task, other than attending to the flashing lights, (b) a first—order manual tracking task, and (c) a grammatical reasoning task. Analysis techniques similar to those applied extensively to manual control analysis were employed here to obtain the frequency response characteristics of the VER. Response metrics consisted of amplitude ratio ("gain") and phase shift, measured at stimulus frequencies, and "remnant" (response components at other than input frequencies) averaged over 1—Hz "windows" centered about each input frequency. Only the gain and phase data are considered here.

Data from seven subjects were considered statistically reliable and were made available for model analysis. Each VER frequency response considered in this paper represents the average of from six to eight 40—second segments of electrocortical recordings. Averaging was performed as described by Levison [2].

Model Analysis

Model analysis was performed as described in Part I. The objectives of this analysis were to unwrap the phase to aid in overall interpretation of the frequency response, and to determine whether or not the independent model parameters would provide a parsimonious and consistent characterization of task—loading effects.

As noted above, preliminary results led us to believe that a lowpass filter of the type defined in Equation 1 would characterize the steady—state visual evoked response at least for a portion of the subject population. Data from all seven subjects were initially modeled in this manner. Positive and negative gains were tested, and whichever sign yielded the smallest matching error was included in the parameterization for a given data set.

Application of the second—order model did not yield consistently useful results, either for phase unwrapping or for interpretation of the evoked response. The resonant lowpass filter provided a good qualitative match to only a portion of the data sets; for data where the match was not qualitatively acceptable, the validity of the resulting phase curve had to be questioned.

1.9

The best—fitting model parameters did not reveal a consistent trend with task loading, and they tended to vary over wide ranges from one data set to the next. Nearly as many data sets were best matched with a positive model gain parameter as with a negative gain. This result implies that the polarity of the recording electrodes was changed from one condition to the next —— a notion at variance with the experimental procedures followed in this study.

Even where a good qualitative match was obtained, the resulting model parameters were often inconsistent with the assumption of a stable linear system. For example, the model fits shown in Figures 3a and 3b were obtained with negative damping ratios —— a characteristic of a system whose oscillatory response grows exponentially with time. Such a result is inconsistent with electrocortical responses obtained with transient stimuli. When subsequent model analysis was performed with the constraint that the damping—ratio and natural—frequency parameters remain positive, substantially greater matching errors were obtained in most cases.

Inspection of the data (specifically, the gain curves) suggested that other model forms would more closely resemble the frequency dependency of the data. Figure 4 shows an example of a data set matched with the following fourth—order bandpass filter:

$$F(s) = K \cdot \frac{s^2}{s^2 + 2\zeta\,\omega_1 s + \omega_1{}^2} \cdot \frac{\omega_2{}^2\, e^{-sT}}{s^2 + 2\zeta\,\omega_2 s + \omega_2{}^2}$$

This model also has four independent parameters: gain, two natural frequencies, and delay. (The damping ratios were fixed at 0.707.)

By constraining the two frequency parameters to be positive, we were able to characterize the data with a stable linear system. Analysis with this model form was not conducted on a large scale, however, because of the sensitivity of the results to the initial parameter selection —— a situation not uncommon when employing gradient search schemes.

The difficulty of obtaining a consistent model—based characterization of the steady—state VER is indicated by inspection of the gain curves shown for two test subjects in Figure 5. For the baseline (no—task) condition, the data for Subject 2 (Fig. 5a) resemble the frequency response of a resonant lowpass filter, whereas the data for Subject 3 (Fig. 5d) resemble an inverted "v" and are perhaps modeled by a tuned bandpass filter. (The data shown in Figure 5a were used for the demonstration of phase unwrapping in Figure 2.)

The curves for the tracking condition (Figures 5b and d) show no consistent effects of task loading: the data from Subject 2 reveal regions of diminished response, whereas the data from Subject 3 show less of a qualitative change from the baseline. For the grammatical reasoning condition, however, both subjects showed gain response curves that appeared to vary less with frequency than the baseline.
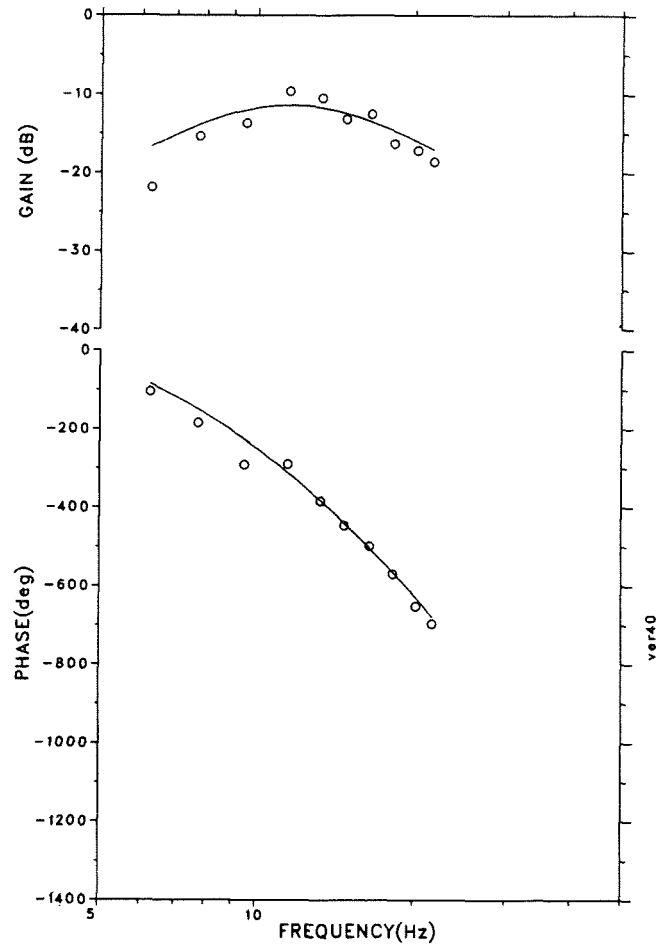
Figure 4. Visual Evoked Response, Example 3
           Fourth-order Bandpass Filter

The trends revealed in Figure 5 suggested the hypothesis that the gain response is "flatter" for the reasoning task than for the baseline condition. Accordingly, data from the first three test subjects providing complete data sets (Subjects 2,3, and 5) were modeled with a simple gain/delay model of the form:

a) Subject 2, No Task

d) Subject 3, No Task

b) Subject 2, Tracking

e) Subject 3, Tracking

c) Subject 2, Reasoning

f) Subject 3, Reasoning
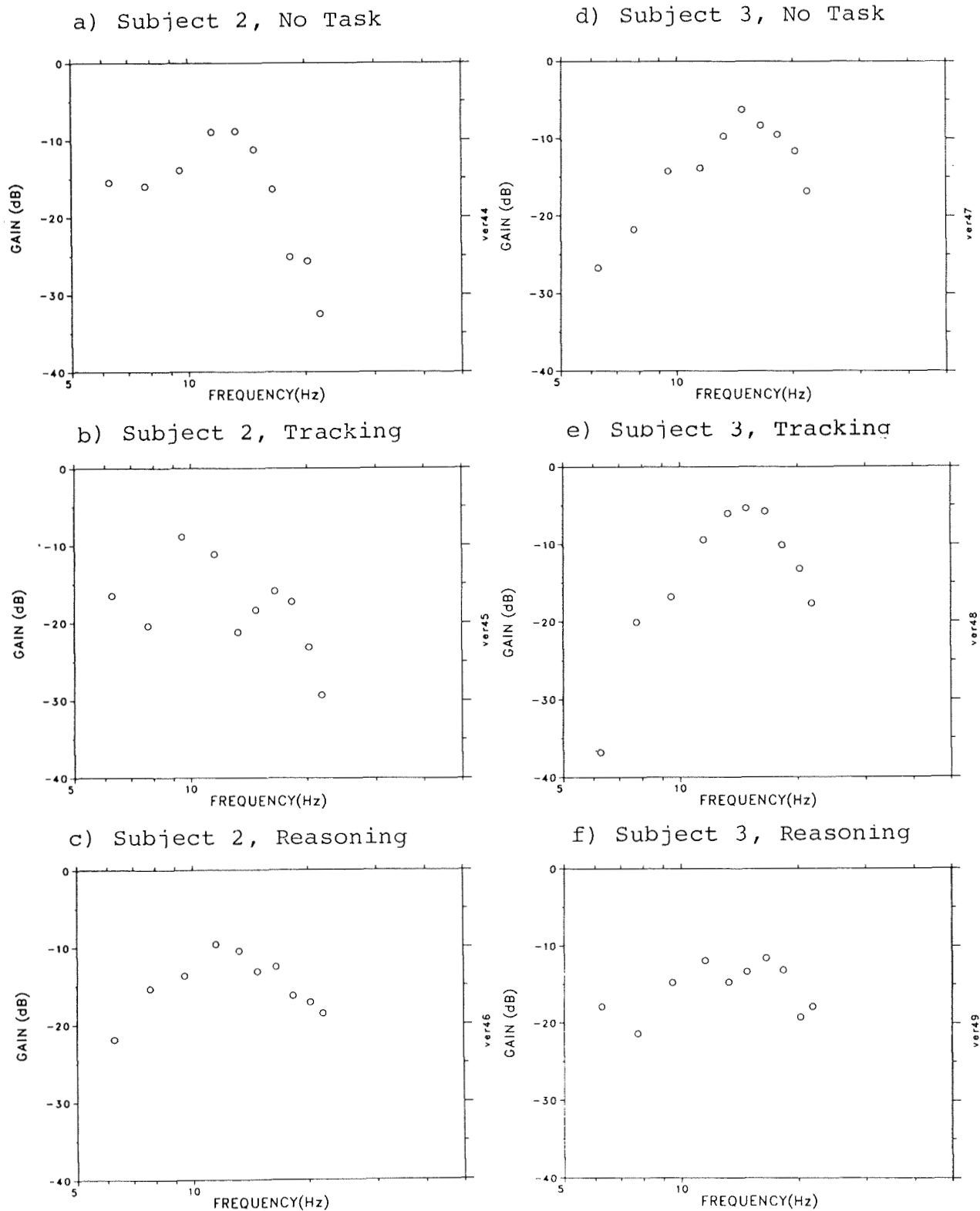
Figure 5. Effects of Task Loading on Visual Evoked Gain Response

1.12

$$F(s) = K e^{-sT}$$

where K and T are the "gain" and delay parameters, respectively. The reasoning behind this test was that, if the flat−response hypothesis were true, this model form would yield lowest matching errors for the grammatical reasoning condition.

Figure 6 (bottom graph) shows that, for the three subjects tested (time did not permit testing of the entire data base), the gain/delay model yielded the lowest matching error for the reasoning task, thereby providing some quantitative support for the qualitative trend suggested above. Testing of the remaining data is required to explore the generality of the hypothesis. Visual inspection of the frequency response yielded by the other subjects (not shown here) suggests that this trend will not hold for the entire subject population.

The top two graphs of Figure 6 show that task loading conditions did not have a consistent effect on the gain and delay parameters across the three subjects. This simple model form, then, appears to be of use only for testing some very general data trends −− not for paramaterizing the VER in a meaningful way.

DISCUSSION

The use of a model to unwrap the phase−shift response is not uncommon, but it is usually informal and implicit. Typically, the individual performing the analysis has an expectation of what the frequency dependency should be, based on previous experience with similar systems, and unwraps the data according to a qualitative "mental model". What we have done here is to suggest that the procedure be made more explicit with the use of a specific mathematical model, with a combined procedure of phase unwrapping and parameter adjustment if need be. Provided a suitable model structure is available, with a solid basis for initial parameter selection, such a procedure provides a means for automated phase unwrapping.

Although preliminary results encouraged the application of linear descriptive models of the VER, modeling of this form has not been demonstrated so far to be a reliable method for characterizing task loading effects. Although model forms can be found to provide a reasonable qualitative match to the data, the appropriate model form appears to vary across subjects and sometimes across tasks, parameter variations do not follow a clear trend, and model parameter values are not always consistent with a stable response mechanism.

It is tempting to conclude that the relative lack of modeling success (in terms of our stated goals) is due, in part, to the fact that we are attempting to model a nonlinear response mechanism with a linear model. We do not think this is a major factor. However nonlinear the VER might be, it does contain a measurable and
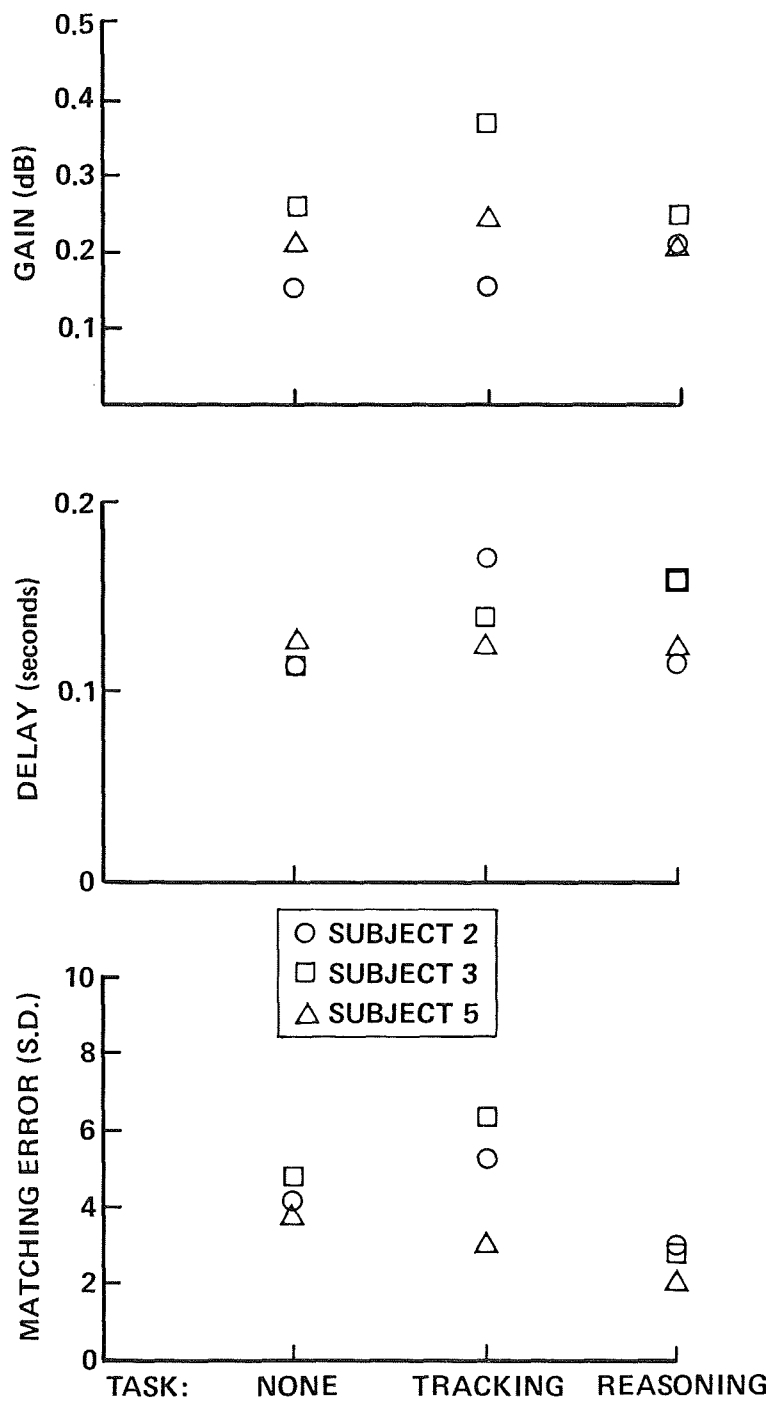
Figure 6. Effects of Task Loading on Parameters of a Gain/Delay Model

generally statistically reliable linear response component. If task loading were to change the response behavior in a consistent manner, we would expect the linear response component to change in a consistent manner.

It is possible that we have not explored the appropriate model forms. To the extent that model analysis is pursued during the remainder of this study, model forms that have a structure based more on theoretical considerations [11,12,13] will be explored. Another avenue to be explored is the effect of task loading on the variability of the VER, rather than the mean [14].

A more likely source of the difficulty is that there is no "reason" for the electrocortical potentials to exhibit a particular pattern, in terms of what the subject is trying to accomplish. To create a situation closer to that of manual control tasks, where generation of a particular response behavior can aid the achievement of task-related goals imposed upon the test subject, it is anticipated that the AFAMRL study will explore the use of the evoked response in a continuous control task employing biofeedback. A task environment of this sort is expected to reduce the variability of the VER and make it more sensitive to task loading. The use of the VER as an "unobtrusive" measure of task loading may be compromised, however, as the VER will now be a component of a secondary task competing for attention with the primary cognitive (or psychomotor) task.

Inspection of the available data base suggests that there may be important inter-subject differences in terms of the linear response behavior. Thus, while not yielding a consistent index of task loading, linear analysis may prove viable as a means for characterizing subject differences. It remains to be established whether such differences, if found to be statistically significant, relate in a consistent manner to behavioral aspects of interest, and not simply to physical characteristics such as differences in the shape of the skull.

Finally, we note that the "remnant" (background eeg) remains to be analyzed. Although the effects of task loading and individual differences appear to be smaller for the remnant than for the main curve, it is possible that remnant changes are statistically more significant.

ACKNOWLEDGEMENT

REFERENCES

1. Junker, A.M.; Kenner, K.M.; Kleinman, D.L.; McClurg, T.D., "Analytical Comparison of Transient and Steady-State Visual Evoked Cortical Potentials", Proc. 21st Annual Conference on Manual Control, June 17-19, 1985, Columbus, Ohio.

2. Levison, W.H., "Some Computational Techniques for Estimating Human Operator Describing Functions", Proc. 21st Annual Conference on Manual Control, June 17-19, 1985, Columbus, Ohio.

3. F-14 Modeling Study, Contract NAS1-17648.

4. Lancraft, R.E., and D.L. Kleinman, "On the Identification of Parameters in the Optimal Control Model", Proc. of the Fifteenth Annual Conference on Manual Control, Wright State University, Dayton, Ohio, March, 1979.

1.15

5. Levison, W.H., "Development of a Model for Human Operator Learning in Continuous Estimation and Control Tasks", AFAMRL-TR-83-088, Air Force Aerospace Medical Research Laboratory, WPAFB, Ohio, December 1983.

6. Regan, David, and Beverly, K.I., 1973. "Relationships Between the Magnitude of Flicker Sensation and Evoked Potential Amplitude in Man." Perception. 2:61-65.

7. Wilson, Glen F., 1979. "Steady State Evoked Response as a Measure of Tracking Difficulty", Final Report contract no. F49620-79-C.

8. Wilson, Glen F. and O'Donnell, Robert D., 1981. "Human Sensitivity to High Frequency Sine Wave and Pulsed Light Stimulation as Measured by the Steady-State Cortical Evoked Response", Technical Report AFAMRL-TR-80-133.

9. Spekreijse, Henk, 1966. "Analysis of HEG Response in Man Evoked by Sine Wave Modulated Light", Dr. W. Junk Publishers. The Hague, The Netherlands, 152 pp.

10. Junker, Andrew M., and Peio, Karen J., 1984. "In Search of a Visual-Cortical Describing Function - A Summary of Work in Progress", Proc. of 20th Annual Conference on Manual Control. NASA Conference Publication 2341, vol II: 37-54.

11. Childers, D.G. and Perry, N.W., "Alpha-like Activity in Vision", Brain Research, 25:1-20, 1971.

12. Childers, D.G. and Mesa, W.; Halpeny, O.S., "A Neuronal Population Model for Simulation of Spatio-Temporal Evoked EEG", IEEE Trans SMC-3:336-349, July 1973.

13. Freeman, W.J., "Parallel Processing of Signals in Neural Sets as Manifested in the EEG", Int. J. Man-Machine Studies 7:347-369, 1975.

14. Robinson, D.N., and Sabat, S.R., "Neuroelectric Aspects of Information Processing by the Brain", Neuropsychologia, 15:625-641, 1977.

ANALYTICAL COMPARISON OF TRANSIENT AND STEADY STATE VISUAL EVOKED CORTICAL POTENTIALS.

Andrew M. Junker, Aerospace Medical Research Lab. WPAFB, Oh.
Kevin M. Kenner, Synergy Inc., Dayton, Oh.,
David L. Kleinman, Univ. of Conn., New Haven Ct.
Terrence D. McClurg, Systems Research Lab. Inc., Dayton, Oh.

To better describe the linear-dynamic properties of the human visual-cortical response system, transient and steady state Visual Evoked Response Potentials (VERP) were observed. The stimulus presentation device provided both the evoking stimulus (flickering or pulsing lights) and a video task display. The steady state stimulus was modulated by a complex, ten frequency, sum-of-sines, wave. The transient VERP was the time-locked average of the EEG to a series of narrow light pulses (pulse width of 10 msec). The Fourier transform of the averaged pulses had properties that approximate band limited white noise, i.e. a flat spectrum over the frequency region spanned by the 10 summed sines. The Fourier transform of both the steady state and the transient evoked potentials resulted in transfer functions that are equivalent and therefore comparable. To investigate the effects of task loading on evoked potentials, a grammatical reasoning task was provided. Results support the relevancy of continued application of a systems engineering approach for describing neurosensory functioning.

INTRODUCTION

A new methodology for analyzing and interpreting the dynamics of the brain's response system based upon sum of sines (SOS) stimulation and systems engineering analysis has been developed (Junker and Peio, 1984). This technology requires that the system being studied possess a significant degree of linearity for the measure to be of descriptive value. The question of linearity is considered in this paper by comparing systemic responses to two types of stimulation.

One of the greatest challenges in examining the brain's electrical potentials is the low signal to noise ratio. Evoked responses are so small in comparison to the background electrical activity of the brain that a method for enhancing the signal to noise ratio must be decided upon. There are two well developed techniques for accomplishing this. Steady state evoked response potentials (SSERP) are based on the frequency following phenomenon in the human response system. Using a repeating stimulus successive ERP's are elicited. It has been shown that the elicited response contains the repeating frequency of the stimulus. The brain does not have a chance to regain its resting, undisturbed state (Regan, 1973, 1975, 1979). With the

aid of the Fast Fourier Transform (FFT) it can be demonstrated that the ERP is at exactly the same frequency as the stimulus. The other method, transient ERP, is based on " hitting" the system, with a pulse or a click, and then measuring the electrical potential. Each response to the pulse is considered a transient response. The amplitude of the transient response can be measured and then a series of responses can be averaged together. This is a time-locked average, that assumes the response always occurs at the same time relative to stimulus onset.

In the field of automatic control systems technology, an input/output relationship for the linear portion of a nonlinear system is defined as a describing function (Kochenburger, 1950). Manipulations of the FFT's of the response potential (output) and evoking stimulus (input) yield a describing function that is a complex measure of the output/input relationship of this sensory-response system (for a detailed description of guidlines for analysis of frequency response data see Levison, 1983). Prior to the development of the ten sine wave stimulation technique (Junker and Peio, 1984) most measurements were made with single sine waves, or at most three sine waves simultaneously (Regan, 1973, Wilson, 1979 and Wilson and O'Donnell, 1981). Construction of comprehensive describing functions were not often undertaken, and there is no data available from researchers showing task loading effects across a broad range of frequencies. Perhaps for this reason no one has taken on the task, until this time, of exploring the actual relationship between steady state and transient VER potentials. Regan (1979) stated that, "For a linear system the transient response has a fixed relationship to the steady-state response. Consequently, transient and steady state descriptions of a linear system's behavior are equivalent and can be regarded as alternative formulations of the same data...therefore, transient and steady- state stimulation can produce responses that provide complimentary information about the sensory system under test." Using our existing stimulus apparatus and computer generating capability, we incorporated into the system the ability to: accurately generate a narrow pulse stimulus, collect and time-lock average the data, FFT the results and compute describing functions from the transforms. These describing functions were used for comparison with steady state describing functions. This analysis has been applied to ERP's in taskloading (workload) and non-taskloading conditions. The task used was grammatical reasoning. This task was selected because it is highly engaging and it only requires a minimal of motor response. An explanation of the stimulation device, EEG data collection, and sum of sines methodology is presented in this paper. In addition the cognitive loading task, the transient stimulation methodology, the transient results, and comparisons between transient and steady state describing functons are given.

METHODS

Apparatus

The test chamber simultaneously delivers the evoking
stimulus (flickering lights) and a video task display (Figure 1).
This presentation was achieved by combining the two images via an
18 cm x 26 cm half-silvered mirror at 45 degrees to the two
images. The evoking stimulus was produced by two 26 cm
xenon/fluorescent light tubes hung horizontally 5 cm apart and
mounted 4 cm behind a 25 cm x 27 cm transluscent, diffusing
screen; which distributed the light, as evenly as possible, over
the visual field. The average intensity of the lights were 40
FL, as measured, by a United Detector,model PIN 10D, high speed
photo cell, placed at the subject's viewing point. This average
intensity was sufficiently low such that a subject could still
comfortably discern the video task-display within the same visual
field. The video task was displayed on an Audiometrix 11 in x 11
in video monitor.

Beckman silver/silver chloride electrodes were used with the
Grass model P511 AC amplifiers, with amplification x50,000 and
bandpass of 0.1 to 300 Hz, to record the EEG. The sum-of-sines
(SOS) wave and transient pulse were generated, and data
collected, on a Digital Equipment Corp.(DEC) PDP 11/60 computer.
Signals from the 11/60 were low pass filtered on a Krohn-Hite
model 3750 filter (cut off at 40 hz) and then fed into a
Scientific Prototype, model GB, tachistiscope/light driver, which
was modified so that average intensity and depth of modulation
could be adjusted. The grammatical reasoning task was generated
by a Commodore model VIC computer. _ The software and the
response-box hardware, for this task, were developed by Systems
Research Laboratories Inc. The two channels of data (photo cell
and EEG) were fed through General Radio low pass filters (cut off
at 25 Hz) to prevent high frequency aliasing. The filtered
signals were then digitized and stored for analysis on the PDP
11/60. The collected data was fast fourier transformed, ensemble
averaged and plotted using a DEC PDP 11/34 computer and a
Printronix model P300 printer.

Stimulus

To better elucidate the linear properties of the visual-
cortical response system the experiment was designed to collect
describing function measures with different forms of inputs. The
modulated light served as the driving stimulus. For steady state
stimulation the lights were modulated using a complex SOS wave
composed of 10 harmonically non-related frequencies. All 10 of
the frequencies were multiples of the fundamental frequency of
0.0244 hz. The component frequencies range from approximately
6.25 to 21.75 Hz, with intermediate frequencies at 7.75, 9.50,
11.50, 13.25, 14.75, 16.50, 18.25, and 20.25 Hz. None of these
component frequencies contained a sum or difference of any of the
other component frequencies; this restriction on sine wave
selection was implemented to avoid the possible corruptions at
the selected frequencies by nonlinearities of the flickering

light generator and possible nonlinear evoked potential responses. Appropriate input selection insured that nonlinear harmonic effects would not occur at the component frequencies.

For every data collecting trial the starting phase values for each of the 10 component sine waves were randomized with a uniform random number generator, insuring that the time sequence of flickering light presentation was random from trial to trial. By utilizing randomized phase with the summing of the 10 sinusoids a maximum depth of modulation of 13%, per sinusoid was possible. The lights were sinusoidally modulated about an average luminance of 40 ft-lamberts. Previous work (Junker and Peio, 1984) had shown 6.5% to be sufficient for obtaining VER's. Regan and Beverley (1973) in looking at the effects of the percent depth of modulation on the VER demonstrated a straight line relationship between VER volts and percent depth of modulation over a limited range (10% to 30%) of modulation. Over 30% a saturation-like effect occurred indicating nonlinear behavior in the VER data. Thus our stimulus depth of modulation minimized nonlinear overdriving while still assuring an adequate VER.

For comparison purposes we created our transient stimulus to have power spectral properties similar to the spectral properties of the sum of sines stimulus. The sum of sines consisted of 10 sine waves ranging from 6.25 to 21.75 Hz, with equal power for each of the component sinusoids. The power spectrum of the transient pulse was adjusted to have a flat spectrum over the same frequency range. The transient stimulus was a narrow (.01 sec duration) computer generated pulse driven through the low pass filter and fed into our light driving circuit. Interstimulus time was varied between 1.28 and 1.38 sec, the variability (0 to 0.1 sec) was generated with a uniform random number generator for each stimulation segment. One run, or trial, consisted of 40 stimulus segments.

Task

The task loading condition used was the grammatical reasoning task from the Criterion Task Set (Shingledecker, et. al., 1983). This task is based on the original grammatical reasoning task developed by Baddeley (1968). The task is designed to impose variable processing demands on resources used for the manipulation of grammatical information. Stimulus items are two sentences of varying syntactic structure accompanied by sets of three symbols. The sentences must be analyzed to determine whether they correctly describe the ordering of the characters in the symbol set. This version used two sentence items worded either actively/negatively or passively/positively and described three symbols. This was considered the high demand level. The object for the subject was to determine whether both sentences match in their correctness. If both sentences correctly described the ordering of the three symbols, or if neither correctly described the symbols, the appropriate response was positive. If one sentence was correct but the other was not

the appropriate response was negative. There was a 7.5 sec time limit for responding. Binary responses were entered manually on two labeled keys, of a four button keypad, placed on the right arm of the subject's chair.

Procedure

Subjects were seated in a darkened IAC chamber facing a 15 cm x 15 cm window. Behind the window was the stimulus presentation device. For the lights only condition the subjects were instructed to "relax and fixate on a small square at the center of the display", for the cognitive loading condition the subjects were instructed to concentrate on the task. Each trial lasted 82 sec. and after every three trials the experimenter entered the booth to inquire about the status of the subject (alertness, fatigue etc); every sixth trial the subjects were given a 3-6 min. break. Sessions were either 12 or 18 trials long. Subjects were advised that the session could be terminated at any time upon their request.

Transient data was collected from subjects at the end of the same sessions in which steady state data was collected. Data was collected for four trials of lights only (no task load) and then for four trials in which subjects performed the grammatical reasoning task (task loading).

Analysis

Manipulations of the fast fourier transforms of the photo cell signal (input) and the evoked response potential signal (output) yields a describing function which is a complex measure of the output-input relationship of this system. The focus of this project was on the amplitude ratio and the phase angle measures obtained from these computations.

For SOS stimulation we were interested in estimates of mean values for the gain and phase computations across replications. For indication of mean variability we calculated the standard error by computing the standard deviation across replications and dividing by the square root of the number of replications. If the data had been normally distributed these computations would have, in fact, been a measure of standard error.

For transient stimulation collected data was analyzed with a DEC 11/34 computer. Time lock averaging of each of the 40 segments for each trial was done first, then averaging across trials for each condition (4 trials per condition) was performed. Time responses were plotted for lights only and task loading conditions. In addition, time responses were Fast Fourier Transformed and describing function gain and phase values were computed. The describing functions were plotted, for comparison, with sum of sines generated describing functions.

Recording

Recording was done with Beckman silver/silver chloride

2.5

electrodes at Oz, with linked mastoids as ground and reference according to the 10-20 International System. The resistance between the electrodes was less than 5 K ohms.

RESULTS and DISCUSSION

Time locked average responses to the pulse stimulus, for both lights only and task loading are presented in Fig 2 for four subjects. Strong effects from task loading, namely overall decreases in response peaks are present for subjects 02, 03, and 05. An opposite trend, a slight increase in response peaks with task loading, can be seen for subject 15. Typically time locked averaged data, such as this, is analyzed using component analysis techniques or principal factor analysis (Regan 1973, John et.al. 1973). It is possible to take this data a step further, into the frequency domain, by computing decribing functions as was done for steady state ERP data.

Corresponding describing function results are given in Fig 3. An important relationship to observe is the mapping between transient time average changes, related to task loading effects (from Fig 2) and corresponding describing function changes, in the frequency domain (Fig 3). Subjects 02, 03, and 05, who exhibited amplitude decrements in their average time responses with task loading, showed a concomitant decrease in describing function gain curves. It is interesting to note where the greatest gain changes occured. For subjects 02 and 05 these changes were in the lower frequency range (centered about the alpha frequency band, 10 Hz), while for subject 03 a reduction in gain, with task loading, occured within a higher frequency range (the beta band, 16 Hz). In contrast the gain curve for subject 15 showed an increase, with task loading, above and below 10 Hz with a noticable decrease at 10 Hz, corresponding to time averaged amplitude increases.

Results of task loading effects for the same four subjects, but with SOS stimulation, are shown in figure 4. Data plotted here represents the averages from six 40-second replications for each condition for each subject. Standard error about the mean is represented by the vertical lines at each data point on the plots. Referring to fig 4, properties of these curves to observe are the uniqueness or 'signature' of the pair of describing functions for each subject. For initial analysis applied to task-loading vs. no task-load conditions we have found the effects of task-load to be related to this signature. Subject 05 exhibits a large resonant peak at 10 Hz (alpha band), which decreases during task loading. There is a commensurate decrease in steepness of the phase curve about 10 Hz, indicating a reduction in resonance. This resonance reduction can be considered, in systems engineering terms, an increase in the damping coefficient of the dynamic system. Subject 02 also

exhibits alpha band resonance properties. Unlike subject 05, however, there is an increase in gain at the higher frequency region (in the 14 Hz, beta band) with task-loading. Subject 03 shows beta band resonance and with task-loading exhibits a gain reduction in the resonance region, similar to subject 05, but in the beta band. Only minor effects from task loading were exhibited by subject 15, in terms of a slight increase in higher frequency sensitivity. Thus it seems that subjects that are alpha responders (subjects that show an alpha-band resonance, e.g. sub 05) show an alpha decrement. Nonalpha responders (those that lack the alpha resonance peak e.g., sub 15) tend not to show this alpha-band decrement with task loading.

Figure 5 shows, combined on each of the four plots, task and no task results for both transient stimulation and steady state stimulation. The thicker solid lines and the thicker dashed lines are the describing functions resulting from transient time averages (repeated from Fig 3). The circles and triangles represent the describing function values at each of the ten component frequencies from the SOS stimulation (from figure 4).

The correspondance between steady state and transient describing function curves is noteworthy. Describing functions for subject 05 show corresponding regions of peak gain sensitivity for transient and steady state stimulation and show similar gain reduction with task loading. Subject 03 shows similar changes across stimuli in the beta range of the gain curve. Thus for both subjects the effects due to task loading, as indicated by describing function changes, are much the same across stimulus conditions. Furthermore the phase curves have a similar shape, across stimuli and across task conditions for all subjects. The overall correspondance between describing function data for transient and steady state stimulation is remarkable for all four subjects.

One condition that seems to be significant, but the effect of which is not yet accounted for, is arousal level. This is suggested by the responses of subject 02. For the transient stimulus, time responses show a marked change between no load and task loading (fig 2) with a correspondingly significant change in the frequency domain in the alpha region (fig. 3). From this data we would conclude that with transient stimulation subject 02 is a strong alpha producer. Past results indicate that this subject is a strong alpha producer with steady state stimulation as well (see data for subject "RP" in Junker and Peio, 1984). Referring to fig. 5 for subject 02, however, this is not indicated by the steady state gain curves. In fact little change occurred in the alpha band with task loading. There was however an increase in steady state ERP gain sensitivity within the beta band. This subject's steady state ERP data with no task loading does not show the usual alpha band resonance (high gain) and has measures with large variability (indicated by standard error) which may be an indication of lowered level of arousal; i.e. high

variability may be an indicator of lowered arousal. Further, given a condition of lowered arousal, it could be argued that task loading was sufficiently engaging to increase the subjects attention level to the task, as indicated by gain increase in the beta region. These hypotheses suggest that general arousal level and/or attention to a specific task may be observed seperately in ERP describing functions.

From working with both transient and steady state ERP's we have found it quite useful to use both stimulation techniques. As hypothesized by Regan, results do in fact compliment one another. Describing functions obtained by transient stimulation span a wide frequency range (0-50 Hz in this study). Thus they provide overall spectral response for modeling and provide clues to phase unwrapping beginning at 0 Hz. In contrast steady state stimulation provides the ability to concentrate stimulus at selected frequencies. As a result steady state stimulation yields ERP measures and background EEG simultaneously.

The most important point of our results, at this time, is the fact that the forms of the describing functions are remarkably similar across stimuli. From this we conclude that we are justified in continuing with the application of a systems engineering perspective in describing neurosensory functioning. In fact, due to observed subject differences, a systems engineering model structure may be the only way to capture the individual differences in a useful and quantitative manner.

We believe the next step in applying our systems engineering methodology will be "closing the loop". By allowing the human operator VERP feedback, issues of attention and arousal could be controlled. This system, without feedback, has no 'reason' to respond. Through the use of feedback displays the full power of systems engineering analysis could be applied to these human response mechanisms.
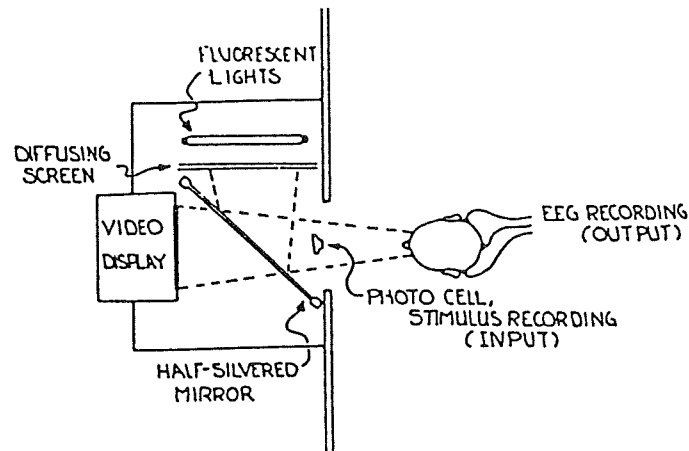
Figure 1. Experimental Setup
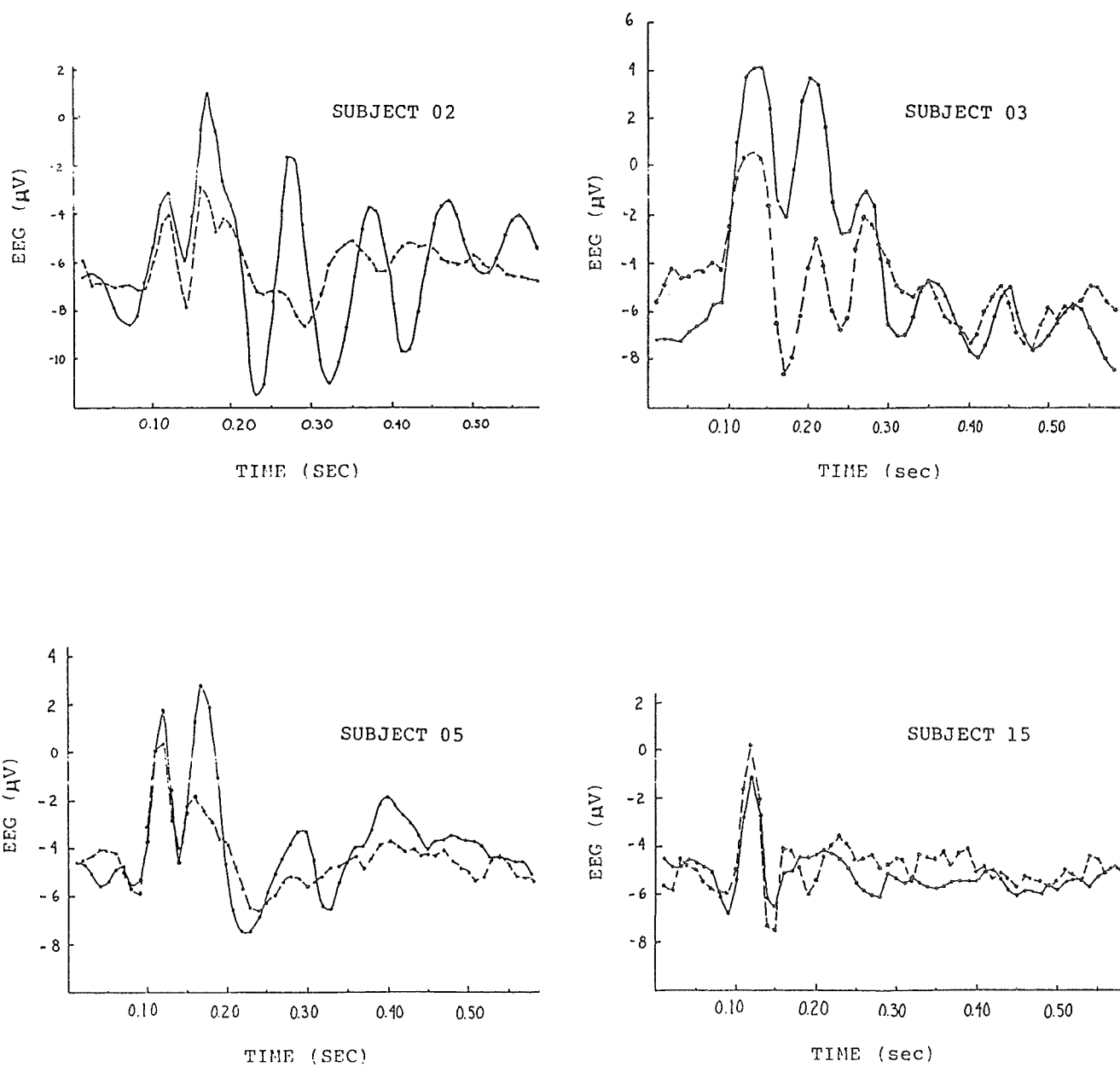
FIGURE 2. AVERAGED TIME RESPONSE TO TRANSIENT STIMULUS.
Solid lines: lights only, dashed line: task.

2.9
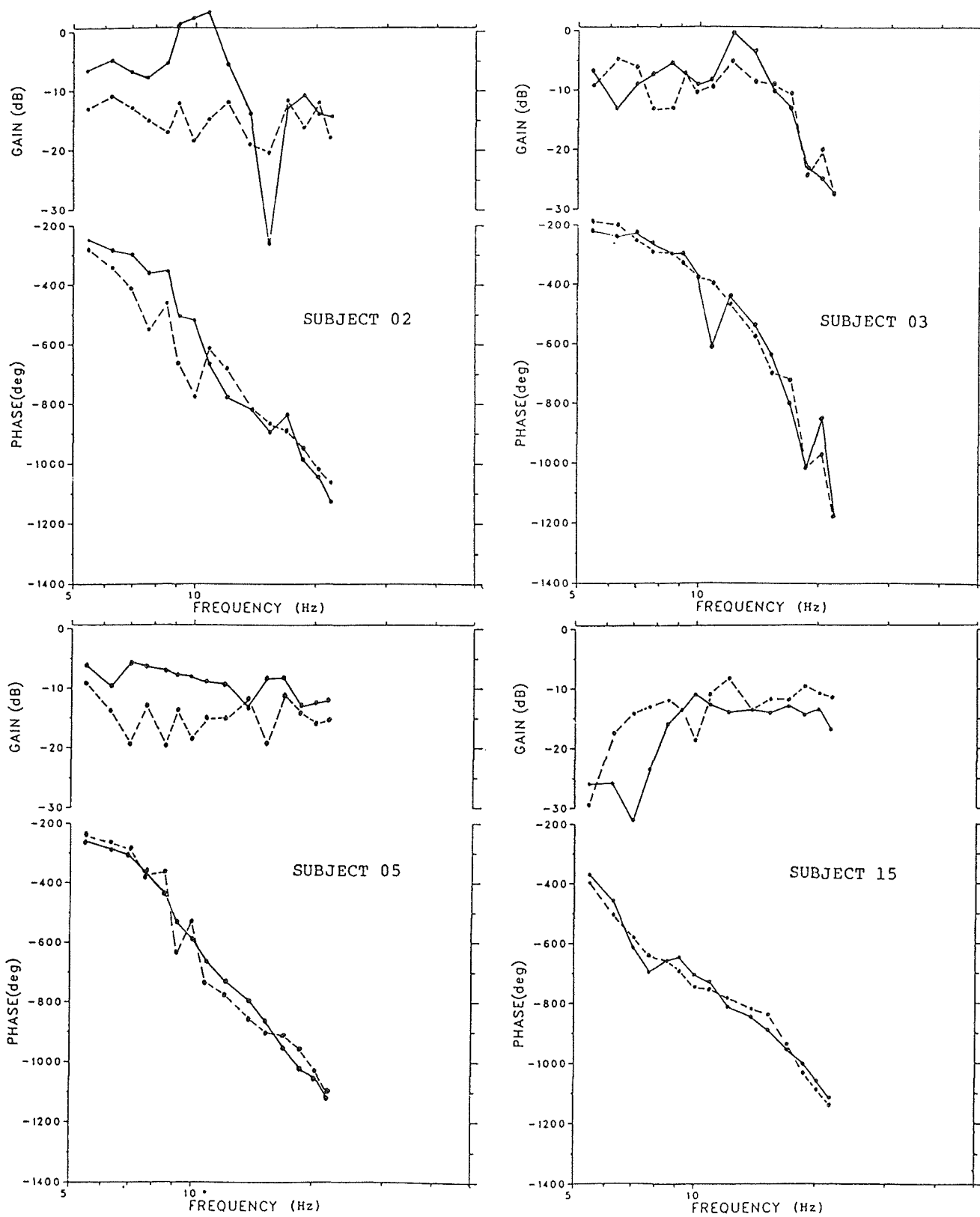
FIGURE 3. DESCRIBING FUNCTIONS FOR TRANSIENT STIMULUS.
Solid lines: lights only, dashed line: task.

2.10
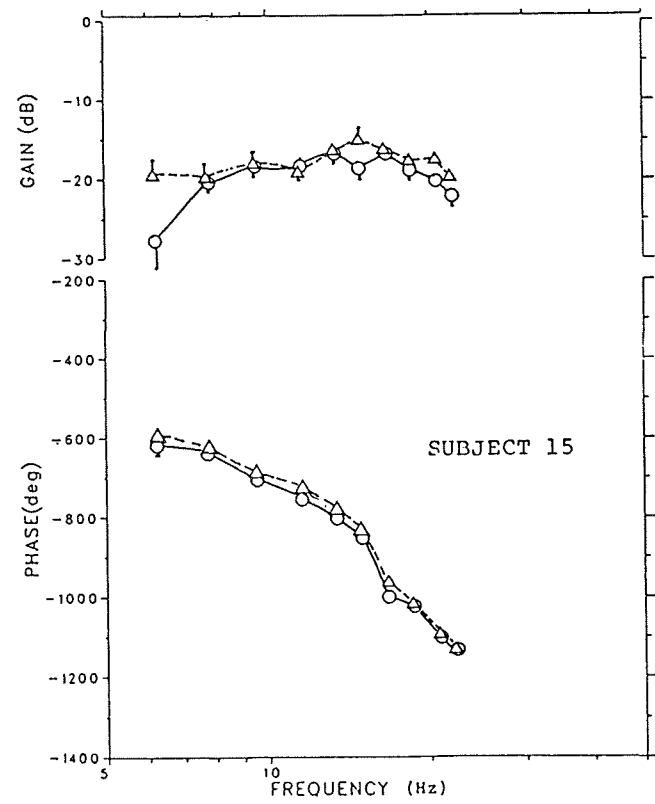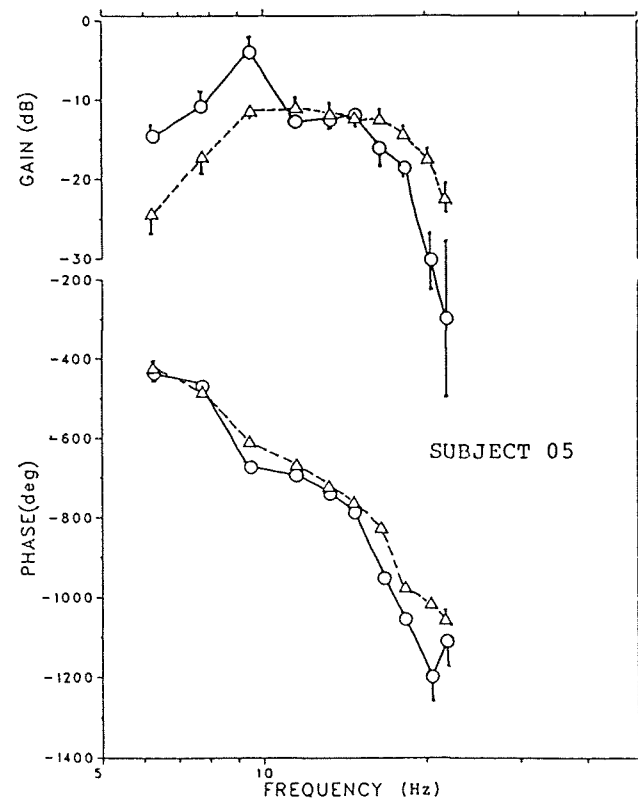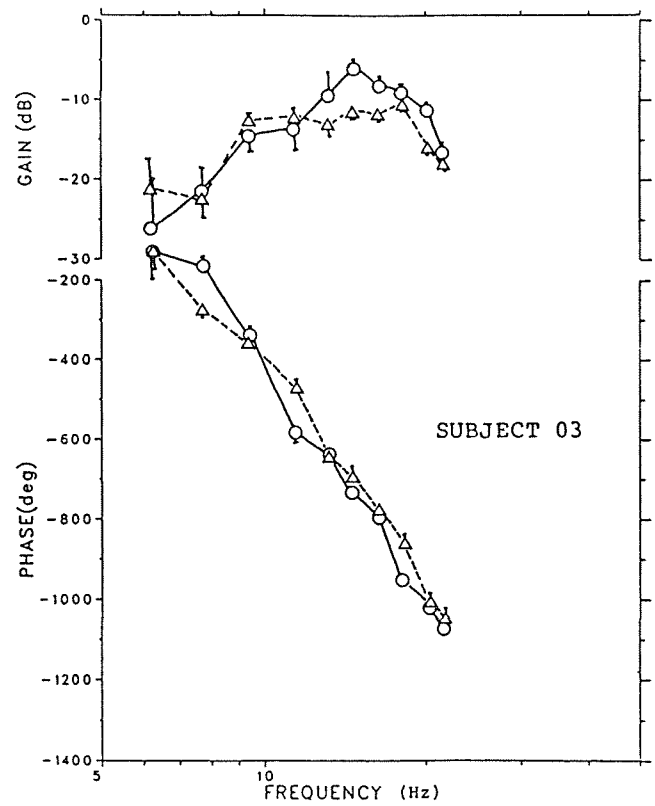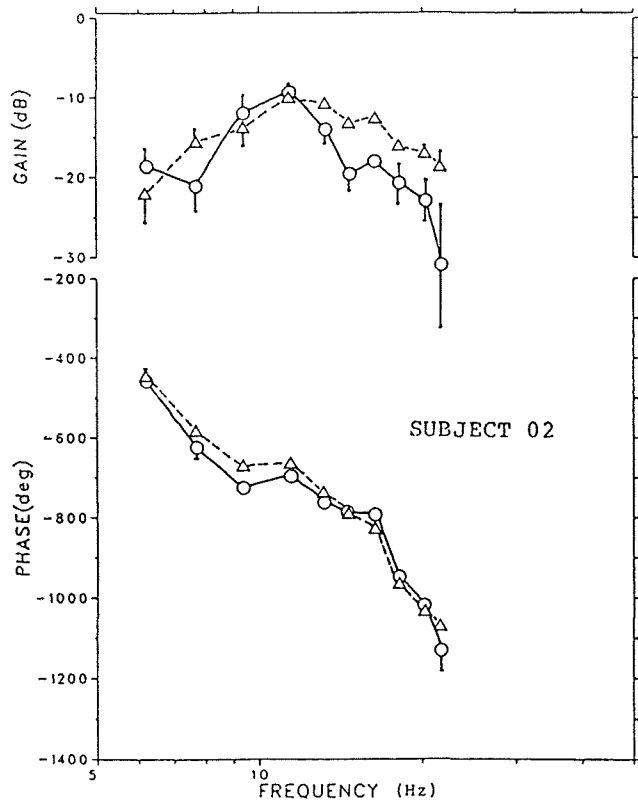
Figure 4. VER Describing Functions. Solid Lines: Lts Only
Dashed Lines: Grammatical Reasoning (Task-Loading)

2.11
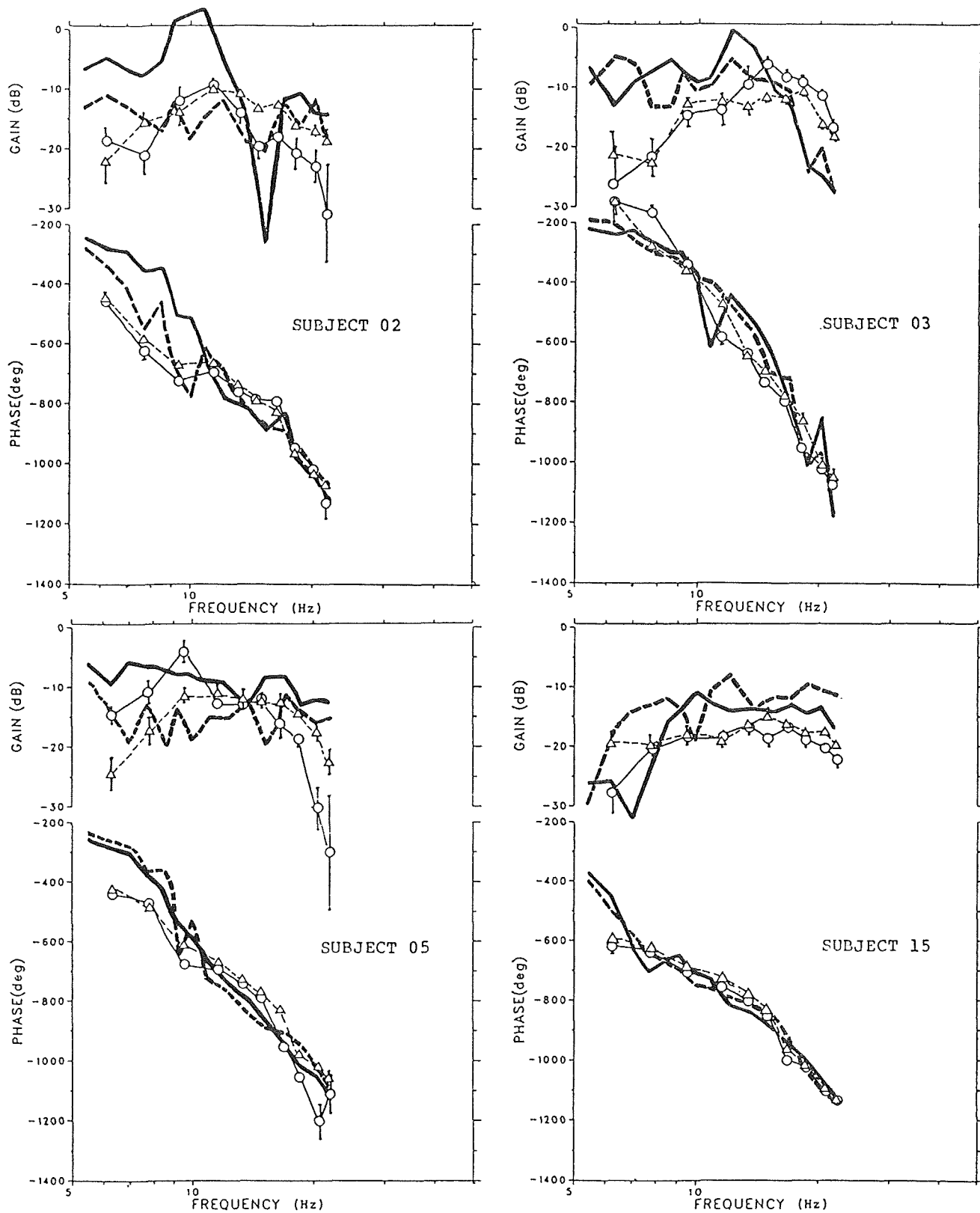
FIGURE 5. DESCRIBING FUNCTIONS FOR TRANSIENT AND STEADY STATE STIMULATION.
Transient:thick solid-no task, thick dashed-task
Steady State: circles-no task, triangles-task

2.12

REFERENCES CITED

1. Baddeley, A.D. 1968. A Three Minute Reasoning Task Based on Grammatical Transformation. Psychonomic Science, vol 10, no 10, 341-342.

2. John, E.R., Walker, P, Cawood, D., Rush, M., Gehrman, J. 1973. Factor Analysis of Evoked Potentials. Electroenceph. and Clin. Neurophys. 34:33-43.

3. Junker, Andrew M. and Peio, Karen J. 1984. In Search of a Visual-Cortical Describing Function - A Summary of Work in Progress. Proc. of 20th Ann. Conf. on Manual Control. NASA Conf. Publication 2341, vol II: 37-54

4. Kochenburger, R.J. 1950. A Frequency Response Method for Analyzing and Syntheszing Contractor Servomechanisms. Trans. AIEE. 69 part 1: 270-284.

5. Levison, William H. 1983. Some Guidlines for Frequency Response Analysis. Technical Memorandum CSD 83-4, Bolt Beranek and Newman,Inc. Cambridge Ma. 22 pp.

6. Levison, William H. 1984. Preliminary Analysis of Steady-State Visual Evoked Response. Technical Memorandum EDP-85-1, Bolt Beranek and Newman, Inc. Cambridge Ma. 14pp.

7. Regan, David. and Beverly K.I. 1973. Relationship Between the Magnitude of Flicker Sensation and Evoked Potential Amplitude in Man. Perception. 2:61-65.

8. Regan, David. 1975. Recent Advances in Electrical Recording from the Human Brain. Nature. 253:401-407

9. Regan, David. 1979. Comparison of Transient and Steady-State Methods. Annals of the NY Acad. of Sciences. 8:45-70

10. Reid, Gary B., Shingledecker Clark A., and Eggemeir F.T. 1981. Application of Conjoint Measurement to Workload Scale Development. Proc. of 1981 Hum. Factors Soc. Ann. Meeting. Oct 1981(a), pp. 522-526.

11. Shingledecker, Clark A., Acton, William H. and Crabtree, Mark S. 1983. Development and Application of a Critereon Task Set for Workload Metric Evaluation. Proc. of 2nd Ann. Aerospace Behavioral Engineering Technical Conf. Aerospace Congress and Exposition Press. Long Beach, Ca.

12. Spekreijse, Henk. 1966. Analysis of EEG Response in Man Evoked by Sine Wave Modulated Light. Dr. W. Junk Publishers. The Hague, Netherlands. 152 pp.

13. Wilson, Glen F. 1979. Steady-State Evoked Response as a Measure of Tracking Difficulty. Final Report AFOSR contract no. F49620-79-C.

14. Wilson, Glen F. and O'Donnell, Robert D. 1981. Human Sensitivitiy to High Frequency Sine Wave and Pulsed Light Stimulation as Measured by the Steady State Cortical Evoked Response. Technical Report AFAMRL-TR-80-133.

Modified Petri Net Model Sensitivity to Workload Manipulations[1]

Stephen A. White, David P. MacKinnon, & John Lyman[2]

University of California, Los Angeles

Introduction

The purpose of this research is to investigate modified Petri nets (MPNs) as a workload modeling tool. This paper describes the results of an exploratory study of the sensitivity of MPNs to workload manipulations in a dual task.

Petri nets have been used to represent systems with asynchronous, concurrent and parallel activities (Peterson, 1981). These characteristics led some researchers to suggest the use of Petri nets in workload modeling where concurrent and parallel activities are common. Petri nets are represented by places and transitions. In the workload application, places represent operator activities and transitions represent events. MPNs have been used to formally represent task events and activities of a human operator in a man-machine system. For example, Madni, Chu, Purcell and Brenner (1983) used MPNs to model the tasks underlying the identification and reaction to a lube oil leak in a ship propulsion system. Madni and Lyman (1983) used a MPN to model the checkout and start-up procedure for a Cessna 182 light aircraft. White, MacKinnon and Lyman (1984) formulated a MPN for POPCORN, a complex computer simulation at NASA-Ames for workload research. These descriptive applications demonstrate the usefulness of MPNs in the formal representation of systems. It is our general hypothesis that in addition to descriptive applications MPNs may be useful for workload

estimation and prediction.

This paper reports the results of the first of a series of experiments designed to develop and test a MPN system of workload estimation and prediction. This first experiment is a screening test of MPN model general sensitivity to changes in workload. Positive results from this experiment will justify the more complicated analyses and techniques necessary for developing a workload prediction system.

Our analytical work with MPNs has exposed three critical issues that are relevant for workload applications of MPNs, viz., task complexity, level of task representation detail, and activity and event classifications.

MPNs differ according to task complexity such as a relatively linear task such as identifying an oil leak and taking appropriate action in comparison to a circular task with many goals and repetitions such as the POPCORN simulation. In POPCORN, An large number of trade offs between offensive and defensive strategies are possible throughout the course of one experimental trial. The critical areas for workload estimation often involve circular type tasks such as the activities of a aircraft or automobile operator. The experimental task to be described below is complex and circular in nature, but is programmed so that the necessary information for MPN development can be obtained.

Level of task representation refers to the level of detail of the task that is being modeled. For example, using Madni and Lyman's (1983) task, the start-up and checkout procedures for the Cessna plane can be simply modeled as follows: enter the plane, start the engine and then take off, (a two activity and three transition MPN). Alternatively each muscle movement of the pilot in each activity leading to take off could be modeled (a very large MPN). The level of task representation issue is important because on the one hand a detailed map of the task is required to obtain adequate sensitivity to workload changes. On the other hand, there are limitations on the measurement techniques that can accurately partition task components for a MPN model with the necessary level of detail. For example, with the current level of technology it is not possible to know precisely when shifts of attention occur between task elements. The MPN derived from

3.2

the task we have devised represents a compromise on the level of detail issue. It is intended to be a task designed to elicit most of the information needed for analysis in MPN terms. Additional information can be obtained via control experiments designed to measure specific mental processes that cannot be measured with the experimental task alone.

Activity and event classification schemes refer to the classification of events and activities in terms of general workload categories. If activities and events can be categorized in terms of workload then it is not necessary to estimate the workload contribution of each individual event and activity. The classification analysis is an area of advanced work and will be conducted in the next stage of this research.

Method

Subjects: Eleven UCLA undergraduate volunteers served as subjects. Each subject participated in a two hour experimental session.

Materials: The entire experimental procedure was conducted on a Televideo 803 computer with a mouse controller.

Procedure: A dual task similar to Derrick and McCloy's (1984) composed of a tracking task and a vowel insertion task was devised. The tracking task was a standard compensatory tracking task with a cursor moving along the horizontal axis driven by a random forcing function. The subject was instructed to try to keep the cursor near the center of the line with a mouse controller. Easy and hard levels of difficulty were introduced by changing the forcing function parameters. A vowel insertion task was incorporated into the tracking task by using the cursor itself as the stimulus letter. It was presented as a consonant that was replaced randomly at intervals of three to seven seconds. The subjects were instructed to mentally insert the letter "A" between the consonant that was currently displayed and the previous one. This consonant-vowel-consonant combination might or might not form an English word. The task of the subject was to indicate whether it was a word or not by means of switches located as part of the mouse controller. The cognitive load of the vowel insertion task was manipulated by increasing the number of vowels that the subject must sequentially insert. For example, with

3.3

three vowels to insert ("A", "E", & "I") the subject was required to make three lexical

decisions and therefore three key press responses. On a comparative basis, one vowel

represented a low cognitive load and three vowels represented a high cognitive load.

The one versus three vowels, and hard versus easy tracking tasks were crossed to form

four conditions. Thus each subject conducted two four-minute trials in all four conditions, viz.,

high cognitive load---low tracking load, high cognitive load---high tracking load, low cognitive

load-low tracking load and low cognitive load---high tracking load. The order of the conditions

was counterbalanced and each subject was given two minutes of practice in each condition.

Subjects performed the task individually and with the CRT screen at eye level. They were

instructed to keep the cursor at the center of the horizontal bar using the mouse controller.

They were told to press one switch on the mouse when the consonant-vowel-consonant (CVC)

was an English word and press another switch on the mouse if the CVC was not a word. After each

condition the subject rated her/his level of workload on ten scales of workload level and task

difficulty that are in use for the POPCORN task at NASA-Ames, with the exception that the

skill-, rule-, and knowledge-based scale was replaced by an scale on automaticity.

Two control conditions were conducted to obtain measurements for certain parameters of

the MPN. The control conditions were used to estimate the length of time necessary for certain

mental process which cannot be derived by the data available from the experimental conditions.

The first experiment obtained a simple reaction time to the change of consonants that are used

in the vowel insertion task. This task generated an estimate of the initial start-up and response

activities involved in the vowel insertion task. The second control task combined the tracking

task and a two choice reaction time. The cursor was displayed as the letter "T" and was replaced

by an "X" or and "O" every three to five seconds. Each "X" or "O" was displayed for one half

second. The subject depressed one key if it was and "O" and another key if it was an "X". This

procedure provided an estimate of letter identification time and the decision processes involved

in selecting the appropriate key to press. The subjects performed each control condition with

both levels of tracking difficulty. The control conditions were randomly mixed with he

experimental conditions.

Each subject generated an individual difference bias rating for the bipolar rating scales. The procedure was the same one as used at NASA-Ames in which subjects rated which of two scales is more important. Each possible comparison of the ten scales was rated. Because the subjective rating scales differ in importance and meaning for each subject, the individual bias information was considered important for accurate workload estimation. This information can be used to weight the ratings. However, only the unweighted rating scores were used in the analyses reported below.

Modified Petri Net of the Dual Task: The MPN for the experimental task is displayed in Figure 1. Figure 1a displays the net for the entire task. Figures 1b, 1c and 1d display the subnets for tracking and vowel insertion. Table 1 presents the activities and events for each experimental task.

Results and Discussion

The preliminary analyses were conducted to verify that the experimental manipulations were effective in changing workload. A 2 (high versus low cognitive load) by 2 (hard versus easy tracking) analysis of variance was conduced on each of the ten ratings, the residual mean square error (RMS) measure and the percent correct on the lexical decision measure.

The anova on the RMS error of the tracking showed a significant main effect for the tracking condition ($F=95.18$, $p<.001$) and vowel insertion condition ($F=8.16$, $p<.05$), with the hard levels of difficulty having the greater RMS error. The anova on the Percent Correct of the vowel insertion showed a significant main effect for the vowel insertion condition ($F=27.40$, $p<.001$). However, the hard level of vowel insertion demonstrated the better performance. This can be explained by the fact that the second and third letters for the vowel insertion ("E" and "I") created much fewer english words compared to inserting the letter "A". Thus, the subjects may have been biased into responding NO for most of the second and third vowel insertions and this strategy paid off. A more appropriate comparison , then, would be to compare the percent

3.5

correct of the first vowel insertion of the hard level (the letter "A") and the percent correct of the easy level (the letter "A"). This analysis showed no main effects.

Table 2 shows the F values of the anovas conducted on the unweighted workload rating scales. Two of the ten scales showed a main effect for the tracking condition, while 8 of the ten showed main effects for the vowel insertion condition. These results indicate that the experimental conditions did indeed manipulate workload.

Anovas were also conducted the output of the MPN simulations of the experimental trials. The data derived was the number of times each transition fired, the total amount of time each place was activated, and the number of times each place was activated.

Because the activities represented by places 6, 7, 8, & 9 were not directly observable, the estimation of these activity times involved the inclusion of data obtained in the control conditions. These derivations were more complicated and were unavailable for the analyses reported below.

Table 3 shows the F values for the anovas on the transitions and Table 4 shows the F values for the anovas on the places of the MPN simulation. Transitions 1 and 4 were not tested since they did not vary across conditions. The main point of these two tables is that the transitions and places that modeled the tracking components showed main effects for the tracking condition, and the ones that modeled the vowel insertion showed main effects for the vowel insertion condition. This indicates that the MPN model appropriately represented the experimental task.

However, a more important question is whether the MPN represented the workload involved in the task. It is possible that other components of the task, which were not possible to model the the MPN, were more important contributers to the workload involved in the task. Thus, it was necessary to demonstrate a relationship between the MPN parameters and the subjective workload ratings. To do this, a canonical correlation was conducted between the MPN parameters and the workload ratings. The results of the canonical correlation showed that the first four eigenvalues were significant. This indicates that four underlying factors of the MPN parameters are highly related to four underlying factors of the workload ratings.

3.6

Summary and Future Directions

The results of the canonical correlation indicated that MPN model of the experimental task represented the task components that influenced subjective workload. Thus, the goal of this experiment was achieved by this demonstration that the MPN model was sensitive to workload changes.

The next stage of this research will involve generating a classification scheme that will group events and activities that are similar in their contribution to task workload. Workload values for each class of events and activities can then be derived. This will allow testing of MPN model simulations for their prediction capability of the workload of a task.

FOOTNOTES

[2] The order of the first two authors is arbitrary.

Table 1

Transitions and places of MPN Model of Experimental Task

Transitions	Places
T_1-Start	P_1-Vowel insertion activity
T_2-Cursor leaves task boundaries	P_2-Tracking activity
T_3-Restart	P_3-Wait (2 seconds)
T_4-Consonant Changes	P_4-Monitor (cursor)
T_5-Lexical decision process complete	P_5-Lexical decision activity
T_6-Letter identification complete	P_6-Letter identification
T_7-Lexicial decision complete	P_7-Lexcial decision
T_8-Response decision complete	P_8-Response decision
T_9-Response complete (continue with next vowel)	P_9-Response
T_{10}-Cursor moves	P_{10}-Monitor (tracking) and calculate move
T_{11}-Cursor passes mark	P_{11}-Move to compensate
T_{12}-Monitor & calculate complete	
T_{13}-Cursor passes mark	
T_{14}-Move completed	

TABLE 2

WORKLOAD RATINGS

F VALUES

	TRACKING	VOWEL INSERTION	INTERACTION
OVERALL WORKLOAD	3.19	15.32**	0.77
TASK DIFF.	3.51	36.22***	0.39
STRESS	9.31*	17.35**	0.01
FRUSTRATION	4.97*	30.93***	0.04
PHYSICAL EFFORT	2.29	0.01	0.49
MENT/SEN EFFORT	3.22	23.14***	0.27
FATIGUE	0.50	0.04	4.41
AUTOMATICITY	4.28	13.27**	0.69
TIME PRESS	1.06	54.50***	1.11
PEFORMANCE	3.69	11.89**	0.34

*p <.05
**p <.01
***p <.001

Table 3

TRANSITIONS

F VALUES

	TRACKING	VOWEL INSERTION	INTERACTION
T_2	6.15*	1.12	16.98**
T_3	6.15*	1.12	16.98**
T_5	1.91	0.60	0.45
T_6	1.91	0.60	0.45
T_7	2.18	107.73***	4.21
T_8	2.18	107.73***	4.21
T_9	2.18	107.73***	4.21
T_{10}	6.18*	2.62	17.30**
T_{11}	9.22*	0.05	3.08
T_{12}	37.81***	0.31	1.60
T_{13}	71.22***	0.08	16.24**
T_{14}	37.63***	0.26	1.57

* $p < .05$
** $p < .01$
*** $p < .001$

3.11

Table 4

PLACES

F VALUES

	TRACKING	VOWEL INSERTION	INTERACTION
TOTAL TIME			
P_1	5.54*	2.17	17.20**
P_2	5.54*	2.17	17.20**
P_3	5.54*	2.17	17.20**
P_4	0.27	31.99***	8.42*
P_5	2.88	26.80***	14.05**
P_{10}	43.34***	0.34	0.20
P_{11}	26.93***	0.09	0.10
FREQUENCY			
P_1	3.40	1.40	13.80**
P_2	3.40	1.40	13.80**
P_3	6.15*	1.12	16.98***
P_4	2.41	0.47	0.51
P_5	1.91	0.60	1.52
P_{10}	36.65***	0.29	1.50
P_{11}	31.38***	0.53	1.20

* $p < .05$
** $p < .01$
*** $p < .001$

3.12

FIGURE 1A

VOWEL INSERTION

TRACKING

CURSOR LEAVES
TASK BOUNDARIES

RESTART

START

P_1

P_2

P_3

T_1

T_2

T_3

MPN OF DUAL TASK

3.13
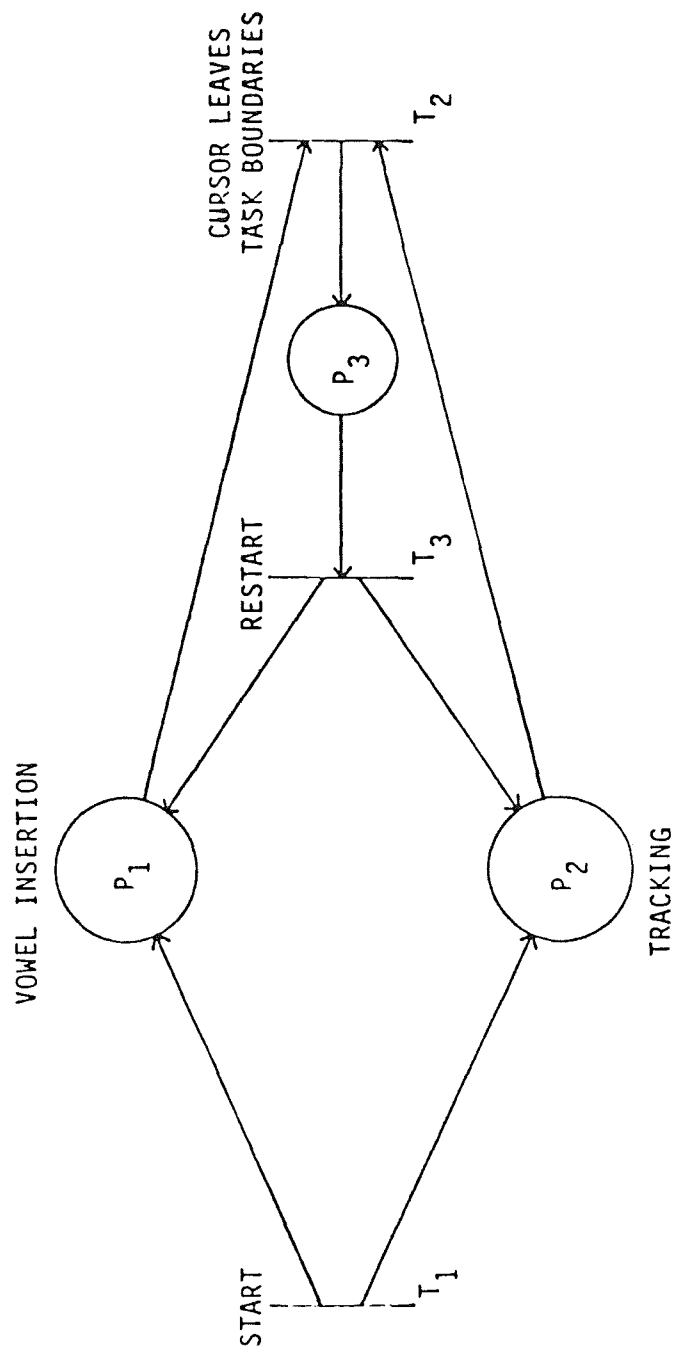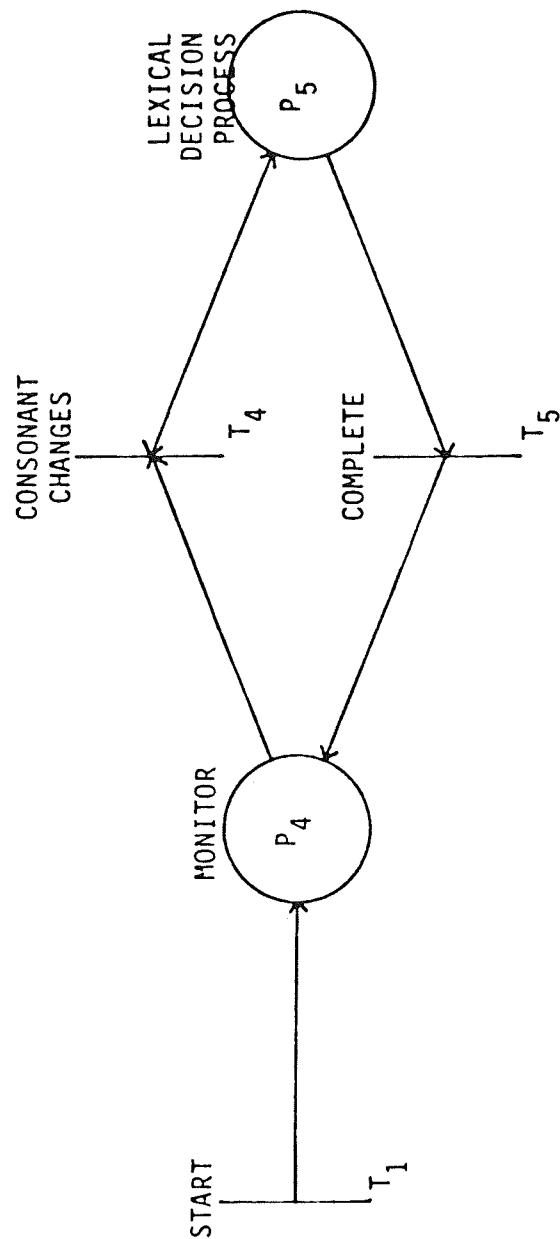
FIGURE 1B

MPN SUBNET OF VOWEL INSERTION ACTIVITY

3.14
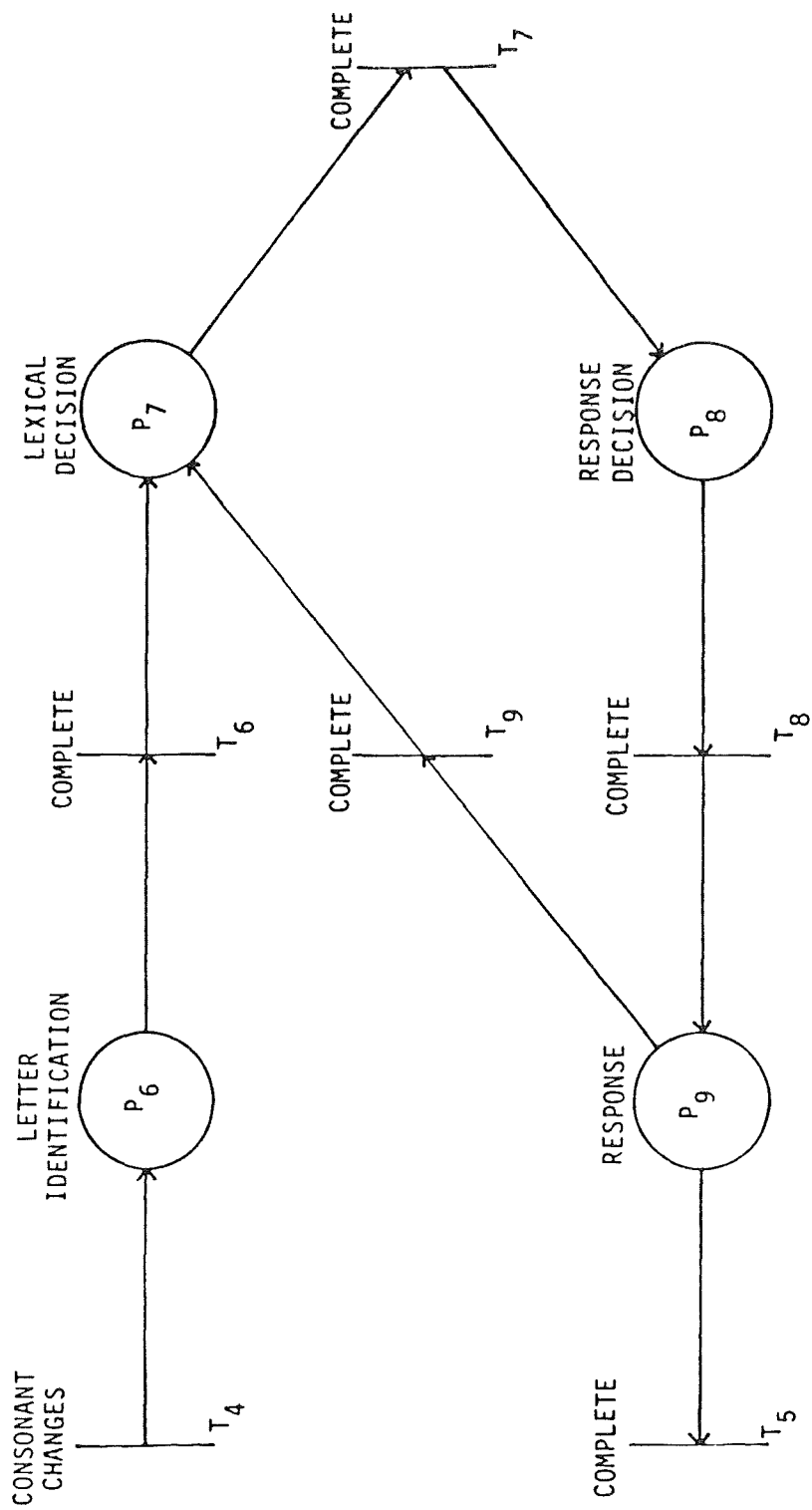
FIGURE 1C

MPN SUBNET OF LEXICAL DECISION ACTIVITY

3.15

FIGURE 1D
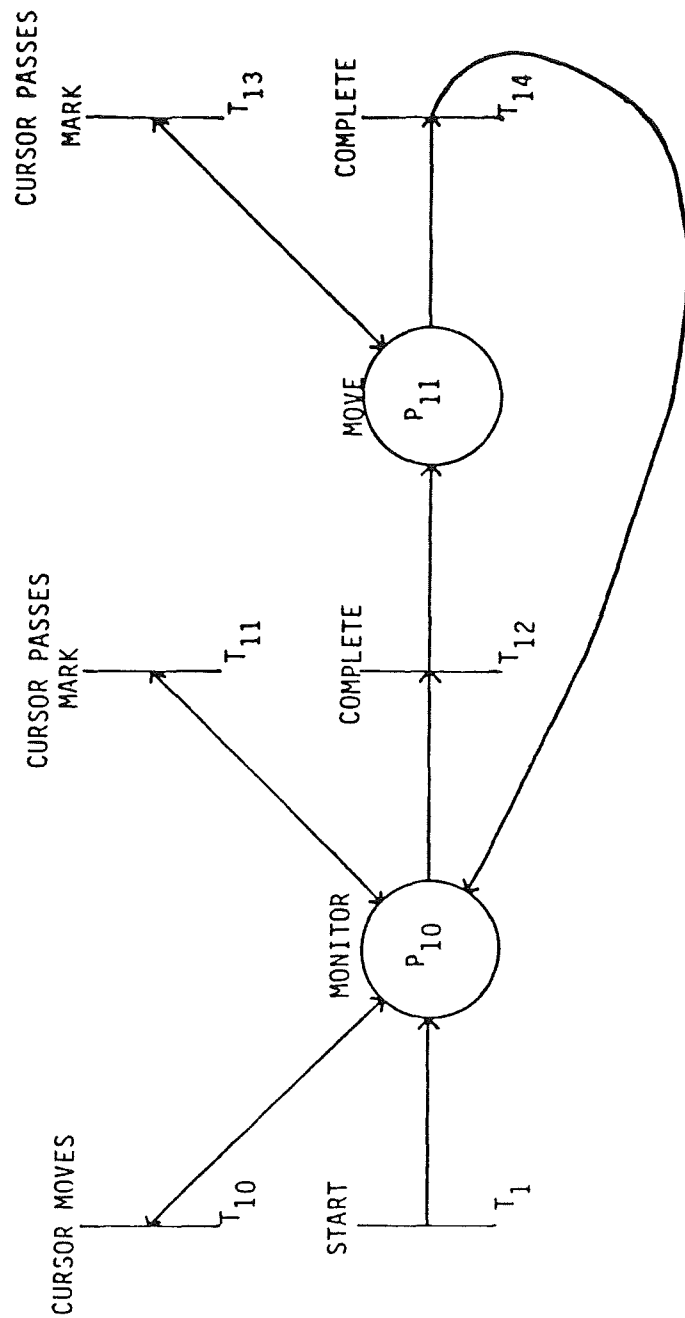
MPN SUBNET OF TRACKING ACTIVITY

3.16

REFERENCES

Derrick, W., & McCloy, T. An empirical demonstration of multiple resources. Proceedings of the Twenty-eighth meeting of the Human Factors Society, October, 1984, 26-30.

Madni, A., Chu. Y., Purcell, D., & Brenner, M. Design for maintainability with modified Petri nets (MPNs): shipboard propulsion system application. Prepared for the office of Naval Research. 1984.

Madni, A., & Lyman, J. Model-based estimation and prediction of task-imposed mental workload. Proceedings of the Twenty-seventh annual meeting of the Human Factors Society, 1983.

Peterson, J. Petri net theory and the modeling of systems. New Jersey: Prentice-Hall, 1981.

White, S., MacKinnon, D., & Lyman, J. Structuring modified Petri net model based assessment of workload components. NASA research report, 1984.

Levels of information processing in a Fitts law task (LIPFitts)

Kathleen L. Mosier
University of California
Berkeley, CA

Sandra G. Hart
NASA-Ames Research Center
Moffett Field, CA

ABSTRACT

State-of-the-art flight technology has restructured the task of human operators, decreasing the need for physical and sensory resources, and increasing the quantity of cognitive effort required, changing it qualitatively. Recent technological advances have the most potential for impacting the contempory pilot in two areas: performance and mental workload. In an environment in which timing is critical, additional cognitive processing can cause performance decrements, and increase a pilot's perception of the mental workload involved. The effects of stimulus processing demands on motor response performance and subjective mental workload are examined in the current study, using different combinations of response selection and target acquisition tasks. The information processing demands of the response selection were varied (e.g., Sternberg memory set tasks, math equations, pattern matching), as was the difficulty of the response execution. Response latency as well as subjective workload ratings varied in accordance with the cognitive complexity of the task. Movement times varied according to the difficulty of the reponse execution task. Implications in terms of real-world flight situations are discussed.

INTRODUCTION

Typical aircraft control tasks require, in some proportion, three types of resources: physical, sensory, and cognitive processing. The job of the contemporary pilot seldom demands strenuous physical effort, other than staying awake and alert on long or fatiguing flights. It requires a small degree of sensory effort, such as reading gauges and listening to warning clackers, etc., and a continually increasing amount of cognitive processing (e.g., calculations, instrument comparisons, decisions) that often must be performed quickly with little margin for error. Flying tasks that were once accomplished by sensory means now demand more sophisticated mental effort, since displays present integrated and refined information rather than raw data. In addition, the quality of cognitive effort required has been redefined. For example, digital readouts are replacing analog gauges, requiring number processing on the part of the operator rather than a quick glance to ascertain that the arrows on several dials are pointing in the same expected direction. Even the task of finding an airport has evolved to a cognitive processing task because of the need to use localizers, instrument approaches, etc. in addition to looking out of the window.

Finally, when timing is critical, extra cognitive processing may increase the time to respond to a signal (Hart, Sellars, & Guthart, 1984),

causing a performance decrement. Even a task as simple as moving left or right in response to a command is more difficult and time-consuming when the information is presented linguistically (e.g., "RIGHT") rather than spatially. For example, Hart et al. (1984) found differences in reaction time (RT) performance based simply on a directional arrow (>) versus a linguistic command (R/L). They also found an additional 40-msec lag in RT when subjects were required to process the size and distance of a target in addition to the directional cue.

Sternberg (1975) and others found RT performance differences depending on the number of items a subject was required to remember and search through (the memory set) before responding as to whether another stimulus (the probe) was or was not a member of the memory set. It is reasonable to expect that these response decrements found in controlled, laboratory experiments that involve relatively minor levels of cognitive processing would, if anything, be exaserbated in a more realistic flight situation, with the potential for life-threatening situations.

Accompanying the demand for a thinking, vigilant, analytical pilot has been a concern over the amount of cognitive load that is placed upon the operator as well as the type of load. Since most of resources currently being tapped are cognitive, it is quite likely that an increase in the complexity of the cognitive demands of a task would have a measurable effect on the pilot's perception of the workload involved. Physical workload is relatively easy to predict and measure, although one is limited by observable behaviors, such as the movement of arms, hands, fingers, and legs, and eyes. Overload results in physical fatigue, injury, or inability to perform a task. Mental workload (i.e., how much a pilot can be expected to process, remember, or analyze in a given time span) is, however, much more elusive. Although mental workload is becoming more and more precisely defined, individual interpretations of the concept itself, as well as its various components, have hindered accurate measurement.

The model for the tasks used in the present study was the "FITTSBERG" paradigm (Hartzell, Gopher, Hart, Dunbar, & Lee, 1983), which combines, serially, a FITTS target acquisition task (Fitts and Peterson, 1964) with a SternBERG memory task (Sternberg, 1975). The decision of which two targets to acquire is based on the results of a Sternberg-type memory search. A series of experiments has been conducted employing variations of this paradigm to investigate the relationship between stimulus processing demands and motor response performance (e.g., Hart et al., 1984). In the original study (Hartzell et al., 1983), subjects were given a choice of two targets, one to the right and one to the left of center. The difficulties of the target acquisitions were indexed (ID) according to Fitts' law (Fitts and Peterson, 1964). The direction of the movement was based on whether or not the probe stimulus was (right) or was not (left) a member of the Sternberg memory set. Memory sets of 1, 2, or 4 letters were used. When compared with performance on a single target task, RT for the combined "Fittsberg" task was sensitive to the additional cognitive processing requirements of the Sternberg memory tasks. As expected, the impact of response selection complexity did not extend into the movement phase (from initiation of response to target capture criterion). Movement times (MT) were not significantly different than for target acquisitions without a response

selection requirement.

In subsequent studies, the workload of the two component tasks (target acquisition and response selection) together was judged to be considerably less than the summed workload of each task done separately. The subtle differences in RT for directional versus linguistic cues continued to be reliable; as was the 40-msec increase in response selection time (RST) with the addition of a target acquisition (TA) task. In a recent study (Staveland, Hart, & Yeh, 1985), it was found that different measures of performance (e.g., RT, RST, MT) selectively reflected different portions of the Fittsberg task, and could be manipulated independently. The workload ratings reflected the average workload within a block of trials (exhibiting no primacy/recency effects of trial difficulty) and integrated the workloads imposed by both selection and execution components.

The present experiment expanded the Fittsberg paradigm to include many other types of information processing, including pattern and rhyme recognition, time estimation, and mathematical problem solving. It also varied the types of information in the memory sets (eg. categories, numerical values, and words, as well as individual letters) and the memory interval (immediate, delayed). The difficulty of the cognitive task that determined movement direction ranged from simple, single-step decisions (e.g., whether or not two simultaneously appearing letters were identical) to relatively complex decisions that required several steps (e.g., solving a complex arithmetical equation and comparing the result to the numerical value of the memory set function).

Current research has focused on the subjective experience of mental workload, either by itself or in combination with performance and physiological measures (Wierwille and Casali, 1983) as the most valuable estimate of load. Multi-dimensional approaches to subjective workload measurement take into account the idea that the experience of workload is a cumulative effect of three (e.g., stress, mental effort, and time pressure) or more factors (Reid, Shingledecker, Nygren, & Eggemeier, 1981), and that the same elements objectively occurring in the same proportions may lead to different estimations of workload from different performers. To account for individual interpretations of factors associated with workload, a system has been devised to combine ratings for each factor with weights reflecting the subjective importance given to that factor (Hart, Battiste, & Lester, 1984). This weighting system, used in conjunction with nine different elements of workload and an overall workload evaluation, was used in the present study.

The goal of the present study was to relate performance and workload changes associated with 10 different information processing tasks. In terms of performance:
    1) The difficulty of a response selection task is reflected in its latency (RST), decision reversals and percent correct. Initiation of a target acquisition is measured by RT.
    2) The difficulty of a target acquisition is reflected in MT, but not in the initial RT (single alternative) or RST (two alternatives).
    3) If the effect of response selection difficulty extends into the movement phase, MTs will increase.
    4) If information processing for response selection and initiating

4.3

response execution are performed serially in the Fittsberg (FB) condition: RST(FB) = RST + RT.
5) If processing is accomplished in parallel: RST(FB) = RST or RT, whichever is greater (implying that no extra time is required for the processing of the additional task).
6) If response selection and initiation of target acquisition overlap, but each requires some unique processing: RST + RT > RST(FB) > RST or RT.

With respect to the subjective ratings of workload (WL):
1) If subjective workload is affected by task complexity, workload ratings will parallel RST and RT differences.
2) If FB imposes more workload than simple response selection tasks, WL(FB) > WL(RS). In this case, either a) workload ratings for the combined tasks will equal the sum of the component task workload ratings [WL(FB) = WL(RS) + WL(TA)]; or b) because of a certain amount of functional overlap, the workload of the combied tasks will be equal to the load imposed by the response selection task plus some non-overlapping part of TA [WL(FB) = [WL(RS) + WL(TA)] * C, where C < 1.0 and C > 0.5].
3) If no additional workload is imposed by the TA task, then FB workload will be equal to the rating of RS or of TA, whichever is greater [WL(FB) = WL(RS) or WL(TA)].

It was hypothesized that: a) RSTs would mirror task complexity; b) information processing for RS and TA would progress essentially concurrently; c) control reversals and percent correct would be affected by response selection task complexity only; d) MTs would reflect target ID only; e) subjective workload ratings would also coincide with task complexity; and f) the extra demands of the TA condition would result in slightly higher, but not additive, workload ratings.

METHOD

Subjects

Nine subjects, ranging in age from 18 to 40, served as paid participants. All of them had been previously trained on different versions of the Fittsberg task that were not used in this experiment (i.e., Sternberg memory sets of one, two, and four with a Fitts target acquisition).

Apparatus

The experimental chamber contained a chair 85cm from a 23-cm monochrome monitor. On the right or left arm of the chair (depending on the handedness of the individual) was a two-axis joystick used for making RT, RST and TA responses. Workload-related ratings were obtained with a slide pot and enter-button on the non-dominant arm rest. An additional switch was mounted on the non-dominant arm rest for response selection in right-target-only and left-target-only conditions. An Apple II computer was used for target generation and data collection (10-msec resolution).

Experimental tasks

Ten response selection tasks, involving several levels of cognitive effort, were presented alone and in combination with a Fitts TA task. The

4.4

pattern match (PM) task was selected as the basic response selection task for the TA control condition, due to its relatively simple processing and memory demands. For most tasks, an answer that was "yes" or "greater" prompted a movement to the right (and acquisition of the target on the right on TA trials). Tasks required no memory, recent (previous trial) memory, or "long term" memory. Each was performed first as a simple response selection task, then as a FB task in combination with a TA. Table 1 illustrates the experimental tasks.

Reaction time (RT) and RST were defined as 2% deflection of the joystick. Three IDs were used for TA, computed in accordance with Fitts' law. Width varied from 5 to 20 pixels, and target distance from 60 to 128 pixels [ID(2.52) = 40/60; ID(4.19) = 7/64 or 14/128; ID(5.67) = 5/128]. Except for the control conditions, the three target IDs were randomly presented within each block of 24 trials. Movement time (MT) was calculated from stick deflection to a steadiness criterion (keeping the cursor in the target).

Feedback

In all tasks (except time estimation) descriptive feedback about correctness and RST was given after each trial, and, where applicable, MT. The time criteria for each feedback phrase remained constant throughout tasks and conditions. Norms for intervals used in providing feedback were derived from earlier studies. Descriptive adjectives comparing current performance to the norms ranged from "truly dismal" to "fantastic".

Subjective rating scales
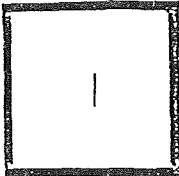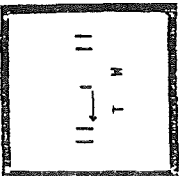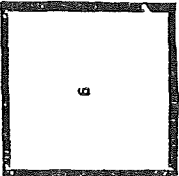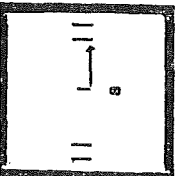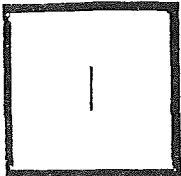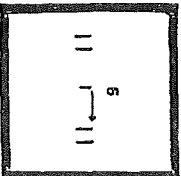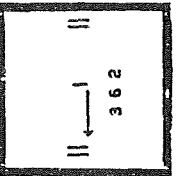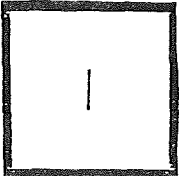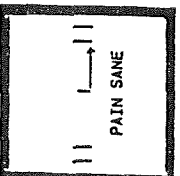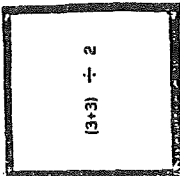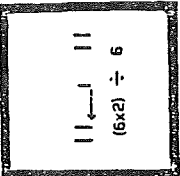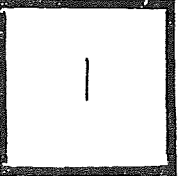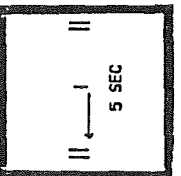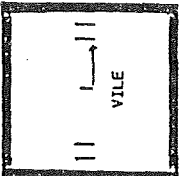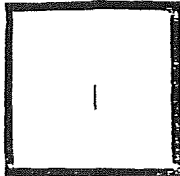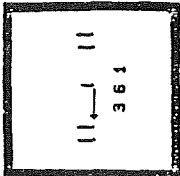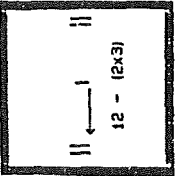
Nine elements of workload were rated: task difficulty, time pressure, own performance, physical effort, mental effort, frustration, stress, fatigue, and activity type (skill- or knowledge-based). Before beginning the experiment, subjects were asked to evaluate the importance of each element to overall workload, compared to every other element, by making 35 pairwise comparisons. The final weight of each factor ranged from 0 (never considered more important than another factor) to 8 (considered more important than any other factor) (Hart et al., 1984). At the end of each experimental block, subjects were asked to rate their experience on each of the nine workload factors, as well as to give an overall workload rating, on 10 bipolar rating scales.

Procedure

After completing the factor weightings, subjects were given an introduction describing the study and the tasks they would be performing, accompanied by demonstration trials. They were given two practice and one experimental block for each task, followed by ratings, in a previously-determined, counterbalanced order. All subjects performed the tasks in the response-selection-only mode first. Prior to performing the TA condition, they were given two practice and one experimental block of trials for each of the control conditions: PM + easy (ID = 2.52) TA (PME); PM + hard (ID = 5.67) TA (PMH); and PM + easy/med/hard TA, right TA only (PMR) or left TA only (PML). A block consisted of 24 trials.

Table 1

Experimental Tasks

EXPERIMENTAL TASK	MEMORY SET	PROBE	MOVEMENT
PATTERN MATCH (PM)	—	T W	MATCH – RIGHT / NO MATCH – LEFT
ODD/EVEN (O/E)	—	9	EVEN – RIGHT / ODD – LEFT
RHYME (RYM)	—	PAIN SANE	RHYME – RIGHT / NO RHYME – LEFT
TIME ESTIMATION (T)	—	5 SEC	WAIT 10 SEC – RIGHT OR WAIT 5 SEC – LEFT
WITTENBORN (W)	—	3 6 1	SUM OF OUTER #S > MIDDLE # – RIGHT / LESS – LEFT

EXPERIMENTAL TASK	MEMORY SET	PROBE	MOVEMENT
RECENT MEMORY CHANGING MS / GREATER/LESS (G/L)	6	8	GREATER THAN PREVIOUS # – RIGHT / LESS – LEFT
NUMBER SET (SET)	5 4 0	3 6 2	ANY #S IN PREVIOUS SET – RIGHT / NONE – LEFT
DELAYED MEMORY / EQUATION (EQ)	(3+3) ÷ 2	(6x2) ÷ 6	GREATER THAN MS – RIGHT / LESS – LEFT
STERNBERG RHYME (SRYM)	TRIAL	VILE	RHYME WITH MS – RIGHT / NO RHYME – LEFT
STERNBERG TIME ESTIMATION (TS)	10 - 2	12 - (2x3)	GREATER THAN MS – WAIT 10 SEC – RIGHT / LESS – WAIT 5 SEC – LEFT

4.6

The data collected for each task were 1) RT, RST, or time duration prior to deflection (for time estimation tasks); 2) MT (where applicable); 3) percent correct; 4) control reversals (e.g., second thoughts about response selection, and 5) bi-polar workload ratings. Several analyses of variance were performed across experimental conditions for each measure: percent correct; RT for TA-only tasks; RST for response-selection-only tasks; and RST, control reversals, and MT for FB tasks. Time estimation tasks were analyzed separately, since RSTs were equal to the duration of 5- or 10-sec time productions. Most of the tasks were also grouped and analyzed by type: 1) control condition (PM, PME, PMH, PMR, PML); 2) math functions (G/L, W, EQ); 3) time estimation (T, TS); and 4) rhyme (RYM, SRYM).

In general, RST was shown to be very sensitive to response selection difficulty, $F(7, 56) = 22.33$, $p<.01$. The addition of the target acquisition task further enhanced this effect (Figure 1). Weighted workload ratings exhibited this sensitivity as well, $F(23, 184) = 8.75$, $p<.01$ (Figure 2). Right/left response differences were not significant, except for tasks in which direction of movement was determined by a yes/no choice. In this case, "no" responses were somewhat slower. Movement time, as expected, was not affected by response selection difficulty or the number of alternative targets. A significant effect was found across all tasks for percent correct, $F(23, 184) = 10.46$, $p<.01$.
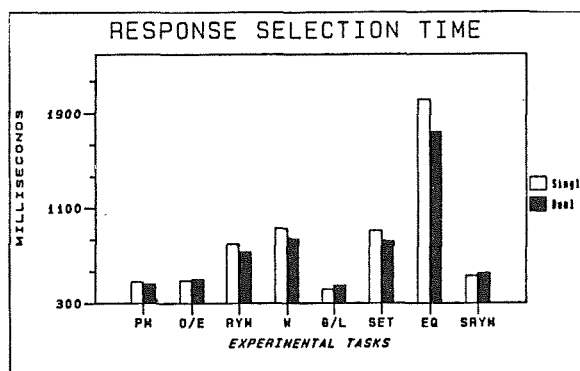
Figure 1. Response selection times for all tasks.
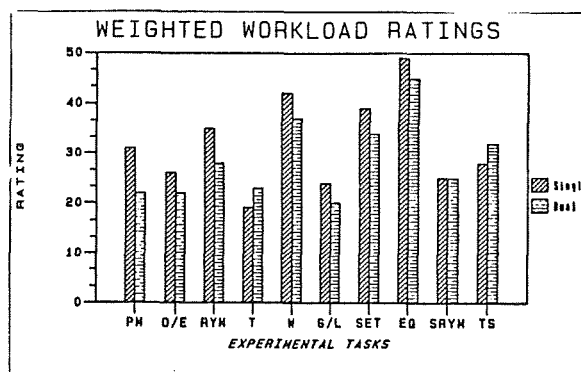
Figure 2. Weighted workload ratings for all tasks.

Control Conditions

Within the pattern match conditions, the effects of several variations of the TA portion of the FB task were examined, i.e., keeping the target ID constant (PME, PMH); keeping the direction of movement constant (PML, PMR); and removing the response selection requirement from target acquisition (PML, PMR). Results of the pattern match condition followed the expected

pattern. No significant differences were found for RT or percent correct, and direction of movement (right versus left) did not have a significant effect.

Movement time differences were found as a function of target ID, as predicted by Fitts' law (Fitts and Peterson, 1964): average MT for easy targets (PME) was .695 msec; for hard targets (PMH), 1.065 msec (Figure 3). A significant interaction was found between PME/PMH and right/left, $F(1,8) = 9.18$, $p < .05$; i.e., the easy/hard MT differences were somewhat more pronounced for right targets than for left targets.
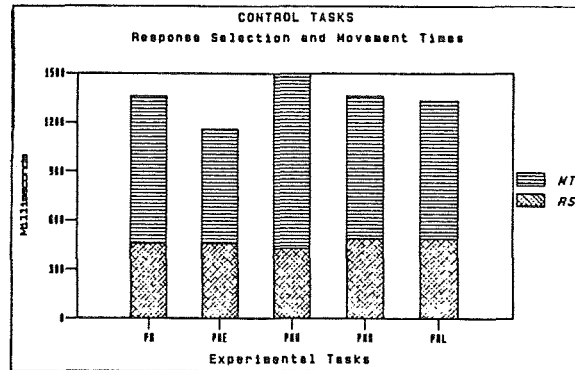
Figure 3. Response selection and movement times for control condition.

In PMR and PML conditions, two RT measures were taken: one for the RS task (a button press); and one for the RT following target appearance (joystick deflection). Responses to the target alone, involving no cognitive processing task, were predictably faster than for any of the cognitive tasks, and were not affected by target difficulty. When one element, either target side (R/L) or ID (E/H), was held constant, and the other was varied, the same RT and MT differences were found that have been indicated in earlier studies. Workload ratings were similar for all of the PM tasks, with the exception that PME was rated as having less workload than PMH.

Math Functions

A significant difference in RSTs was found due to the complexity of the different mathematics tasks, following the expected trend: the RSTs were shortest for the G/L task, followed by the Wittenborn task (W), and the EQ task, $F(2, 8) = 24.00$, $p < .01$. There was a significant effect of task on percent correct as well, $F(2, 16) = 16.5$, $p < .01$.

Response selection times for the Math + TA condition were slightly faster than for the math tasks alone. This could be an effect of training, since all of the TA tasks were presented after the response-selection-only tasks. Two other findings were of interest: There was a significant interaction between task and right/left responses, $F(2, 16) = 10.69$, $p < .01$. Right RSTs were faster than left RSTs, $F(2, 8) = 10.73$, $p < .01$, due primarily to the EQ task, in which left movements (less) were twice as slow as right movements (greater). Also, an effect was found for task on MT, $F(2, 16) = 6.31$, $p < .01$; however, since the conditions having the most control reversals also had the longest MTs, the extra time taken by the reversals accounts for this effect.

Workload ratings mirrored task complexity, with EQ $>$ W $>$ G/L, F(2,8) = 21.11, p<.01 (Figure 4). Single (RT only) task workload was not significantly different than dual task workload.

Time Estimation

In the time estimation tasks, no effects were found for percent correct, number of reversals, or MTs. Left (5-sec) and right (10-sec) responses were examined separately. For left responses: time estimates were significantly longer for the TS task than for T, in both the single and dual task modes, F(1, 8) = 11.46, p<.01. Estimates in both TS and T were also longer in the single task condition than in the dual task condition, F(1, 8) = 8.41, p<.05.

Figure 4.   Weighted workload ratings for math tasks.
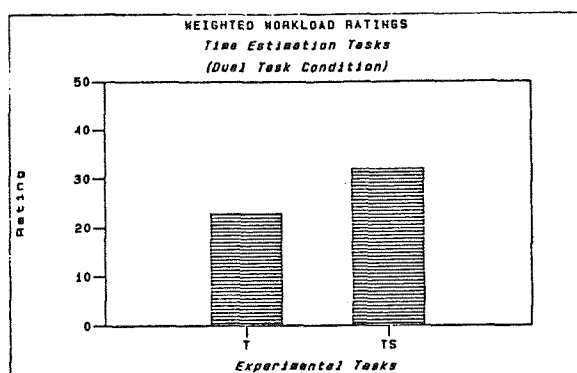
Figure 5.   Weighted workload ratings for time estimation tasks.

Right (10 sec) responses showed somewhat similar results. Estimations were longer in the single-task condition for both TS and T, but there was no difference in estimates between the T and TS tasks. Overall, in the time estimation tasks, the 5-sec estimations were more accurate than 10-sec estimations, which were generally too short.

Workload ratings ranked TS as harder than T, F(1, 8) = 7.2, p<.05, and showed no difference between the single and dual task conditions (Figure 5). A somewhat surprising finding was that many subjects considered the TS task, which involved estimating time as well as solving an equation, to be easier than the EQ task. Reportedly, this was because they did not feel as much time pressure in solving the equation, since the solution to the equation could be completed at any time up to the end of the shortest of the two estimation intervals.

Rhyme Tasks

The delayed rhyme task (SRYM) resulted in significantly faster RSTs, F(1, 8) = 25.35, p<.01 (Figure 6), and a greater percent correct (49% vs. 47%), F(1, 8) = 13.3, p<.01, than the immediate rhyme task. No difference was found between single and dual task conditions in RSTs; however, there was a significant difference in percent correct in favor of the dual task condition, F(1, 8) = 8.4, p<.05, probably due to training. No differences

for MT or reversals were found. "No" (left movement) responses in both tasks were much slower than "yes" (right movement) responses, $F(1, 8) = 19.01$, $p<.01$, a common finding.

There were several interactions: the difference in RST between RYM and SRYM decreased with training, as illustrated by a Task x Condition (RYM vs. SRYM x RS vs. FB) interaction, $F(1, 8) = 12.01$, $p<.01$; the "yes/no" effect was more pronounced in the RYM task than in the SRYM task, as shown by a Right/Left x Task interaction, $F(1, 8) = 13.78$, $p<.01$; and practice reduced this "yes/no" effect, illustrated by a Right/Left x Condition interaction $F(1, 8) = 7.78$, $p<.05$.
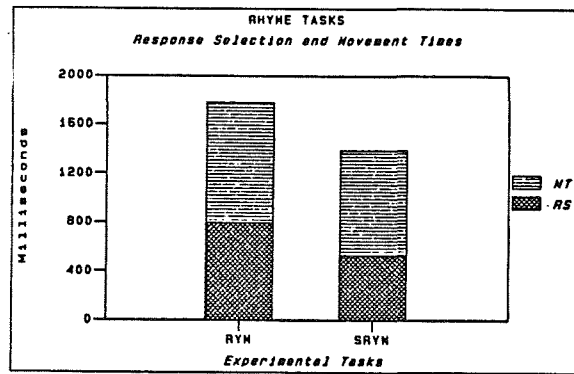
Figure 6. Response selection and movement times for rhyme tasks.
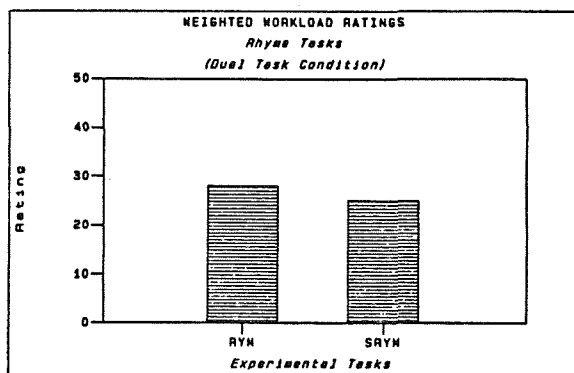
Figure 7. Weighted workload ratings for rhyme tasks.

The RYM task was rated as having greater workload than the SRYM task, F (1, 8) = 6.65, p <.05 (Figure 7) both in RT and RT/Fitts TA conditions. Reasons for this are discussed in the following section.

DISCUSSION

In general, task complexity had the predicted effect on response latency; that is, the more complex the required cognitive processing, the longer it took before a response was selected. The additional processing demanded by the response execution task, however, was not reflected in RST. In fact, for some tasks (e.g. math tasks), RSTs in the dual-task modes were somewhat faster than in the response-selection-only mode. The probable cause for this counterintuitive result could be training; by the time subjects began performing the dual condition, they were familiar with all of the cognitive tasks, and, since they had previously participated in a Fittsberg study, were practiced in target acquisition.

Workload ratings also reflected task complexity, with a few unforeseen results. Several of the response-selection-only tasks were rated as having somewhat higher workload than the same task in the Fitts TA mode. Since all of these tasks were performed first, this again could be the result of

training. The simultaneous rhyme task (RYM) was seen as being more loading than the Sternberg rhyme task (SRYM), even though SRYM involved no memory and used the same type of words. The equation task (EQ) was perceived as being much more difficult than the combination of equation and time estimation (TS), even though the latter involved an additional processing step. The apparent reduction in time pressure mentioned earlier seems to have been an overriding factor here. The equation task took the longest to complete; thus, removing the pressure of having to do it immediately served to greatly reduce its perceived workload (TS was one of the tasks rated lowest in workload, even though its EQ task component was rated as highest).

Performance

Since MTs were not affected by the complexity of the response selection task, it is reasonable to assume that any decision making was completed prior to the movement phase—or, at least, that whatever processing did carry over was sufficiently minimal to be accomplished simultaneously with movement, causing no detriment to MT.

Task complexity did have an observable effect on percent correct; it was largest with the easiest tasks (96%), and smallest with the most difficult tasks (82%). Control reversals did not follow the same pattern, as there were relatively high numbers of reversals for some of the less complex tasks (O/E, RYM). One possible explanation for this is that these tasks were so simple that they were performed "enroute"; that is, subjects may have "jumped the gun" by starting movement in one direction before they had completely processed the stimulus, then finished processing, changing direction if necessary once the stimulus was fully absorbed.

The results of this study indicate that information processing of the response selection task and of the target are done concurrently, since dual task RSTs were, in general, equal to or only slightly greater than the single task RSTs. This is in keeping with previous findings.

Workload

Workload ratings for the response selection tasks paralleled almost exactly response latencies, especially at the extremes: for the immediate response tasks, G/L was considered to be the easiest task (WL = 22) and resulted in the shortest mean RST; EQ was considered to be most loading (WL = 47), and had the longest mean RSTs. The time estimation task (T), in which there was no pressure for a fast response, was also considered to be very low in workload (WL = 23). This indicates that subjects were very sensitive to the relative amounts of required processing and in their perception of the time pressure imposed on them. These were each reflected in their evaluations of the tasks.

Dual task workload was not consistently greater than the same task presented in the single-task mode. This replicated, in general, findings of earlier studies (e.g., Hart, Shively, Vidulich, & Miller, 1985). A tentative explanation for this would, again, be training effects, negating the perception of additional load. Subjects had had enough practice on the basic TA task and the single-task response-selection tasks that the combined task might have imposed no extra load. Another possibility is that most of the perceived workload was in the response selection phase; therefore, the

Fitts TA was experienced as an equivalent task, even though RST's indicated that additional processing was required. Since their existed a functional relationship between the response selection task and the target acquisition task, the latter may have been viewed as merely an extension of the former.

With regard to various specific tasks, the type of memory involved did not appear to have as much impact on workload and RST as did the specific design of the task. That is, the concurrent memory tasks were not, as a group, faster or slower than the recent memory or long-term memory tasks, with some interesting anamolies. For example, examining the RYM and SRYM tasks, it would seem logical that the concurrent processing task, RYM, would have been at least as easy, if not easier, than a long-term memory task; however, the immediate comparison (RYM) resulted in longer RSTs, more errors, and higher workload ratings that SRYM. A factor that may have contributed to this was that, in SRYM, the same word was compared with each other word continuously through the block of trials; in RYM, however, two completely different words were presented on each trial.

The major direction of movement differences were found for tasks in which right or left signified a yes/no response. The lag in RST for a "no" response is of consequence in the real-world cockpit environment in that the discovery that instrument readings (e.g., altitude, heading, fuel supply) are not as they are supposed to be usually signifies trouble--and this may be a situation that calls for the quickest possible action. Also of interest was the fact that the three tasks with the longest RT's--EQ, W, and SET--all involved dealing with numbers. The solution of a simple function in EQ took, on the average, one minute longer than the next slowest task and resulted in more mistakes. This has important operational implications as well.

There were many incorrect responses for the SET and EQ tasks. The SET task is similar to those performed in flight; headings, altitudes, radio frequencies (i.e., sets of numbers), are continually being updated, and the operator is often required to compare current sets of values to previous sets. The design of SET made this activity particularly difficult because subjects could not "chunk" the three numbers; each digit had to be tested against the previous values, and remembered individually.

A key issue in these findings is the difference between the actual Fittsberg RST and WL ratings that were observed in the present study, and what might be predicted on the basis of simply adding the levels of the two component tasks. If RST and WL are cumulative, that is, if each additional task imposes its own requirements on top of those of the previous task, then one would predict that RST(FB) = RST + RT(TA) and WL(FB) = WL(RST) + WL(TA). Table 2 illustrates the RST and WL that would be expected if this were the case. However, the actual figures are much less than this sum; in fact, in some cases, the obtained RST or WL was equal to or only slightly greater than that of either the response selection or response execution task alone.

Table 2

Predicted versus Observed WL and RST

TASK	RS	RE	RST SUM	OBS	RATIO	WL RS	RE	SUM	OBS	RATIO
PM	.479	.45	.929	.456	.49	31	20	51	22	.43
G/L	.423	.45	.873	.454	.52	24	20	44	20	.45
O/E	.485	.45	.935	.495	.53	26	20	46	22	.48
RYM	.803	.45	1.253	.729	.58	35	20	55	28	.51
SRYM	.528	.45	.978	.553	.56	25	20	46	25	.54
SET	.910	.45	1.360	.817	.60	39	20	59	34	.58
T	7.413	.45	7.863	6.612	.84	19	20	39	23	.59
W	.932	.45	1.382	.838	.61	42	20	62	37	.60
EQ	2.016	.45	2.466	1.744	.71	49	20	69	45	.65
TS	7.837	.45	8.287	7.087	.85	28	20	48	32	.67

In this study, the response selection tasks that required only one processing step (e.g., PM, O/E, G/L) were most easily integrated with the TA task, and evidenced the largest discrepancy between the additive prediction of RST and WL and the actual figures. The cognitive processing of these tasks was simple enough to be accomplished in parallel with TA, without additional cost; and dual task WL ratings and RSTs are essentially equivalent to those of the response selection tasks alone. In keeping with this, WL ratings for all of the dual tasks were found to be highly correlated with RST. The processing tasks requiring more than one step (e.g., W required addition + comparison; SET required memory + comparison; EQ required arithmetic problem solving + memory + comparison) were less easily integrated, and the observed WL and RST in these tasks came much closer to the additive predictions. If the tasks were not at all functionally related, the expected ratio of observed to predicted WL would be >1.

Perceived time pressure, rather than experimental manipulation of time pressure, contributed significantly to rated workload, with unforeseen results. For example, the EQ task was rated as having the most workload, and resulted in the largest number of errors and the longest RSTs. However, the TS task (which contained the same equations with the additional task of time estimation), was rated as one of the easiest tasks and resulted in minimal errors — because subjects were able to perform the mental arithmetic calculations at their leisure during the time estimation interval. Since the TS task was a combination of two cognitive tasks, time estimation and arithmetic problem solving, TA actually imposed a third requirement. The predicted WL for TS in the "dual" task condition would be 19(T) + 49(EQ) + 20(TA) = 88. The obtained WL rating for this task, however, was 32 — less than half the prediction. This would seem to indicate that reducing or removing the significant elements contributing to WL, as well as increasing the functional relatedness of tasks, can greatly reduce exp̃rienced workload.

The results of this study have implications for laboratory as well as operational tasks. In functionally related tasks, processing for response

selection and execution appear to be done in tandem. The cognitive complexity of the task profoundly affects the response selection part of the task, but only the physical properties of the target affect the difficulty of its acquisition. Subjects can measurably differentiate the cognitive complexity of tasks - both in terms of performance (actual motor responses) and in terms of perceived workload. Also, the more functional overlap that exists among tasks that are to be performed concurrently or serially, the more the operator can mentally integrate the tasks, and the less the cost in terms of performance and experienced load.

In view or this, human factors engineers must concentrate on keeping cognitive complexity to a level that is manageable and has acceptable consequences in terms of response latencies. Additionally, since the cost of imposing more tasks can vary widely, the nature and relatedness of the simultaneous or serial tasks required of the human operator must be taken into account.

Indications were present on some tasks that training can have the effect of not only improving performance, which is intuitively predictable (as shown in the math tasks, which had the longest RTs, and possibly the most room for improvement); but can also function to reduce perceptions of workload in an equivalent or objectively more difficult task. This was illustrated by the several tasks in which the single task, presented first, was rated as being higher in workload than the same task in the dual condition. One of the possible effects of training is to facilitate integration of the tasks being performed. Therefore, training apparently can, to a certain extent, compensate for increased task loading.

REFERENCES

Fitts, P. M., & Peterson, J. R. Information capacity of discrete motor responses. Journal of Experimental Psychology, 1954, 47, 381-391.

Hart, S. G., Battists, V., & Lester, P. T. POPCORN: A supervisory control simulation for workload and performance research. Proceedings of the 20th Annual Conference on Manual Control, Sunnyvale, CA., 1984.

Hart, S. G., Sellers, J. J., & Guthart, G. The impact of response selection and response execution difficulty on the subjective experience of workload. Proceedings of the 28th Annual Meeting of the Human Factors Society, San Antonio, TX., 1984.

Hart, S. G., Shively, R. J., Vidulich, M. A., & Miller, R. C. The effects of stimulus modality and task integrality: Predicting dual-task performance and workload from single task levels. Paper presented at the 21st Annual Conference on Manual Control, Columbus, OH., 1985.

Hartzell, E. J., Gopher, D., Hart, S. G., Dunbar, S. L., and Lee, E.

Moray, N. Attention: Selective Processes in Vision and Hearing. New York: Academic Press, 1970.

Polich, J., McCarthy, G., Wang, W. S., & Donchin, E. When words collide: Orthographic and phonological interference during work processing. Biological Psychology, 16, 1983, 155-180.

Reid, G. B., Shingledecker, C. A., Nygren, T. E., & Eggemeier, F. T. Development of multidimensional subjective measures of workload. IEEE Proceedings of the International Conference on Cybernetics and Society, 1981, 403-406.

Staveland, L., Hart, S. G., & Yeh, Yei-Yu. Memory and subjective workload assessment. Paper presented at the 21st Annual Conference on Manual Control, Columbus, OH, 1985.

Sternberg, S. Memory scanning: New findings and current controversies. Quarterly Journal of Experimental Psychology, 27, 1-32.

Wierwille, W. W., and Casali, J. G. A validated rating scale for global mental workload measurement applications. Proceedings of the 27th Annual Meeting of the Human Factors Society, Norfolk, VA, 129-133.

THE EFFECTS OF STIMULUS MODALITY AND TASK INTEGRALITY: PREDICTING DUAL-TASK PERFORMANCE AND WORKLOAD FROM SINGLE-TASK LEVELS

Sandra G. Hart   Robert J. Shively   Michael A. Vidulich
NASA-Ames Research Center
Moffett Field, CA

Ronald C. Miller
Informatics General Corporation
Palo Alto, CA

ABSTRACT

The influence of stimulus modality and task difficulty on workload and perfor-mance was investigated in the current study. The goal was to quantify the "cost" (in terms of response time and experienced workload) incurred when essentially serial task components shared common elements (e.g., the response to one initiated the other) which could be accomplished in parallel. The experimental tasks were based on the "Fittsberg" paradigm; the solution to a SternBERG-type memory task deter-mines which of two identical FITTS targets are acquired. Previous research suggested that such functionally integrated "dual" tasks are performed with substan-tially less workload and faster response times than would be predicted by sum-ming single-task components when both are presented in the same stimulus modality (visual). In the current study, the physical integration of task elements was varied (although their functional relationship remained the same) to determine whether dual-task facilitation would persist if task components were presented in different sensory modalities. Again, it was found that the cost of performing the two-stage task was considerably less than the sum of component single-task levels when both were presented visually. Less facilitation was found when task elements were presented in different sensory modalities. These results suggest the impor-tance of distinguishing between concurrent tasks that compete for limited resources from those that beneficially share common resources when selecting the stimulus modalities for information displays.

INTRODUCTION

The current experiment is one in a series that investigated the rules by which single task estimates of workload or performance can be used to predict the results of different task com-binations. Theoretically, some task combinations should be simply additive; the workload of two tasks performed concurrently should be equal to the sum of component task levels. This was found, for example, by Gopher and Braune (1984). In this study, as in many others, however, performance on one or both of the component tasks suffered when they were presented con-currently. Numerous experiments have been conducted with a dual-task paradigm in which a variety of tasks are presented and learned individually and then different combinations are performed concurrently. It is assumed that subjects' resources can be allocated, up to their limit, in graded quantities among separate activities. The fact that some tasks appear to interfere with each other more than others led to the formulation of a multiple resources model that postulated that different amounts and types of resources are required for different tasks and task combinations (Navon & Gopher, 1979). Performance limitations arise from insufficient resources in one or more processes that might be differentiated by the

modality of input, output, or type of central processing (Wickens & Kessel, 1979). In many cases, the difficulty levels of one or both tasks are varied to determine the limits of capacity (Kantowitz & Knight, 1978). In addition, the required performance levels or task emphasis may be specified (Gopher, Brickner and Navon, 1982) to shift the relative priorities among dual-task components. It was found that subjects can dynamically allocate their attention to achieve the required levels of performance (Tsang & Wickens, 1984).

The dual-task paradigm has been used to identify the causes and magnitudes of dual-task performance decrements and subjective workload experiences with different combinations of input and output modalities, levels of loading, and requirements for stages of cognitive processing. In general, it has been found that performance on one (or both) tasks suffers to the extent the demands for resources exceeds the system capacity (Wickens, Sandry and Vidulich, 1983). For example, the decrement in performance for a visual/manual spatial transformation task was found to be greater than for the same task presented with auditory input and speech output when each was performed with a visually displayed manual control task (Vidulich & Tsang, 1985a; 1985b). This occured even though the auditory/manual version of the spatial transformation task was performed more slowly and imposed more workload when presented as a single task. Subjective workload ratings for the dual-task combinations were somewhat less than the sum of the single- task levels. However, the cost (in terms of subjective workload experience) was significantly greater for dual-task combinations with the same input and/or output modalities, than for those that were presented in different sensory modalities or required responses in different output modalities. Dual-task workload ratings were equal to 60% of the sum of single task levels for tasks with different input or output modalities, and 75% of the sum of single-task levels for tasks that competed for the same resources.

The results of dual-task experiments, particularly those within the general structure of multiple resources theory, have provided ideas and guidance for design engineers faced with the problem of off-loading visually (or manually, vocally, etc) overloaded operators with alternative information sources or response modalities. For example, voice input or synthesized voice output has become an almost universal proposal for off-loading pilots whose ability to process additional visual information has been exceeded (Vidulich and Wickens, 1985). In addition, graphic display alternatives have been proposed to replace digital displays of instruments and the need for information integration has been recognized in order to reduce the physical number of sources and formats of information (National Research Council, 1983). Not all concurrent task components can be divided among different sensory modalities with the same improvements in performance and workload, however. It is possible that tasks elements that are functionally related by the structure of the task or their temporal relationship should be presented or performed in the same input or output modalities, while unrelated but concurrent tasks should be displayed or performed in different sensory modalities. The former might promote subjective integration, thereby reducing workload (Wickens & Yeh, 1982; 1983), whereas the latter can reduce competition for limited resources, also reducing workload.

In the typical dual-task paradigm, the two tasks must be performed within the same time period (thereby competing for an operator's limited resources), yet the component tasks are unrelated either functionally or subjectively. An alternative paradigm would be one in which component tasks are functionally related; the output or response to one serves to initiate or provide information for the other. This type of task is common in operational environments where the decision to initiate a change in a system's state requires preliminary information gathering, processing, and decision making, which is followed by one or more discrete or continuous control actions. The sources of information, processing requirements, response modality, and workload levels of the first stage are independent of those of the second stage. Nevertheless, the two tasks are functionally related and some or many processing stages may either be performed in parallel, or the activities required for one may simultaneously satisfy

some of the requirements of the other. For example, mental anticipation and physical response preparation for a control input can begin while instruments are monitored to determine the correct value or time for the control input. For these types of tasks, it is possible that presenting information in the same sensory modality would result in reduced workload and dual-task performance time, which is in direct opposition to the typical dual-task finding.

The tasks selected for the current study were based on the "Fittsberg" paradigm (Hartzell, Gopher, Hart, Dunbar, & Lee, 1983) in which a target acquisition task based on FITTS Law (Fitts & Petersen, 1964) was combined with a SternBERG memory search task (Sternberg, 1969). Two identical targets are displayed equi-distant from a centered probe stimulus. Subjects acquire the target on the right if the probe is a member of the memory set and the target on the left if it is not. Performance on the response selection portion of the task is evaluated by measures of speed (reaction time - RT) and accuracy (percent correct and decision reversals). Response execution is accomplished by moving the control stick in the selected direction (right or left) and acquiring the target on the selected side of the display. Target acquisition performance is evaluated by measuring movement time (MT), which is the total time required to acquire the target less RT. Target acquisition difficulty is manipulated within blocks of trials by varying the width (W) of the target area and its distance from the home position of the cursor (A) according to Fitts' Law (MT = a + b(ID)) where:

$$\text{Index of Difficulty (ID)} = \log_2(2A/W)$$

MT, but not RT, increased as the difficulty of the target acquisition task was increased. RT but not MT increased as the cognitive load of the response selection task was increased. Subjects rated the workload of the combined "Fittsberg" task as slightly greater than the workload of the response selection task by itself. Workload ratings for a block of trials in which different levels of target acquisition difficulty were imposed integrated the load levels imposed by both the response selection and response execution components.

In subsequent experiments (Hart, Sellers & Guthart, 1984; Mosier & Hart, 1985; Staveland, Hart & Yeh, 1985), response selection was accomplished by responding to directional commands presented symbolically or with linguistic abbreviations, identifying a stimulus with or without the additional task of comparing it to a remembered value, computing the results of mathematical equations, performing matching tasks, and time estimation, among others. The response selection demands ranged from none (in the single-target Fitts baseline condition) to stimulus identification, short-term or long-term memory search, prediction, computation, comparison, and estimation. Again, the two-stage "Fittsberg" tasks were performed with approximately the same performance and rated workload as the response selection tasks performed alone. A small "concurrence cost" (Navon and Gopher, 1979) of 40 msec in RT was again found for the combined tasks, as well as a slight increase in rated workload over single task levels (from 33 to 43). Dual task RTs were equal to 63% of the sum of single task levels and dual task workload ratings were equal to 64% of the sum of single task levels. MT was never affected by response selection difficulty manipulations. Again in opposition to the results of traditional dual-task experiments, performance decrements for the response selection (measured by RT) or response execution components (measured by MT) were not found as the difficulty of the other component was increased. Rather, the two components appeared to impose independent (or at least parallel) demands that did not increasingly degrade performance as load levels of one or both was increased.

Although this could be considered a dual-task paradigm, the response selection and execution elements can be performed sequentially and their difficulty manipulated independently, in

5.3

keeping with the assumptions of serial models of memory scanning (Sternberg, 1969) and information theoretic models of choice reaction time and target acquisition (Fitts & Petersen, 1964). In addition, the types of activities that are represented are typical of many operational environments in which operators must decide what to do (response selection) and then accomplish the desired function (response execution). The results of earlier studies suggest that the addition of automation to accomplish one or more functions might have limitations in effectiveness to moderate the demands placed on busy operators. If the execution of control inputs is automated, this might simply reduce the response execution load, leaving the demands of response selection (e.g., when and how to initiate the system) unchanged and providing little real savings in performance time or workload for functionally integrated tasks.

The current experiment was designed to address one of the issues raised earlier: For functionally integrated tasks, is the savings (measured in terms of workload, response time, or accuracy) found for functionally related tasks presented in the same sensory modality also present when the same tasks are presented in different sensory modalities? Four response-selection tasks were presented individually (in the single-task baseline experiment) and in combination with a target acquisition task (in the dual- task, Fittsberg experiment): (1) right/left decision based on spatial (Spatial); (2) or linguistic (Right/Left) information; (3) Sternberg memory search with a memory set size of one (Memory-1); and (4) Sternberg memory search with a memory set size of four (Memory-4). Each response selection task was presented visually and auditorially in both baseline and Fittsberg experiments. In the Fittsberg experiment, each response selection task was coupled with visually displayed target acquisition tasks.

The goal was to determine the rules by which dual-task performance and workload levels might be predicted from single-task levels. The spatial and linguistic command conditions were included to determine whether the large RTs found for a Right/Left condition in two earlier studies (Hart et al, 1984; Hartzell, et al, 1983) occurred because a directional command presented with a verbal code (R or L) was more difficult to translate into a directional movement than a spatial command or because additional time was required to translate the abbreviation (R or L) into its linguistic representation (right or left). The two levels of memory task difficulty were included to investigate the possibility of an interaction for measures of performance and workload between stimulus modality and the subsequent processing requirements for probes that were identical in meaning but not physical representation.

The specific experimental predictions were:

1. For simple right/left decision tasks, spatial stimuli will result in faster RTs and lower workload ratings, replicating earlier studies.

2. For memory search tasks, RT and workload will be directly related to memory set size, replicating earlier studies.

3. MT will be unaffected by the difficulty or modality of the response selection task, replicating earlier studies.

4. For both single- and dual-task presentations, the auditory display modality will result in slower RTs and higher workload ratings

5. When response selection and response execution task components are presented in the same sensory modality, substantially more dual- task facilitation will be found than when they are presented in different modalities.

METHOD
(Single-task and Dual-Task Experiments)

Subjects

Eight subjects, five men and three women participated in the single-task baseline study. None of them had served in earlier Fittsberg experiments. Eight different subjects, six men and two women served as paid participants in the dual-task experiment. All of them had served previously in an experiment in which they had received extensive training on the target acquisition task coupled with many different response selection tasks.

Apparatus

The experiment was conducted in a small experimental booth. Subjects were seated in a chair located 85 cm from a 23-cm monitor where the experimental tasks were displayed. The visual angle subtended by the most extreme targets was 11 degrees. A two-axis joystick was mounted on the right arm of the chair for the response selection and target acquisition responses. Workload-related rating values were selected with a slide-pot and entered with a button mounted on the left arm of the chair. The experiment was performed with an Apple II+ microcomputer and a Cyborg ISAAC interface modified to allow rapid and accurate recording of responses (to the nearest 10 msec). Subjects wore stereo headsets to receive stimulus information for the auditory response selection conditions. Tones were generated by the ISAAC. Linguistic information for the Right/Left and Memory tasks was generated by a Votrax Type n' Talk.

Experimental conditions

The basic task involved a binary decision to move to the right or left. The stimulus for the visual response selection tasks was a single symbol (< or >), alphabet letter (e.g., "A", "D", etc), or word ("Right" or "Left") presented in the center of the display. Stimuli for the auditory response selection tasks were presented via stereo headphones. Tones for the spatial task were presented monaurally to either the right or left ears. Right/Left commands, the memory set item(s), and memory task probes were presented binaurally. For the Fittsberg experiment, two identical targets were symmetrically presented on either side of the screen at the onset of the response selection task. (Figure 1) Their distance from the center (A) was determined by the ID for that trial. (Figure 1) The targets were two 1.25 cm vertical lines separated by the distance (W) specified by the ID for the trial. A 1.25 cm vertical line (the cursor) was controlled by movement of the joystick.

Response selection Tasks

The baseline experiment provided single-task performance and workload comparisons for the dual-task experiment. Each response selection task was presented as a choice reaction time task in both auditory and visual modalities. There were four levels of response selection difficulty: (1) Spatial command; (2) Right/Left command; (3) Memory-1; and (4) Memory-4. For the dual-task experiment, the cursor and targets were presented visually at the same time that either auditory or visual response selection stimuli were initiated.

A/Spatial information was generated by the ISAAC system. A short tone burst (1000 Hz) was presented for 1000 msec in either the right or left ear cuff. V/Spatial information was presented immediately beneath the centered cursor: "<" and ">" for left and right movement respectively. A/Right/Left commands were generated by a Votrax Type n'Talk speech synthesizer. The word "Right" or "Left" was presented binaurally at the beginning of each trial. Utterance durations were 400 and 500 msec respectively. For V/Right-Left trials, the word "Right" or
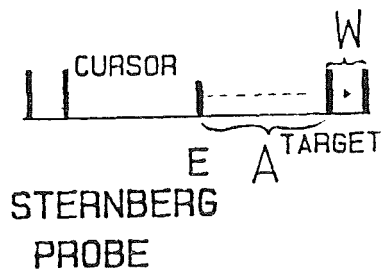
Fig. 1 Typical Fittsberg display
Memory set = 4
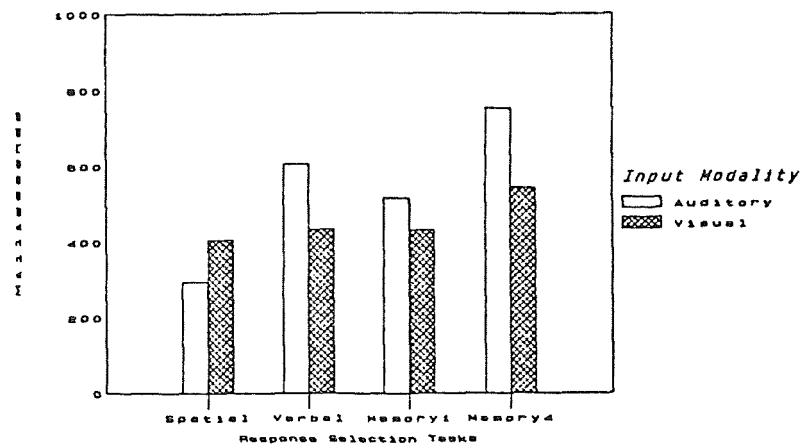
Fig. 2 Reaction times - Single Task Conditions

"Left" was displayed centered beneath the cursor. A/Memory trial blocks were preceded by binaural presentation of the memory set item(s) for the entire block of trials (e.g., "A" might be presented for Memory-1; and "B", "M", "T" and "R" for Memory-4) generated by the Votrax. Single-letter probes, also generated by the Votrax, were presented at the onset of each trial. The average duration of the alphabet-character stimuli was 300 msec. For V/Memory trials, letters were displayed on the CRT for 2000 msec before each block of trials and centered beneath the cursor at the beginning of individual trials. In the visual modality response selection stimuli remain on the display until the trial is completed.

Response execution

Response Selection component. The interval between onset of the response selection stimulus and a 2% stick deflection to the right or left was recorded as the RT. RT intervals were computed from stimulus onset for both auditory and visual presentations, as the total time required to process information is the most operationally relevant measure to use in comparing alternative stimulus presentation modalities.

Target acquisition component. The combinations of target widths and amplitudes used were all that were possible within the limited precision of the display (widths ranged from 5 to 20 pixels, amplitudes from 60 to 128 pixels). Three IDs were created (2.52 (40/60), 4.19 (either 7/64 or 14/128), and 5.67 (5/128)) in accordance with Fitts' Law. They were the same IDs that were used in earlier experiments. They were randomly presented within each block of 24 experimental trials (mean ID = 4.15). MTs were recorded as the interval between the end of the response initiation portion of the task (RT) until the steadiness criterion for keeping the cursor within the selected target had been satisfied. Single-task baseline levels for the target acquisition tasks were obtained by randomly presenting one of the four possible target configurations on the right or left.

Knowledge of results

Immediately after each trial ended (either by the selection of a response or by target acquisition), the experimental display was replaced for 2 sec by a verbal evaluation of RT and MT performance (if the subject had made a correct decision) or the word "WRONG" (if the subject selected an incorrect direction of movement). The verbal evaluations (e.g, "Fantastic", "Good", "Truly Dismal", etc.) were based on norms obtained in earlier studies.

Rating Scales

Workload experiences were evaluated by computing a derived score (Hart, et al, 1984) based

on evaluations of nine workload-related factors obtained after each experimental condition, weighted to reflect the importance placed on the factor by individual subjects. The nine factors were considered to be representative of the dimensions considered relevant to different individuals' definitions of workload: task difficulty (TD), time pressure (TP), own perfor-mance (OP), physical effort (PE), mental effort (ME), frustration (FR), stress (ST), fatigue (FA), and activity type (AT).

The relative importance of the nine factors to each subject (e.g. the weights) was deter-mined by a pretest. All possible pairs of the nine factors were presented on the computer display in a different random order to each subject. The member of each pair selected as most relevant to workload was recorded and the number of times each factor was selected was computed. The resulting values could range from 0 (not relevant) to 8 (more important than any other factor).

Subjects rated their experiences after each experimental condition on the same nine workload-related dimensions and a single global rating of workload. Each scale was presented on the experimental display as a 11-cm vertical line with a title (e.g. "MENTAL EFFORT") and bipolar descriptors at each end (e.g. "EXTREMELY HIGH/EXTREMELY LOW"). Numerical values from 0 to 100 were assigned to the selected scales positions during data analysis.

Procedure

A brief introduction was read to familiarize subjects with the purpose of the study and the types of tasks they were to perform. Then, the workload weights were obtained. The eight experimental conditions were presented in a counter-balanced order to the subjects in both experi-ments. Each condition consisted of 72 trials; two blocks of 24 practice trials presented immediately before a block of 24 experimental trials. For all conditions, half of the correct responses were "right" and half were "left", and were presented in random order. The bipolar rating scales were presented after completion the third block of experimental trials. The base-line study required one, two-hour session. The Fittsberg experiment required two three-hour sessions.

RESULTS AND DISCUSSION

Single-Task Baseline Experiment

The following data were obtained: percent correct, average RT, and bipolar ratings for each block of experimental trials. Individual 2-way and 1-way analyses of variances for repeated measures were performed between experimental conditions to determine if the predicted changes in performance and workload occurred due to response selection difficulty and stimulus modality. Selected correlations were performed among the raw bipolar ratings, weighted workload scores, and RT.

Percent Correct

Responses were made relatively accurately; average values ranged from 84% to 98% across sub-jects and from 87% (V/Memory-4) to 98% (V/Spatial) across experimental conditions. The difference in accuracy for the four response selection tasks was statistically significant ($F(3,21)$ = 6.18, $p<.01$). Although slightly more correct responses were made for the auditory display modality, the difference was not significant ($F(1,7)$ = 3.99, $p>.10$). These differences were in the same direction as the reaction times, thus ruling out the possibility of a speed-accuracy trade-off (Pachella, 1974).

Reaction Time

There were highly significant differences in RT among the response selection tasks ($F(3,21) = 50.44$, $p<.001$) and stimulus modalities ($F(3,21) = 45.74$, $p<.001$). (Figure 2) However, there was a significant interaction between the two variables ($F(1,7) = 28.10$, $p<.001$). RTs were 170 msec faster for the spatial tasks than for any other conditions. For this task, RT was considerably faster for the auditory mode of presentation than for the visual mode. A tone presented in one ear or the other is an imperative stimulus having immediate directional connotations that apparently required a minimal level of processing for a directional decision to be completed. For the Right-Left and Memory tasks, however, RTs were as much as 200 msec faster for the visual mode of presentation than for the auditory mode. The same difference occurred in RT between spatial and linguistic presentation of a directional command that was found in the earlier studies, suggesting that the earlier results were not due to difficulty in translating an abbreviation (e.g., R for right) into the word it represented. Rather, the increase in RT reflected difficulty in translating a linguistic command into a spatial movement.

It is unlikely that the presentation time for auditory stimuli influenced the modality differences. Not only was the RT shorter for the Spatial condition, but the magnitude of the differences for the remaining conditions was great enough, that the effect could not be explained by stimuli durations, although a potential confound exists. RT was recorded from the onset of the stimulus presentation. Thus, while the visual information was immediately available, the temporal nature of the auditory stimuli does not allow immediate information extraction. However, identification of information does not require the entire stimulus interval to be completed Remington (1977).

Relative importance of workload-related factors (Weights)

Subjects' initial biases about the factors they would consider in evaluating workload were obtained in a pre-test. Figure 3) Even though there was considerable diversity among the subjects' opinions, as expected, there was a small but statistically significant difference in the average importance placed on the nine factors ($F=(8,56)= 3.41$, $p <.01$). Mental Effort was the most important factor, while Physical Effort and Fatigue were the least. There was the most disagreement about the importance of Frustration and Activity Type. A multiple correlation was performed on the weights. The only statistically significant positive correlations found were for Stress (with Time Pressure and Fatigue). The only significant negative correlations found were for Activity Type (with Frustration, Time Pressure, and Stress). These results suggest that, not only do subjects disagree about the relative importance of different factors to workload, but there are few consistent relationships among the factors themselves.

Workload Ratings/Derived Score

Bipolar ratings obtained after the third replication of each experimental condition varied widely in average values and standard deviations across subjects and experimental conditions: TD (24/16); TP (40/24); OP (41/24); ME (32/16); PE (8/11); FR (38/23); ST (35/20); FA (32/22); AT (15/18); and OW (26/15). Not only did subjects disagree about what factors were relevant to workload, but they also disagreed about the degree to which each of the factors were imposed by or experienced during different experimental conditions (e.g., standard deviations were occasionally greater than the average values).

Following the procedure used in earlier studies, (Hart et, al, 1984; Vidulich & Tsang, 1985) a derived workload score was computed that reflected the subjective importance of each factor for each subject. Factors that were essential to an individual's concept of workload might be entered many times whereas others, considered less important, might be entered few times, or not at all. The averaged combination of the weighted ratings was used as the primary measure of subjective workload. As has been found in every other application of this technique,

5.8

significant relationships among experimental variables (estimated by previous research, Overall Workload ratings, and performance measures) were maintained or increased, while average between-subject variability within each experimental condition was decreased. In this experiment, between-subject standard deviations, within experimental conditions, were reduced from 14 to 11.
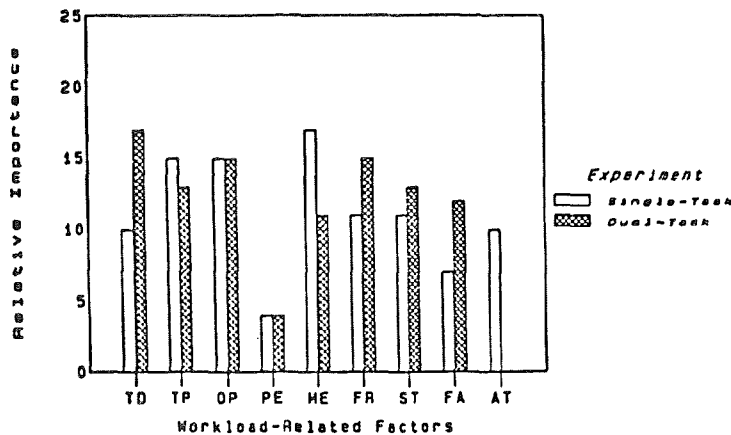
Fig. 3 Relative importance of workload related tasks
Single vs. Dual Tasks
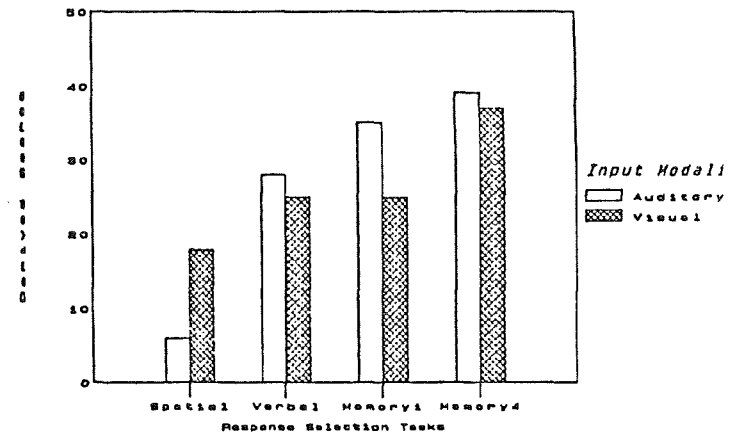
Fig. 4 Weighted Bipolar Workload Ratings
Single- Task Conditions

The Derived Workload Score reflected a pattern of statistical significance similar to that obtained for RT. There was a significant difference among the four experimental tasks ($F(3,21)$ = 15.52, p <.001), and a significant interaction between display modality and response selection task ($F(3,21)$ = 3.19 ,p<.05). The spatial decision task was less loading in the auditory modality whereas the visual versions of the other tasks were more loading. (Figure 4) As expected, the spatial decision task was considered less loading than the Right/Left decision task ($F(1,7)$ = 9.65, p<.05) and a memory set size of one for the Sternberg task was experienced as less loading (F $(1,7)$ = 5.51 p <.05) than a memory set size of four.

The relationships among the individual scales, and their association with overall workload (the weighted workload score) and RT were determined by a multiple correlation. The nine workload-related factors were not independent, suggesting a potential source of problem for multi-dimensional rating scale techniques that require statistical independence among the dimensions. Task Difficulty and Stress were related to many other factors whereas Activity Type was not. Task Difficulty, Own Performance, Mental Effort, Frustration, and Stress were significantly correlated with Overall Workload ratings and all of the factors were significantly correlated with the Derived Workload score. Although the latter result may be an artifact of the weighting procedure, it possibly reflects the fact that the derived score represents a composite of factors relevant to each subject, providing a common denominator across subjects (regardless of the factors that each considered) and measuring the workload imposed by a specific task. Few rating scales were significantly correlated with RT, even though both measures were significantly influenced by experimental manipulations. In fact, Task Difficulty and Overall Workload were the only scales that even approached a significant relationship. This finding again points out the importance of obtaining independent measures of workload and performance, as they may reflect different phenomenon.

Dual-Task, Fittsberg Experiment

The following data were analyzed: percent correct, number of decision reversals, average

5.9

RT and MT, and bipolar ratings for each block of experimental trials. Preliminary one-way analyses of variance for repeated measures were performed within blocks of trials to examine differences in performance attributable to the direction of response. Two- way analyses of variance for repeated measures were performed between experimental conditions to determine whether the predicted changes in performance and workload occurred, and multiple correlations were performed to assess the associations among bipolar ratings, derived workload scores, and performance measures.

Direction of Movement

There were no significant differences in correct selections or MT between targets presented on the right or left. There was a significant right/left differences in RT for the memory tasks (but not the other response selection tasks), as expected; "yes" responses (to the right) were made significantly more quickly $(F(1,7) = 8.02, p< .05)$ than "no" responses (to the left). This is a common finding with the Sternberg paradigm (Sternberg, 1969). Since there was an equal number of right and left conditions and because it did not interact with any of the other experimental variables, subsequent analyses were performed without regard for the direction of movement.

Percent Correct

The number of incorrect response selections did not vary significantly across experimental conditions $(F < 1.0)$ or stimulus modalities $(F < 1.0)$ Since errors were made on less than 2% of the trials, there appears to be no evidence of a speed/accuracy tradeoff. Somewhat more reversed decisions were found. A reversed decision is one in which initial decision (identified by the direction of movement recorded for RT) was made in a different direction than the target that is acquired subsequently. The differences were statistically significant for memory set size $(F(1,7) = 10.66, p < .01)$ and spatial versus linguistic directional command $(F(1,7) = 17.14, p< .01)$. Spatial commands resulted in 2.5 times fewer control reversals (less than 1 per block of 24 trials) than linguistic commands (2.5 per block). Finally, a significant interaction was found between Stimulus Modality and Method of Presentation for the direction command tasks $(F(1,7) = 7.00, p > .05)$. There were more reversals for V/Right-Left than A/Right- Left (4 versus 2 per block) whereas both A/Spatial and V/Spatial conditions were performed with consistently few reversals (less than 1 per block), regardless of stimulus modality. Subsequent analyses for performance measures included non-reversed trials only, to eliminate very long MTs for trials in which reversed decisions occurred.

Reaction Time

RTs for the dual-task conditions were generally lower $(F (1.14) = 20.75, p <$ reflecting differences in abilities between the two groups of subjects. However, there was no interaction between experiment and response selection manipulations.

RT differences within the dual-task experiment were similar to those obtained for the baseline experiment, providing sensitive indicators of response selection manipulations. (Figure 5) There was a highly significant difference in RT among the four response selection tasks $(F(3,21) = 34.83, p < .001)$. The expected differences were found between the spatial and linguistic presentation modes for the direction tasks (345 msec vs 442 msec) and between the difficulty levels of the memory task (422 msec vs 528 msec). In addition, there was a significant difference between stimulus modalities: responses to visual stimuli were generally made more quickly than to auditory stimuli $(F (1,7) = 11.62, p < .05)$. There was, however, a significant interaction between stimulus modality and response selection task $(F (3,21) = 43.73, p < .001)$, as was found in the Baseline experiment. RTs were slower for the V/Spatial than for the A/Spatial tasks, whereas the other tasks were performed more quickly with visual information than auditory.

5.10

RT for the target acquisition task presented in its single-task configuration was 421 msec, virtually the same time required to perform the simplest response selection/target acquisition task presented in the dual-task mode (413 msec), and within 100 msec of the most difficult task (Memory-4). Since the response selection tasks required at least 296 msec (A/Spatial) and as much as 754 msec (A/Memory-4) to complete by themselves, it is clear that some of the processing required to complete the response selection portion of the Fittsberg task and the initial preparation for target acquisition must have progressed in parallel. In every case, the obtained performance was equal to one half or less of the levels that would be predicted by simply adding the single task levels. This finding replicates that of earlier studies.

To adjust the reaction time distributions for the two different population samples (Experiment 1 versus Experiment 2) the following transformation was performed. Each distribution was converted to z-scores based upon its own mean and standard deviation. A grand mean was then computed on both distributions and the variances were pooled. The original z-scores were then multiplied by the square root of the pooled variance and added to the grand mean. This produces a single distribution with a mean based on all data, while retaining the shapes of the original distributions. When this transformation was applied, significant overall differences were found for response selection and stimulus modalities (as found for the experiments individually), but no interaction was found between either of these factors and experiment. When RT for the dual-task was predicted with these transformed scores, obtained RTs were 49% of the sum of single task levels for the visual modality and 60% of the sum of single task levels for the auditory modality; a significant difference in the cost of performing complex but functionally related tasks.

Movement Time

Although MTs were not analyzed within each block of trials to determine whether or not the linear relationship predicted by Fitts Law between ID and MT held, it was assumed that it did, as the same set of target configurations had been used in all of the earlier experiments, where this relationship was found. MTs for the three IDs were combined within each trial block for subsequent analyses, as each ID occurred the same number of times and no interaction between target ID and response selection difficulty manipulations was found in any of the earlier studies. No significant differences in MTs due to direction of movement were found for any of the experimental conditions.

Single-task baseline MTs averaged 888 msec. In contrast, average MTs for the Fittsberg, dual-task conditions, ranged from 834 to 874 msec across experimental conditions, 100 to 150 msec faster than were obtained in earlier studies and within 48 msec of the baseline level. (Figure 6) As predicted, there was no significant difference among MTs due to response selection load. There was, however, a significant difference in MT due to the modality of the response selection task $(F(1,7) = 11.41, p< .01)$; MTs were significantly longer when the decision of which of two targets to acquire was presented auditorially than when it was presented in the same visual modality as the target acquisition task itself. These differences were observed for every response selection task, ranging from 10 to 100 msec. Thus, there was no interaction between response selection tasks and modality $(F < 1.0)$. This is the first time that MT differences have been found due to response selection manipulations for any of the Fittsberg experiments. It is also the first time that the response selection tasks were presented auditorially as well as visually. It is possible that there is an extra cost (in MT) for processing and responding to information presented in one modality and then completing a subsequent task presented in another. This increase in MT following auditory presentation of a response selection task occurred even though the output for the response selection task (which initiated movement toward the correct target) was completed before the MT interval began. These results were based on correct and non- reversed decisions and, therefore, did not occur as a result of inaccuracy or indecision.

5.11
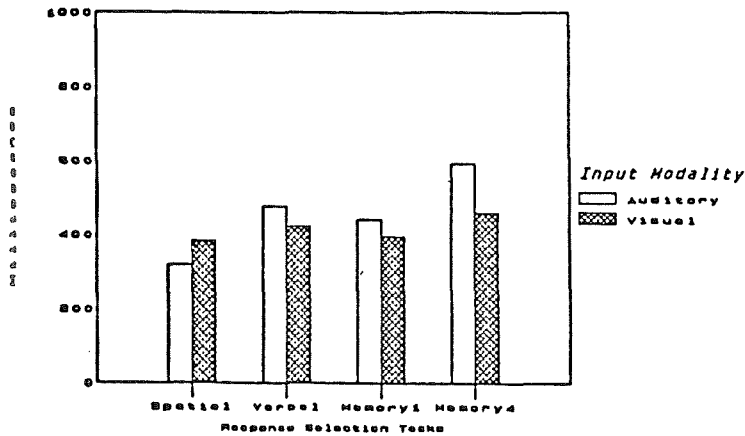
Fig. 5 Reaction Times
"Fittsberg Conditions"
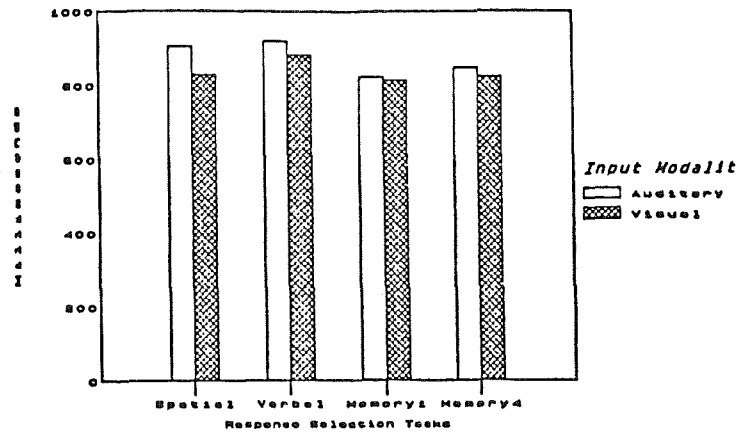
Fig. 6 Movement Times
"Fittsberg Conditions"
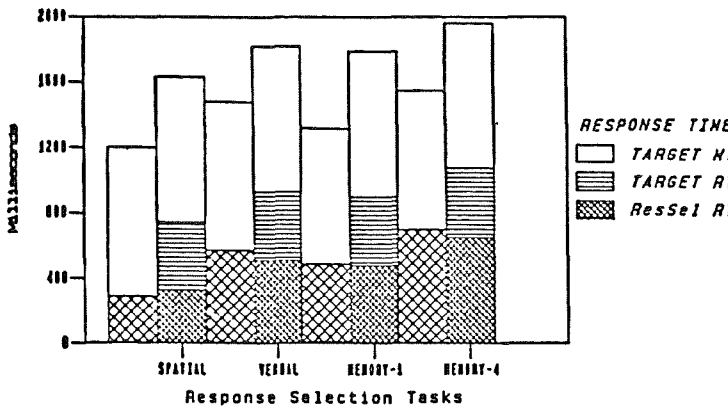
Fig. 7a Auditory/Visual
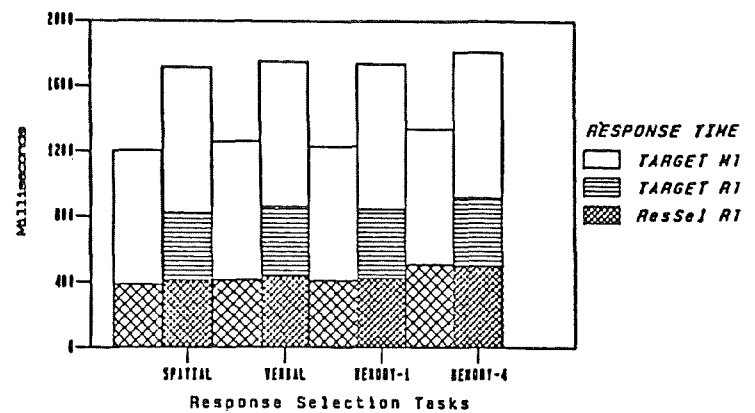
Fig 7b Visual/Visual

Total Time (RT(RS) + (RT(RE) + RE
Leftmost Bars - Obtained Total Times
Rightmost Bars - Predicted Times (Sum of Single Tasks Levels)

Because MTs were influenced by response selection modality, it is not clear whether all of the initial preparation required to perform a visual target acquisition (as estimated by RT in the single-target baseline condition) was completed in parallel with and by the end of a response selection decision if it was based on auditory information. Although target acquisition preparation could have been transferred to the beginning of the MT interval, given the design of the Fittsberg paradigm, this does not appear to have occurred in earlier studies, nor did it occur in the current study. Single-task baseline levels for MT were only 888 msec, 45 msec slower than the average dual-task MTs. Thus, this can not account for a significant portion of the 300-500 msec difference in predicted dual-task RTs compared to the sum of the single task levels and the obtained dual-task values.

Total Response Time

The total response time is the interval between stimulus presentation and target capture (the sum of RT and MT). Total times ranged from 1200 to 1440 msec across experimental conditions.

These values ranged from 70 to 81% of the levels that would be predicted by combining single-task target- acquisition RT, MT, and response selection RT for each condition. The predicted and obtained total times may be seen in Figure 7a,b. As you can see, there was a significant difference due to stimulus modality $(F(1,7) = 20.75, p<.001)$ and response selection task $(F (3,21) = 12.89, p<.001)$ when the two measures of performance were combined. There was no significant interaction. Obtained levels for the visual modality were 71% of the predicted levels and 77% for the auditory modality; again a reliable difference in the cost of performing complex but functionally integrated tasks presented in the same or different modalities.

Relative importance of workload-related factors (Weights)

The importance placed on eight of the workload-related factors may be seen in Figure 3. The Activity Type scale was not used, since it had demonstrated so little relationship with experimental manipulations in the earlier study. For this reason, only 28 pairwise combinations of factors were evaluated and the maximum value that any factor could assume was 7 (rather than 8). As you can see, there were large difference among subjects, although Task Difficulty, Own Performance, and Frustration were selected significantly more often than the rest $(F (7,49) = 3.04, p < .01)$. There was the greatest agreement among subjects about Physical Effort and the least agreement about Time Pressure and Fatigue. Again, a correlation matrix was obtained to determine the relationships among the individual factors. No statistically significant correlations were found. The weights for the eight factors in common between the two experiments were compared to determine the degree of similarity between the two groups of subjects. The two groups were not found to be significantly different. They agreed that Physical Effort and Fatigue were relatively unimportant and that Frustration, Task Difficulty, Stress, and Own Performance were important. Although the differences were not statistically significant, the two groups disagreed about the importance of Frustration, Fatigue, and Mental Effort.

Workload Ratings/Derived Score

Again, there were large differences among subjects in the degree to which subjects that felt different factors were present in specific experimental conditions. The grand mean and overall standard deviations for the nine scales were: TD (24/17); TP (22/13); OP (29/17); ME (25/18); PE (10/12); FR (21/19); ST (20/18); FA (13/18); and OW (22/18).

Following the procedure used in the first experiment, bipolar ratings were weighted to compute a derived workload score. The weighted bipolar ratings were compared to those obtained in the baseline experiment. There was a highly significant difference $(F (1,14) = 26.63, p < .001)$ between the magnitudes of ratings in the two experiments; they were consistently larger in the single- task experiment (33) than in the dual-task experiment (21), although between-subject standard deviations were identical. This may either reflect fundamental differences in the two groups of subjects, or a difference in the level of experience each had with the Fittsberg paradigm. The dual-task subjects had many hours of practice with the target acquisition tasks and a variety of response selection conditions. Thus, their perception of the workload imposed by the specific conditions included in this study could have been influenced by their previous experiences. Despite this difference, there were no significant interactions between experimental group and experimental manipulations $(F < 1.0)$.

Workload ratings followed the same pattern obtained in the baseline experiment and for RTs. As you can see in Figure 8, there was a significant difference in experienced workload among the response selection tasks $(F(3,21) = 7.13, p<.01)$. The most demanding task was the Memory-4 task (29). The least demanding task was the Spatial task (14). In addition, there was a significant difference due to stimulus modality $(F (1,7) = 13.18, p < .01)$; auditory was generally rated as more loading 23) than visual (19). In addition, a significant

interaction between stimulus modality and response selection task was found (F(3,21) = 13.34, p<.001), in agreement with the first experiment and RT performance; the Spatial task was less loading when presented auditorially, whereas the other tasks were more loading. As expected, the spatial presentation of the directional task was significantly less loading than the Right/Left version (F (1,7) = 9.52 , p <.01) and the Memory-1 was significantly less loading than Memory-4 (F (1,7) = 5.29, p <.05), replicating earlier results.
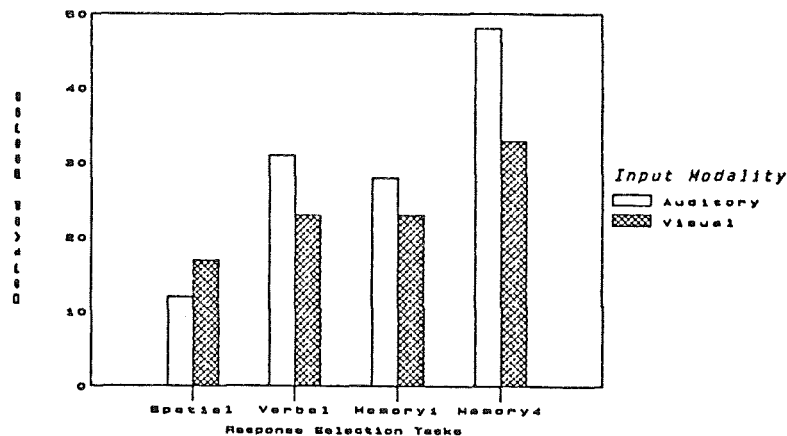
Fig. 8 Weighted Bipolar Workload Ratings
"Fittsberg Conditions"

The correlations among the nine bipolar ratings, weighted workload ratings, and total response time were obtained. Again, there was wide variation in the degree to which the different scales covaried with each other. Most of the individual scales were significantly correlated with each other with the exception of Own Performance, which was independent of the other scales. The dimensions were not, obviously, orthogonal. Every scale except Own Performance was significantly correlated with Overall Workload and all scales were significantly correlated with the derived workload scores, as was found before. None of the subjective measures were significantly correlated with total time, although they had each reflected many of the same experimental manipulations individually. This finding provides additional support for the suggestion that there may be a dissociation between measures of workload and performance (Wickens & Yeh, 1982; 1983).

Because the basic levels of ratings in the two experiments were so different, they were transformed employing the technique described earlier for RTs. When this transformation was applied, the ratings from the two experiments could be compared more directly. No significant interactions between experiment and experimental manipulations were found. Dual-task work-load levels were equal to approximately half of the sum of single-task levels for the Spatial tasks (A and V). For the remaining, tasks, visual/visual conditions were equal to 49% of the baseline task sum while auditory/visual conditions were equal to 61% of the baseline task sum. This suggests that there was greater savings (in workload experienced) with tasks presented in the same sensory modality than for those presented in different modalities (Figure 9a,b).

CONCLUSIONS

This experiment succeeded in answering a number of questions about the influences of response selection and response execution difficulty and modality on measures of performance and workload As has been found in earlier experiments with the Fittsberg paradigm, response selection load significantly affected RT but not MT. Both RT and MT were significantly longer
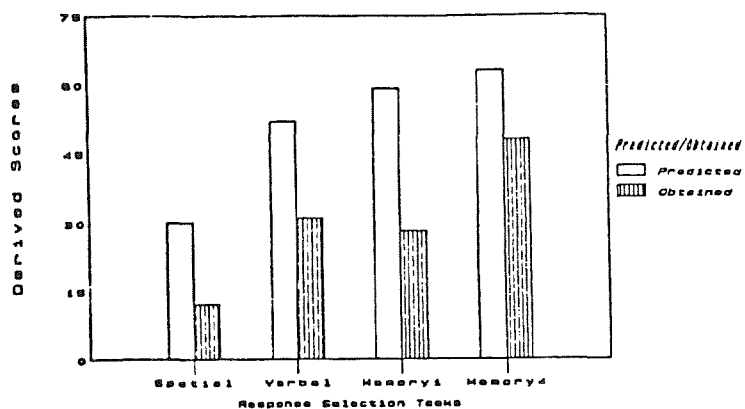
Fig. 9a Auditory/Visual
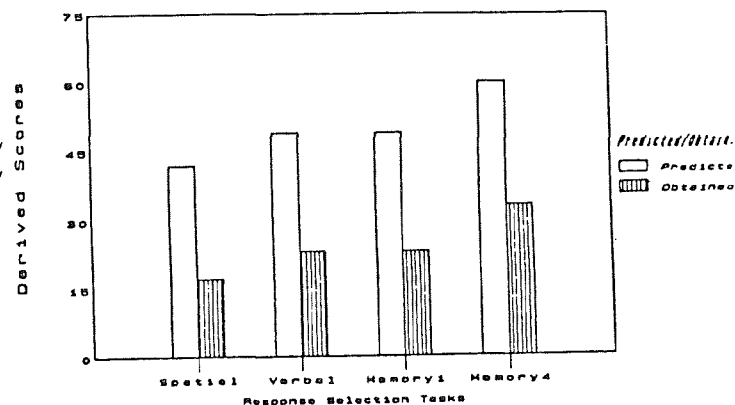
Fig. 9b Visual/Visual

Obtained vs Predicted Workload Ratings
Obtained = Dual-Task "Fittsberg" Ratings
Predicted = Sum of Single-Task Ratings

when linguistic information required for response selection was presented in a different sensory modality than the subsequent response execution task. The number of correct responses did not discriminate between any of the response selection tasks, however the frequency of reversed decisions did. The weighted averaged bipolar ratings were significantly influenced by both response selection and response execution difficulty manipulations and the stimulus-modality compatibility of the two task components.

Even though there were significant and consistent patterns of performance and workload changes as a result of all experimental manipulations, the correlations among the different measures were not statistically significant. This reinforces the point made by Wickens and Yeh (1982; 1983) that measures of workload and performance may dissociate as each is particularly sensitive to different, often subtle, aspects of experimental manipulations. For example, in the current study, both measures were sensitive to the modality of input and the response selection load, although there was an interaction between stimulus modality and difficulty for workload ratings but not for total response time or percent correct. These factors were independently influenced by each experimental manipulation. For this reason, subjective evaluations as well as multiple measures of performance are desirable to obtain a complete understanding of task demand characteristics.

Difficulty manipulations for one or both task components did not result in an interaction for any measures of performance or workload between single-and dual-task presentations. Such an interaction might have been expected with a traditional dual-task paradigm. This could have occurred because the capacity of the subjects was not exceeded by the task requirements (although there was a small RT and workload cost for putting the two tasks together), but this concurrence cost was consistent across difficulty manipulations and did not interact with level of difficulty. This provides additional support for the assertion that specific types of task combinations result in different patterns of performance and workload (e.g., either interference or facilitation).

Workload ratings integrated all task elements; both response selection and response execution sources of loading were both represented in subjective evaluations. In addition, ratings were sensitive to differences in the workload imposed by the alternative stimulus modalities, as were measures of speed and accuracy. This occurred even though there was considerable disagreement among the subjects about which dimensions were considered when evaluating workload and about the absolute magnitudes of these factors during any specific task.

5.15

As expected, the visually presented response selection tasks were well integrated with the visual target acquisition components. This physical stimulus compatibility enhanced the functional integration inherent in the Fittsberg design (e.g., the output for one served to initiate the other). The result was a considerable savings in response time and experienced workload over what might have been expected by combining single task load or duration levels. In general, RTs were 49% of the predicted additive levels, total times were 71% and workload ratings were only 46%. Response preparation for the Fitts target acquisition portion of the task was either performed in parallel with (or was replaced by) the response selection requirements of the combined tasks.

For the auditory display modality, however, the savings were not as great. RTs were 60% of predicted levels, total response times were 77%, and workload ratings were 56%. In addition, the requirement to switch from processing an auditory stimulus (in the response selection task) to acquiring a visually presented target imposed an additional cost of as much as 100 msec that was reflected in increased MTs. This could have occured because of a modality switching cost. Alternatively, the fact that the visual stimuli remained on the display during target acquisition allowed reconfirmation of response selection during this phase, whereas auditory stimuli ended before target acquisition began, thereby requiring echoic memory for reconfirmation. Although all of these values were still less than the sum of single task levels, the savings in performance time and workload were not as great. For response selection tasks that shared the least processing requirements with the response execution task (e.g., the Memory-4 task), the obtained values approached 80% of the levels predicted by adding single task levels. For this task, the additional requirement of a four-item memory search task (particularly when conducted with auditory stimuli) required a significant amount of time and effort on the part of the subjects, yet only the final decision of "yes" or "no" was directly related to the subsequent target acquisition.

These results would not be predicted in traditional dual-task paradigms where it is commonly found that concurrent tasks presented in different sensory modalities impose less interference and workload, and those in the same modalities, more. Instead, it was found that both functional and physical integration of task components resulted in a facilitation of performance and a reduction in rated workload that were often less than either single-task level. These results suggest the importance of evaluating the relationships among task components when considering display modalities in operational environments. It would appear that concurrent but independent tasks would be best presented in different sensory modalities to reduce the competition for resources if stimulus/response compatibility is not grossly violated. For task elements that are functionally related, however, the opposite might be true. Task components should be presented in the same sensory modality to enhance an operator's ability to perceive them as an integral unit (thereby reducing the perception of workload) and to reduce the need to switch information obtained from one sensory modality to subsequent activities displayed in another.

REFERENCES

Fitts, P. M. & Peterson, J. R. (1964) Information capacity of discrete motor responses. *Journal of Experimental Psychology, 67,* 103-112.

Gopher, D. & Braune. R. (1984) On the psychophysics of workload: Why bother with subjective measures? *Human Factors, 26,* 519-532.

Gopher, D., Brickner, M. & Navon, D. (1982) Different difficulty manipulations interact differently with task emphasis: Evidence from multiple resources. *Journal of Experimental Psychology: Human Perception and Performance, 8,* 146-157.

Hart, S. G., Sellers, J. J. & Guthart, G. (1984) The impact of response selection and response execution difficulty on the subjective experience of workload. In *Proceedings of the 28th Annual Meeting of the Human Factors Society*. Santa Monica, CA: Human Factors Society, 732-736.

Hartzell, E. J., Gopher, D., Hart, S. G., Dunbar, S. & Lee E. (1983) The Fittsberg Law: The joint impact of memory load and movement difficulty. Paper presented at the 27th Annual Meeting of the Human Factors Society. Seattle, WA.

Kantowitz, B, H. & Knight, J. L. (1978) Testing tapping, time-sharing: Attention demands of movement amplitude and target width. In G. E. Stelmach (Ed.), *Information Processing in Motor Control and Learning*. New York: Academic Press, 205-227.

Navon, D. & Gopher, D. (1979) On the economy of the human processing system. *Psychological Review, 86*, 214-255.

Mosier, K. & Hart, S. G. (1985) Levels of Information Processing in a Fitts Law Task (LIP-Fitts), This volume, in press.

National Research Council (1982) *Automation in Combat Aircraft*. Washington, D. C. National Academy Press.

Pachella, R. (1974) The interpretation of reaction time in information processing in research. In B. H. Kantowitz (Ed.), *Human Information Processing: Tutorials in Performance and Cognition*. : Lawrence Erlbaum Associates, Hillsdale NJ. 41-82

Remington, R. (1977) Processing of phonemes in speech: A speed-accuracy study. *Journal of the Acoustical Society of America, 62*, 1279-1290.

Staveland, L., Hart, S. G. & Yeh, Y.- Y. (1985) Memory and Subjective Workload Assessment, This volume, in press.

Sternberg, S. (1969) The discovery of processing stages: Extension of Donder's method. *Acta Psychologica, 30*, 276-315.

Tsang, P. S. & Wickens, C. D. (1984) The effects of task structures on time-sharing efficiency and resource allocation optimality. *Proceedings of the 20th Annual Conference on Manual Control NASA CP-2341*. Washington, D. C.:National Aeronautics and Space Administration, 305-318.

Vidulich, M. A. & Tsang, P. S. (1985) Techniques of subjective workload assessment: A comparison of two methodologies. *Proceedings of the Third Biannual Symposium on Aviation Psychology*. Columbus, OH: Ohio State University, in press.

Vidulich, M. A. & Wickens, C. D. (1985) Stimulus-central processing- response compatibility: Guidelines for the optimal use of speech technology. *Behavior Research Methods, Instruments, and Computers*, in press.

Wickens, C. D. & Kessel, C. (1979) The effects of participatory mode and task workload on the detection of dynamic system failures. *IEEE Transactions on Systems, Man, and Cybernetics, SMC-9*, 1080-1082.

Wickens, C. D., Sandry, S. L., & Vidulich, M. A. (1983) Compatibility and resource competition between modalities of input, central processing, and output. *Human Factors, 25*, 227-248.

Wickens, C. D. & Yeh, Y.- Y. (1982) The dissociation of subjective workload and performance. *Proceedings of the International Conference on Cybernetics and Society.* New York: Institute of Electrical and Electronics Engineers, 584-587.

Wickens, C. D. & Yeh, Y.- Y. (1983) The dissociation between subjective workload and performance - A multiple resources approach. *Proceedings of the Human Factors Society 27th annual Meeting.* Santa Monica, CA: Human Factors Society, 244-248.

THE IMPACT OF PHYSICAL AND MENTAL TASKS
ON PILOT MENTAL WORKLOAD

Scott L. Berg and Thomas B. Sheridan
Man-Machine Systems Laboratory
Room 3-346
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

Abstract

Seven instrument-rated pilots with a wide range of backgrounds and
experience levels flew four different scenarios on a fixed-base simulator.
The Baseline scenario was the simplest of the four and had few mental and
physical tasks. An Activity scenario had many physical but few mental
tasks. The Planning scenario had few physical and many mental tasks. A
Combined scenario had high mental and physical task loads. The magnitude of
each pilot's altitude and airspeed deviations was measured, subjective
workload ratings were recorded, and the degree of pilot compliance with
assigned memory/planning tasks was noted.

Mental and physical performance was a strong function of the manual activity
level, but not influenced by the mental task load. High manual task loads
resulted in a large percentage of mental errors even under low mental task
loads. Although all the pilots gave similar subjective ratings when the
manual task load was high, subjective ratings showed greater individual
differences with high mental task loads. Altitude or airspeed deviations
and subjective ratings were most correlated when the total task load was
very high. Although airspeed deviations, altitude deviations, and
subjective workload ratings were similar for both low experience and high
experience pilots, at very high total task loads, mental performance was
much lower for the low experience pilots.

6.1

C - 2

I. INTRODUCTION

Cockpit design practices of the last 15 years share a common thread: the degree and complexity of automation is increasing and accelerating. Current state-of-the-art designs such as the Boeing 757, 767, and Airbus Industries A310 have radically changed flight deck activities. Future designs, such as the U.S. Air Force's proposed Advanced Technology Fighter and the Navy's Advanced Combat Aircraft will demand far greater levels of automation because of the requirement to operate in an extremely hostile, changing environment.

Expert systems and artificial intelligence will reduce or eliminate certain types of pilot workload. However, in some instances they may simply change the type of workload. Pilots are operating less as manual controllers and more as supervisory controllers.

The increased time and effort expended in monitoring aircraft equipment has raised concerns that in automating aircraft we may be raising the pilot's mental workload to unacceptable levels (or conversely, lowering it to undesirable levels). Thus, there is great interest in measuring this mental workload and its effects. However, measuring mental workload has been a difficult problem to solve.

Different researchers and different segments of the engineering and design communities have defined mental workload differently. Systems engineers, psychologists, and physiologists all have their own models of mental workload and their own methods of measuring it.

However, over the last decade, there has been a growing consensus that: a) mental workload is multidimensional in nature; and b) because of this multidimensionality, the "best" approach to measuring mental workload is to combine objective performance measures and subjective rating measures.

II. OBJECTIVES

This research examines several issues relating to mental workload. First, how does automation affect pilot mental workload? Since mental workload is multidimensional, automation may affect each dimension differently. Second, how does the level of mental workload affect physical and mental performance? Third, is the magnitude of a pilot's mental workload a function of the time between receiving instructions and executing them?

III. SIMULATOR CONFIGURATION

Figure 1 pictures the laboratory flight simulator environment for this project. The volunteer pilot subjects manipulate controls and switches on a control box while getting aircraft state information from a MEGATEK high resolution cathode ray tube (CRT) display (Figure 2). The MEGATEK displays flight instruments, aircraft and equipment configuration, and a forward

Figure 1: The Laboratory Environment

Figure 2: Diagram of the Simulator's high resolution MEGATEK CRT display

Figure 3: Layout of the Control Box for the flight simulator

Figure 4: Navigation Chart for the Experiment
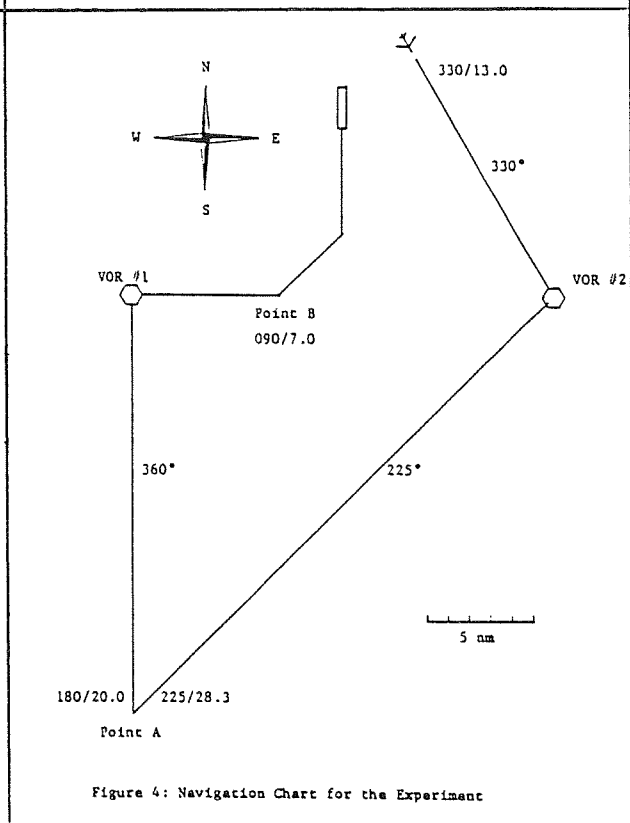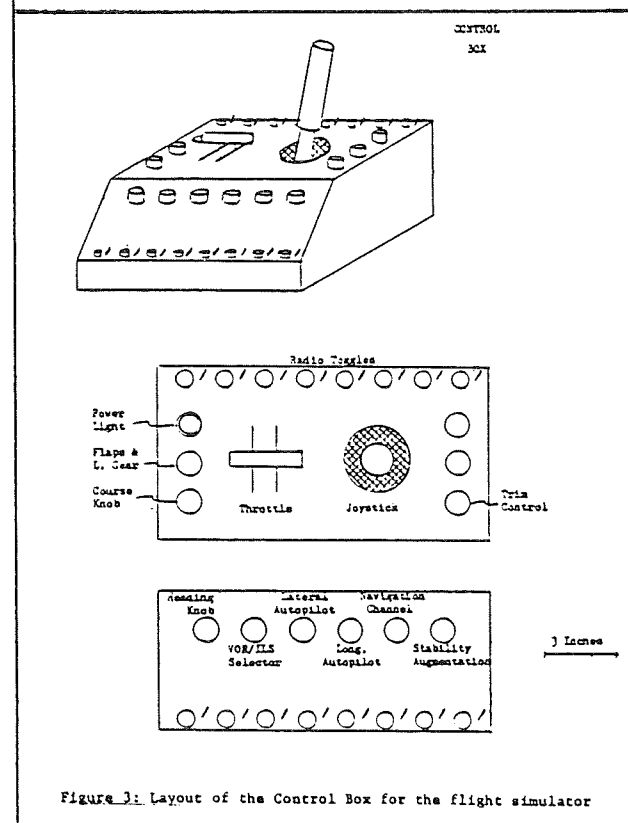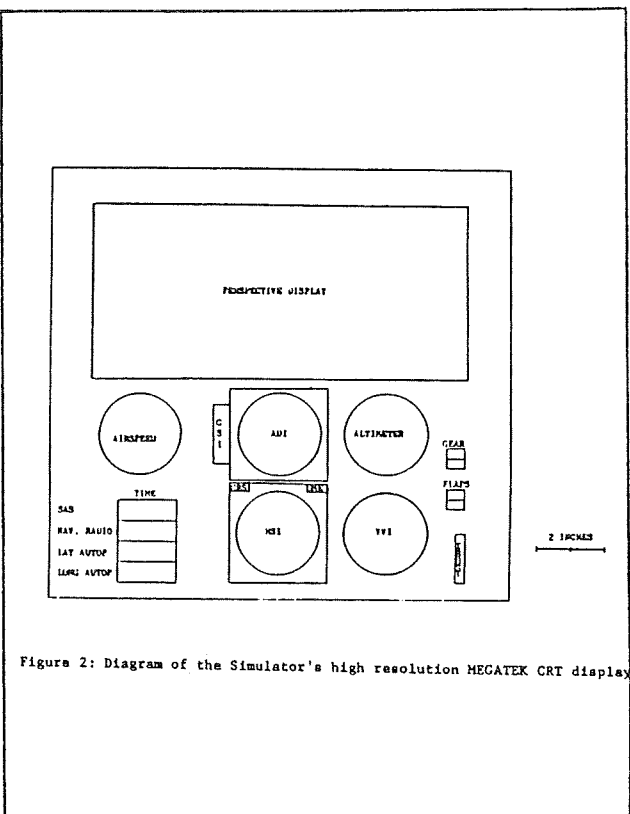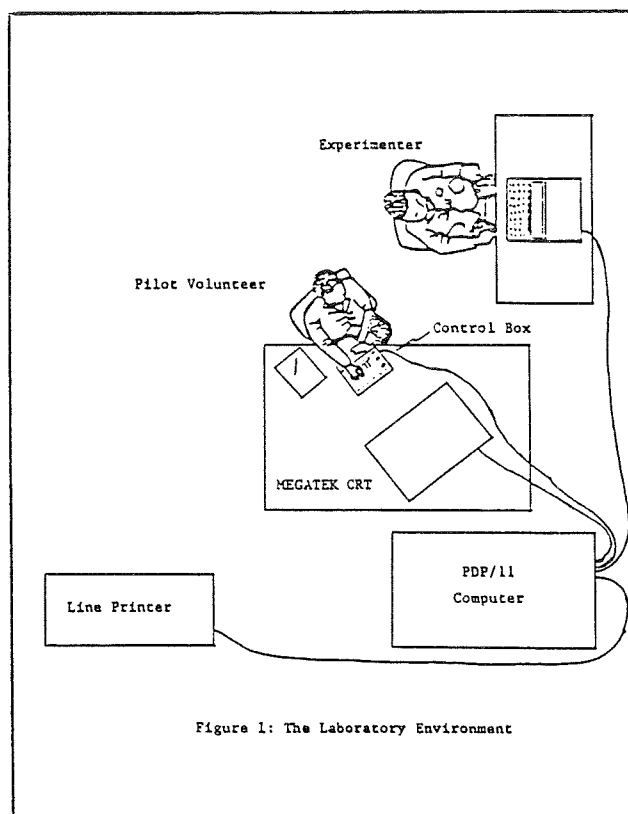
perspective view. The investigator has his own video display terminal (VDT) and keyboard for controlling the system.

A drawing of the Control Box is shown in Figure 3. The subject interprets the flight information displayed on the MEGATEK and manipulates the controls and switches on the Control Box to make the "aircraft" respond in a desired fashion. Control Box signals are fed to a PDP/11 Computer. The Computer's simulation program takes the present aircraft state information, Control Box inputs, and the investigator's Keyboard commands to determine aircraft dynamics and a new aircraft state. The information is used to update the MEGATEK and VDT displays.

A great deal of experimental trial and error went into making the simulator's response as close as possible to the response of an actual aircraft. A number of pilots came to the lab, flew the simulator, and evaluated its handling qualities. Eventually, the simulation fidelity was brought to a high level, including realistic stall and landing characteristics.

The Computer stores all Control Box switch or control manipulations and stores aircraft state data every 10.0 seconds. This data can be displayed on the investigator's VDT or printed out on a Line Printer.

IV. SUBJECTS

Initially, approximately 30 pilots volunteered to participate. Although we had hoped to use at least a dozen pilots of varied background, the list of 30 was eventually reduced to 7. Unfamiliarity with the flight characteristics of high performance aircraft and the simulator's ADI/HSI display, and the inability to devote the time needed for qualifying on the simulator and flying the data runs eliminated most of the pilots.

All seven subjects were good pilots, and there was a good mix of experience. Three subjects were Air Force pilots with 2400 to 3200 hours of flight time. Two pilots were Certified Flight Instructors with instrument ratings. The four civilian pilots ranged in experience from 300 total hours to 3000 total hours and had between 50 and 250 hours of instrument time.

V. EXPERIMENTAL DESIGN

Four different scenarios were flown using one basic route, illustrated in Figure 4. The four scenarios were labeled Baseline, Activity, Planning, and Combined. The Baseline scenario was the easiest. It simulated a "normal" flight and the pilots were encouraged to use the autopilot to keep workload at a minimum. There were no directed deviations from the basic course, and airspeed and altitude changes were rare. Also, there were very few assigned memory or planning tasks.

A data session consisted of a Baseline run followed by one of the other scenarios. The Baseline scenario was used as a warm-up data run and as a calibration run. Each second run's data was compared to that session's Baseline run. Baseline performance and ratings for different sessions could then be compared to adjust the data for variations due to day-to-day differences such as fatigue, stress, emotional state, et cetera.

The Activity scenario was loaded with a large number of manual-control tasks, but like the Baseline scenario, had a light planning task load. The pilots flew this scenario without using the autopilot.

The Planning scenario was very different from the Activity scenario. It was almost identical in manual activity to the Baseline scenario, (and thus, had a low activity level) but instead of being directed to perform actions immediately, the pilots were directed to perform these actions at a certain time in the future. These instructions often involved overlapping time periods, and the requests were not ordered chronologically. For example, prior to 2:00 minutes the pilot might be told to descend 1000 feet at 5:00 minutes, then told to turn to 300 degrees heading at 13:30 minutes, then to slow to 190 knots at 8:00 minutes. Therefore, the pilots had to "plan ahead".

The Combined scenario was designed to be the most difficult of all. It combined the manual activity of the Activity scenario with the planning requirements of the Planning scenario. This was an effort to saturate the pilots. The pilots were allowed to use the autopilot for help, but the pace of this scenario usually limited its use.

Figure 5 lists the order in which each pilot flew each of the non-Baseline scenarios. Different pilots flew the various scenarios in different orders. However, they all began each session's data runs with a Baseline run. The other three scenarios were not truly order randomized, but they were mixed. No pilot flew the Combined scenario in the first session. It was so unusually difficult, it was felt that this scenario might create an impossible workload for any pilot flying it first.

A Navigation Chart (Figure 4) and a note pad were provided for each pilot's use. Also, special placards were displayed beneath the instrument display to give configuration/airspeed data and help the pilots with the various lateral and longitudinal autopilot modes.

Ground tracks, altitude profiles, and airspeed profiles provided in Figures 6 through 9, clearly illustrate some of the differences and similarities of the various scenarios. Those three items were nearly identical for the Baseline and Planning scenarios, and for the Activity and Combined scenarios. Figure 6 shows the ground track for the Baseline and Planning scenarios while Figure 7 shows the ground track for the Activity and Combined scenarios. Note the large number of heading changes for the Activity/Combined scenarios. In the Activity and Combined scenarios the subjects were given new headings, altitudes, and airspeeds each 2 minutes for the first 5 minutes, each minute for the next 10 minutes, and each 30

SCENARIO	PILOT						
	A	B	C	D	E	F	G
Activity	1	2	3	3	1	2	1
Planning	2	1	1	1	3	1	2
Combined	3	3	2	2	2	3	3

Figure 5: Session number in which pilots flew each scenario
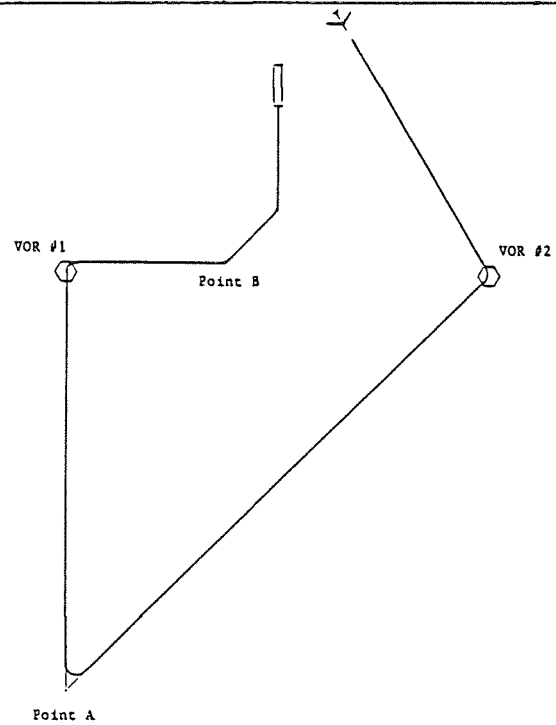
Figure 6: Nominal ground track for the Baseline and Planning scenarios
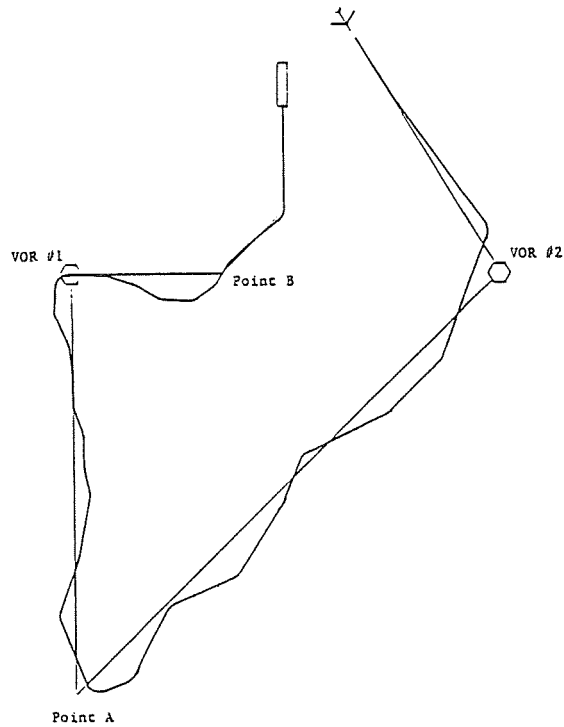
Figure 7: Nominal ground track for the Activity and Combined scenarios
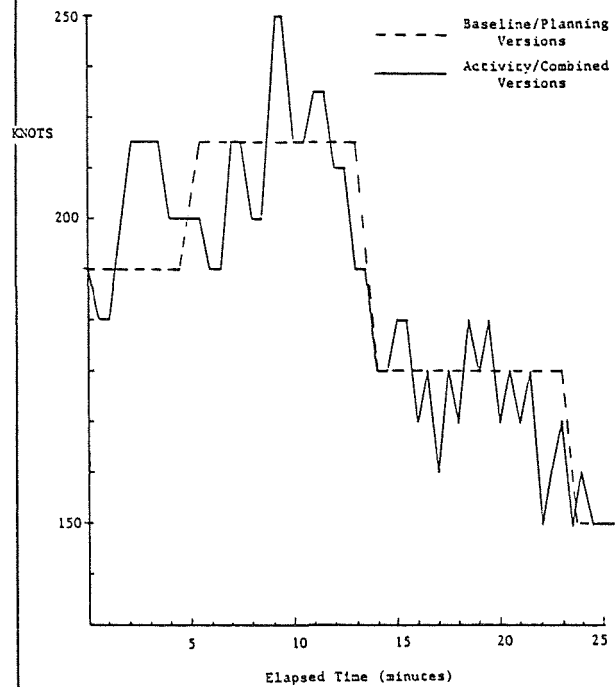
Figure 8: Planned airspeed versus elapsed time
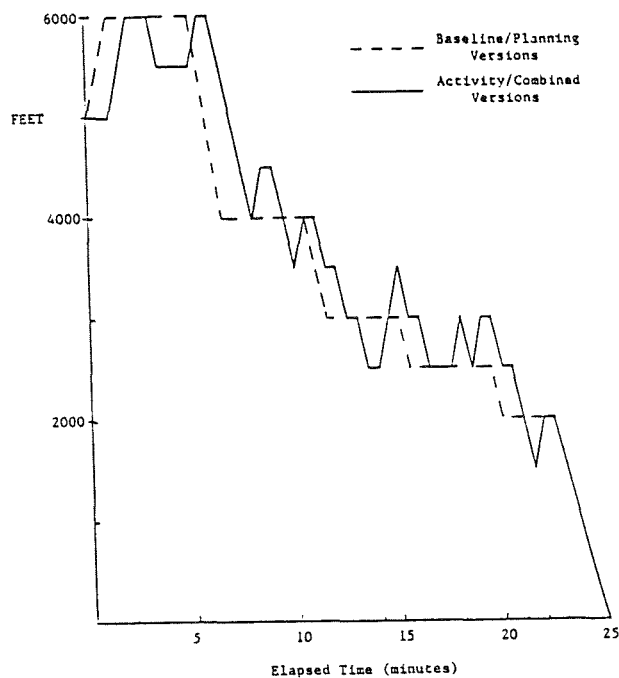
Figure 9: Planned altitude versus elapsed time

	Scenario			
	Baseline	Activity	Planning	Combined
Total Planning WU's	43	47	253	254
Total Number of Planning Tasks	3	3	23	24
Short-term Planning Tasks	0	0	14	16
Medium-term Planning Tasks	3	3	6	5
Long-term Planning Tasks	0	0	3	3
Total Activity WU's	28	150	29	142

Figure 10: Scenario characteristics
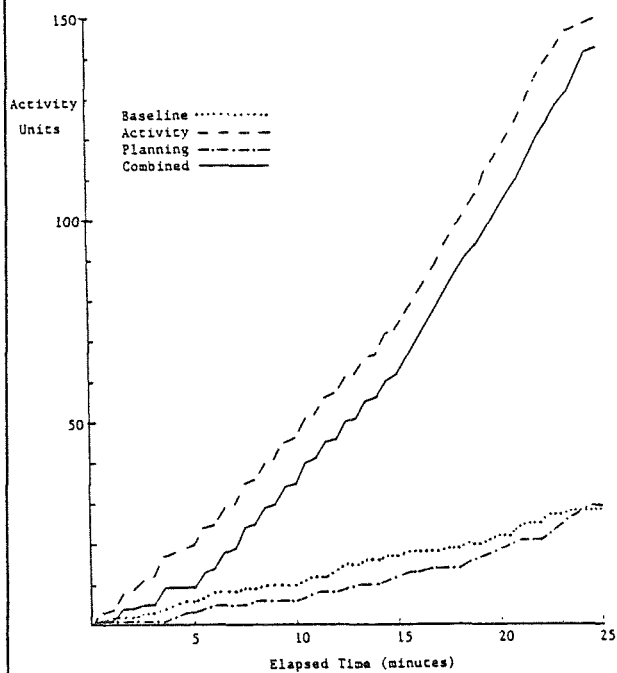
Figure 11: Accumulated Activity Workload Units versus elapsed time
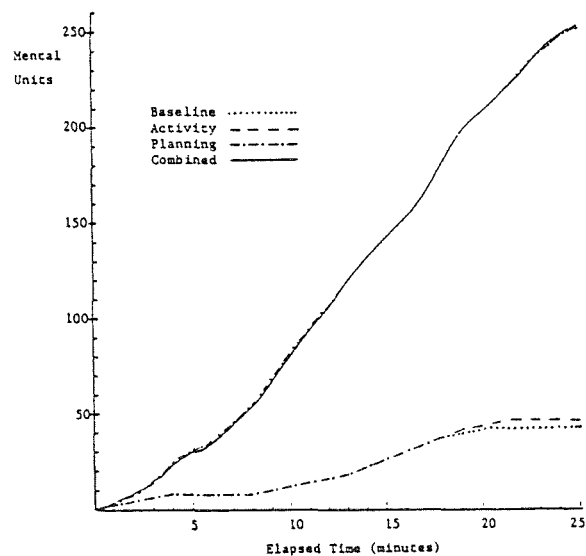
Figure 12: Accumulated Mental Workload Units versus elapsed time

6.7

seconds for the final 10 minutes. At several points, pilots were given instructions to contact ARTCC rather than perform some task. Figure 8 is an airspeed versus time plot for the various scenarios. There are 31 airspeed changes for the Activity and Combined scenarios and 3 for the Baseline and Planning scenarios. Finally, Figure 9 shows altitude versus time. The Activity and Combined scenarios have 21 directed altitude changes to 5 for the Baseline and Planning scenarios.

Each mental or physical task was evaluated and assigned a number of "workload units". The total number of workload units (WU's) and the workload unit rate were used to compare the four scenarios. An extensive explanation of the method used to calculate these workload units can be found in Berg and Sheridan, 1984.

Each scenario had a number of planning tasks. These planning tasks were categorized as either Short-term, Medium-term, or Long-term. We arbitrarily defined a short-term planning task as lasting from 0 to 4 minutes, a medium-term task lasting from 4 to 12 minutes, and a long-term task lasting over 12 minutes. The average short-term task was 2.6 minutes long, the average medium task was 7.2 minutes, and the average long-term task was 16.6 minutes.

Figure 10 summarizes the information for all four scenarios. Note that the Planning and Combined scenarios have about 5 times as many planning WU's as the Baseline and Activity scenarios. Also, the Activity and Combined scenarios have roughly 5 times as many activity WU's as the Baseline and Planning scenarios. Finally, the Planning and Combined scenarios have almost 8 times as many planning tasks as the Baseline and Activity scenarios.

In recognition of Miller's (1956) findings about human limits on immediate memory, the number of simultaneous planning tasks never exceeded 9. Although the Planning and Combined scenarios had what seemed to the subjects to be an intense level of simultaneous planning tasks, the mean number of simultaneous planning tasks was only 5.0, with a standard deviation of 1.8.

Figures 11 and 12 portray some of this workload data graphically. Figure 11 is a plot of the accumulated number of activity WU's as a function of time. Figure 12 is a plot of the accumulated number of planning WU's as a function of time. Note not only the difference between dissimilar scenarios, but also the similar workload rates for scenarios with similar types of workload.

VI. TRAINING AND INSTRUCTIONS

In addition to the initial screening sessions, each pilot participated in 4 to 10 hours of additional training. Three of the four pilots had flown the simulator before, but had never used the autopilot. They required about 4 hours of additional practice.

This autopilot is different from most commercial equipment. Longitudinal and Lateral modes must be engaged separately, adding one additional step in

selecting some autopilot functions.

Before a session's data runs, pilots "warmed up" by flying instrument approaches, turns to headings, etc., for 20 to 30 minutes. After this warm up period, the pilots were handed an Instruction Sheet, the Subjective Ratings/Comments Sheet shown in Figure 13, and a sheet which explained the scale to be used in making the subjective ratings.

In the instructions, pilots were told to fly "as well as you can" and follow all directions "to the best of your ability". They were also told that they would be scored on their ability to "follow instructions and comply with requests". Thus, they had no idea which parameter(s) would be measured. Any or all might be scored.

As explained in the instructions, the simulation was "frozen" for subjective ratings at 5:00, 16:00, and 27:00 minutes elapsed time. The desired method for scoring subjective ratings was explained, and the subjects warned that only one minute would be allowed for making the ratings during each break. Preliminary experiments had shown that the pilots only required about 20 to 30 seconds to make these ratings.

After each run, the pilots were debriefed and asked to put any comments or explanations on the rear of the Rating Sheet.

VII. DATA

Every 10 seconds, the computer recorded the aircraft's airspeed and x, y, and z position. This data yielded a ground track, and by comparing position and elapsed time, desired altitudes and airspeeds were determined. This information was then compared with the actual airspeeds and altitudes to derive altitude and airspeed error. Altitude errors were not computed during directed climbs and descents and airspeed errors were not computed during directed airspeed changes. Pilots were expected to climb or descend at a minimum of 1000 feet per minute and accelerate or decelerate to the desired airspeed within 30 seconds or at a rate of at least 50 knots per minute for airspeed changes greater than 25 knots. These rates of change are consisted with recommended piloting techniques.

Ground tracks were plotted for reference, but deviations from the nominal ground track were not scored.

Altitude deviations seemed to be the "best" objective measure to use. However, with only one objective measure, it was possible that pilots might give higher priority to one aspect of aircraft control than another. Thus, airspeed deviations were scored to serve as a check. Both variables were scored with mean absolute and RMS deviations.

Five experimentally proven subjective ratings were used in order to examine the multi-dimensionality of the mental workload. These ratings were ACTIVITY LEVEL, COMPLEXITY, DIFFICULTY, STRESS, and WORKLOAD. Ratings were

Figure 13: Subjective Ratings/Comments Sheet for Primary Experiment

SCENARIO	SEGMENT	MEAN	STD DEV	RMS
Baseline	I	39.1	18.7	50.6
	II	41.4	24.0	51.0
	III	30.6	13.8	41.4
	Overall	37.0	19.0	47.7
Activity	I	114.4	110.5	147.8
	II	97.7	24.8	153.3
	III	138.0	36.6	199.0
	Overall	116.7	67.3	166.7
Planning	I	19.5	23.0	23.5
	II	47.8	19.1	56.6
	III	55.6	34.0	60.4
	Overall	41.0	29.5	46.8
Combined	I	93.5	85.6	131.7
	II	122.4	77.5	198.2
	III	154.8	81.8	204.9
	Overall	123.6	81.7	178.3

Figure 14: Overall mean absolute and rms altitude deviations (feet)

Figure 15: Average altitude deviations for the Baseline (B), Activity (A), Planning (P), and Combined (C) scenarios

made at three points during each run. Subjects were not asked to make an overall rating because overall ratings made during previous experiments were nearly identical to the arithmetic mean of the segment ratings and we believed the same would be true here.

The distance from the left edge of each scale to each pilot rating was measured, divided by the total scale length, and multiplied by ten. This gave subjective ratings with a possible range of 0 to 10.

An integral aspect of this set of experiments was an investigation into not only the degree of mental workload, but also the effect this effort had on observable pilot behavior. Thus, in addition to the aircraft control measures and subjective ratings just discussed, other aspects of pilot behavior were also measured.

During each run, notes were made on the pilot's compliance in carrying out assigned planning or memory tasks. All pilots were assigned specific elapsed times (clearly displayed on the instrument panel) at which to perform these tasks. Each pilot was given $\pm$ 15 seconds from the designated time in which to begin the task. If a task was begun outside these limits, it was noted. When a task was performed improperly, for example climbing to a wrong altitude or accelerating 10 knots instead of climbing 1000 feet, this was also noted. A third type of mental error was forgetting or missing an item entirely.

A final source of information was post-run debriefings. The pilots had many interesting and useful insights into mental workload, stress, and their affect on performance.

VIII. RESULTS

Learning effects
The Objective and Subjective data was examined for "learning effects".
Using Student t-test and F-test techniques, we found no significant learning effect for altitude or airspeed deviations for any of the four scenarios.

Each session's Baseline run acted as a "warm up" run and served as a day-to-day metric for the Subjective ratings. For each Subjective rating, the Baseline run ratings were averaged across all seven pilots and all three runs for each pilot. This yielded an overall mean baseline rating. This mean rating was added to the difference of a session's Baseline rating and second run (Activity, Planning, or Combined) rating. This gave an "adjusted" second run rating. The intent was to compensate for day-to-day differences in emotional state, stress, fatigue, et cetera.

Using these adjusted subjective ratings, there was no "learning effect" for any of the ratings for the Activity scenario. For the Planning scenario, only the WORKLOAD ratings showed a learning effect (80 percent confidence level).

6.11

So, the extensive training, the modified counterbalancing of scenarios and subjects, and "adjusting" the subjective ratings appears to have minimized learning effect for the Activity and Planning scenarios.

However, there was some evidence of learning effect for the Combined scenario. Three subjective ratings were lower for the third sessions than the second sessions. The effect was at an 80 percent confidence level for COMPLEXITY ratings. Since post-run debriefings showed that COMPLEXITY ratings were closely tied to the pilots' ease with the autopilot, this may be due to greater familiarity with the device. Learning effect was at a much stronger 95 percent confidence level for the DIFFICULTY and WORKLOAD ratings. None of the practice rounds were nearly as intense as the Combined scenario. Furthermore, the Combined scenario was a combination of the Activity and Planning scenarios. Thus, subjects who had seen both the Activity and Planning scenarios before flying the Combined scenario had an advantage over those who flew the Combined scenario after flying only one of the others.

Finally, an analysis of variance showed no statistically significant difference for planning task performance for any scenario.

Objective activity performance results
Altitude and Airspeed error data was synthesized from the computer's output. Altitude error data is summarized in Figure 14. Note the standard deviation data in Figure 14. The bulk of pilot deviations tended to lie near the mean. However, there was usually some pilot whose deviations took an extreme, isolated jump, inflating the standard deviation for the group.

In general, just as the WU rate increased from Segment I to Segment III, so did altitude deviations (see Figure 15). Segment-to-segment mean absolute error differences were significant at a 90 percent confidence level for the Combined scenario, 95 percent for the Baseline and Activity scenarios, and 99 percent for the Planning scenario. The larger spread of individual performance in the Combined scenario was responsible for its lower confidence level.

As Figure 15 shows, there was a considerable difference (99 percent confidence level) between the manually controlled Combined and Activity scenarios and the autopilot controlled Planning and Baseline scenarios. The average deviation was 3.1 times greater (120.2 feet versus 39.0 feet) under manual control, and the rms deviation was 3.6 times greater (172.5 feet versus 47.3 feet). However, it should be noted that the manually controlled Combined and Activity scenarios also had much more difficult altitude profiles than the autopilot controlled scenarios. (See Figure 9)

Interestingly, the magnitude of mental tasking had no significant impact on the magnitude of the altitude deviations. The Baseline scenario's altitude deviations were statistically similar to those of the Planning scenario, the latter differing from the former solely in having a large number of mental planning tasks. Similarly, the mentally easy Activity and mentally demanding Combined scenarios were statistically identical.

6.12

Airspeed error data was also synthesized from the computer's output and is summarized in Figure 16. Like the altitude deviation data, some of the large standard deviations in Figure 16 are due to some pilot's momentary lapse. Most of the deviation data was fairly consistent in magnitude.

Segment-to-segment differences were significant for all four scenarios (See Figure 17). For mean absolute airspeed errors, the segments differed at a 90 percent confidence level for the Activity scenario and a 99 percent level for the Baseline, Planning, and Combined scenarios. RMS airspeed errors differed at a 95 percent confidence level for the Baseline and Activity scenarios and a 99 percent confidence level for the Planning and Combined scenarios.

Like the altitude deviation data, the magnitude of airspeed errors was a strong function of the mode of aircraft control. As shown in Figure 17, when airspeed was under manual control, deviations were much greater than when airspeed was under autopilot control. The difference was statistically significant at a 99 percent confidence level for mean absolute error and a 98 percent level for rms errors. Again, part of this result may be due to the much more difficult airspeed profile for the manually controlled scenarios (See Figure 8). This airspeed deviation data also showed little mental tasking effect. There was no significant difference between scenarios which had similar manual activity levels but different planning workloads.

Both altitude and airspeed deviations were similar for all the pilots. In general, the low experience pilots had slightly higher deviations than the most experienced pilots. However, there was enough scatter in the data to keep the differences statistically insignificant.

This objective data showed only a hint of performance degradation due to pilot workload saturation. During the Activity scenario runs, only two pilots out of seven had average mean altitude deviations greater than 150 feet in Segment III, and two other pilots had average mean airspeed deviations greater than 15 knots in Segment III. For the Combined scenario, the number of saturated pilots rose to three for the altitude deviations and remained at 2 for the airspeed deviations.

Within each scenario, there was no significant correlation between airspeed and altitude deviations because different individuals traded-off airspeed and altitude control during all four scenarios. However, overall scenario airspeed and altitude control were correlated. The Baseline and Planning scenarios had low deviations for each score and the Activity and Combined scenarios had high deviations for both scores.

Subjective ratings results
The Subjective Rating data was useful because it illustrated the impression these scenarios were making in the minds of the pilots. Thus, although only an indirect measure, one would expect these ratings to provide a better indication of mental workload than objective performance data.

SCENARIO	SEGMENT	MEAN	STD DEV	RMS
Baseline	I	1.9	0.7	2.9
	II	3.9	0.7	5.0
	III	3.4	1.9	4.4
	Overall	3.1	1.4	4.1
Activity	I	7.9	6.6	9.2
	II	9.5	4.3	12.5
	III	11.9	5.9	15.5
	Overall	9.8	5.7	12.4
Planning	I	0.7	0.4	1.0
	II	3.7	2.4	4.0
	III	3.3	1.9	3.9
	Overall	2.6	2.2	3.0
Combined	I	5.2	2.4	6.2
	II	11.0	4.5	14.3
	III	9.6	4.2	13.2
	Overall	8.6	4.4	11.2

Figure 16: Overall mean absolute and rms
airspeed deviations (knots)
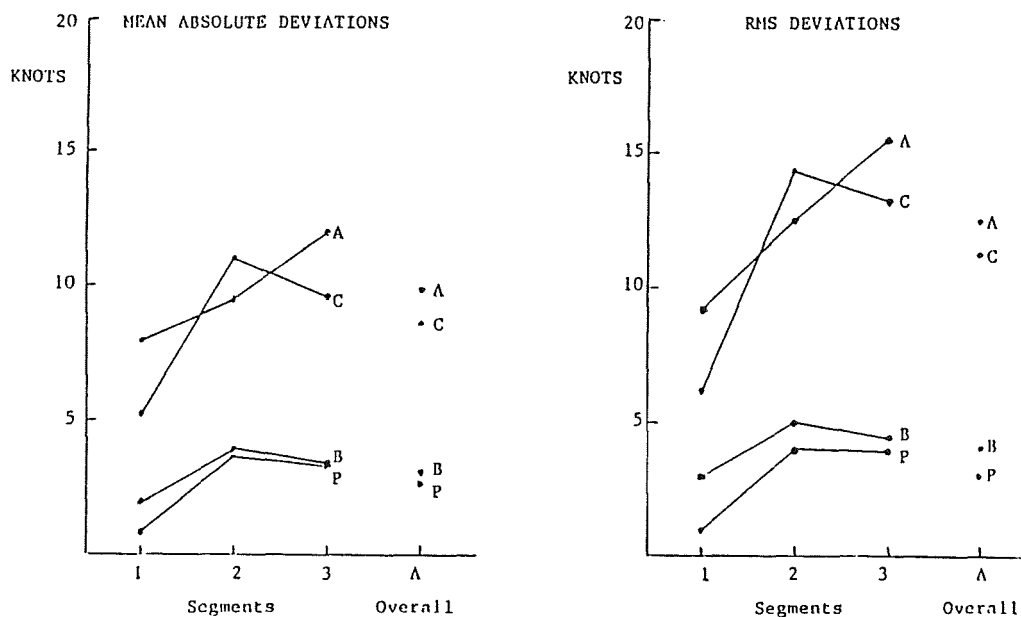
Figure 17: Average airspeed deviations for the Baseline (B), Activity (A),
Planning (P), and Combined (C) scenarios

6.14

Figure 18 gives the subjective rating data averaged over all the pilots for each segment, scenario, and category. Note that the standard deviation data is very consistent from rating to rating and scenario to scenario. Individual ratings did not exhibit the wide variations present in the altitude and airspeed deviation data.

In general, subjective ratings for the five categories were similar for the Activity and Planning scenarios, but statistically different for those two scenarios and the Combined scenario. The Combined scenario ratings were statistically different from the Activity and Planning scenarios at a 90 percent confidence level for the WORKLOAD and DIFFICULTY ratings, a 98 percent confidence level for the ACTIVITY LEVEL ratings, and a 99 percent confidence level for the COMPLEXITY and STRESS ratings. The averaged ratings for each scenario, segment, and subjective category are plotted in Figures 19, 20, 21, 22, and 23.

The Planning scenario was essentially a Baseline scenario with an added mental task load component. The Activity scenario was a Baseline scenario complicated by a great deal of manual control work. The Combined scenario was a combination of the Activity and Planning scenarios. Therefore, the construction of the scenarios and the results plotted in Figures 19 to 23 led us to investigate whether this construct was reflected in the subjective ratings.

For all five ratings, we found the incremental difference between the Baseline scenario and each of the other three scenarios. We then examined how the sum of these increments for the Activity and Planning scenarios compared with the incremental Combined ratings. For example, suppose that the Baseline rating for DIFFICULTY was 3.0 and the DIFFICULTY ratings for the Activity, Planning, and Combined scenarios were 5.0, 5.3, and 7.5 respectively. The incremental ratings for the Activity, Planning, and Combined ratings would then be 2.0, 2.3, and 4.5. The sum of the Activity and Planning scenario increments would be 4.3. This increment (averaged with the increments for all the other pilot's increments) was compared with the Combined scenario's increment of 4.5 (averaged with the other pilot's Combined scenario increments).

For all five subjective ratings, the sums of the Activity and Planning increments were not statistically different from the incremental Combined ratings.

In view of the well established fact that the magnitude of subjective perception is logarithmically related to stimulus magnitude, this nearly linear response was somewhat surprising. At no point were the pilots ever told that the Combined scenario contained the sum of manual and mental tasks from the Activity and Planning scenarios. However, although this result may be useful when going from low or moderate workloads to high workloads, this linearity must obviously break down when trying to go from high workloads to even greater workloads.

How difficult did the pilots think the three non-Baseline scenarios were?

SCENARIO	SEGMENT			Overall	Std Dev
	I	II	III		
BASELINE					
Activity Level	2.6	2.8	3.5	3.0	0.9
Complexity	2.3	2.5	3.4	2.7	1.0
Difficulty	2.2	2.4	3.1	2.6	0.8
Stress	2.0	2.1	3.0	2.4	0.7
Workload	1.8	2.2	2.8	2.3	0.5
ACTIVITY					
Activity Level	5.4	6.7	7.3	6.5	1.2
Complexity	3.4	5.0	5.7	4.7	1.3
Difficulty	4.5	6.0	6.7	5.7	1.1
Stress	3.7	4.9	6.1	4.9	1.1
Workload	3.9	5.5	7.0	5.5	1.4
PLANNING					
Activity Level	4.1	5.1	7.0	5.4	1.4
Complexity	4.1	4.6	5.9	4.8	1.3
Difficulty	3.3	4.0	6.3	4.6	1.1
Stress	3.3	3.9	5.3	4.2	1.2
Workload	3.9	4.7	6.2	4.9	1.2
COMBINED					
Activity Level	5.9	8.3	9.8	8.0	1.1
Complexity	5.4	6.9	8.5	6.9	1.6
Difficulty	5.9	7.8	9.1	7.6	1.7
Stress	5.5	7.6	8.9	7.3	1.3
Workload	5.7	7.7	9.6	7.7	1.6

Figure 18: Average Subjective Ratings for each Segment
(Adjusted)
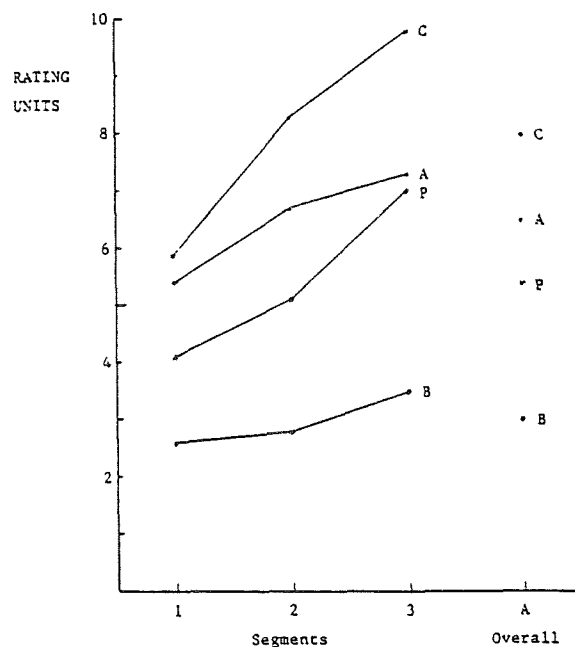
Figure 19: Average subjective ACTIVITY LEVEL ratings for the
Baseline (B), Activity (A), Planning (P),
and Combined (C) scenarios
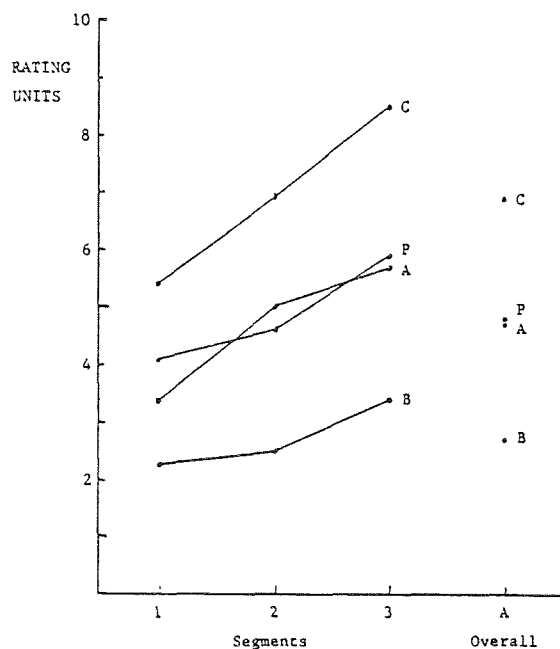
Figure 20: Average subjective COMPLEXITY ratings for the
Baseline (B), Activity (A), Planning (P),
and Combined (C) scenarios
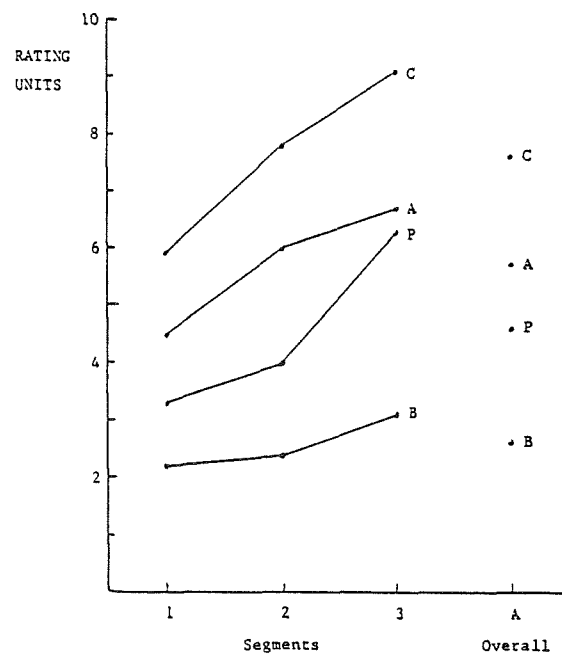
Figure 21: Average subjective DIFFICULTY ratings for the
Baseline (B), Activity (A), Planning (P),
and Combined (C) scenarios

6.16
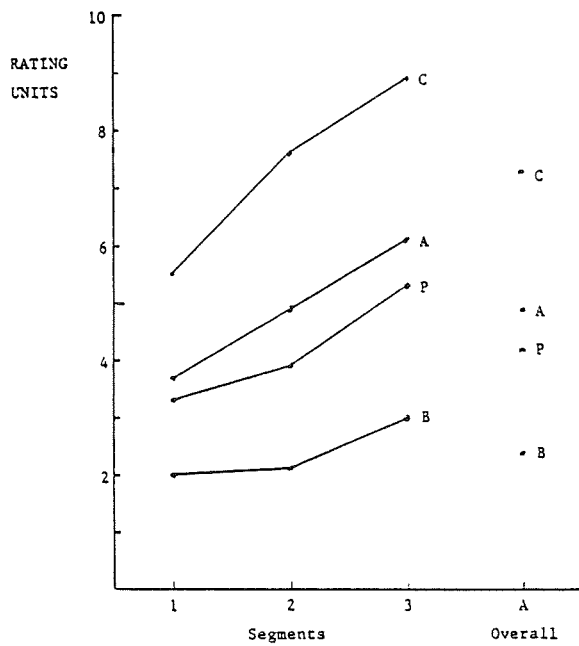
Figure 22: Average subjective STRESS ratings for the
Baseline (B), Activity (A), Planning (P),
and Combined (C) scenarios
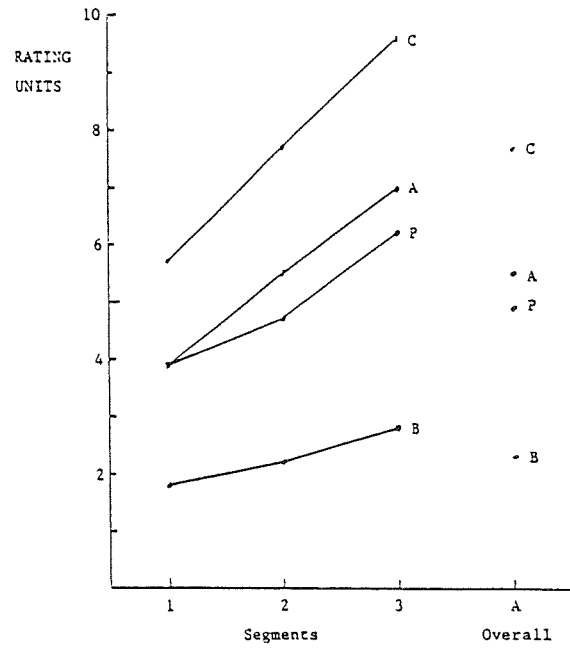
Figure 23: Average subjective WORKLOAD ratings for the
Baseline (B), Activity (A), Planning (P),
and Combined (C) scenarios

SCENARIO	Activity		Planning		Combined	
DEVIATION TYPE	mean	rms	mean	rms	mean	rms
Activity Level	.401	.805	.880	.782	.986	.953
Complexity	.389	.797	.843	.777	.999	.896
Difficulty	.403	.807	.817	.746	.990	.945
Stress	.583	.911	.428	.792	.986	.954
Workload	.568	.903	.882	.823	.999	.911

Figure 24: Pearson Product-Moment Correlation Coefficient
for aggregate Altitude Deviations and
Subjective Ratings

		SEGMENT	
	I	II	III
Activity Level	5.8	7.4	9.6
Complexity	6.5	6.8	8.3
Difficulty	4.5	4.1	11.0
Stress	1.1	3.0	3.1
Workload	5.5	5.6	10.0
Altitude Error: Mean	11.9	59.8	110.6
RMS	12.7	60.3	110.6
Airspeed Error: Mean	1.2	4.1	7.3
RMS	2.1	4.2	7.3

Figure 25: Example of related performance deterioration and
subjective saturation: Pilot C; Planning Scenario

The only scenario which consistently "saturated" pilots was the Combined scenario. If one defines a "saturated" pilot as one who scores a subjective rating category at 9.0 or higher, the Activity scenario was least likely to saturate pilots. This is interesting because when there were significant differences between the Activity and Planning scenario ratings, the Activity scenario rating was always slightly higher. Thus, certain individuals found the Planning scenario very difficult, while the pilots as a group, found the Planning scenario slightly less demanding than the Activity scenario.

For the Activity scenario, there was one saturated rating for WORKLOAD. For the Planning scenario, there were two saturated ratings for ACTIVITY LEVEL, and one each for DIFFICULTY and WORKLOAD. For the Combined scenario, there were five saturated ratings for ACTIVITY LEVEL and WORKLOAD, four for DIFFICULTY and STRESS, and two for COMPLEXITY.

These experiments verified that on a subjective level, a difficult, purely mental task load can equal a difficult, purely manual task load. In general, all the subjective category ratings were similar for the Planning and Activity scenarios.

There was no consistent correlation between subjective ratings and a pilot's experience level. This is not surprising since there is no universal subjective mental metric. Two persons working equally hard may rate their workloads very differently. They have different utilities, and one person may use a linear scale while another uses a logarithmic, and still another, an exponential scale.

Objective activity performance versus subjective ratings
We looked for a correlation in altitude or airspeed deviations with each pilot's subjective ratings. On an individual basis, objective activity performance data and subjective ratings were uncorrelated. This result was not unexpected, and had been reported previously. See, for example, the short discussion in Kantowitz, Hart, and Bortolussi, 1983.

Nevertheless, in the aggregate, objective performance data was correlated with subjective ratings. Using Pearson's Product-Moment Correlation Coefficient, "r", rms altitude errors weakly correlated with the corresponding subjective ratings for the Activity scenario (See Figure 24). ACTIVITY LEVEL, COMPLEXITY, and DIFFICULTY correlated with an "r" of 0.8 (.805; .797; .807). For the STRESS and WORKLOAD ratings, "r" was about 0.9 (.911; .903).

Correlations were slightly better for the Planning scenario. Mean absolute altitude deviations and ACTIVITY LEVEL had an "r" of .880. COMPLEXITY, DIFFICULTY and WORKLOAD had "r's" of .843, .817, and .882. Mean altitude errors did not correlate with STRESS, but rms errors did: .792. The ability of the rms error data to correlate with STRESS ratings better than the mean deviation data did might be due to the fact that the rms data weights large errors more heavily than small errors. Intuitively, beyond a certain point, stress should be an exponential function of the magnitude of deviations. Thus, large deviations would be better reflected in the rms values and

STRESS ratings.

There was excellent correlation between mean absolute error data and all five ratings for the Combined scenario. The lowest "r" was for STRESS, (.986) with COMPLEXITY having an "r" of .9999. Because the pilots were heavily loaded during the Combined scenario, they may have been operating near their personal limits. This may have lessened differences in proficiency resulting in the good correlation between objective performance data and the subjective ratings.

Tulga and Sheridan, 1980, reported that once a subject passed "saturation", performance deteriorated sharply. While flying the Planning scenario, Pilot C crashed during Segment III. Figure 25 lists relevant data for Segments I, II, and III for this pilot. Although he reported only low STRESS, the other four subjective factors sharply increased from Segment II to Segment III. Likewise, note that his mean absolute and rms altitude errors increased by 85 percent and 83 percent, and the corresponding airspeed errors increased by 78 percent and 74 percent from Segment II to Segment III. Although one can argue about which was cause and which was effect, mental saturation accompanied a severe performance degradation.

Planning/memory task performance
As workload increased, there were a number of ways that each pilot could respond to these requests for some action at a future time. They could fail to perform a task, choosing not to do it or simply forgetting to do it. They could also perform the task incorrectly, do some unrequested task, or perform the required task at some time other than the directed time. Overall planning task error percentages for each scenario are plotted in Figure 26.

Although the planning task load for the Baseline and Activity scenarios was the same, the overall error percentage was much higher for the Activity scenario. Similarly, although the Planning and Combined scenarios had similar planning task loads, the Combined scenario percentage was much higher (and differed at a 99 percent confidence level). The Planning and Activity scenarios had similar Subjective ratings, but their mental task performance data was very different. A high manual workload had a profound effect, increasing errors.

The standard deviations for the overall error percentages varied widely from scenario to scenario. For the Baseline and Planning scenarios where the error percentages were low, standard deviations were only 8.8 and 13.4 percent respectively. The difficult Combined scenario had a standard deviation of 27.2 percent, indicating more variability among the pilots. The Activity scenario showed the greatest variability. The low number of mental tasks and the high error percentages for some pilots resulted in a standard deviation of 51.4.

Figure 27 illustrates the error percentages for each segment and scenario. The performance for the Planning and Combined scenarios was virtually identical for Segment I. However, for Segments II and III, the difference
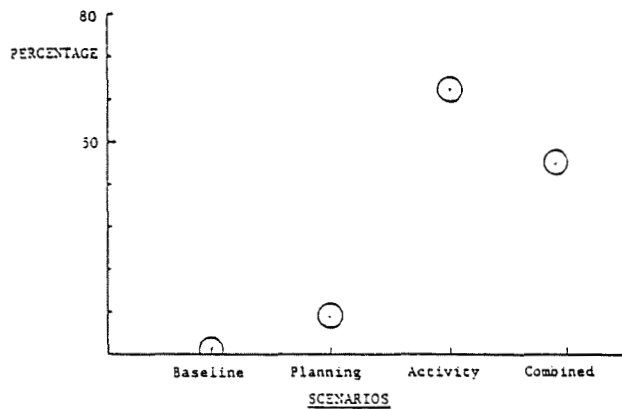
Figure 26: Overall percentage of planning/memory task errors
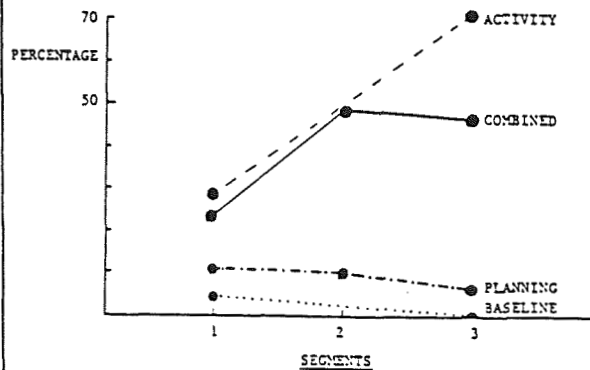
Figure 27: Percentages of planning/memory task errors per segment
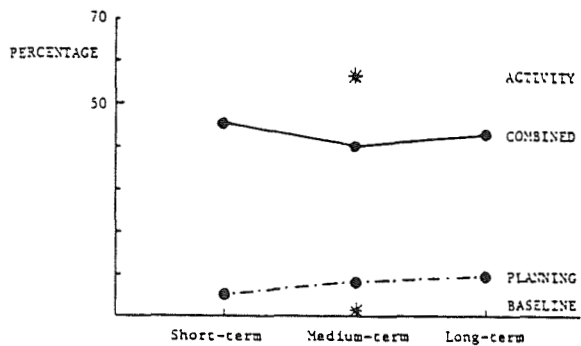
Figure 28: Error percentages for three different planning task time spans
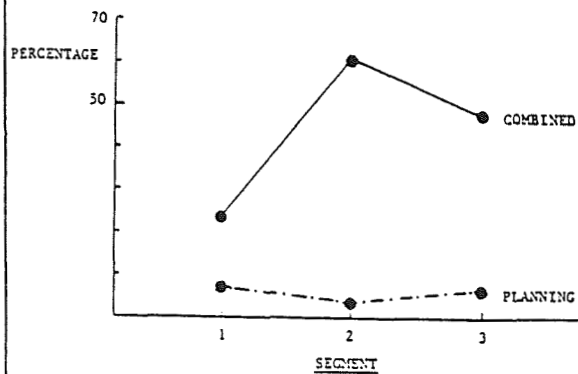
Figure 29: Short-term planning/memory task error percentages by segment

between the two scenarios was significant at the 99.9 percent confidence
level. Although individual performance differed a great deal, the data
suggests that at low or moderate levels, manual control workload does not
affect mental performance. Sufficient cognitive reserve exists to handle
all tasks. However, at relatively high manual control levels, cognitive
reserves disappear and mental performance deteriorates. Figure 26 suggests
that this mental deterioration may even be evident for low levels of mental
tasking, such as in the Activity scenario.

The various planning tasks were categorized as Long-term, Medium-term, or
Short-term based upon the length of time the pilot had from receiving the
task assignment to performing it. When aggregated for each scenario, the
data yields the plot shown in Figure 28. Analyzing the error percentages
for each scenario, there was no statistically significant difference between
the three different task time spans. This was probably because the pilots
were allowed to take notes. Additional errors probably arose in the
Short-term tasks when the pilots struggled to plan and perform these tasks
in a very busy environment. Thus, they would miss some tasks or perform
them late. This balanced the errors engendered in the Long-term tasks by
the pilots forgetting about tasks.

An analysis of the data supports this hypothesis. There were no Long-term
planning errors due to performing an action at the wrong time. However, 33
percent of the Short-term and 53 percent of the Medium-term errors were due
to performing an action at the wrong time.

Planning task errors for all three time spans were affected by
manual-control activity. Note in Figure 28 that the two low manual workload
scenarios (Baseline and Planning) had low error percentages while both high
manual workload scenarios (Activity and Combined) had high error
percentages. The Activity scenario had a high error percentage even though
its planning task load was low.

Looking only at the two scenarios (Planning and Combined) with a high
planning task load, the differences between the scenarios was statistically
significant for all three time spans. Differences were significant at an 80
percent confidence level for medium-length tasks, at a 95 percent level for
long-term tasks, and 98 percent level for short-term tasks. Thus, the level
of manual control was again decisive in determining mental performance. The
data was too coarse and individual pilot performance was too variable to
make standard deviation data useful.

Only the Planning and Combined scenarios had Short-term planning tasks.
Examining Figure 29, differences between the Planning and Combined scenarios
for Short-term planning tasks were not statistically significant in
Segment I. However, the differences were at a 98 percent confidence level
for Segments II and III, when workloads were higher.

All four scenarios had Medium-term planning tasks. Looking at Figure 30,
there was no statistically significant difference between the scenarios in
Segments I or II. However, in Segment III, the highest workload segment,

6.21
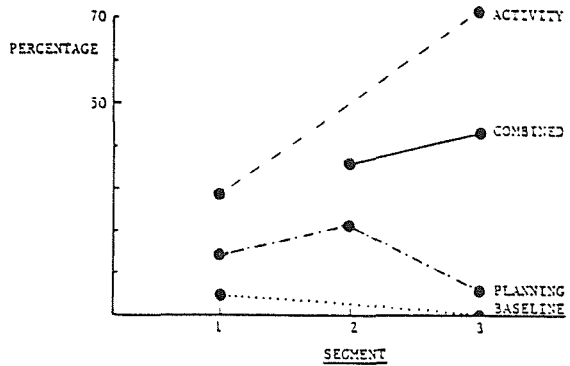
Figure 30: Medium-term planning/memory task error percentages by segment
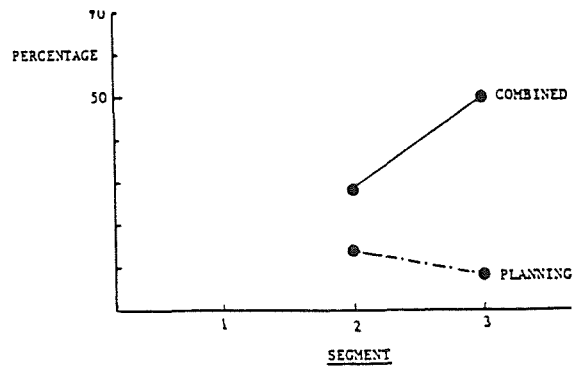
Figure 31: Long-term planning/memory task errror percentages by segment

the Combined scenario errors were higher than the Planning scenario errors (90 percent confidence level). The Planning and Activity difference was even greater (at a 95 percent confidence level). The Activity and the Combined scenarios, and the Planning and Baseline scenarios were statistically similar. Once again, at high overall workload levels, the presence of a high manual task load made a significant difference.

Figure 31 is a plot of the Long-term planning task results. In Segment II, the Planning and Combined scenarios were statistically indestinguishable. However, at the higher workload level of Segment III, the error percentage for the Combined scenario was clearly greater (90 percent confidence level).

The Activity and Planning scenarios had moderate manual or mental workloads, respectively. At these levels, error percentages were similar for all of the pilots. However, some differences arose in the high workload Combined scenario. The low experience pilots averaged 14.0 task errors while the high experience pilots averaged 7.3 task errors. Thus, there were signs of experience related saturation in this mental performance data which was much less obvious in the objective performance data and subjective rating data. This difference was verified at a 95 percent confidence level.

The number of individual planning errors and individual altitude or airspeed deviations were not correlated. Nor were planning errors and subjective ratings. However, in the aggregate, altitude and airspeed deviations, subjective ratings, and the number of planning errors all increased with increasing task loads.

Pilot comments
The planning task instructions given to the pilots were seldom in chronological order. This was done to make the planning function more difficult. This strategy apparently worked, since several subjects mentioned that instructions "mixed in time" were difficult to organize.

Some pilots considered the autopilot a hindrance while others found it a useful aid. Several pilots stated that when things "really got busy", the autopilot was the only thing which kept workload at a manageable level. But, several pilots reported that having to plan how to use the autopilot was worse than the demanding manual control work. An oft-reported result is once again clear: if the initial set-up or programming of a "pilot aid" is difficult or unduly time consuming, pilots will use manual procedures and avoid its use.

A number of the pilots stated that planning and memory items tended to get second priority to immediate task demands. This is consistent with the finding that a high activity workload significantly increased planning task errors. Pilots were obeying the prime directive taught every student pilot: "First, fly the aircraft!" These statements and results are also consistent with Tulga and Sheridan's (1980) finding that subjects don't plan ahead when they're very busy.

Finally, the pilots mentioned four items which increased their mental stress and workload. One was the "annoyance" factor caused by having too many things to do or by being interrupted before completing a task. This type of problem is common on final approach when the need to fly and/or monitor equipment, clear for other aircraft, look for the runway, interact with ATC, and run aircraft checklists, combine to make the flight deck a busy, stressful environment. A second item was the effect of "getting behind". Again, this is most likely to occur when things get very busy. The stress generated by a lengthening "mental queue", combined with the possible need to modify a former plan, increases the perceived workload. Similarly, abnormal events significantly increase workload, disrupt concentration, and increase the frustration level. These effects have been discussed in the open literature. See, for example, Hart and Bortolussi (1983), Jensen and Chappell (1983), and Tanaka, Buharali, and Sheridan (1983). The fourth item concerned the effect of adding an increment of workload when the workload is already high. As the pilot becomes task saturated, additional tasks must be prioritized, added to a mental queue, or ignored. This increases stress, frustrates the pilot, and increases his mental manipulations. These factors result in lower performance, increased mental workload, and lower safety margins.

IX. FINDINGS AND CONCLUSIONS

1. The number of assigned mental tasks had no statistically significant impact on the degree of aircraft control. The level of manual workload was the decisive factor. When mental tasking was high but manual tasking was at a low level, altitude and airspeed deviations were small. When mental tasking was low but manual tasking was high, altitude and airspeed deviations were large. The level of mental activity affected aircraft control only when mental workload reached "critical" levels.

2. Incremental subjective ratings were calculated relative to the ratings for a Baseline scenario. The incremental rating for a high manual workload scenario added to the incremental rating for a high mental workload scenario was equal to the incremental rating for a scenario which combined both types of workloads.

3. Subjective ratings given by individual pilots during the high manual tasking scenario were very similar. However, there were individual differences in the subjective ratings for the high mental tasking scenario. Some pilots were not stressed by the mental tasks while others significantly increased their subjective ratings. Subjective ratings were more sensitive than aircraft deviation measures in indicating individual mental workloads.

4. At low or moderate levels of manual and mental task loads, aircraft deviations and memory task performance did not correlate with the subjective ratings. At high workload levels, the correlation was very good. It's possible that at lower task loads, there is reserve mental capacity which varies from pilot to pilot, affecting performance and ratings. At high workload levels, all pilots may be tapping most or all of their mental

capacity, resulting in much greater consistency between performance and the subjective ratings.

5. The magnitude of manual task loads was decisive in determining the ability of the pilots to handle mental tasks. A mentally difficult, manually easy scenario resulted in a low percentage of mental errors. A mentally easy, manually difficult scenario resulted in a high percentage of mental errors. The manual activity was presumably consuming a great deal of the pilots' mental processing capacity, even when they were not aware of it. This finding was equally valid for long-term, medium-term, and short-term mental tasks. Thus, pilots flying a highly automated flight control system might be able to more easily handle high mental workloads.

6. Under conditions of high manual and mental workload, the low experience pilots did not perform mental tasks as well as the high experience pilots did. However, objective aircraft performance and subjective ratings were similar for the two groups. Thus, these experiments suggest that monitoring and measuring mental performance might be a more sensitive indicator of mental workload and reserve mental capacity than objective aircraft performance data or subjective ratings.

X. RECOMMENDATIONS FOR FOLLOW-UP STUDIES

1. In future studies of this type or in a re-examination of this study, it might be enlightening to "filter" the data by only considering altitude deviations greater than $\pm$ 50 or $\pm$ 100 feet, or airspeed errors greater than $\pm$ 5 or $\pm$ 10 knots. This might compensate for individual pilots' tolerance boundaries.

2. Subjective Ratings should be used in future studies of mental workload. They provide a useful, if imprecise, measure of the pilot's mental state.

3. The only significant difference found between the low experience and high experience pilots was in their performance of mental planning tasks. This should be further investigated in future studies.

XI. REFERENCES

1.   Berg, S. L.; and Sheridan, T. B.. Measuring workload differences
        between short-term memory and long-term memory scenarios in a
        simulated flight environment. Proceedings of the twentieth annual
        conference on manual control; 1984 June 12-14: 397-416.

2.   Berg, S. L.; and Sheridan, T. B.. Effects of time span and task load on
        pilot mental workload. Cambridge, Massachusetts: Massachusetts
        Institute of Technology; 1985. Master's Thesis.

3. Hart, S. G.; Bortolussi, M. R.. Pilot errors as a source of workload. Paper presented at the second symposium on aviation psychology; 1983 April 25-27; Columbus, Ohio.

4. Jensen, R. S.; Chappell, S.. Pilot performance and workload assessment: an analysis of pilot errors. Report submitted to NASA Ames Research Center in 1983 February.

5. Johannsen. G.. Workload and Workload Measurement. N. Moray, ed. Mental workload; theory and measurement. New York: Plenum; 1979.

6. Kantowitz, B. H.; Hart, S. G.; Bortolussi, M. R.. Measuring pilot workload in a motion-based simulator: asynchronous secondary choice-reaction task. Paper submitted to the IEEE Transactions on Systems, Man, and Cybernetics. 1983.

7. Katz, J. G.. Pilot workload in the air transport environment: measurement, theory, and the influence of Air Traffic Control. Massachusetts Institute of Technology Flight Transportation Laboratory. FTL Report R80-3. 1980 May.

8. Leplat, J.. Factors determining workload. Ergonomics. 21: 143-149; 1978.

9. Miller, G. A.. The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychological Review. 63: 81-97; 1956.

10. Rehmann, J. T.; Stein, E. S.; Rosenberg, B. L.. Subjective pilot workload assessment. Human Factors. 25(3): 297-307; 1983.

11. Sheridan, T. B.; Simpson, R. W.. Toward the definition and measurement of the mental workload of transport pilots. Massachusetts Institute of Technology Flight Transportation and Man-Machine Laboratories, Technical Report No. DOT-OS-70055. 1979 January.

12. Tanaka, K.; Buharali, A.; Sheridan, T. B.. Mental workload in supervisory control of automated aircraft. Proceedings of the nineteenth annual conference on manual control; 1983 May 23-25: 40-58.

13. Tulga, M. K.; Sheridan, T. B.. Dynamic decisions and workload in multitask supervisory control. IEEE Transactions on Systems, Man, and Cybernetics. SMC-10 (No. 5): 217-232; 1980.

14. Walden, R. S.; Rouse, W. B.. A queueing model of pilot decision making in a multi-task flight management situation. IEEE Transactions on Systems, Man, and Cybernetics. SMC-8 (No. 12): 867-875; 1978.

15. Williges, R. C.; Wierwille, W. W.. Behavioral measures of aircrew mental workload. Human Factors. 21(5): 549-574; 1979.

Memory and Subjective Workload Assessment

Lowell Staveland - San Jose State University Foundation, San Jose, CA
Sandra Hart - NASA-Ames Research Center, Moffett Field, CA
Yei-Yu Yeh - University of Illinois, Urbana-Champagne, ILL

Abstract

Recent research suggested subjective introspection of workload is not based upon specific retrieval of information from long-term memory, and only reflects the average workload that is imposed upon the human operator by a particular task. These findings are based upon global ratings of workload for the overall task, suggesting that subjective ratings are limited in ability to retrieve specific details of a task from long-term memory. To clarify the limits memory imposes on subjective workload assessment, the difficulty of task segments was varied and the workload of specified segments was retrospectively rated. The ratings were retrospectively collected on the manipulations of three levels of segment difficulty. Subjects were assigned to one of two memory groups. In the Before group, subjects knew before performing a block of trials which segment to rate. In the After group, subjects did not know which segment to rate until after performing the block of trials. The subjective ratings, RTs, and MTs were compared for within group, and between group differences. Performance measures and subjective evaluations of workload reflected the experimental manipulations. Subjects were sensitive to different difficulty levels, and recalled the average workload of task components. Cueing did not appear to help recall, and memory group differences possibly reflected variations in the groups of subjects, or an additional memory task.

Introduction

Much attention is being focused on the utility of subjective evaluations to measure mental workload and human performance. The potential for subjective ratings to reflect a human operators sensitivity to varying task demands, has been validated in several experiments (Yeh, Wickens & Hart 1985; Hart, Sellers, & Guthart, 1984; Arbak, Shew, & Simons 1984). These findings, however, are based on global ratings of workload for a group of similar tasks, or segments of a continuously changing task (Bortolussi, Kantowitz, Hart, 1985), which measure the overall loading on cognitive processes, irregardless of when they were obtained. Global ratings obtained while performing a task are highly correlated with the global ratings obtained retrospectively (Bortolussi et al, 1985), even though they may not reflect moment-to-moment variations in cognitive loads that operators experience while performing a task. Yeh et al, (1984) found that "...subjective introspection of workload is not based on specific retrieval of information from working memory and only reflects the average workload imposed on human operators by a particular task".

The tasks selected for their study were based on the 'Fittsberg' paradigm (Hartzell et al) which was originally based on the serial combination of FITTS target aquisition tasks following selection among the alternative locations based on a STERNberg memory search decision. For this application, two response selection tasks were used: pattern match and arithmetic equations. For each response selection task and target aquisition task, three levels of difficulty were imposed. Difficulty levels of the two task components were

consistent within a block of trials, and either both were increased or decreased in difficulty, or the difficulty of one component was increased while the other decreased. Measures of performance independently reflected task difficulty manipulations within trial blocks; RT varied with RS difficulty, whereas MT varied with RE difficulty. Workload ratings accurately reflected the integrated workload of all tasks within a block, displaying no primacy/recency effect, or greater influence by one task component than another. Since ratings were consistently equal to the average workload of a blck of trials, the question remained whether subjects were simply insensitive to task manipulations, or in fact accomplished the summary evaluation that was required by the design of the experiment. In either case, it was not clear whether subjects would have been able to provide more selective evaluations of trial block segments had they been required to do so. Such global ratings are fine where the goal is to evaluate differences between tasks (e.g. comparing the difficulty of one flight to another). In many circumstances though, the difficulty of specific segments within a flight need to be evaluated. In this case global ratings do not suffice. More detailed evaluations are required to reflect the varying difficulty levels experienced by operators during a flight.

Previous research suggested that delaying retrospective evaluations of task segments does not significantly alter the relationships among reflective ratings, even though the absolute values might be somewhat different (Eggemeier, Melville, & Crabtree 1984; Notestine 1984). Even interevening task performance does not significantly effect workload ratings (Eggemeier, et al 1984). These results have direct implications for this study, considering subjects had to reflectively rate different segments of a task after a block of segments. If a subject is asked to rate the first segment out of three in a block of trials, the intervening segments should not significantly effect their retrospective rating. This means the workload ratings obtained in this study should reflect specific retrieval of a particular segment from long-term memory, independent of the other segments influence on ratings. Delays in rating the first or second segments while performing the second or third segments also should not influence subjective experience of workload. This rules out delay as a confounding variable, and increases the confidence in the obtained ratings as being indicative of an operators workload and cognitive loading for a particular segment.

The current study addressed the limits memory imposed on subjective ratings. Subjects were divided into two memory groups: Before and After. Subjects in the Before group knew in advance the segment-to-be-rated. Subjects in the After group did not know in advance the segment-to-be-rated, they were told after completing the block of trials which segment to rate. The purpose was to elicit answers to the following questions: (1) How sensitive are subjects to task component manipulations? (2) Is the information about different segments in a task available retrospectively? Or is the average workload all that can be recalled (3) Does knowing in advance the segment-to-be-rated aid recall? And (4) Do all task components contribute equally to workload? This experiment follows up Yehs findings that subjective ratings are limited in their capacity to retrieve specific details from working memory.

The task selected for this experiment was based on a version of the Fittsberg paradigm used by Yeh et al (1985), and Hartzell et al, (1983). It involved two components: response selection and response execution. The response selection component was based on completing arithmetic equations. As

the equations complexity increased from one operator to three, difficulty increased as well. The response execution component was a target acquisition task based on Fitts law (Fitts & Petersen, 1964). Its difficulty was manipulated by varying the targets index of difficulty (ID). The two components were combined to form three categories: <u>Consistent:</u> The RS/RE components had a consistent difficulty level across the three segments within a conditon; (2) <u>Changing-consistent:</u> RS/RE components difficulty levels were positively correlated, either increasing or decreasing in difficulty from segment to segment within a condition; and (3) <u>Changing-inconsistent:</u> RS/RE components difficulty levels were negatively correlated (the RS component increased while the RE component decreased, or vice-versa). Cognitive loading was expected to vary as a function of the response selection component, whereas response execution would influence MTs. Workload ratings were expected to vary as a joint function of the difficulty levels of both components within each trial-block segment.

Method

Subjects

Eighteen male and two female subjects served as paid volunteers. None had any prior experience with Fitts tasks, but all had served as subjects in other experiments at NASA-Ames Research Center. Thus, most had experience with the use of the bipolar rating scales. All subjects had competent arithmetic skills.

Apparatus

The experiment was conducted in a sound-attenuated chamber. The subject was seated in a chair located 85 cm from a 23-cm monitor where all experimental tasks were displayed. The visual angle subtended by the most extreme targets was 11 deg. A two-axis joystick was mounted on the right arm of the chair for response selection and target aquisition responses. Subjective ratings were entered with a slide pot and button mounted on the left arm of the chair. The experiment, data acquisition, and reduction were performed with an Apple II+ microcomputer, modified to allow rapid recording of response (10 msec resolution). The data were analyzed with a Dec 11/70, and a Vax 11/750.

Task Components

Each task had two components: response selection and response execution. The outcome of the response selection task served as input to the response execution task. Thus, the two task components could be performed serially and were functionally related. There were three levels of difficulty for each component: easy (E), medium (M), and hard (H). The two components were combined to form seven conditions: EE, MM, HH, II, DD, ID, DI. The first letter of each pair represents the response selection component, and the second letter for the respomse execution component. 'I' indicates that the difficulty of that component was increased from the beginning to the end of that trial block; 'D' indicates that it decreased.

<u>Response Selection</u> The solution to an equation performed mentally

determined the direction of movement. Each equation involved one, two or three mathematical operations which determined the level of difficulty. The easy condition required one operation, (e.g. 2+3.), medium required two (e.g. 3*2/1), and hard required three (e.g. (4-1)*3). The solutions were always whole numbers, either greater or less than a single digit memory set presented prior to each block of trials. These were similar to three of the RS tasks employed in the previous study (Yeh et al, 1985). Subjects were told to move the joystick right if the solution was greater than the remembered digit (7, 8, or 9), or left if it was less. The interval between stimulus onset and a 2% joystick deflection was recorded as reaction time (RT).

Response execution. The response execution component was a target aquisition task. Two identical target areas were displayed symmetrically on either side of the stimulus at a distance determined by the index of difficulty computed according to Fitts law (ID=log2(2A/W)). The targets were two 1.25 cm lines separated by a distance appropriate for the ID of that condition. The same ID levels used in earlier studies were selected for the three levels of difficulty: Easy = 2.52, Medium = 4.19, and Hard = 5.67. The interval between a 2% joystick deflection and satisfaction of the steadiness criterion for keeping the cursor within the target, was recorded as movement time (MT).

Condition Characteristics

Each of the seven experimental blocks of trials (EE, MM, HH, II, DD, ID, DI) were divided into three equal segments of twelve trials each. The eight equations within a segment had the same difficulty level as the eight IDs, but the difficulty levels from one segment to the next depended on the condition. For EE, MM, and HH conditions, all three segments within a block had the same response selection and target aquisition difficulty levels (consistent). For two other conditions (changing-consistent), the difficulty of both components either increased (II) or decreased (DD). For the last two conditions, (changing-inconsistent), the difficulty of the two components, (ID, and DI), changed in opposite directions. The six equations that transitioned between segments were randomly mixed so that the divisions between segments was less evident. Capture time (RT+MT), was the total response time for each trial, averaged across all trials, and was presented as feedback at the end of each condition along with the number of correct responses.

Subjective Ratings

Two types of ratings were collected in this study:

(1) Individual differences in definition. The relative importance of nine factors to each subject's definition of mental workload was determined. These nine factors were: task difficulty, time pressure, own performance, physical effort, mental effort, frustration, stress, fatigue, and activity type (Yeh et al, 1985). Each factor was paired with every other factor (36 pairs) in a pretest. Subjects, selected the member of each pair that was most related to their definition of workload. Each factor could be selected from 0 (never considered relevant) to 8 (more important than any other factor) times. The number of times a factor was selected was its weight.

(2) Bipolar ratings. Ratings on nine bipolar rating scales plus an overall workload scale were collected at the end of each condition. Each scale was presented on the experimental display as an 11 cm vertical line with a title (e.g., "OVERALL WORKLOAD) and bipolar descriptions at each end (e.g., "EXTREMELY HIGH/EXTREMELY LOW"). The cursor was positioned at the desired point on the scale with a slide pot, and entered with a button. Each selection was assigned a value from 1 to 100 during data reduction.

Procedure

Each subject participated in the experiment two hrs per day, for three days. The first day, and the first 30 min on subsequent days were used for practice.

The subjects read a brief explanation of the experiment to familiarize themselves with the objectives and experimental tasks. After the workload weights were collected, the subjects practiced the target aquisition task: 20 blocks of 24 trials each.The basic response execution task entailed acquiring a target displayed on either the right or left side of the display; there was no response selection task. Following this, they performed the three difficulty levels of the response execution task (E,M,H), the response selection task (E,M,H): no targets were displayed, and the combined tasks (E,M,H). The response selection task entailed solving an equation, and moving the joystick right if the solution was greater than the remembered digit, or left if the solution was less. The practice trials at the beginning of each subsequent day were combined tasks involving changing-consistent (II,DD), and changing-inconsistent (ID,DI) conditions.

Each of the seven conditions were presented three times, so subjects could rate the workload of the first twelve trials after one block, the second twelve trials after another, and the third twelve after the third block. Subjects in the before group were told the segment-to-be-rated before performing each block of 36 trials. Subjects in the after group were told the segment-to-be-rated after performing each block of 36 trials. A total of 21 experimental conditons were rated. The segments-to-be-rated were presented to each subject in counterbalanced order, and the seven different conditions were presented in random order.

Results

General Comparison of Memory Groups

ANOVAs of mean RTs and MTs, percent correct, and bipolar ratings were collected for each of the three segments for the seven conditions in the three categories: consistent, changing-consistent, and changing-inconsistent. As shown in Figure 1a, the RTs for the Before group were less than for the RTs of the After group. RTs reflected the response selection difficulty, and were not affected by response execution difficulty. MTs for the Before group

Figure 1a. RT-Before vs After for all conditions.
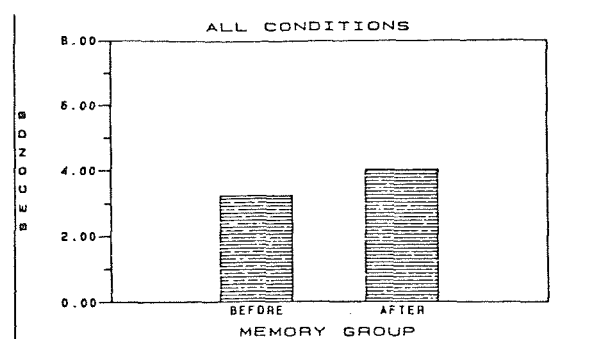
were greater than the After group, and reflected response execution difficulty, but did not reflect response selection difficulty (Figure 1b). The MT, and RT results were consistent across all conditions for both experiments. RTs were always greater than MTs. The average levels of workload ratings were similar for the two groups. However, differences in response to experimental manipulations were observed.

Percent Correct

There were no significant speed-accuracy trade-offs. In the consistent condition, there was a trend for both speed and accuracy to decrease, as the difficulty increased from conditions 'EE' to 'MM' to 'HH'. For the changing-consistent, and changing-inconsistent conditions, this trend is not apparent between conditions, or between segments. Overall, the subjects were highly accurate across all conditions and segments, $F(1,9) = 534.03$, $p<.001$.

RTs and MTs.

The ANOVA results for the Before and After groups are presented in Figures 2a-2c, 3a-3c, and 4a-4c.

Consistent. RTs and MTs reflected the relevant RS or RE difficulty manipulations, (Figure 2a) . The Before RTs were less than the After ($F(1,486) = 27.95$, $p<.001$) (Figure 2b). The Before MTs were greater than After ($F(1,486) = 35.52$, $p<.001$), (Figure 2c).

Before group. RT increased as the math equations increased in complexity (EE to MM to HH) ($F(2,18) = 32.1$, $p<.001$), reflecting an increase in cognitive loading. MTs also reflected these results, increasing in duration as RE difficulty increased from (EE to MM to HH) ($F(2,18) = 68.51$, $p<.001$).

After group. The results followed the same pattern as the Before group. RTs increased as RS difficulty in creased across the three conditions (EE,MM,HH) ($F(2,18) = 87.88$, $p<.001$).

Figure 1b. MT-Before vs After for all conditions.
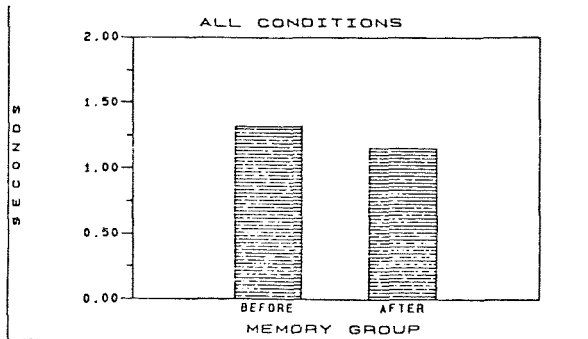
Figure 2a. Capture time-RT vs MT for consistent conditions.
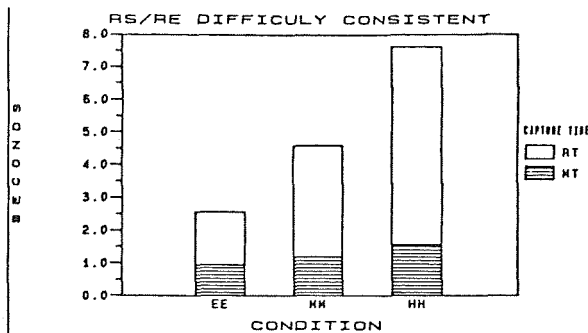
Figure 2b. RT-Before vs After for consistent conditions.
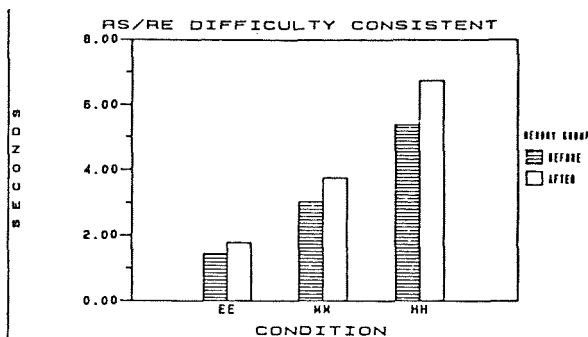
Figure 2c. MT-Before vs After fo consistent conditions.
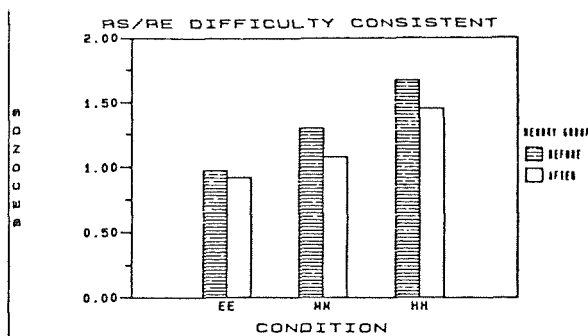
MTs increased as RE difficulty in-
creased (F(2,18) = 28.67, p<.001).

Changing-consistent. As the RS/RE com-
ponents increased in difficulty in the
'II' condition, and decreased in the
'DD' condition, RTs and MTs reflected
the changing difficulty levels (Figure
3a). Before RTs were less than After
RTs (F(1,324) = 22.32, p<.001), (Figure
3b), while their MTs were greater
(F(1,324) = 25.87, p<.001), (Figure
3c).

Before groups. For this group,
there was a significant interaction
between conditions (II,DD) and segment
for RT (F(2,18) = 43.84, p<.001). As RS
difficulty increased across segments
in the 'II' condition, and decreased in
the 'DD' condition, the RTs increased
or decreased respectively. MTs reflect-
ed the same interaction for the RE
component (F(2,18) = 52.16, p<.001).

After groups. There was a signifi-
cant interaction between conditions
(II,DD) and segment (F(2,18) = 62.76,
p<.001). As the RS difficulty in-
creased across segments in the 'II'
condition, and decreased in the 'DD'
condition, RT increased or decreased
respectively. Again, MT reflected the
same interaction in the RE component
(F(2,18) = 29.67, p<.001).

Changing-inconsistent. The difficulties
of the RS and RE components for the
'ID', and 'DI' were varied in opposite
directions. For the 'ID' condition, as
the RS component increased in difficul-
ty across segments within the condi-
tion, the RE component decreased in
difficulty. The converse was true for
the 'DI' condition. RT reflected the RS
manipulations and the MT reflected the
RE manipulations independently (Figure
4a). As in the previous two conditions,
Before RTs were less than After
(F(1,324) = 24.92, p<.001), (Figure
4b), while their MTs were greater
(F(1,324) = 28.89, p<.001), (Figure
4c).

Before group. There was a signifi-

Figure 3a. Capture time-RT vs MT for
changing-consistent conditions.
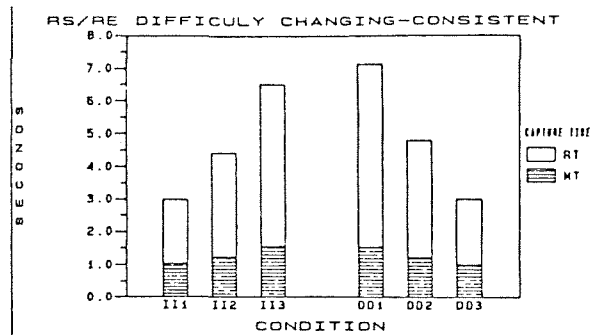
Figure 3b. RT-Before vs After for
changing-consistent conditions.
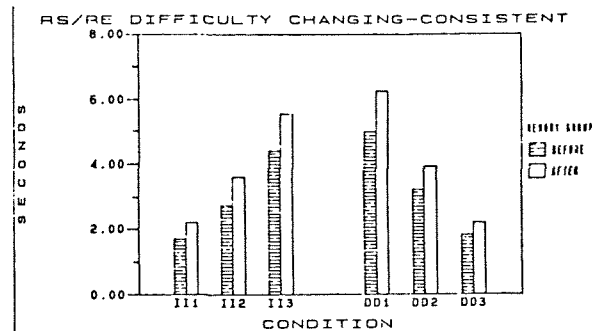
Figure 3c. MT-Before vs After for
changing-consistent conditions.
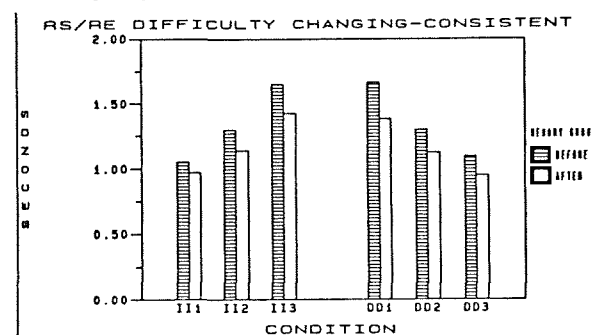
Figure 4a. Capture time-RT vs MT for
changing-inconsistent conditions.
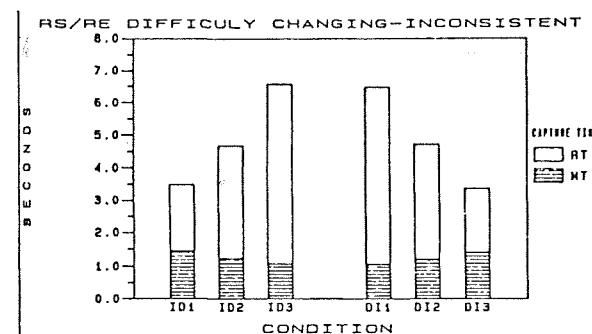
cant interaction between conditions (ID,DI), and segment. For the 'ID' condition, RTs increased as the RS component increased in difficulty, or decreased as the RS component decreased in difficulty $(F(2,18) = 36.60, p<.001)$. Conversely, MTs decreased as the RE component decreased in difficulty in the 'ID' condition, and increased in the 'DI' condition $(F(2,18) = 37.98, p<.001)$.

Figure 4b. RT-Before vs After for changing-inconsistent conditions.

Figure 4c. MT-Before vs After for changing-inconsistent conditions.

After group. The interaction between conditions and segment for RTs and MTs followed the same pattern as found in the Before group. RTs in the 'ID' and 'DI' conditions were inversely related $(F(2,18) = 104.74, p<.001)$, as were the MTs in the same two conditions $(F(2,18) = 17.13, p<.001)$.

Subjective Ratings

Relative importance of workload-related factors. There were large differences in the importance that subjects placed on the nine factors. Due to this variability in subject biases, there were no significant differences between memory groups in the relative importance each subject placed on the workload-related factors (Figure 5). These results follow widespread findings of variabilty in subjects biases, substantiating the importance of using weights to reduce between-subject variability in subjective evaluations of workload.

Figure 5. Relative importance of workload-related factors.

Weighted bipolar ratings Weighted bipolar ratings were weighted workload. Their means ranged from 19 to 49 for the Before group, and 8 to 50 for the After group. The workload involved in performing the 21 experimental conditions was evaluated at the end of each block of trials. These ratings were combined with the weights to calculate the weighted workload of the experimental tasks. This reduced between-subject variability by 32%. Once weighted workload was calculated, ANOVAs were conducted for the same three categories: (1) Consistent, (2) Changing-consistent, and (3) Chang ing-inconsistent. Separate ANOVAs were conducted for the Before and After groups. Weighted workload generally reflected the results obtained for the performance data.

Consistent (Figure 6). The Before group rated the RS/RE difficulty in the 'EE', 'MM', and 'HH' conditions as having significantly more workload than the After group did $(F(1,162) = 7.59, p<.01)$.

7.8

<u>Before</u> <u>group</u>. Workload increased
from the 'EE' to 'MM' to 'HH'
conditions as the RS/RE difficulty
increased (F(2,18) = 23.45, p<.001).
There was a small but significant ef-
fect between rated segments for the
'EE' condition (F(2,18) = 4.97, p<.05),
but there were no significant effects
between rated segments for the 'MM',
and 'HH' conditions.

<u>After</u> <u>group</u>. Workload increased
across conditions similarly to the
increase in the Before group (F(2,18) =
19.04, p<.001). Within the 'EE' condi-
tion, there was a significant effect
between rated segments (F(2,18) = 4.05,
p<.05), but there were no significant
effects between rated segments for the
'MM', and 'HH' conditions.

<u>Changing-consistent</u> (Figure 7). Sub-
jects in the Before group rated the
workload in the 'II' and 'DD' condi-
tions higher the After group did,
Figure 8. Weighted workload-Before but
the differences were not vs After for
changing-inconsistent significant.
conditions.

<u>Before</u> <u>group</u>. Workload ratings
increased across rated segments within
the 'II' condition (F(2,18) = 4.09,
p<.05), and decreased across conditions
within the 'DD' condition (F(2,18) =
5.79, p<.05).

<u>After</u> <u>group</u>. The results for the
After group parallel those of the Be-
fore group. Ratings increased across
rated segments within the 'II' condi-
tion (F(2,18) = 6.01, p<.05), and de-
creased across rated segments within
the 'DD' condition (F(2,18) = 3.07,
p<.05).

Figure 6. Weighted workload-Before
vs After for consistent conditions.
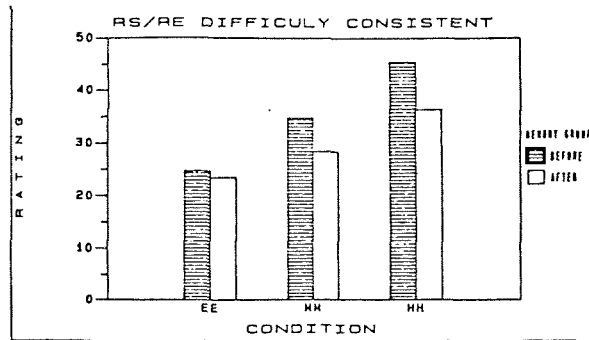
Figure 7. Weighted workload-Before
vs After for changing-consistent
conditions.
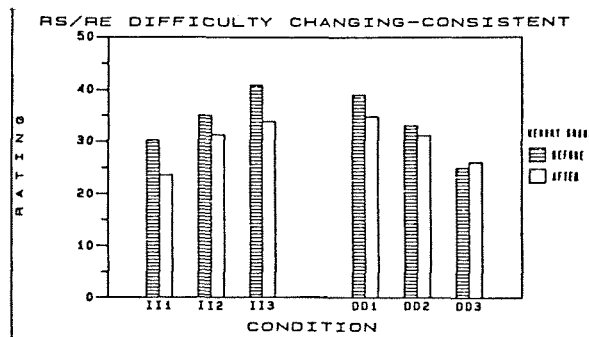
Figure 8. Weighted workload-Before
vs After for changing-inconsistent
conditions.

<u>Changing-inconsistent</u>(Figure 8). There was a significant difference in
workload ratings between groups for the 'ID', and 'DI' conditions. Across
segments, the Before groups ratings were greater (F(1,162) = 4.25, p<.05).

<u>Before</u> <u>group</u>. There were no significant effects, or interactions between
conditions or segments for the 'ID', and 'DI' conditions. Workload ratings in
the 'ID' condition did not reflect increased RS difficulty or decreased RE
difficulty.

7.9

After group. Although marginally significant, the differences between rated segments in 'ID', and 'DI' conditions did not clearly reflect both RS and RE difficulty manipulations in an orderly way. For these conditions, workload ratings were more influenced by RS than RE.

Correlations among workload ratings and performance measures. Table 1 shows the correlations among the bipolar ratings, weighted workload, RT and MT, obtained with BMDP 6R. There were large variations in the correlations between raw bipolar ratings, and did not correlate very highly with RT and MT. With the exception of activity type, raw bipolar ratings highly correlated with weighted workload.

Table 1. Correlations among bipolar ratings, weighted workload, RT, and Mt.

	TD	TP	PF	ME	PE	FR	ST	FA	AT	OW	WW	RT	MT
Task Difficulty	-												
Time Pressure	.77	-											
Performance	.35	.35	-										
Mental Effort	.59	.56	.16	-									
Physical Effort	.73	.58	.28	.46	-								
Frustration	.70	.71	.46	.42	.64	-							
Stress	.71	.79	.15	.45	.55	.72	-						
Fatigue	.37	.44	.14	.13	.34	.44	.60	-					
Activity Type	.24	.13	.00	.23	.27	.12	.15	-.12	-				
Overall Workload	.86	.74	.23	.55	.69	.67	.72	.39	.25	-			
Weighted workload	.89	.83	.53	.62	.78	.81	.76	.52	.26	.79	-		
RT	.26	.11	.19	.13	.24	.14	-.05	-.03	.19	.18	.21	-	
MT	.42	.43	.40	.29	.29	.44	.35	.15	.08	.40	.43	.23	-

Discussion

The results of this experiment support the findings of previous experiments (Yeh, et al, 1985; Hart, et al, 1984; 1985) that subjects ratings are sensitive to task manipulations. Performance measures (RTs, and MTs) accurately and consisistently reflected the difficulty manipulations in RS and RE components across the consistent conditions (EE,MM,HH). This supports earlier views that as cognitive loading increases as a function of increasing difficulty, performance measures increase. Performance measures also reflected the different difficulty levels in RS and RE when the difficulty within conditions was positively correlated, as in the 'II' and 'DD' conditions, or when the difficulty within conditions was negatively correlated, as in the 'ID' and 'DI' conditions. In all the conditions, RTs were driven by RS components, and MTs were driven by RE components. This is evident in the changing-inconsistent condition (ID, DI), where RTs varied with MTs the same way RS components varied with RE components. The fact that RTs were slower than MTs, suggests that the RS component, solving math equations, loaded cognitive processes more heavily than the RE component. These performance results hold true for the Before group, as well as the After group.

Subjective ratings also were sensitive to cognitive loading (Yeh et al, 1985; Hart et al, 1984), and reflect task manipulations. A major concern of this experiment was to look at the degree to which introspective subjective ratings were sensitive to specific variation in cognitive loading of segments

within a block of trials. Yeh demonstrated that subjects could integrate all the information in a block of trials, and could differentiate between different levels of cognitive loads with retrospective workload ratings. This experiment demonstrated that subjects are also sensitive to different cognitive loads within blocks of trials, although the degree to which retrospective workload ratings reflect manipulations in cognitive loading depends on the difficulty levels within conditions.

In the consistent (EE, MM, HH), and the changing-consistent condition ('II', 'DD'), the information about RS/RE difficulty levels was still available retrospectively, and workload ratings selectively reflected the difficulty of individual segments. In the consistent conditions, subjects rated segments of the same difficulty level as having the same workload. In the changing-consistent conditions, subjects rated segments of different difficulty levels as having significantly different workload. In this case, difficulty segments were rated as being more loading than medium difficulty segments, which were rated as being more loading than segments of easy difficulty. Knowing in advance did not appear to increase subjects sensitivity to task manipulations. Possibly subjects in the Before group gave higher workload ratings than subjects in the After group due to individual differences rather than increased sensitivity to the magnitude of difficulty manipulations, because the interactions between subject, group, experimental condition, and trial block segments were not significant. However, this difference may be due to a perceived additional memory task for the before group.

These results suggest workload ratings are a good indicator of the direction of RS/RE component difficulty manipulations rather than absolute magnitude, but only so long as the difficulty levels of the RS components and RE components were consistent and varied in the same direction. When this occurred, performing the RS/RE task components facilitated recall of the average difficulty of the task components for each segment of a different difficulty level. These findings are unlike dual-task results which reflect interference between tasks due to direct competition for limited resources. Since the output from the RS component serially fed into the RE component, and had to be completed prior to RE, the pairing of these processes did not lead to competition for common resources. Therefore, workload ratings reflecting the differences in difficulty between segments were reinforced.

In the changing-inconsistent condition (ID,DI), the difficulty levels of the RS and RE components were varied in the opposite directions. In this case, performing the RS/RE task components facilitated recall of the average difficulty of the task components across segments of different difficulty levels. It may be that more resources were allocated for integrating task components as in the changing consistent conditions, However, since the task components had opposing difficulty levels, recall of the average workload of the difficulty levels experienced across segments was facilitated. Consequently, workload ratings did not significantly reflect the direction of either the RS or RE component. This suggests that the workload ratings were not driven exclusively by the response selection component (which had a higher cognitive load than the response execution component), as RTs were, but by an integration of the two components. Although, in the 'ID', and 'DI' conditions for the After group, the workload ratings of the third segment reflected the difficulty level of the RS component, while the workload ratings of the first two segments reflected an integration of the two components. This appears to be a small recency effect,and suggests that the when integrating two task compon-

ents, RS may carry more weight, (i.e. the component that loads heavier on cognitive processes may weigh heavier in evaluating workload).

Conclusion

This study succeeded in determining some of the limits memory imposes on subjective ratings. Subjects appear to be sensitive to task component manipulations, and their ratings reflect the specific retrieval of information from long-term memory about the workload of particular segments, but only in certain conditions. Task components need to be stimulus/response compatible and well integrated for a human operator to accurately recall segments of a task that vary in difficulty, as all were in this study. If the task components vary in difficulty, human operators integrate them and recall the average workload of the difficulty levels. It appears that knowing in advance which segment should be rated may not additionally facilitate recall. Finally, the results from the changing-inconsistent condition indicate that the response selection component may load on cognitive processes more heavily, and consequently contribute more to workload ratings than the response execution component. Thus, the degree to which the response selection component drives workload ratings may be greater under some circumstances and not under others, and requires further research.

Acknowledgements

References

Arbak, C. J., Shew, R. L., and Simons, J.C. (1984) The use of reflective swat for workload assessment. Proceedings of the 1984 Human Factors Society Annual Meeting, 952-962.

Bortolussi, M. R., Kantowitz, B. H., & Hart, S. G. (in press). Measuring pilot workload in a motion base trainer: a comparison of four techniques. Proceedings of the 1985 Human Factors Society Annual Meeting.

Eggemeir, F. T., Melville, B. E., & Crabtree, M.S. (1984). The effect of intervening task performance on subjective workload ratings. Proceedings of the 1984 Human Factors Society Annual Meeting, 954-958.

Hart, S. G., Sellers, J. J., & Guthart, G. (1984). The impact of response selection and response execution difficulty on the subjective experience of workload. Proceedings of the 1984 Human Factors Society Annual Meeting, 732-736.

Hartzell, E. J., Gopher, D., Hart, S. G., Lee, E. & Dunbar, S. (1983). The Fittsberg law: the joint impact of memory load and movement difficulty. Proceedings of the 1983 Human Factors Society Annual Meeting.

Notestine, J. C. (1984). Subjective workload asessment and effect of delayed

ratings in a probability monitoring task. <u>Proceedings of the 1984 Human Factors Society Annual Meeting</u>, 685-689.

Yeh, Y. (in press). The effect of varying task difficulty on subjective workload. <u>Proceedings of the 1985 Human Factors Society Annual Meeting</u>.

THE EFFECTS OF ACCELERATION STRESS ON
HUMAN WORKLOAD AND MANUAL CONTROL

Richard T. Gill, Department of Engineering Sciences
University of Idaho, Moscow, Idaho 83843

William B. Albery and Sharon L. Ward
Harry G. Armstrong Aerospace Medical Research Laboratory
Wright-Patterson AFB, OH 45433

ABSTRACT

This study assesses the effects of +Gz stress on operator task performance and workload. Subjects were presented a two-dimensional maze (via a CRT) and were required to solve it as rapidly as possible (by moving a light dot through it via a trim switch on a control stick) while under G-stress at levels from +1 Gz to +6 Gz. The G-stress was provided by a human centrifuge. The effects of this stress were assessed by two techniques; (1) objective performance measures on the primary maze-solving task, and (2) subjective workload measures obtained using the subjective workload assessment technique (SWAT). It was found that while neither moderate (+3 Gz) nor high (+5 Gz and +6 Gz) levels of G-stress affected maze solving performance, the high G levels did increase significantly the subjective workload of the maze task.

INTRODUCTION

Technological advances in recent years have considerably increased the complexity of fighter aircraft cockpits. The number of individual parameters which the pilot must monitor and control has increased dramatically. A fundamental consequence of these advances has been to significantly alter the pilot's role from primarily a skilled manual control operator to that of an executive manager or decision maker.

Such alterations in the pilot's tasks have created an additional constraint for design engineers, namely pilot workload. As pilot workload increases, not only does he become fatigued more readily, but his performance begins to deteriorate. Excessive pilot workload can result in some piloting tasks not even being performed, with potential catastrophic consequences. Clearly, pilot workload is a crucial factor which must be addressed in the design of modern fighter aircraft.

Concurrent with the advances in aircraft avionics have been major advances in propulsion, aerodynamics and airframe materials. As a result, the modern fighter aircraft is capable of maneuvers with considerably higher G-levels and G-onset rates than those of its predecessors. The effects of G-forces will become even more severe with the next generation fighters which are expected to exceed the current G-capabilities.

Although the increased G-environment of modern fighter aircraft may appear to be independent of the pilot workload problem, it probably is not. First, in order to maintain an adequate field of vision and

8.1

consciousness at higher G-levels, pilots must perform an M-1 or L-1 straining maneuver. This coordinated grunting and isometric muscular straining is an additional piloting task which requires some mental attention; thus the potential for increasing pilot workload. In addition, the G-induced reduced blood flow to the brain could impair higher level cognitive activity. This in turn would decrease the pilot's mental processing efficiency, making it more difficult for him to complete all necessary tasks adequately.

Pilot workload and pilot G-stress are two very important issues in the design of modern fighter aircraft. Although the physiological effects of G-stress have been studied for years by the aerospace medical community, little is known about the psychological effects associated with G-stress. Therefore, the objective of this effort was to investigate the impact of G-stress on pilot workload.

BACKGROUND

There is an enormous amount of research that has been conducted on the effects of acceleration on humans. Most of it deals with the physiological effects; none of it specifically addresses the issue of G-stress on pilot workload. In fact, very little work has been done which addresses the effects of G-stress on the pilot's cognitive abilities. In Collyer's very thorough 1973 review (1), he cited several studies (2,3,4,5,6) which suggest increased G-stress has a negative effect on pilot's cognitive abilities. However, he concluded that knowledge in this area was quite incomplete and little has been done since then. Given the increased cognitive demands and increased G-stress being placed on the pilots of modern fighter aircraft, this area demands thorough investigation.

Similarly, there has been a plethora of research conducted in the general area of assessing operator workload. In a comprehensive review of the workload literature, including over 400 references, Wierwille and Williges (7) identified twenty-eight specific techniques for assessing operator workload. In the same paper, they also presented a method for selecting the most appropriate technique for a given context. Following their guidelines, it was decided to employ two alternative techniques: (1) an objective measure of performance (the specific task selected was 2-dimensional maze solving); and (2) a subjective measure of workload (the specific technique being the Subjective Workload Assessment Technique). Each of these techniques will be discussed in detail.

To be consistent with the referenced literature, the generic term of workload is used throughout this paper. However, the correct interpretation of workload as used here is that it is a combination of both internal and external workload, as well as processing capacity. Alternately, workload can be thought of as being inversely related to the amount of unused processing resources. That is, as the amount of available or unused processing resources decreases, the workload, by definition, has increased.

8.2

PERFORMANCE MEASUREMENT - MAZE SOLVING

The performance measurement technique used here was the maze-solving technique developed by Ward and her colleagues (10,11). Subjects were presented with an unfamiliar two-dimensional maze on a CRT and were required to move a dot through the maze from one side to the other as rapidly as possible (Fig. 1). The dot moved at a constant speed and direction, but subjects could change the direction with discrete control inputs of either up, down, left, or right. The score was defined as the ratio of the optimum solution time to the actual solution time.

This task was selected primarily because it required considerable cognitive resources, yet only minimal response resources. Thus, if an increase in G-stress resulted in a decrease in performance, it could not be attributed solely to a decrease in motor coordination. Rather, it would be primarily a consequence of a decrease in cognitive processing capabilities. Alternatively, if there was no change in performance, it could be the case that the maze-solving task did not provide sufficient task loading. This would leave reserve processing resources to be expended under G-stress, and the measured performance would not change.

WORKLOAD ASSESSMENT - SWAT

In general, subjective workload assessment consists of requiring subjects to estimate the workload imposed by a given experimental manipulation via introspection. Although such a technique can be useful, it has been criticized for being easily biased and rather insensitive to changes in workload. Furthermore, it only produces a rank ordering of workload rather than a more desirable ratio or interval scale of workload.

Recently, however, Reid, Eggemeier, and their colleagues at the AFAMRL have developed a generic subjective technique called the Subjective Workload Assessment Technique or SWAT (12,13,14,15). It combines subjective ratings on three different scales, via the mathematical technique of conjoint measurement, to produce an interval scale of workload. It has been shown to be both a reliable and sensitive measure of workload. One significant advantage of SWAT is that it is a relatively simple and unobtrusive technique that could be easily implemented jointly with the maze-solving technique in the high G environment. Thus, SWAT was used in conjunction with the performance measure obtained via the maze-solving scores.

METHODOLOGY

The objective of this study was to assess the effects of +Gz stress on pilot workload. It was conducted in three phases: Phase I - Static Training, Phase II - Dynamic Training, and Phase III - Data Collection. Pilot performance and workload were measured using primary task performance, via the two-dimensional maze-solving task, and subjective ratings via SWAT. AFAMRL's Dynamic Environment Simulator (DES) provided the G-stress. Special equipment included a modified ACES II or F-16 seat,

side-arm controller, flight suit, gloves, anti-G suits, and Doppler
temporal artery flow meter.

PHASE I - STATIC TRAINING

There were two tasks which were accomplished in the static training
phase. Subjects performed a card sort to rank-order the subjective
ratings that were used in Phase III, and they practiced solving two dimen-
sional mazes similar to the ones which were used as the primary task in
phase III. All work conducted in Phase I was in a normal +1 Gz environ-
ment. Each subject participated in four one-hour training sessions with
each session occurring on a different day.

The purpose of the first session was to perform a card sort for the
SWAT portion of Phase III. The subjects were provided with a deck of 27
cards placed in random order. Each card represented one of the possible
combinations of three categories (time load, mental effort load, and
psychological stress load) with each category at three different levels
(low, medium, and high, Fig. 2). The subject's task was to sort these
cards so that all 27 combinations were rank-ordered with respect to the
degree of subjective workload imposed by each. These rank-orderings were
then used to develop an interval scale of workload for evaluating the
subjective ratings that were obtained in Phase III.

In the remaining three static training sessions, subjects practiced
solving two-dimensional mazes. Figure 1 depicts a maze typical of those
used throughout the experiment. All mazes consisted of the basic 10 x 10
grid as shown, but differed in the placement of the maze barriers. For
each trial, a given maze was displayed on a CRT with a dot at the entrance
of the maze. The subject's task was to solve the maze as rapidly as
possible. The dot moved at a constant speed and the subject could change
its direction (left, right, up, or down) by moving the trim tab button on
a joystick controller in the appropriate direction.

The trial concluded as soon as the dot was successfully guided
through the maze to the goal. The shortest possible solution time divided
by the actual time required to complete the maze was used as the measure
of performance, and this score, multiplied by 100, was displayed to the
subject immediately after completion of each trial. Typical solution
times were approximately one minute or less. If the subject failed to
complete the maze within two minutes, the trial was terminated and a
message was displayed indicating that the subject had run out of time.
The displayed score was then computed as the ratio of the shortest
possible solution time to actual solution of time, multiplied by the
percent of the maze solved.

PHASE II - DYNAMIC TRAINING

The purpose of the dynamic training phase, which was conducted
entirely on the DES, was to reduce the experimental variance in Phase III
by permitting subjects to practice maze solving while under G-stress.
Each subject practiced in two daily sessions of approximately one-half

hour each. The specific G-profile (number of runs per session, duration of each run, etc.) was identical to those which were used in the data collection phase and are discussed in detail in the following paragraph. Different mazes were used in each of the three phases to prohibit learning effects due to subjects becoming familiar with any particular maze.

PHASE II - DATA COLLECTION

The data collection phase was also conducted on the DES. Each subject participated in five daily sessions of approximately one-half hour each; each session was comprised of eight trials. A trial was comprised of four parts.

(1) Positive Onset: Starting at a baseline level of +1.5 Gz, the subject's Gz level increased (or decreased) at the rate of .25 Gz/sec until the desired level of Gz for that trial was attained. Four levels of +Gz (1,3,5, and 6) were employed. The slow onset rate and the baseline level of 1.5 Gz were chosen to minimize the problems associated with vertigo.

(2) Test: Once the desired level of +Gz was attained, the subject was presented with a maze on the CRT and asked to solve it as rapidly as possible.

(3) Acceleration Offset: Two different rules for determining the time of acceleration offset were used in this study. For the first two subjects, offset started when the subject solved the maze or after two minutes at +Gz, whichever occurred first. Data from these two subjects indicated a moderate improvement in performance between +3 Gz and +5 Gz. It was believed that increased motivation to terminate the trial as quickly as possible at the higher +Gz level may have caused this performance improvement. To remove this possible confounding effect, the determination of acceleration offset time was changed. For the last two subjects, offset started after one minute at +Gz, regardless of maze completion.

For both offset rules, Gz was decreased (or increased if a 1 Gz trial) at the rate of .25 G/sec. until the baseline level of +1.5 Gz was attained.

(4) Rest: The subject then rested at the +1.5 Gz level for a minimum of one minute before initiating the next trial. However, this rest period could be extended for as long as desired by either the subject or the medical monitor. During this rest period, the subject was required to rate his perceived workload during the previous trial, using SWAT. After the SWAT rating was completed, the subject was informed of his maze-solving score.

Each daily session consisted of eight trials, with the first and second half of the session separated by a rest period of at least three minutes. Within each half of a session, the order of presentation of the +Gz levels followed an incomplete 5 x 4 (5 sessions x 4 trials per half-

session) Latin square. The eight different mazes used in Phase III were randomly assigned to each trial, with the following constraints: (1) each maze was used exactly once during a session, and (2) each maze was used once in combination with each +Gz level during the first four sessions of Phase III.

The first four daily sessions of Phase III were identical to that of Phase II. On the fifth day, each maze had the optimum solution path identified on the CRT. Thus the subject's only task was to maneuver the dot along the path. Comparisons of solution times between mazes with and without the solution path shown served as a direct measure of the cognitive effects of G-stress since the same motor coordination task was required for all conditions.

Heart rate (EKG), temporal artery blood flow (Doppler flow meter), and anti-G suit pressures were recorded for all trials in both Phases II and III (Fig. 3).

RESULTS

Figures 4 and 5 summarize the effect of +Gz stress on the maze-solving scores and the SWAT ratings respectively. Separate results are shown for conditions in which the solution paths were not shown (the first part of Phase III) versus those in which they were. Each figure shows the means by +Gz level, averaged across mazes, subjects and acceleration offset rule. The confidence intervals shown are based on the error terms obtained from the analyses of variance.

Univariate analysis of variance (ANOVA) was used to test the effects of the independent variables on maze-solving scores and on SWAT ratings obtained during trials in which the solution paths were not shown. The factors used and their levels were as follows:

1. +Gz stress (+1.5, 3, 5, or 6 Gz).
2. Maze used (eight different ones).
3. Offset rule for acceleration (variable length duration with a maximum of 2 min, or fixed length duration of 1 min at +Gz).
4. Subjects nested within offset rule.

The factors of maze and subject were treated as random factors, and the other two were fixed. All main effects and interactions of the first three factors were tested. For some of these hypotheses, an exact F-test was not available because of the constraints imposed by the presence of random factors; the procedure outlined by Scheffe (16) for approximate F-tests was used for these cases.

The distribution of residuals in the analyses of variance were found to be approximately normal, with the exception of one unusually small value for one of the SWAT ratings. This observation was omitted from the formal analysis.

The data obtained from the first part of Phase III are summarized in Tables 1 and 2, for the maze-solving scores and the SWAT ratings respectively. The results of the analysis of variance for these two measures are shown in Tables 3 and 4. The latter two tables also show results for the analysis performed on data obtained during the last session of Phase III, when the solution paths were shown.

Task performance as measured by the maze-solving score was not affected by the level of +Gz stress ($F=1.94$, $df=3,6$, $p<.10$). The only significant differences in maze-solving scores were attributable to a main effect for the differences among mazes ($F=4.35$, $df=7,14$, $p<.01$), a main effect for the differences among subjects ($F=3.48$, $df=2,14$, $p<.10$), and an interaction effect of maze and offset rule ($F=3.13$, $df=7,14$, $p<.05$). The main effects of subject and maze were expected. The interaction effect for maze by offset rule was somewhat anomalous. Of the eight mazes, higher scores were obtained on four of them with one offset rule, and on the other four with the other offset rule. Which offset rule yielded the higher score did not appear to be related to performance or structural characteristics of individual mazes.

+Gz stress had a significant effect on the SWAT ratings obtained ($F=12.41$, $df=3,6$, $p<.01$). Even though the performance measure did not detect a difference among +Gz levels, this subjective measure showed a dramatic increase in workload as a function of increased +Gz. The only other significant effects on SWAT ratings were due to a main effect for differences among mazes ($F=2.65$, $df=7,14$, $p<.10$), a main effect for differences among subjects ($F=21.96$, $df=2,14$, $p<.01$), and an interaction effect of +Gz with subject ($F=5.43$, $df=6,41$, $p<.01$). As for the maze-solving scores, the main effects of maze and subject were expected. The interaction of +Gz with subject was due to one subject who gave relatively low workload ratings to conditions under acceleration; the remaining three subjects were quite consistent with each other.

Linear and quadratic functions of +Gz level were used to fit the SWAT ratings obtained. The linear effect was statistically significant ($F=31.5$, $df=1,9$, $p<.01$) and the quadratic effect was not ($F=2.45$, $df=1,9$, $p<.10$). There is, however, some suggestion that the SWAT rating for +6 Gz is slightly higher than a linear trend would predict, so there may be a quadratic relationship for higher +Gz levels.

Analysis of variance was also used to assess the effects of the independent variables for conditions in which the solution path was shown, during the last session of Phase III. A simplified model was used for these hypothesis tests, since all mazes were not observed at all levels of +Gz. Only the main effects of +Gz, maze, and subject were tested, and the error term used was the residual. The effect of offset rule was not included, since it seemed unlikely to have any impact on trials in which the mazes were solved very quickly, as these were.

The maze-solving score was affected moderately by all three of these factors. The largest difference was among subjects ($F=30.3$, $df=3,18$, $p<.05$). Differences among +Gz levels were smaller ($F=17.1$, $df=3,18$, $p<.10$),

8.7

as were differences among mazes (F=14.5, df=7,18, p<.10). Differences among subjects and mazes were expected. The differences among +Gz levels were small; a decrement in average score from 94.4 at +1.5 and 3 Gz to 92.4 at +5 and 6 Gz. This change may reflect the extent of additional physical difficulty in performing the maze-solving task at the higher levels of +Gz. The maze-solving scores obtained with the solution paths shown were much higher than those obtained when they were not, indicating that maze-solving is primarily a cognitive rather than a motor response task.

SWAT scores were affected by both +Gz levels (F=12.15, df=3,18, p< .01) and mean differences among subjects (F=4.31, df=3,18, p<.05). As can be seen in Figure 5, the changes in SWAT scores as a function of +Gz level paralleled the changes found for conditions in which the solution path was not shown. The mean difference in SWAT scores obtained when a solution path was shown versus when it was not was an increase of approximately 14 points on the rating scale, regardless of +Gz level. This indicates that the increase in subjective workload imposed by the cognitive aspects of the maze-solving task was independent of the amount of +Gz stress.

DISCUSSION

Performance on the maze-solving task was not affected by +Gz stress, although subjective ratings of workload were. The level of demand presented by the maze-solving task appears to have been such that subjects were able to accommodate the additional demand imposed by acceleration stress, and maintain their performance. Results of this study show that SWAT ratings may precede performance decrements, and be important "leading indicators" of task performance degradation. The increase in SWAT ratings was linear with +Gz, and there was some indication that the increase may become quadratic at higher acceleration levels.

The offset rule for acceleration was modified after data from the first two subjects had been obtained, because it was felt that the small improvement in performance from +3 Gz to +5 Gz was due to the variable duration of acceleration. The hypothesis was that when subjects were exposed to the higher +Gz level and the duration of the acceleration depended on how quickly they solved the maze, their motivation to complete it as quickly as possible increased. However, with the second offset rule for acceleration (a fixed duration of 1 min.) the same small improvement in performance at +5 Gz was obtained. This increase in score, if it is repeatable, may be due to overall motivational factors unrelated to the rule for acceleration offset.

The comparison between trials in which the solution paths were not shown versus those in which they were demonstrates that maze-solving is primarily a cognitive rather than a motor-response task. This is apparent from both the performance scores and the subjective workload ratings.

8.8

CONCLUSIONS

There are two important conclusions to be drawn from these results. First, it is evident that increased +Gz stress produced a significant increase in perceived workload. Second, the discrepancy between the performance and workload measures suggests that the demand imposed by the maze-solving task did not force subjects to work at capacity, and allowed them sufficient processing resources to compensate for the effects of the +Gz stress.

REFERENCES

1. Collyer, SC. Testing psychomotor performance during sustained acceleration. USAF School of Aerospace Medicine, SAM-TR-73-52. 1973.

2. Comrey, A.L. et al. The effect of increased positive radial acceleration upon perceptual speed ability. J. Aviat. Med 22: 60-69. 1951.

3. Frankenhaeuser, M. Effects of prolonged gravitational stress on performance. Acta Psychol 14: 92-108. 1958.

4. Chambers, R.M. Long term acceleration and centrifuge simulation studies. N63-19313. Aviation Medical Accel Lab, US NADC, PA. 1963.

5. Ross, B.M. and R.M. Chambers. Effects of transverse g-stress on running memory. Percept Mot Skills, 24:423-435, 1967.

6. Ross, B.M., R.M. Chambers, R.R. Thompson. Effects of transverse acceleration on performance of two running matching memory (RMM) tasks. NADC-MA-6309. 1963.

7. Wierwille, W.W. and R.C. Williges. Survey and analysis of operator workload assessment techniques. Naval Air Test Center, S-78-101. 1978.

8. Gartner, W.B. and M.R. Murphy. Concepts of workload. AGARDOGRAPH No. 246, Survey of Methods to Assess Workload, AGARD-AG-246, ADA078319. 1979.

9. Crabtree, M.S. and C.A. Shingledecker. Secondary task workload assessment methodology. IEEE Nat Aerosp & Elect Conf, NAECON Proceedings, May 17, 1983.

10. Poturalski, R.J. and S.L. Ward. Maze solving as a tool for measuring performance under stress. Proceedings of Annual Mtg of the Aerospace Medical Association, 1982.

11. Ward, S.L. and R.J. Poturalski. Changes in maze solving errors due to stress. Annual Mtg of Human Factors Society. 1982.

12. Reid, G.B. et al. Application of conjoint measurement to workload scale development. Annual Mtg of Human Factors Society. 1981.

13. Reid, G.B. et al. Development of multi-dimensional subjective measures of workload. IEEE Conf on Cybernetics & Society. 1981.

14. Eggemeier, F.T. et al. Subjective workload assessment in a memory update task. Annual Mtg of Human Factors Society. 1982.

15. Eggemeier, F.T. et al. The effects of variations in task loading on subjective workload rating scales. IEEE Nat Aerosp & Elect Conf, NAECON Proceedings, 1983.

16. Scheffe, H. The Analysis of Variance. John Wiley & Sons, New York. 1959.

Fig. 1 TYPICAL 2-D MAZE WITH SOLUTION PATH (DASHED LINE)

8.11

Fig. 2 SWAT

PSYCHOLOGICAL
STRESS LOAD

MENTAL EFFORT
LOAD

TIME
LOAD

8.12

6.0

1.5

G_z

7.4
psi

Anti G-suit pressure

0

Doppler temporal artery flow meter (cm/sec)

EKG

| 0 | 10 secs | 20 | 30 | 40 | 50 | 60 secs | 70 | 80 | 9C |
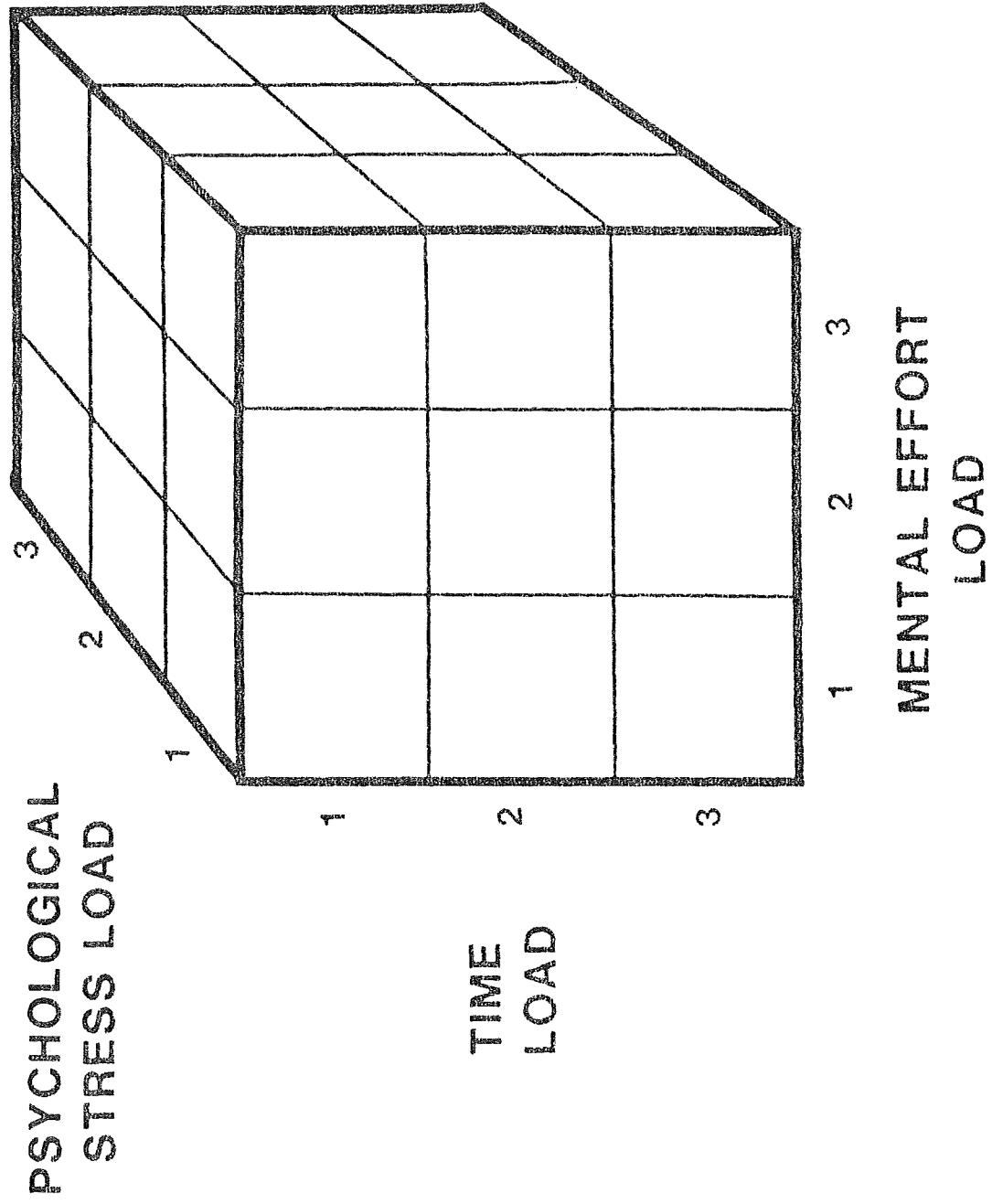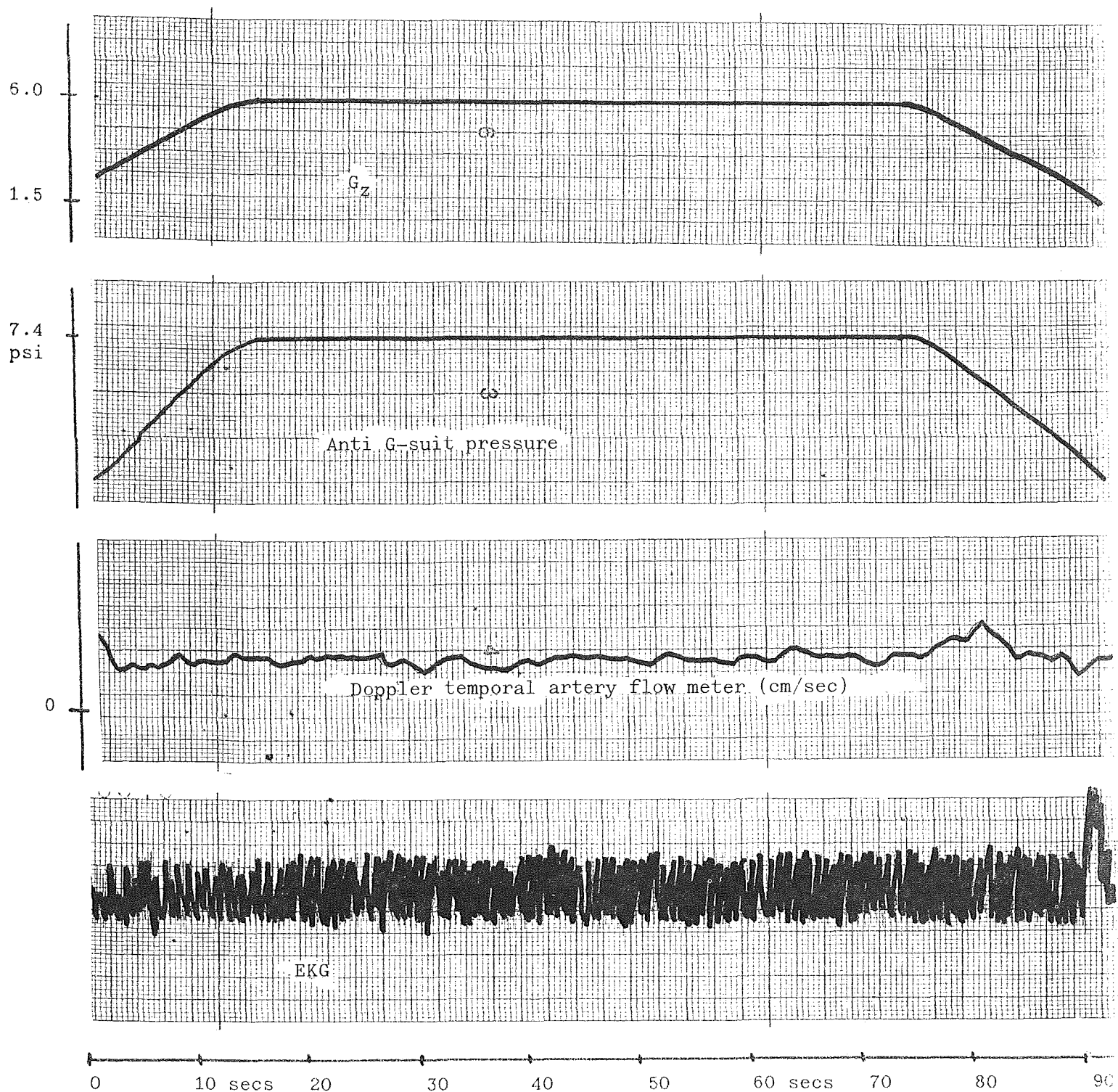
Fig 3. Typical Strip Chart Recordings from Maze Experiment, 6 G run

8.13

EFFECT OF +G$_z$ ON MAZE-SOLVING PERFORMANCE

MEANS AND 95% CONFIDENCE INTERVALS
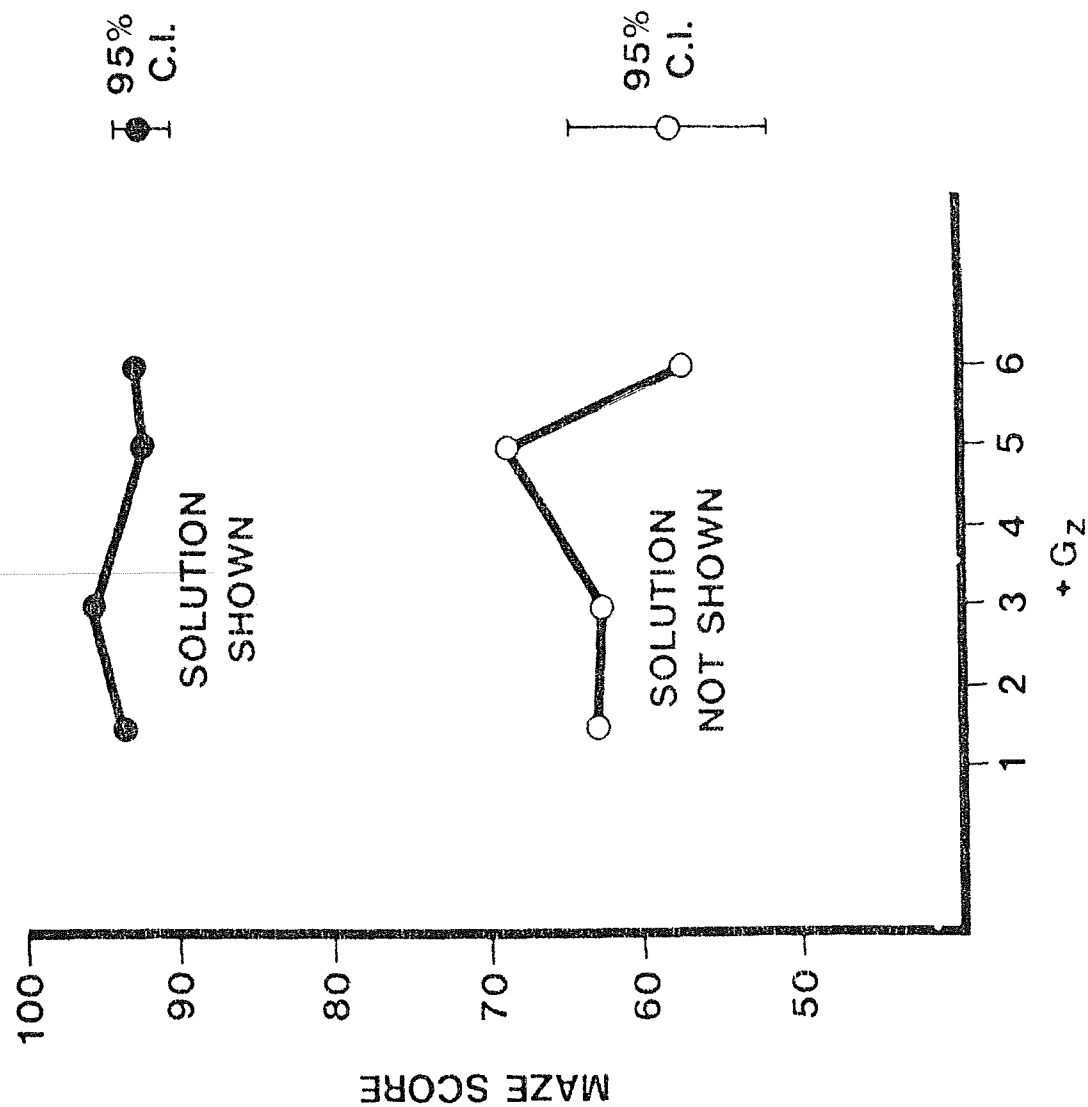
Figure 4

8.14

EFFECT OF +G$_z$ ON SUBJECTIVE
WORKLOAD ASSESSMENT

MEANS AND 95% CONFIDENCE INTERVALS
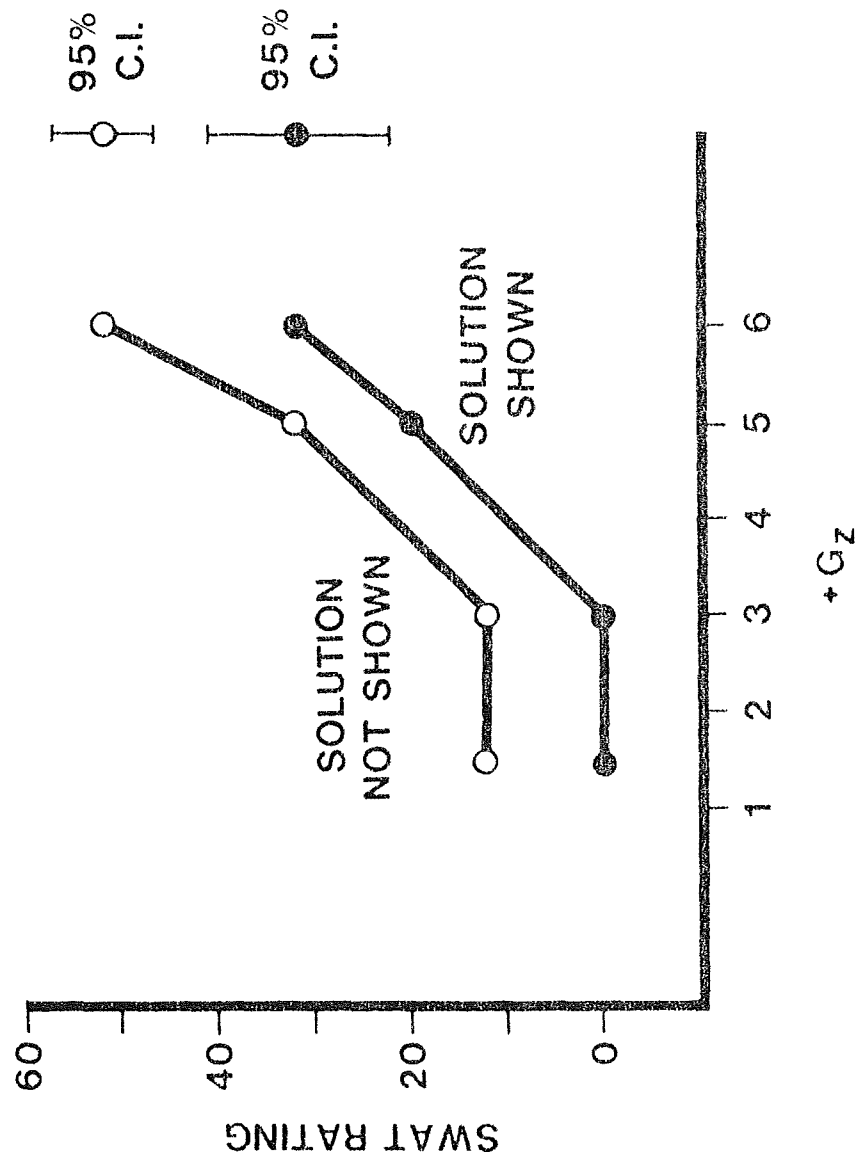
Figure 5

8.15

Knowledge-Based Load Leveling and Task Allocation in
Human-Machine Systems

M.H. Chignell and P.A. Hancock

Department of Industrial and Systems Engineering
Safety Science and Human Factors Departments
University of Southern California
Los Angeles, CA 90089-1452

ABSTRACT

Conventional human-machine systems use task allocation policies
which are based on the premise of a flexible human operator. This
individual is most often required to compensate for and augment the
capabilities of the machine. The development of artificial intelligence
and improved technologies have allowed for a wider range of task
allocation strategies. In response to these issues a Knowledge-Based
Adaptive Mechanism (KBAM) is proposed for assigning tasks to human and
machine in real time, using a load leveling policy. This mechanism
employs an online workload assessment and compensation system which is
responsive to variations in load through an intelligent interface. This
interface consists of a loading strategy reasoner which has access to
information about the current status of the human-machine system as well
as a database of admissible human/machine loading strategies.
Difficulties standing in the way of successful implementation of the
load leveling strategy are examined.

Introduction

Since the industrial revolution, human-machine systems of
increasing complexity have been developed. Initially, machine
capability was both limited and inflexible and human operators were
required to adapt themselves to the needs of the machine, often carrying
out boring and repetitive tasks in cramped quarters and under hazardous
conditions. The development of a human factors orientation, coupled
with advances in technology, has improved working conditions and human-
machine performance. As human-machine systems become more complex,
however, the division of labor between human and machine becomes less
clear-cut. Ideally, tasks should be allocated to the system component
best suited to perform them. Machines tend to be superior in
calculation, rote memory and coordination of simultaneous activities
while humans excel in creative problem solving, pattern recognition, and
decision making under uncertainty.

The resolution of the task allocaction problem is not always
straightforward. While many tasks as yet can be performed
satisfactorily only by humans, advances in machine intelligence and
automated systems have increased the number of tasks which may be
performed both by human and machine. Consider the task of regulating
the speed of an automobile. On an open freeway, control might be passed

to an automated system (cruise control), whereas the human operator (the driver) should be operating the accelerator and brake in city traffic. Changing task definition and environment may necessitate a revision of the task allocation policy. This policy will also be affected by changes in the state of the individual. When the operator is fatigued, under stress, or overloaded there is a tendency to focus on restricted elements of the task as exhibited by attentional narrowing (Hancock & Dirkin, 1983). To continue with the automobile example, part of a cab driver's task might involve carrying out a conversation with the passenger, but the driver may avoid this when fatigued or overloaded (Brown, 1967). The task of entertaining and informing might then be carried out by the radio, although perhaps not as well as by a talkative cab driver.

In complex systems where both task definitions and system capabilities vary over time, task allocation should be viewed as a dynamic rather than static process. Adaptive mechanisms are required which can diagnose the state of the machine and operator in order to reallocate subtasks accordingly and thereby optimize performance. This paper will consider how these adaptive mechanisms can be designed and implemented and will discuss some of the problems which may be encountered.

Human-Machine Cooperation

Conventional views of human-machine systems have the human controlling, or being controlled by, the machine component. In machine-paced assembly, for instance, the human is effectively controlled by the machine, whereas in driving a car, the human appears to be in control, at least under normal operating conditions. We follow an alternative perspective of human-machine systems in regarding them as a cooperative enterprise. In this view, human and machine work together to ensure successful system performance and the satisfaction of task demands. The assumption of a synergistic and cooperative relationship between the components of a human-machine system leads to a new type of system design. Firstly, cooperation presupposes communication between intelligent entities. Expert system consultants (Hayes-Roth, Waterman, & Lenat, 1983) provide low-level examples of this communication. Secondly, an interface (translation process) is required for the communication of needs, requests and ideas between the entities.

Early human-machine systems forced the human to fit in with the requirements of the machine, i.e., the human bent while the machine was straight. Human factors engineers designed machines and tools which fitted human capabilities, in keeping with the maxim "bend the tool, not the person" (McCormick & Sanders, 1982, Chapter 10). The ideal situation would be one where both the person and the machine stood straight (i.e., performed according to their design principles) while a translating interface adapted inputs and outputs so as to render them compatible.

Task Structuring

The type of adaptive interface necessary for genuine human-machine cooperation would be capable of restructuring the task in accordance with system goals and environmental constraints, and of reallocating task components between human and machine for a given task structure.

Figure 1 shows the overall control structure for such an adaptive interface. Task structuring would be carried out by a task definition supervisor. As technology advances and human–machine systems develop, the task will no longer be set as a fixed entity. Instead, the definition of the task will change in accordance with the higher level goals set by some external agency and with changes in the number and type of environmental constraints acting on the human–machine system. Given a particular definition of the task, allocation of tasks to human and machine would be carried out by an intelligent interface.
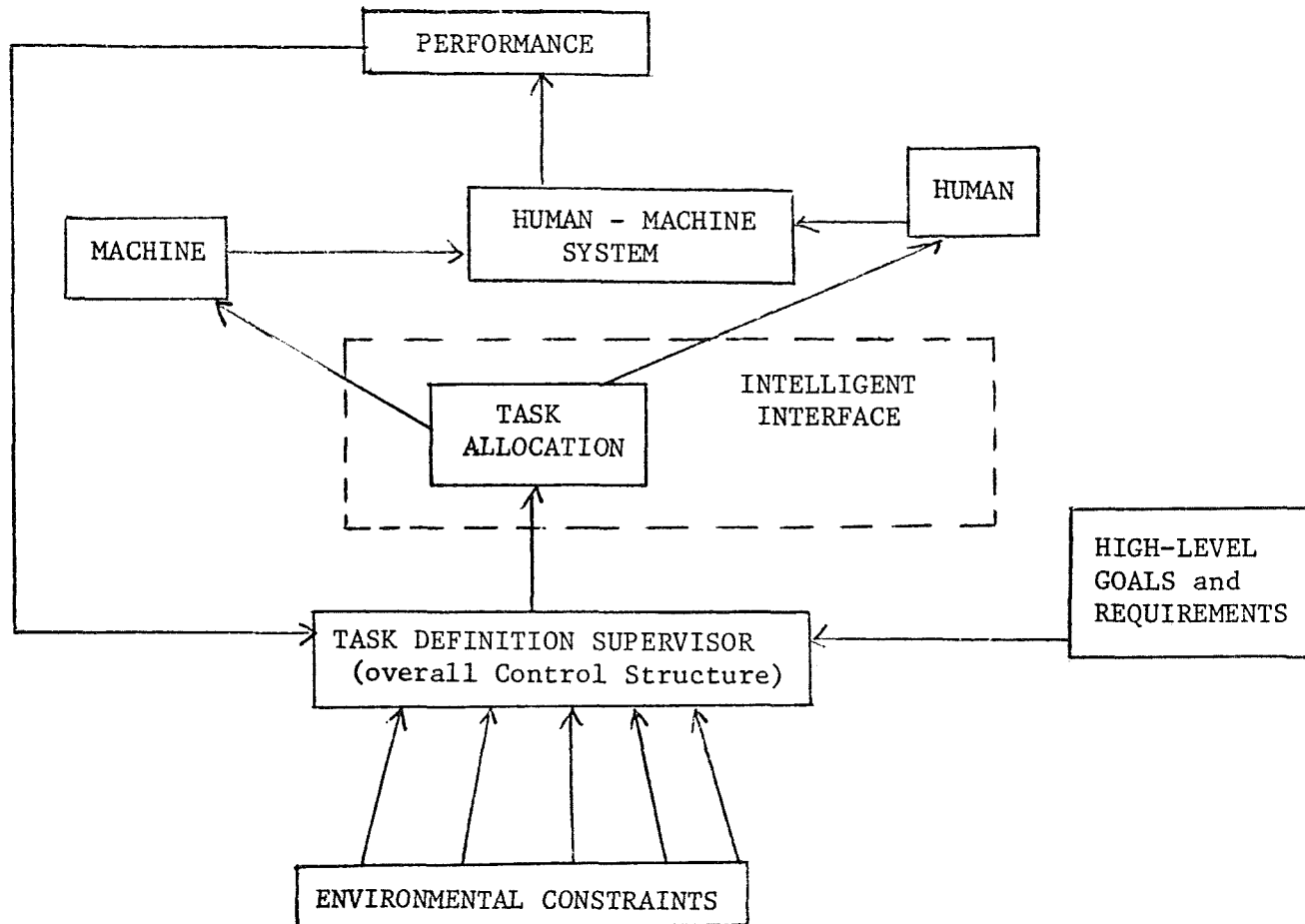
Figure 1. An overall control structure for a human–machine system which allow flexible restructuring and reallocation of tasks.

Task Allocation

It is clear that optimal allocation of task functions between operator and system requires a knowledge of human versus machine capabilities. Some of the task functions will not be involved in the allocation decision because they are clearly suited only for the machine or only for the human component of the system. Allocation will be

9.3

relevant for tasks which can be switched between human and machine without having a detrimental impact on overall performance (Figure 2).

Price (1985) has developed a method for allocating functions between humans and machines which requires human designers to construct a preset and fixed task allocation. The optimal allocation will not be fixed (see above), however, but will be conditional on the task definition, working environment, and current capabilities of system components. Given a particular task definition and system with currently specified capabilities, optimal task allocation requires a detailed understanding of human cognition and capability. At present, knowledge about the human abilities that are relevant to the performance of various tasks is incomplete, but there are several major characteristics which should be taken into account These include human sensitivity to relative rather than absolute change, limitations in attention and memory, compatibility of various input/output modalities for different tasks, performance variability, error correction capabilities, fatigue and reactions under stress (Chignell & Hancock, 1985, Hancock & Chignell, 1985).
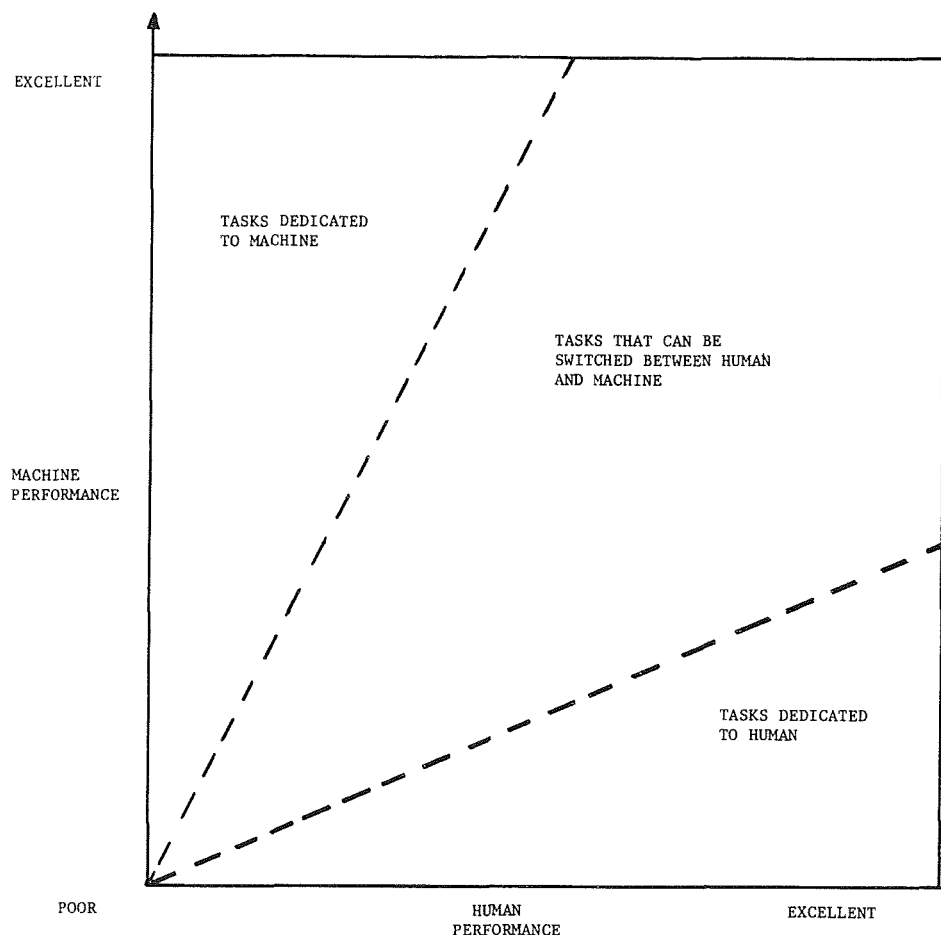
Figure 2. Decision space for allocation of tasks within a human-machine
          system.

9.4

Optimal, or close to optimal, task allocation will require a sophisticated reasoning process based on models of human and machine capability and a detailed analysis of the current operating environment. The previous discussion has outlined an ideal approach to synergistic and cooperative human-machine systems. This approach presupposes an intelligent interface which will allow communication between human and machine. The interface will then act in much the same way as an intermediary helps the user communicate with an online information retrieval system or as an interpreter translates the communications of two people who speak different languages. Intelligent systems are a powerful tool for improving the cooperation between human and machine. Building such systems into the machine allows it to function as an intelligent entity. Designing the interface as an intelligent system allows effective communication between human and machine. For a given system, it becomes debatable as to whether the intelligence resides in the machine or the interface. In this paper we shall focus on augmentation of the interface.

Static task allocation policies ignore intrinsic variability in the nature of the task and the human-machine system's response. Taking the human's point of view, the perceived difficulty of the task is a reflection of the mismatch between task demands and available resources (capacity). This mismatch will vary over time, as illustrated in Figure 3, with the human tolerating the variation in most cases. At times, however, the mismatch may be so great as to produce inadmissible overload or underload with consequent decrements in performance.
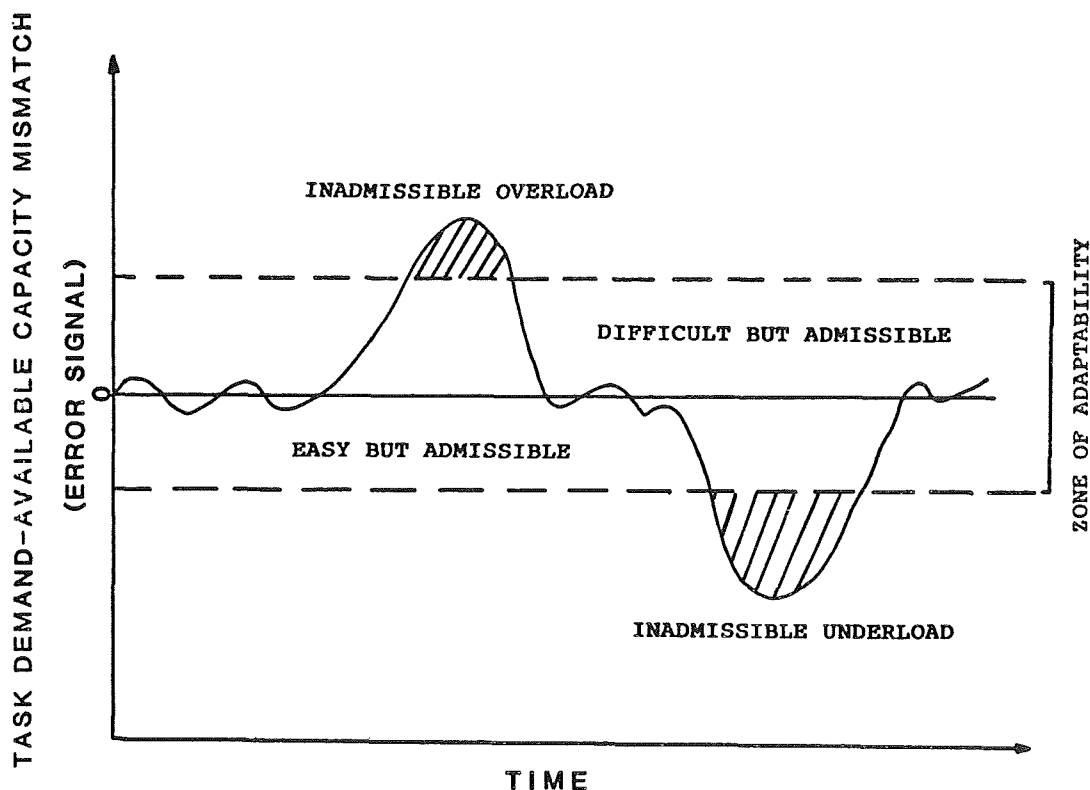
Figure 3. Schematic representation of the time-varying mismatch between task demands and available capacity. Shaded regions indicate inadmissible loading conditions.

9.5

In order to achieve a consistently high level of performance, the human-machine system will need to cope with large task and person fluctuations. In many cases, humans will be able to adapt to mismatches between current task demands and their available capacity. In some situations, however, the mismatch between task demands and available capacity will be so great as to preclude sufficient adaptation by the human operator. In such cases, return to the zone of adaptability to the task demand-resource mismatch (see Figure 3) requires dynamic reallocation of the task components, and possible, restructuring of the task. The process of dynamic reallocation requires an adaptive interface which is outlined below.

A Knowledge-Based Adaptive Mechanism (KBAM)

Although the concept of an adaptive interface has been discussed (e.g., Edmonds, 1981; Morris, Rouse & Ward, 1984), adaptive interfaces which allocate tasks dynamically have yet to be implemented. Part of the reason is that adaptive interfaces will be useful only with certain types of task. Firstly, the task must be performed by a human-machine system and, secondly, the task must be of moderate complexity, being neither so difficult as to tax the system as a whole, nor so easy as to allow the system to perform adequately whatever allocation policy is adopted, as depicted in Figure 4. In general, there will be some combination of task complexity and variability where an adaptive reallocation policy will improve performance significantly. Personnel training and selection will tend to shift the boundaries of the region upwards, as shown in Figure 4.

Adaptive interfaces have also been conspicuously absent in the past because the technology was not available. Recent developments in computer hardware, mental workload assessment (MWL) and artificial intelligence now make dynamic task reallocation technically feasible, although difficult to implement.

Dynamic task reallocation requires an adaptive mechanism which can assess the mismatch between task demands and available capacity (Figure 3) and redefine the task so as to reduce this mismatch. This adaptive mechanism will represent more than a simple reflexive action (e.g., table lookup) in many cases, since the complexities of human capabilities, task flexibility (i.e., the extent to which, and conditions under which, it can be redefined) and physiological variability will require some degree of knowledge-based reasoning.

The first step in developing the knowledge-based adaptive mechanism (KBAM) for load-leveling is the identification of the error signal representing the mismatch between the current task demands and the available capacity of the human operator. Task demands can be assessed either by examining task characteristics directly, or indirectly through assessment of MWL. MWL assessment will be necessary in many situations because of variability in response to specific task characteristics, both within and between individuals. Theoretically, the error signal can be expressed 1), as a mismatch between global attentional capacity and task requirements (cf. Kahneman, 1973), or 2), as a mismatch involving multiple attentional resources (Wickens, 1980).

9.6

TASK COMPLEXITY

HIGH — MEDIUM — LOW

LOAD REDUCTION

ADAPTIVE INTERFACE
USAGE WHERE TASK /
PERSON VARIABILITY
IS SUFFICIENTLY
LARGE

LOAD INCREASE

SELECTION AND TRAINING

INCREASE DUE TO SELECTION AND TRAINING

MANUAL TASKS
(OPERATOR)

HUMAN-MACHINE
INTERFACE

AUTOMATED TASKS
(SYSTEM)
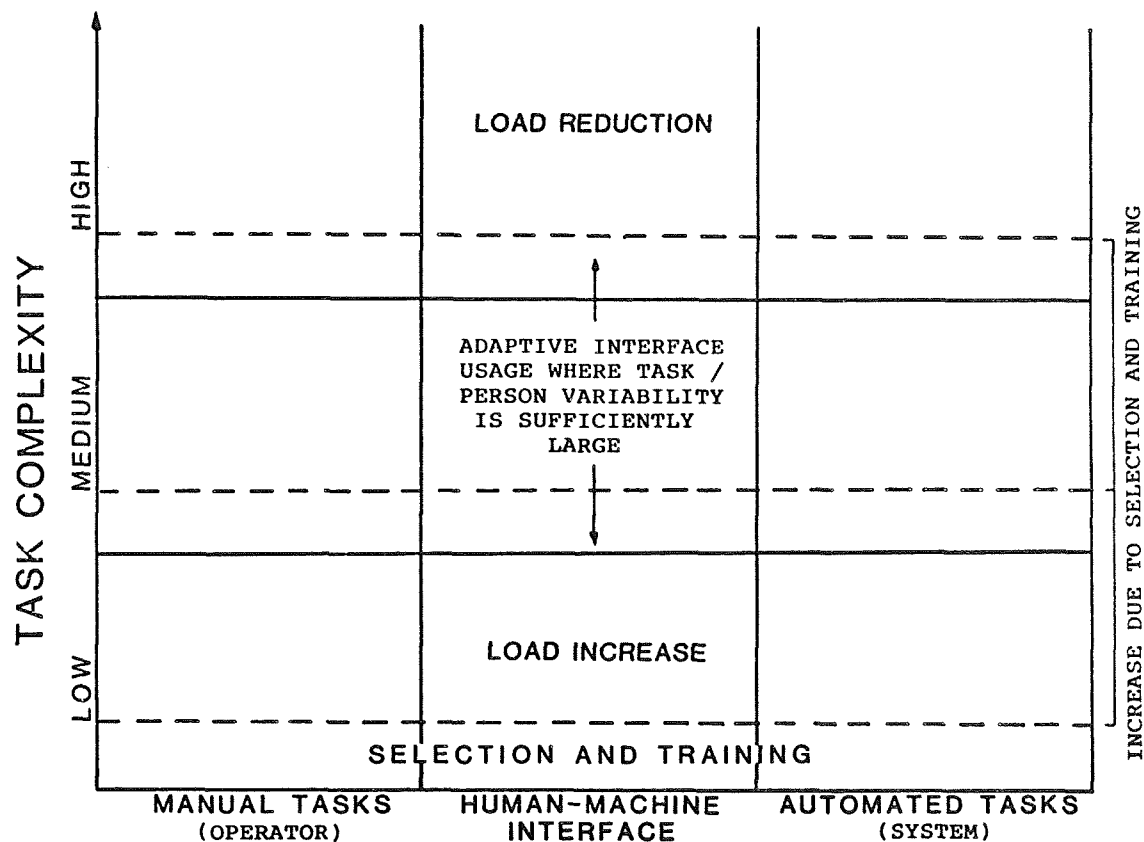
Figure 4.   The effects of task complexity and degree of automation on
the applicability of adaptive interfaces.

The formation of the error signal is problematic. MWL measures may
be used to derive the error signal directly, but they are likely to be
confounded by a number of factors, not the least of which is emotional
response to the task situation and performance feedback.   Ideally, the
error signal would be based on a theory of attentional resource
utilization and the various relationships between task demands, workload
and physiological response (see Hancock, Meshkati, & Robertson, 1985).
At present, we favor redundancy in error signal derivation.   Direct
measurement of the error signal via MWL assessment would be augmented by
calculations of the mismatch between task demands and attentional
resources.   Indirect assessment of the mismatch by calculation requires
an understanding of the demands generated by different tasks and
estimates of resource capacity based, possibly, on performance measures.
Specification of how the error signal should be derived under different
circumstances is a subject that is being investigated in our laboratory.
Once the error signal is derived, it is input to the adaptive
mechanism. Models of the task, system, and person then allow prediction
of the effect of alternative task redefinitions, while lookup of a
database of admissible loading strategies will enable a quick check of
whether or not a proposed strategy violates guidelines relating to
minimum task performance and imposed safety standards. Only tasks which
can be switched reasonably between the human and machine components of

9.7

the system (Figure 1) will be considered for reallocation, except under
emergency conditions which demand continued performance, with the result
of system failure having fatal consequences.

The output of the adaptive mechanism will be a reallocation and,
possibly, restructuring of the task which alters (where necessary) the
loading of task components between the human and machine so as to reduce
the error signal. Thus a human-machine interface is developed which
acts as a servomechanism minimizing the difference between current
demands and available capacity. The overall structure of this interface
is shown in Figure 5. It is referred to as an intelligent interface
because of the extensive use of knowledge and reasoning in formulating
the load leveling strategy. Central to the reasoning process is a
knowledge base containing information about all the facts deemed to be
relevant to the performance of the task. Figure 5 shows only one of a
number of ways in which a KBAM might be designed, although we expect any
variation to contain the components identified here, albeit in different
configurations.
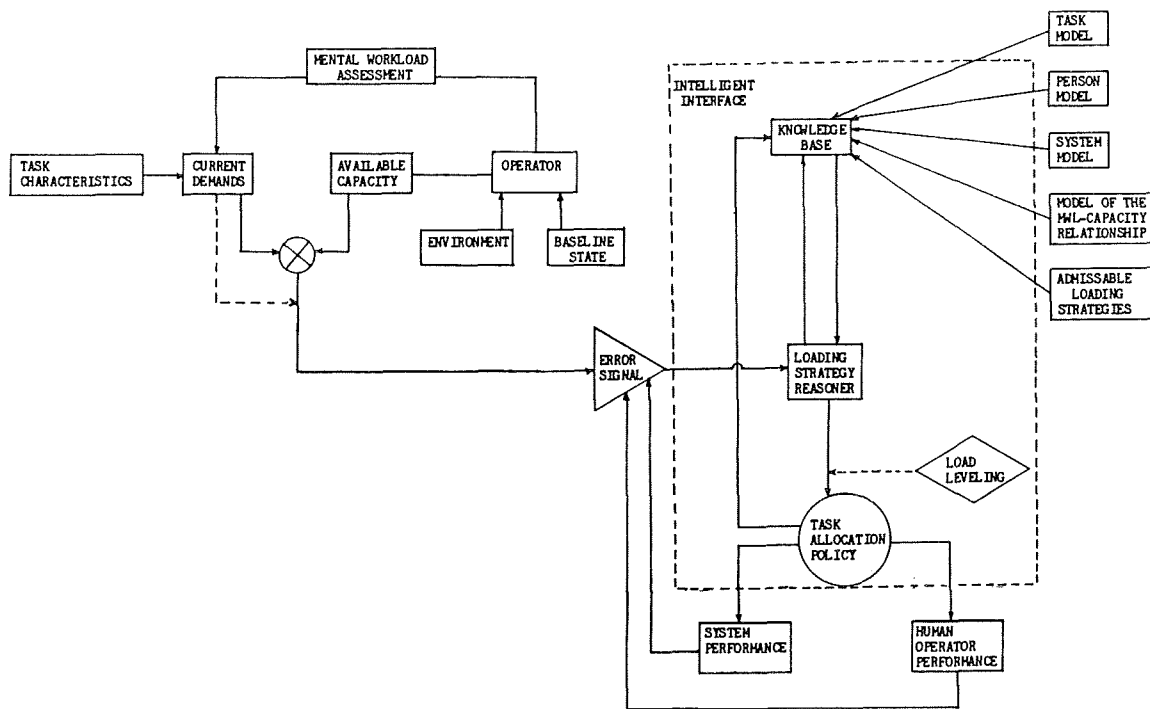
Figure 5. The overall structure of a knowledge-based adaptive mechanism
that could be used as an adaptive interface for dynamic task
reallocation.

Load Leveling and Task Allocation

Adaptive mechanisms of varying sophistication could be developed, differing in the amount and complexity of the reasoning used in load leveling. The simplest system would have a lookup table which assigned a loading level to each task definition. The tabulated task loading would be adjusted on the basis of the apparent effort being used by the person (measured either physiologically or as a subjective rating) in performing the task.

Additional complications would be introduced if a multivariate error signal were used, as would occur if the error signal was constructed using a multiple resource model of attentional capacity and multicomponent analysis of the task. Even so, the reasoning process might not be too complex providing that the redefinition of the task were regarded as a classification task. Conventional expert systems are well equipped to handle the classification problem and existing techniques using a well defined knowledge base and a production rule inference engine (e.g., Hayes-Roth, Waterman & Lenat, 1983) might be sufficient to accomplish this purpose.

The task redefinition process can be viewed as one of classifying a given error signal in terms of a fixed set of task definition choices, as occurs in the table lookup methods. In the more complex versions of KBAM, however, the table lookup procedure of classification is replaced by rule based inference. The success of this strategy will depend to a large extent on the quality of the information stored in the knowledge base, i.e., whether or not the set of allowed task definitions or the specific rules are appropriate. One factor which will affect the complexity of reasoning required will be the variability of the environmental contingencies affecting the task. Tasks can be classified as open, i.e., subject to variable environmental contingencies, or closed, where environmental contingencies do not vary and task restructuring will not be required, except when satisfactory allocation is not possible within the current task structure. An example of this open/closed task distinction occurs in flying where good weather and safe flying conditions will provide a closed task, whereas poor weather and intermittent hazards will result in an open task which requires more flexible responses. In general, the more closed a task is, the easier it will be to model it and build an appropriate adaptive mechanism.

The purpose of KBAM is an allocation of task functions between human and machine which maximizes performance outcome. The process of task reallocation should not be disruptive. A large number of sudden, discrete changes might lead to a worsening, rather than an improvement, in performance. It is likely that smooth and relatively continuous changes in task definition will be preferable.

The manner in which the change in task demands is best communicated to the human operator will depend to some extent on the task being performed. One can distinguish between insidious systems which reallocate tasks without directly warning the human, and conversational systems which signal explicitly each change in the task definition and allocation. Alternatively, the adaptive interface may be consultative, suggesting a better task allocation policy, while allowing the human operator to decide whether or not the suggested task reallocation should take place. In order to minimize a variety of stresses associated with the lack of autonomy, it is suggested that the latter proposal will generally be more useful for dynamic human-machine systems.

Implementation Problems

Development of the type of adaptive mechanism outlined here entails a number of difficulties which are summarized briefly below. It is not clear what task should be used in building a KBAM prototype. A suitable task will have components which can be switched between human and machine, well defined parameters (for task definition and analysis), will test a variety of human resources, and possess continuous measures of success and failure. Flying tasks appear to be appropriate, but they are not readily amenable to experimental manipulation without access to a sophisticated ground-based simulator. Video games are likely to be among the first tasks for which a KBAM prototype is developed.

MWL assessment is a controversial topic. Since KBAM is designed to deal with complex tasks where overloading may be a problem, dual task assessment methods may not be appropriate. Similarly, reliance of subjective ratings would be unwise in situations where the person is fully occupied by the task. We favor using physiological measures in this instance, supplemented with direct assessment of performance. More research is required before MWL measures can be used with confidence in generating an error signal. The development of physiological recording systems which are unobtrusive and reliable is a technical problem that remains to be solved (Hancock, Meshkati, & Robertson, 1985).

As specified earlier, KBAM should operate in close to real time. In the prototype that we are developing, processing tasks are divided between three computers. The first computer carries out data acquisition, while the second does the automated reasoning and the third presents the task. Given current laboratory facilities, it is likely that there will be a delay of approximately half a minute between the physiological response and the resulting task reallocation, with up to 10 seconds being required for each of the three major steps in the process. This may be acceptable in a laboratory demonstration of the concept, but this lag will be excessive and impractical in applications such as flying. While improvements can be made in the speed of data acquisition and task presentation, automated reasoning is likely to remain a bottleneck for the forseeable future. Thus there will be a tradeoff between sophistication of reasoning and response latency with the hardware and techniques available now and in the immediate future.

The final difficulty considered here (although there are others) is that of putting a number of complicated components together into a working system. A working prototype is necessary to demonstrate the feasibility of the concept. Each of the components of KBAM, MWL assessment, modeling of attentional resources, task analysis, automated reasoning, and task reallocation, is a major technical challenge.

Summary

The knowledge-based adaptive mechanism (KBAM) is a powerful method for implementing dynamic reallocation of task components between human and machine. Implementation of this method requires models of attention, cognition and physiological response, as well as expert systems and related techniques. Despite potential problems, the benefits of the KBAM technology justify a large scale research and development effort. As technology advances, particularly in aerospace applications, KBAM systems may be the only way of preserving harmonious, cooperative and successful human-machine relationships.

REFERENCES

Brown, I.D (1967) Measurement of control skills, vigilance and performance of a subsidiary task during 12 hours of car driving. Ergonomics, 10, 665-673.

Chignell, M.H. & Hancock, P.A. (1985) Intelligent human-computer interfaces and their implementation using a knowledge-based systems approach. Unpublished manuscript, Dept. of Industrial & Systems Engineering, University of Southern California.

Edmonds, E.A. (1981) Adaptive man-computer interfaces. In M.J. Coombs and J.L. Alty (eds.), Computing Skills and the User Interface. N.Y.: Academic Press.

Hancock, P.A. & Chignell, M.H. (1985) The principle of maximal adaptability in setting stress tolerance standards. In R. Eberts and C.E. Eberts (eds.), Trends in Ergonomics/Human Factors II. N.Y.: North-Holland.

Hancock, P.A. & Dirkin, G.R. (1983) Stressor induced attentional narrowing: Implication for design and operation of person-machine systems. Proc. Human Factors Association of Canada, 16, 19-21.

Hancock, P.A., Meshkati, N., & Robertson, M.M. (1985) Physiological reflections of mental workload. Aviation, Space and Environmental Medicine, in press.

Hayes-Roth, F., Waterman, D.A. & Lenat, D. (1983) Building Expert Systems. Reading, Mass.: Addison-Wesley.

Kahneman, D. (1973) Attention and Effort. Englewood Cliffs, N.J.: Prentice-Hall.

McCormick, E.J. & Sanders, M.S. (1982) Human Factors in Engineering and Design. N.Y.: McGraw-Hill.

Miller, G.A. (1956) The magical number seven, plus or minus two: Some limits on our capacity for processing information. Psychological Review, 63, 81-97.

Morris, N.M, Rouse, W.B., & Ward, S.L. (1984) Human-computer interaction: A conceptual model. In Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, 178-183.

Peterson, L.R. & Peterson, M.J. (1959) Short-term retention of individual verbal items. Journal of Experimental Psychology, 58, 193-198.

Price, H.E. (1985) The allocation of functions in systems, Human Factors, 27, 33-45.

Wickens, C.D. (1980) The structure of attentional resources. In R.S. Nickerson (ed.), Attention and Performance VIII. Hillsdale, N.J.: Erlbaum.

The Effects of Stress on Attentional Resources

P.A. Hancock and M.H. Chignell

Departments of Safety Science and Human Factors and
Department of Industrial and Systems Engineering
University of Southern California
Los Angeles, CA 90089

Abstract

This paper presents a new perspective from which to view the action of stress on human behavior. At a behavioral level, the action of stress is related to contemporary notions of human attention and an indication of an isomorphic relationship between modes of control at a physiological and behavioral level is presented. Examples of this phenomenon are extracted from performance under heat stress, since this is one of the most simple stress circumstances. We suggest that stress sufficient to overcome adaptive capability, that is efficient homeostasis, acts to drain attentional resources. The manner in which such resources fail approximates that function typical of a positive feedback system, which also characterizes the breakdown of physiological response under severe environmental stress. The end point of this draining sequence is the absence of all attentional resources, which we take to be unconsciousness, to be rapidly followed by the failure of physiological adaptability upon which life sustaining functions depend. This overall picture preserves the inverted-U shaped relationship between stress and performance, yet is in distinct contrast to the traditional arousal account of such behavior. The theoretical and practical ramifications of these observations are explored (see also Hancock & Chignell, 1985).

Reference

Hancock, P.A. & Chignell, M.H. (1985) The principle of maximal adaptability in setting stress tolerance standards. In. R. Eberts and C. Eberts (Eds.). Trends in Ergonomics/Human Factors II, North-Holland, Amsterdam.

Christopher D. Wickens   University of Illinois
John Zenyuh   Institute for Aviation
Victor Culp   Savoy, Illinois 61874
William Marshak

The Effects of Voice and Manual Control Mode on Dual Task Performance

Two fundamental principles of human performance--compatibility and resource
competition, are combined with two structural dichotomies in the human in-
formation processing system--manual versus voice output, and left versus
right cerebral hemisphere--in order to predict the optimum combination of
voice and manual control with either hand, for time-sharing performance of
a discrete and continuous task.

Eight right handed male subjects performed a discrete first-order tracking
task, time-shared with an auditorily presented Sternberg Memory Search Task.
Each task could be controlled by voice, or by the left or right hand, in all
possible combinations except for a dual voice mode.

When performance was analyzed in terms of a dual-task decrement from single
task control conditions, the following variables influenced time-sharing
efficiency in diminishing order of magnitude, (1) the modality of control--
discrete manual control of tracking was superior to discrete voice control
of tracking and the converse was true with the memory search task (2) response
competition--performance was degraded when both tasks were responded manually
(3) hemispheric competition--performance degraded whenever two tasks were
controlled by the left hemisphere (i.e., voice or right handed control). The
results confirm the value of predictive models invoice control implementation.

EVALUATION OF TWO COGNITIVE ABILITIES TESTS
IN A DUAL-TASK ENVIRONMENT

Michael A. Vidulich
NASA-Ames Research Center
Mail Stop 239-3
Moffett Field, CA 94035

and

Pamela S. Tsang*
NASA-Ames Research Center
Mail Stop 239-21
Moffett Field, CA 94035

ABSTRACT

Most real-world operators are required to perform multiple
tasks simultaneously. In some cases, such as flying a high-
performance aircraft or trouble-shooting a failing nuclear
power plant, the operator's ability to "time-share" or "pro-
cess in parallel" can be driven to extremes. This has
created interest in selection tests of cognitive abilities.
Two tests that have been suggested are the Dichotic Listening
Task and the Cognitive Failures Questionnaire. Correlations
between these test results and time-sharing performance were
obtained and the validity of these tests were examined. The
primary task was a tracking task with dynamically varying
bandwidth. This was performed either alone or concurrently
with either another tracking task or a spatial transformation
task. The results were: (1) An unexpected negative correla-
tion was detected between the two tests. (2) The lack of
correlation between either test and task performance made the
predictive utility of the tests scores appear questionable.
(3) Pilots made more errors on the Dichotic Listening Task
than college students.

INTRODUCTION

Many complex operational tasks, such as flying high-performance
aircraft, air-traffic control, or controlling a nuclear power plant in
an emergency, can be very unforgiving of errors. Therefore, it is
highly desirable that the operators in charge of such tasks be as
unlikely to commit an error as possible. Traditionally, the operator's
training was expected to minimize error probability. However, training

* National Research Council Research Associate at NASA Ames Research
Center

alone is often not the most cost-effective solution. Most notable is the problem that some people seem to be less able to learn a task than others. This is evident from the high wash-out rate found in many training programs. The resources spent on individuals who ultimately do not finish the training program are unavailable to those that do. Consequently, it is highly desirable to identify individuals who are likely to successfully complete the training program before training commences.

Of course, there are many factors that could be involved in failing to complete a training program. A few obvious examples are poor motivation, inadequate sensory acuity, inability to cope with stress, or insufficient cognitive capacity. Motivation is difficult to test in a laboratory, but since many of the jobs that have high wash-out rates are highly sought after, it seems likely that the typical trainee is well motivated. Sensory acuity can generally be measured quite accurately to insure that trainees meet an acceptable level. In general then, the most pressing need appears to be in the identification of individual differences in cognitive capacities and ability to cope with stress.

In view of the fact that cognitive control is likely to be related to performance on complex tasks, the present paper examines the relationship between time-sharing performance and two tests of cognitive abilities that have been proposed: the Dichotic Listening Task and the Cognitive Failures Questionnaire. The Dichotic Listening Task was developed by Gopher and Kahneman (1971) and is intended to test how well individuals can focus and switch attention to dichotic stimuli (i.e., different auditory stimuli simultaneously presented to each ear). The Dichotic Listening Task score (error) has been found to correlate negatively with success in flight training, to discriminate between transport pilots and fighter pilots (Gopher, 1982), and to correlate with accident proneness in bus drivers (Kahneman, Ben-Ishai, & Lotan, 1973). The Cognitive Failures Questionnaire (Broadbent, Cooper, FitzGerald, & Parkes, 1982) is a series of questions concerning the frequency of failures in perception, memory, and motor function. Through their own research as well as reviews of other's, Broadbent et al. found that the Cognitive Failures Questionnaire score is fairly stable over time. More relevant to the present paper is their finding that the various kinds of failures (i.e., perceptual, memory, or motor) all seem to occur in the same person and need not be treated as separate categories. Broadbent et al. argued that this would support the notion of some deficiency existing in overall cognitive control and that the Cognitive Failures Questionnaire score seemed to be a measure of a general likelihood of failures. So far, they have not found any significant relationships between the Cognitive Failures Questionnaire score and short-term memory, long-term memory, or dual-task performance. What they did find, suggested that the Cognitive Failures Questionnaire score would be a good indicator of how resistant an individual is to stress.

In the present experiment the two tests were administered to a group of pilots and a group of students. Performance measures on a variety of single- and dual-tasks at various levels of difficulty were

12.2

obtained. Discussion of the results will focus on the degree to which scores on the two tests are related to each other and the extent to which either test's scores is related to the single-task and dual-task performance. Inasmuch as both tests had been found to be related to the tendency of an individual to commit errors, it was expected that scores on the Cognitive Failures Questionnaire and the Dichotic Listening Task would be positively correlated. This expectation was based on the assumption that the attentional abilities evaluated by the Dichotic Listening Task might underly the "overall cognitive control" postulated by Broadbent et al. as responsible for the Cognitive Failures Questionnaire scores.

Both tests were also expected to correlate with performance, especially time-sharing performance. Correlations between good cognitive abilities, as tested by these procedures, and single-task performance would not be a problem in and of itself. But since the dual-task trials employed in the present experiment involved dynamically changing difficulty in a high workload task (and hence a potentially stressful situation), it was expected that the propensity towards cognitive failure or attentional misdirection evaluated by the tests would be manifested in the dual-task performance scores. Individuals with better (lower) scores on these tests were therefore expected to show better time-sharing performance on the experimental tasks. Although, Broadbent et al. (1982) did not obtain any correlations between the Cognitive Failures Questionnaire score and dual-task performance, a replication seemed justified. First, very little procedural detail was provided in Broadbent et al.'s review. Second, the results from the present experiment could be compared with a another objective measure - that provided by the Dichotic Listening Task.

METHOD

Subjects

Twenty-four male subjects served as paid participants. Half of the subjects were pilots (with an average age of 28.8 years) and half were college students (with an average age of 21.3 years). All but two of the pilots were instrument rated and all but one had a commercial pilot's license, an instructor pilot's license, or both. Total flight time for the pilots varied from 120 hr to 2000 hr with a mean of 863 hr.

Apparatus

The experimental tasks were implemented on a PDP 11/34 minicomputer. Visual displays were presented on a CRT screen in front of the subjects and auditory stimuli were presented through stereo headphones. A joystick was mounted on the right armrest of the chair. Either another joystick or a set of eight microswitches arranged in a circle could be mounted on the left armrest. Subjects' vocal responses were processed via a Votan speech recognition device.

Tasks

Two basic tasks were used in this experiment: a tracking task and a transformation task. The tracking task was a one-dimensional compensatory tracking task with first-order control dynamics. Three levels of constant bandwidth were used: .3 Hz, .5 Hz, and .7 Hz. Also, the bandwidth could vary dynamically within a trial; ranging from .3 Hz to .7 Hz. In a given trial, the right-hand tracking task could be any one of the four levels (i.e., .3 Hz, .5 Hz, .7 Hz, or variable within the trial), but the left hand tracking always had a constant .5 Hz bandwidth.

The second task was a spatial transformation task. Stimuli designating one of eight compass directions (north, northeast, east, etc.) were presented one at a time. Subjects were required to respond with the next direction in a clockwise direction. The initial direction could be indicated either visually by the appearance of a tick mark on the CRT or auditorily by a tone of specific pitch and channel (ear). The subjects' responses could be either manual, via the microswitches on the left armrest, or vocal, via the voice recognition device. Therefore, the transformation task could be presented in any one of four possible input/output (I/0) configurations: visual/manual (VM), auditory/manual (AM), visual/speech (VS), or auditory/speech (AS).

The tracking task and the transformation task were first performed as single-tasks. In the dual-task conditions, the right-hand tracking tasks (either .5 Hz or variable bandwidth) was paired with either the left-hand tracking or one of the four transformation tasks. After some initial single- and dual-task training, a secondary task technique was adopted and the right-hand task was designated as the primary task. Subjects were instructed to maintain the primary task performance constant at the single-task level. This was to be achieved by allocating the appropriate amount of resources to the concurrent tasks according to the changes in the difficulty in the primary task. There were 10 experimental sessions. A more detailed experimental design is described in Tsang (1985).

Cognitive Ability Tests

The Dichotic Listening Task consisted of a series of 48 trials recorded on a cassette tape. Each trial consisted of two simultaneous messages, one presented to each ear. The messages were made up of simple words with a few digits embedded in each message. Each trial was divided into two sections. The first section of the trial was intended to evaluate the ability to focus attention; the second section the ability to switch attention. In section I, the subject's task was to focus on the ear indicated by a tone. Upon detecting any digits in the appropriate ear, the subject wrote it down on the appropriate line of a prepared form. A second tone indicated the beginning of section II. The subject was required to switch attention to the other ear if the second tone was different from the first. The subject's task was to record the

digits presented to the relevant ear throughout the two sections of the trial. Any deviations from the correct sequence were recorded and categorized as omissions and intrusions in the first section and as switching errors in the second section. The two sections of each trial were scored separately. Poor performance was indicated by a high total error score. The first 12 trials of the 48 run were not included in the final scores, because several subjects showed extreme practice effects during this period. The stimulus tape was an English version (Braune & Wickens, 1983) of the original Hebrew Dichotic Listening Task (Gopher & Kahneman, 1971). The entire tape, including the instructions, took approximately 35 min to complete.

The Cognitive Failures Questionnaire was taken directly from Broadbent et al. (1982). The questionnaire consisted of 25 questions describing common cognitive failures most people experience (e.g., "Do you find you confuse right and left when giving directions?"). On a five-point scale of frequency, ranging from Very Often (4) to Never (0), the subject simply circled a response to indicate how often the described event had happened in the previous 6 months. The frequency score for each response was totaled to generate the subject's score. A high frequency of cognitive failures was indicated by a high score. The Cognitive Failures Questionnaire was administered immediately after the subject had completed the Dichotic Listening Task. Both tests were administered individually.

RESULTS

The results of the present study will focus on three issues: (1) the correlation between the two tests, (2) the relationship between the test scores and performance on the experimental tasks, and (3) the effect of the background of the subjects on the test scores. Each of these topics will be dealt with in turn.

Inter-Test Correlations

Three Pearson's product-moment correlations were obtained for the ability tests; the inter-test correlations listed in Table 1 are the correlation between the two sections of the Dichotic Listening Task (Dichotic I and Dichotic II), and each with the Cognitive Failures Questionnaire (CFQ) scores. The critical r for two-tailed test (df = 22, p < .05) is .404 (Edwards, 1984). As shown in Table 1, the only positive correlation that met this criterion was the correlation between the two sections of the Dichotic Listening Task. This implies that the ability to focus attention and the ability to switch attention may be related or that focusing attention plays an important role in both sections. The unexpected negative correlation between the Cognitive Failures Questionnaire score and the Dichotic I score was significant at .05 level and that with the Dichotic II score at .1 level. These results show that the individuals who reported themselves more likely to experience cognitive failures were able to perform the Dichotic Listening Task better.

Table 1

Inter-Test Correlations

Tests	r
Dichotic I/Dichotic II	.582
Dichotic I/CFQ	-.423
Dichotic II/CFQ	-.344

Correlations between the Test Scores and Performance

Performance measures obtained included Root Mean Square Error (RMSE) for the tracking task, reaction time (RT) and percent error for the transformation task. Decrement scores, generated by subtracting the corresponding single-task performance score from any given dual-task score, were used to the dual-task analyses. Decrement scores were used to remove the effects of difficulty differences present in the single-task conditions and to isolate the magnitude of interference caused by the performance of the concurrent task. The single-task data reported here were obtained in Session 4 (last session before any dual-tasks were introduced); dual-task data in Session 7 (last dual-task session before the secondary task technique was adopted) and Session 10 (last session of the experiment).

Correlations were performed to assess the relationship between whatever abilities that are assessed by the tests and the abilities required to perform the experimental tasks. Any use of the test scores as a predictor variable presupposes that such a relationship exists. Table 2 displays the findings of these analyses: correlations between test scores and single-task performance are on top and correlations between test scores and dual-task performance are at the bottom. A positive correlation represents better performance being associated with superior (i.e., lower) test scores. In contrast, a negative correlation indicates that more reported cognitive slips on the Cognitive Failures Questionnaire or more errors on the Dichotic Listening Task are associated with better performance. In Table 2, correlations that are significantly different from zero are marked with an asterisk (2-tailed critical r (df=22) = .404, $p < .05$).

Four significant correlations between the Cognitive Failures Questionnaire and performance were obtained. However, three of the four significant correlations were negative. The subjects who reported experiencing more frequent cognitive failures tended to perform better on the single-task trials. The sole significant positive correlation occurs with the RT decrements of the dual-task trials. None of the remaining four dual-task correlations approached significance.

The two sections of the Dichotic Listening Task both show the same trends. In neither case, is there a significant positive correlation between the test scores and any dual-task performance measures. In fact, Dichotic I correlates negatively with the transformation task's RT decrements. The only strong positive correlation is with the single-task transformation task RTs. The unexpected lack of correlation with dual-task performance is problematic. Neither test demonstrated a reliable relationship with time-sharing performance in the present experiment.

Table 2

Correlations between Test Scores and Performance

Performance Measure Type	Test Score Type		
	CFQ	Dichotic I	Dichotic II
Single-Task Performance			
Left-Hand RMSE	-.415*	.236	-.148
Right-Hand RMSE	-.512*	.312	-.076
Transformation RT	-.543*	.675*	.635*
Transformation % Error	-.140	.305	.028
Dual Tracking Performance			
Left-Hand RMSE Decrement	-.187	-.225	-.077
Right-Hand RMSE Decrement	-.008	.148	.108
Transformation/Tracking Performance			
Right-Hand RMSE Decrement	-.092	-.098	-.205
Transformation RT Decrement	.410*	-.420*	-.173
Transformation % Error Decrement	.156	.296	-.028

*$p < .05$.

Background Effects

Table 3 displays the mean errors on the two sections of the Dichotic Listening Task obtained from the students and the pilots separately. The students committed significantly fewer omissions or intrusions in Section I ($t(22) = 1.82$, $p < .05$) and made fewer errors on Section II of the Dichotic Listening Task ($t(22) = 1.64$, $p < 0.1$). No significant difference was found between the students and the pilots on the Cognitive Failures Questionnaire (Student Mean = 38.6; Pilot Mean = 38.2). Previous results (Tsang, 1985) also indicated that there were no substantial difference in performance between these two groups.

12.7

Table 3

Students vs. Pilots
on the Dichotic Listening Task

Section	Student Errors	Pilot Errors
I	6.08	10.92
II	1.83	4.33

DISCUSSION

There are three issues to be discussed: (1) the negative
correlation between the Cognitive Failures Questionnaire and the two
sections of the Dichotic Listening Task, (2) the correlations between
the test scores and performance, and (3) the student vs. pilot
difference in the Dichotic Listening Task score.

Negative Inter-Test Correlations

Part of the inspiration for this study arose from Broadbent et
al.'s (1982) suggestion that an objective correlate of Cognitive
Failures Questionnaire would be useful. Hopefully, such a correlated
test would be free of the "problem of defensive unwillingness to admit
error," (Broadbent, et al., 1982, p. 12). Broadbent et al. reviewed
several attempts to find such a correlate. Most attempts centered
around some test of memory performance; none achieved very promising
results. The present study was undertaken to see if the cognitive
failures reported in the questionnaire were related to a subject's
attentional control capabilities as detected on the more objective
Dichotic Listening Task.

Surprisingly, not only were the correlations not significantly
positive, they tended to be negative. These results caution against
relying heavily on either of these tests as a selection tool or a
classification criterion of performance on complex tasks. Replications
of these results will, of course, be required and if negative
correlations persist, reinterpretation of one or both tests may be
unavoidable.

Correlations between Test Scores and Performance

The general paucity of positive correlations between the test
scores and the performance measures is troublesome. It is tempting to
explain the overall lack of positive correlations in this experiment as
a result of insufficient statistical power to detect small, but

important, effects. However, this explanation does not account for the disturbing presence of the negative correlations. Nor does it account for the fact that the strongest correlations obtained with the test scores were with various measures of single-task performance.

Both tests had been expected to correlate better with the dual-task performance measures. The Dichotic Listening Task was expected to correlate better with dual-task performance because the continuous control of attention allocation was believed to be a major determinant of the dual-task performance in the present experiment. However, it is conceivable that the mechanism required for continuous attention division may be independent from attention switching. The latter being postulated to be highly related to the Dichotic Listening Task score. The Cognitive Failures Questionnaire was expected to correlate better with dual-task performance because the dual-task conditions were expected to induce higher levels of stress. But as in Broadbent et al.'s findings, no significant relationship between the Cognitive Failures Questionnaire score and dual-task performance was obtained here.

Taken as a whole, the results of this investigation suggest that the utility of these tests as predictor variables of performance in dual-task laboratory research is quite limited. It is possible that the Dichotic Listening Task will correlate with other tasks which emphasizes the switching of attention rather than its sharing. However, the present findings suggest that the predictability of the Dichotic Listening Task scores on dual-task performance may be highly task specific.

Student/Pilot Differences

One possible explanation for the difference between the students and the pilots on the Dichotic Listening Task may concern the pilots' hearing. It is possible that the pilots' hearing may have been suboptimal due to exposure to the noisy aviation environment. Whatever the explanation, the present finding suggests the possibility that experience as a pilot may be disruptive to good performance on the test. The implication is that caution must be exercised when the Dichotic Listening Task is used as a pilot trainees selection tool, especially when the pool of applicants have different levels of piloting experience and possibly various degrees of hearing damage. Again, this is a result that requires replication and careful consideration before application of the test should be taken for granted.

REFERENCES

Braune, R. J. & Wickens, C. D. (1983). The functional age profile: An objective decision criterion for the assessment of pilot performance capacities and capabilities. In R. S. Jensen (Ed.), Proceedings of the Second Symposium on Aviation Psychology (pp. 437-444). Columbus: Ohio State University Aviation Psychology Laboratory.

Broadbent, D. E., Cooper, P. F., FitzGerald, P., & Parkes, K. R. (1982). The Cognitive Failures Questionnaire (CFQ) and its correlates. British Journal of Clinical Psychology, 22, 1-16.

Edwards, A. L. (1984). An Introduction to Linear Regression and Correlation (2nd edition). New York: Freeman.

Gopher, D. (1982). A selective attention test as a predictor of success in flight training. Human Factors, 24, 173-183.

Gopher, D. & Kahneman, D. (1971). Individual differences in attention and the prediction of flight criteria. Perceptual and Motor Skills, 33, 1335-1342.

Kahneman, D., Ben-Ishai, R., & Lotan, M. (1973). Relation of a test of attention to road accidents. Journal of Applied Psychology, 58, 113-115.

North, R. A., & Gopher, D. (1976). Measures of attention as predictors of flight performance. Human Factors, 18, 1-14.

Tsang, P. S. (1985). Can pilots time-share better than non-pilots. In Proceedings of the Third Symposium on Aviation Psychology. Columbus: Ohio State University Aviation Psychology Laboratory.

At Least Some Errors are Randomly Generated
(Freud was Wrong)

Abigail J. Sellen and John W. Senders

Dept. of Industrial Engineering,
University of Toronto,
Toronto, Ont., Canada.
M5S 1A4

An experiment was carried out to expose something about human error generating mechanisms. In the context of the experiment, an error was made when a subject pressed the wrong key on a computer keyboard or pressed no key at all in the time allotted. These might be considered, respectively, errors of substitution and errors of omission.

Each of seven subjects saw a sequence of three digit numbers, made an easily learned binary judgement about each, and was to press the appropriate one of two keys. Each session consisted of 1000 presentations of randomly permuted, fixed numbers broken into 10 blocks of 100. One of two keys should have been pressed within one second of the onset of each stimulus.

These data were subjected to statistical analyses in order to probe the nature of the error generating mechanisms. Goodness of fit tests for a Poisson distribution for the number of errors per 50 trial interval and for an exponential distribution of the length of the intervals between errors were carried out. Given the resulting Chi-square values, we cannot reject the hypothesis that a constant probability generator is operating. Thus, there is evidence for an endogenous mechanism that may best be described as a random error generator. Furthermore, an item analysis of the number of errors produced per stimulus suggests the existence of a second mechanism operating on task-driven factors producing exogenous errors. Some errors, at least, are the result of constant probability generating mechanisms with error rate idiosyncratically determined for each subject.

PILOT INTERFACE WITH FLY BY WIRE CONTROL SYSTEMS

William W. Melvin
Chairman, Airworthiness & Performance Committee
Air Line Pilots Association
1101 West Morton
Denison, Texas 75020

ABSTRACT. Aircraft designers are rapidly moving toward full fly by wire control systems for transport aircraft. Aside from pilot interface considerations such as location of the control input device and its basic design such as side stick, there appears to be a desire to change the fundamental way in which a pilot applies manual control. A typical design would have the lowest order of manual control be a control wheel steering mode in which the pilot is controlling an autopilot. This deprives the pilot of the tactile sense of angle of attack which is inherent in present aircraft by virtue of certification requirements for static longitudinal stability whereby a pilot must either force the aircraft away from its trim angle of attack or trim to a new angle of attack. Whether or not an aircraft actually has positive stability, it can be made to feel to a pilot as though it does by artificial feel. Artificial feel systems which interpret pilot input as pitch rate or G rate with automatic trim have proven useful in certain military combat maneuvers, but their transposition to other more normal types of manual control may not be justified.

INTRODUCTION. The purpose of this paper is to describe what may be a problem of pilot interfacing with fly by wire control systems. To do so it is necessary to describe some diferences between manual and automatic control of aircraft which explains the evolution of the problem and why some would so easily accept a change in the basic way a pilot flies an airplane. This paper should not be construed as an argument against fly by wire. It isn't. Rather, it is an argument for a fly by wire control system that interfaces with the pilot in a manner similar to current airplanes with good flying characteristics.

MANUAL CONTROL. During the evolution of the modern airplane many examples of unstable aircraft have brought considerable grief to their pilots. Consequently, the FARs and MILSPECs specify stability requirements for aircraft which produce a consistency of feel that pilots have learned to depend upon.

Perkins and Haig (1) clearly identified the fact that "...the airplane's speed is determined by the value of the

equilibrium lift coefficient, while its rate of climb or descent is regulated principally through the throttle control." for climbing and descending flight. Others (2-12) agree and have expanded upon the conclusion that net thrust over drag defines climb/descent angle while angle of attack defines airspeed except that small temporary changes in angle of attack are required to make temporary changes in lift required to change an aircraft's inertial trajectory.

With static longitudinal stablilty, a pilot knows his aircraft will not diverge far from its trimmed angle of attack without his input which provides an important tactile feedback of angle of attack. If a pilot tilts the lift vector by banking, he can depend upon the fact that he will need to increase the angle of attack to increase total lift so that the vertical component will equal weight. Lateral stability requires that a pilot hold an aircraft into a bank. Increasing angle of attack tilts the lift vector aft increasing induced drag which requires an increase in thrust to maintain equilibrium. When induced drag is a large percentage of total drag (low speed flight), then larger thrust increases are required. Although pilots in flight do not analyze these factors any more than birds do, the above characteristics are ones that pilots have learned to depend upon. They make up the feel of an aircraft which a pilot learns to balance much like a bicycle rider rides a bicycle. Pilots do not have to know anything about the forces that make up this feel any more than a bicycle rider has to know why he doesn't fall off when he goes around a corner.

It is very important to understand how pilots interface with current aircraft. Primarily we fly with visual feedback of our control inputs which is complemented with tactil feedback. An example of a similar task is that of steering a car. A driver has immediate visual feedback of control inputs which are complemented by a tactile feel of the steering wheel because it is self centering to the straight line condition.

An airplane is controlled by a pilot in a similar manner. He has instant visual sensing of control inputs with a self centering action to the wings level, trimmed angle of attack, condition which can be climbing or descending depending upon thrust.

Concerning static longitudinal stability, MIL-F-8785C states "For levels 1 and 2 there shall be no tendency for airspeed to diverge aperiodically when the airplane is disturbed from trim with the cockpit controls fixed and with them free." and "Alternatively, this requirement will

be considered satisfied if stability with respect to speed is provided through the flight control system, even though the resulting pitch control force and deflection gradients may be zero." Moorhouse and Woodcock in discussing the above (13) say "...it should be noted that zero speed stability removes an airspeed cue that pilots sometimes find valuable, particularly at low speed; and that automatic trimming has been known to lead to an insidious slowdown of some aircraft to stall when the pilot holds a small back force."

Despite the above, there have been many attempts to change these basic characteristics. Sometimes the flying qualities have been bad. Instead of improving the flying qualities, attempts have been made to aid the pilot by relieving him of some perceived workload such as holding a desired pitch or bank angle, thus removing the pilot from the control loop. Pilots generally do not fly by attempting to hold a desired pitch, but autopilots do.

AUTOMATIC CONTROL. The issue of instrument flight brings into focus the differences between manual and automatic control. Under visual meterological conditions with manual control, the pilot feels the balance of his aircraft better than under any other circumstance. Under instrument conditions the pilot is hampered by inadequate displays which do not give him the intuitive assessment of aircraft trajectory and position that occurs with visual reference and some improved instrument displays.

If it can be accepted that pilots generally have no problem flying manual approaches to very good landings under visual conditions, the obvious question is why can't they perform equally well under instrument conditions? The equally obvious answer has to be the difference in information presented to the pilots and how they interpret it. The primary reference for instrument flight is a gyro horizon which displays airplane bank angle and pitch angle. Pitch does not tell a pilot where the airplane is going. Under visual conditions he not only perceives pitch, but trajectory. When this vital information is provided to him in a intuitively assessable manner he begins to perform under instrument conditions similar to visual conditions.

Approach couplers were designed to fly in the manner of their parent autopilots, i.e., they waved the elevator up and down at the glide slope, leaving airspeed control to the throttles. Pitch command on flight directors was designed to tell the pilot to do what the approach coupler was doing.

14.3

As pointed out by Perkins and Haig (1) and others (2-12),
If an aircraft is descending too steeply but at the proper
airspeed, its basic need to correct its descent path is an
increase in thrust with a temporary increase in angle of
attack to temporarily increase lift to redirect the
inertial vector. Thereafter it will fly at its newly
defined descent angle at its trimmed angle of attack. An
approach coupler will not function in this manner, but will
instead increase pitch to some precomputed value dependent
upon its glideslope deviation. Then the airspeed will
decrease and the pilot or autothrottle will attempt to
restore the airspeed, except it will take more thrust than
is necessary for the stable condition and a subsequent
change will have to be made. Direct lift control was
invented to help the autopilot fly in this manner.

Pilots at first had difficulty using flight directors and
approach couplers but they soon learned to anticipate the
thrust changes that accompany the pitch changes and so came
to accomodate the manner in which an approach coupler flies
and a flight director directs. Autothrottles were not
satisfactory until they incorporated pitch anticipatory
circuits which told them ahead of time that the pitch was
being changed so begin a thrust change in anticipation of
the new requirements. Thus the automatic system learned to
coordinate its inputs just like the accomplished pilot had
been doing. Nevertheless, pitch changes which are adequate
for normal conditions are very inadequate for the adverse
conditions of strong wind shear (14).

Instrument displays with flight path angle and flight path
command instead of pitch command provide the pilot with
intuitive assessment of his trajectory relative to a
desired value, an instant recognition of his departure from
a desired flight path and the proper command to return.
With a properly designed system it is not possible to
center the flight path command unless the aircraft is
actually correcting to the desired flight path at the
desired rate. Such is not the case with current flight
directors in the presence of strong wind shear. In
addition, the difference between pitch and flight path
angle is geometric angle of attack which is a highly
desirable performance parameter.

The practice of providing superior information to an
automatic system and leaving the pilot to monitor the
performance of the automatic system is undesirable as
humans are very poor monitors of automatic systems. An
important causal factor in a recent accident (15) was the
pilots' over reliance upon an automatic system to perform
its intended function. A contributing factor was the very

high workload the pilots were faced with. Automatic systems intended to reduce pilot's workload do not necessarily perform this function especially when they remove the pilot from the control loop (16). Automatic systems have advanced toward a goal of total automatic control of the aircraft while little attention has been paid to the needs of pilots to enhance their performance under instrument conditions with manual control.

STABILITY AUGMENTATION SYSTEMS. Stability augmentation systems can be used on any aircraft and can be as simple as a yaw damper. However, recently they have been incorporated with fly by wire automatic systems to enhance stability of aircraft which are inherently unstable. The major use has been for military aircraft but they are now being considered for commercial aircraft. With a fly by wire system it would be possible to tailor the aircraft response to pilot inputs in a variety of ways. For instance the aircraft could be tailored to respond like an aircraft which perfectly complied to the certification regulations. It could fly a trimmed angle of attack with an appropriate stick force gradient for deviation. Phugoid dampening could be incorporated if desired or the phugoid could be left as in conventional aircraft for the pilot to dampen.

Aircraft simulators are in fact fly by wire systems which have been designed to duplicate a particular aircraft's flying qualities, bad as well as good, and sometimes not too accurately. However, with a fly by wire aircraft it should be possible to have the aircraft behave very much like conventional aircraft. Unfortunately the fly by wire systems are being designed by the same people who formerly designed autopilots and there is an indication they think pilots should fly and think like autopilots. In fact there is some indication of a desire not to let the pilot have an aircraft with synthetic stability like a conventional aircraft but instead to have him control the aircraft through an autopilot. The lowest order of control that will be possible will be a control wheel steering (CWS) mode which removes the pilot from the feel of the stability (real or synthetic) of his aircraft. CWS uses the normal controls of an aircraft to make inputs to an autopilot instead of using separate turn knobs and climb/descent wheels. In this mode the pilot does not have the tactile feedback of centering to the wings level position nor the feedback of trim stability, i.e., a tactile feel of angle of attack. Trim is done automatically by the autopilot. This could cause serious problems under adverse conditions as well as transition problems to and from other aircraft.

One reason for the above may be the fact that designers are tending toward side stick control for transport aircraft, the avowed purpose of which is to open up space on the instrument panel. The primary reason for a side stick in military aircraft is for pilot control in high G maneuvers. If space on the instrument panel is the primary incentive for transport aircraft, a center stick would accomplish this objective as well as the broilly handles used by Boeing in part of their SST development, both of which are controllable with either hand. The side stick will not be easily controllable with either hand. That could be one reason some designers want it to act as an autopilot controlling device instead of a conventional control system.

In certain combat maneuvers such as bombing, straffing, fighter tactics, etc., it has been found desirable for military aircraft to use a control mode where the aircraft automatically trims. In these maneuvers the pilot usually has a Heads Up Display (HUD) with flight path angle and some type of flight path command. Even with automatic trim the control system can still exhibit the other characteristics of a conventional aircraft or it can act as an autopilot with control wheel steering.

CONCLUSION. New aircraft are being designed which do not comply with the basic certification requirements regarding stability. Automatic systems are being designed to cause the aircraft to meet the stability requirements, but there appears to be little concern from regulatory authorities to require such systems to provide a pilot interface similar to that which is inherent with aircraft having natural stability. The manufacturers apparently desire to change the basic manner in which a pilot interfaces with the control system. This objective has not been validated with sufficient research to prove the desirability from a piloting standpoint. Problems may occur with pilots transitioning to and from other aircraft; pilots will probably not want to give up the tactile feel of angle of attack and wings level centering which is present with conventional control systems; and automatic trim could be dangerous in low speed flight.

Research should be done with a variety of systems which should include at least one which attempts to duplicate the best current conventional control systems incorporated with improved instrument displays expected to enhance a pilots performance with current control systems. Also the use of a side stick with a conventional system (not CWS) should be validated; and if fouund objectionable, alternate solutions

should be implemented.

REFERENCES.

1. PERKINS, C. & HAGE, R., Airplane Performance Stability and Control.

2. HURT, H., Aerodynamics for Naval Aviators.

3. KERMODE, A., Introduction to Aeronautical Engineering, Vol. 1, Mechanics of Flight.

4. TOWER, Basic Aeronautics.

5. UNIVERSITY INSTITUTE OF AVIATION, IL., Fundamentals of Aviation and Space Technology.

6. DIEHL, Engineering Aerodynamics.

7. LANGEWIESCHE, W., Stick and Rudder.

8. KERSHNER, W., Private Pilot's Flight Manual.

9. REITHMAIER, L., Pilot's Handbook of Instrument Flying.

10. HOLLAND, J., Learning to Fly.

11. COWLEY, Aeronautics in Theory and Experience.

12. MELVIN, W., The Bastard Method of Flight Control, Pilot Safety Exchange Bulletin, Flight Safety Foundation, March/April 1976.

13. MOORHOUSE, D. & WOODCOCK, R., Background Information and User Guide for MIL-F-8785C, Military Specification - Flying Qualities of Piloted Airplanes, AFWAL-TR-81-3109.

14. MELVIN, W., Effects of Wind Shear on Approach With Associated Faults of Approach Couplers and Flight Directors, AIAA 69-796.

15. SAS Flight 901, JFK, February 28 1984.

16. MELVIN, W., A Philosophy of Automation, SAE 831501.

INVESTIGATION OF LIMB-SIDESTICK DYNAMIC INTERACTION WITH ROLL CONTROL

Donald E. Johnston
Duane T. McRuer
Systems Technology, Inc.
13766 S. Hawthorne Blvd.
Hawthorne, CA 90250
(213)679-2281

ABSTRACT

A fixed-base simulation was performed to identify and quantify interactions between the pilot's hand/arm neuromuscular subsystem and such features of typical modern fighter aircraft roll rate command control system mechanizations as

● force sensing side-stick type manipulator

● vehicle effective roll time constant

● flight control system effective time delay

The simulation results provide insight to high frequency PIO (roll ratchet), low frequency PIO, and roll-to-right control and handling problems previously observed in experimental and production fly-by-wire control systems. The simulation configurations encompass and/or duplicate several actual flight situations, reproduce control problems observed in flight, and validate the concept that the high frequency nuisance mode known as "roll ratchet" derives primarily from the pilot's neuromuscular subsystem. The simulations show that force-sensing side-stick manipulator force/ displacement/command gradients, command prefilters, and flight control system time delays need to be carefully adjusted to minimize neuromuscular mode amplitude peaking (roll ratchet tendency) without restricting roll control bandwidth (with resulting sluggish or PIO prone control).

The results further demonstrate that roll ratchet tendency, which is difficult to detect in fixed-base simulations, is readily apparent from application of frequency response spectral analysis techniques. Consequently the application of appropriate spectral measurement techniques during flight control system design/development piloted simulation phases promise to reduce later and more costly flight test investigation.

INTRODUCTION

Almost every new aircraft with fly-by-wire or command augmentation (Fig. 1) in the roll axis has encountered either Pilot-Induced Oscillations (PIO) or roll ratcheting (or both) in early flight phases. PIO has typically been associated with high gain, neutrally stable closed-loop pilot-vehicle control oscillations with a frequency of about 1/2 Hz. The "roll ratchet" has been somewhat more obscure and idiosyncratic, appearing most often in rapid rolling maneuvers. Ratchet frequencies are typically 2-3 Hz.

EFFECTIVE AIRFRAME

Command Gradient Shaping

Prefilter $\frac{K_F}{s + 1/T_F}$

P_c

Equalization and Flex Mode Filtering Y_b

Actuators Y_a

Airframe Y_δ^ϕ

Rate Gyro s

K_p

$$\frac{p}{p_c} \doteq \frac{K_v e^{-\tau_v s}}{(T'_R s + 1)}$$
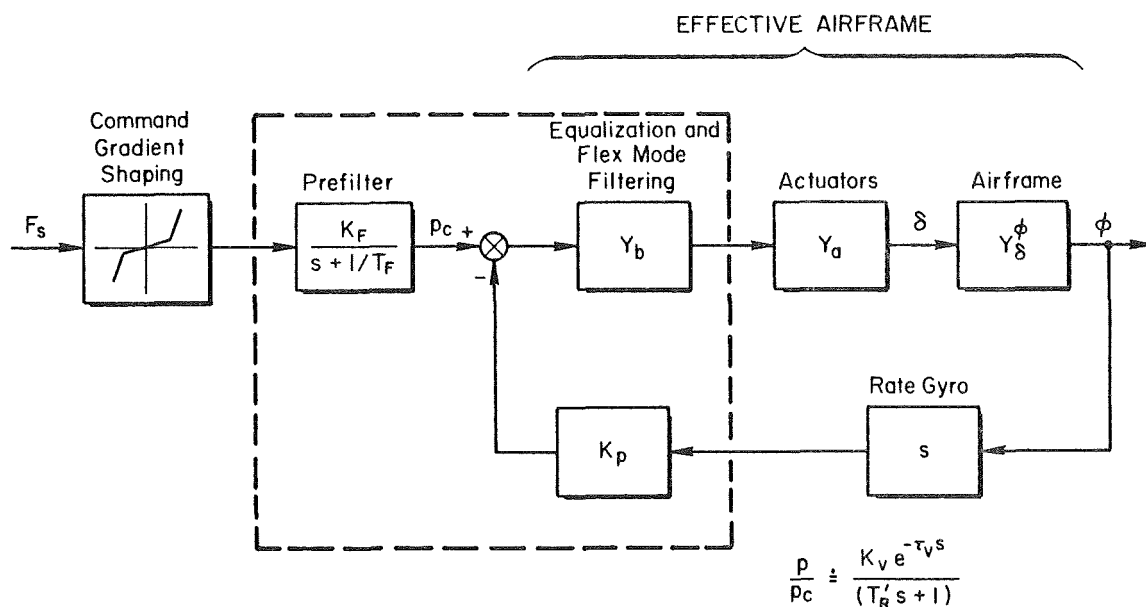
Figure 1. Typical Fly-by-Wire Roll Control System

Figure 2 illustrates this oft-remarked but seldom recorded phenomenon. The frequency difference alone indicates that the PIO and ratchet situations are different phenomena, yet both clearly involve the closed-loop pilot vehicle system.

An interesting set of roll ratcheting phenomena has been observed in variable stability NT-33 flight.[2-4] Chalk[5] speculates that the oscillations were due to the near K_c/s character of the effective controlled element. He used a rudimentary $(K_p e^{-\tau s})$ non-adaptive pilot model with τ ranging from 0.09 to 0.13 sec to show that one can get the observed instability (at about 12-17 rad/sec) with a K/s-like aircraft and high pilot gains. This effective time delay must account for all the open-loop system lags, i.e., controller, actuator, filters, etc., plus the effective latency of the pilot. So, if this explanation of the roll ratchet is to be reasonable the total τ value must be appropriate. The 0.09 - 0.13 second range is remarkably low for the pilot alone, and is very low indeed when aircraft plus control system effective lags are also considered.

Mitchell and Hoh[1] also examined some of the same data. They cite sinusoidal vibration data in which a simple lateral tracking task was performed (using a center stick) while under the influence of high frequency lateral accelerations.[6] Frequencies from 1 to 10 Hz were employed and an oscillatory arm/stick "bobweight" mode occurred at about 12 rad/sec. They note that this higher frequency mode of the pilot-aircraft systems is near the frequencies of the observed ratcheting in F-16 and Calspan flight experiments and cite it as a possible cause.

15.2

a) F-16   b) Aircraft A, CAS I
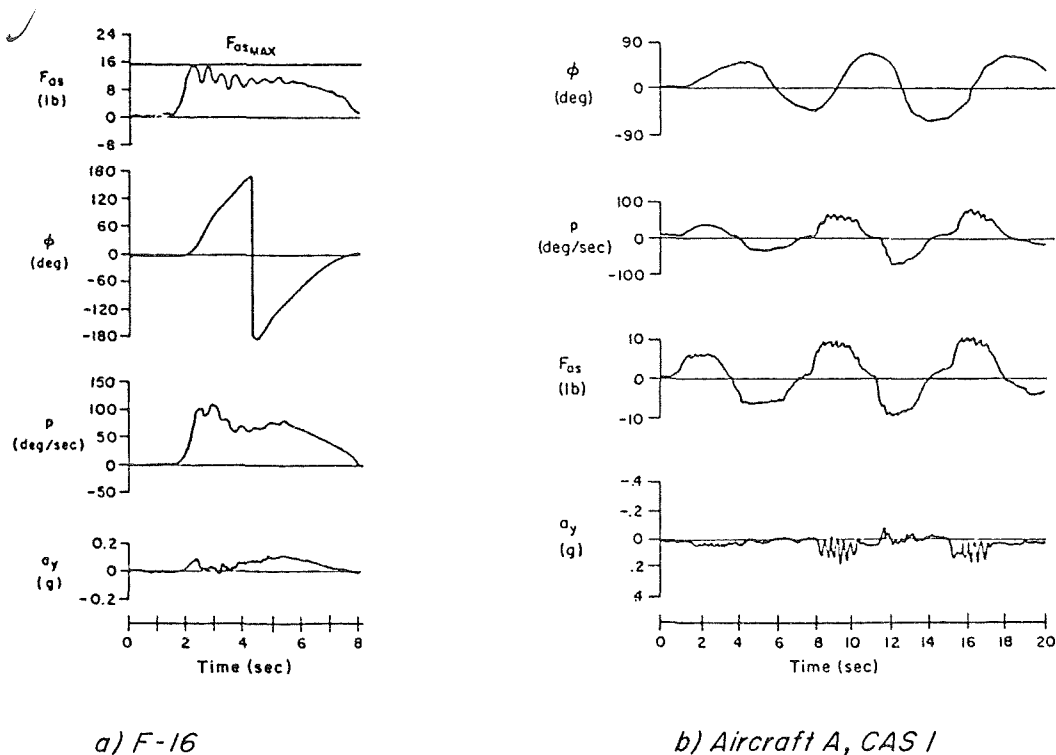
Figure 2. Roll Ratchet During Banking Maneuver

From the earliest studies on the interaction between the human pilot's neuromuscular system and aircraft control devices,[7,8] the presence of a neuromuscular system limb-manipulator dynamic resonance peak at 14-19 rad/sec has been well known. Neuromuscular system characteristics are cited[9] as "exceptionally important and critically limiting in such matters as

- ● control precision where limited by the pilot's neuro-muscular system.

- ● effects of control system nonlinearities, including their connections with control system sensitivity requirements."

Other summaries place great stress on the importance of considering these characteristics even though this frequency range of major activity may be well above bandwidth associated with the "usual" control task.[10]

It is becomming more and more apparent that modern, high performance, high gain, response command flight control system bandwidths may be encroaching on the neuromuscular system. Advances in flight control system fly-by-wire technology permit new manipulation devices, for example force sensing side-sticks, at the pilot output/effective-vehicle interface. These have thus far been generally successful in application, but have introduced or enlarged some pilot-vehicle flying qualities problems. Particular problems include:[2-4,11,12]

15.3

- high roll control sensitivity and PIO's in precision maneuvering;

- roll ratchet in otherwise steady rolling maneuvers;

- sensitivity to the way the pilot grips the stick or to location of his hand/arm support;

- effective time delay associated with stick filters, with attendant increase in pilot remnant;

- biodynamic interactions, e.g., hand/arm stick bob-weight effects.

Attempts to alleviate these effects have involved adjustments in stick force gradients, filtering, and sensitivity. These have included introduction of various nonlinear elements such as command gain reduction as a function of pilot input amplitude or frequency, filter time constant changes with sense of input (increase vs. decrease), and different force gradient for right and left roll commands. These adjustments have generally involved ad hoc empirical modifications in the course of the aircraft development. Much of this has been accomplished in flight test with correspondingly large cost.

The purposes of this paper are to

- explore the origins of the roll ratchet phenomenon;

- develop insights about the tradeoffs involved in adjusting the properties of force-sensing sidesticks;

- present guidelines to minimize roll control problems.

HUMAN PILOT-DYNAMIC SYSTEM CONSIDERATIONS

Ideal Crossover Model (and Its Implications)

The prescription for K/s-like controlled element dynamics in the region of pilot-vehicle system crossover as an often desirable form stems from the fundamental feature of human dynamics that no pilot lead is then required to establish good closed-loop system dynamics over a wide range of pilot gains. The basic recipe is almost invariably conditioned by such statements as "in the frequency region about crossover." Such statements are made to restrict the form of the pilot model to that required only in the crossover region. In particular, the cases covered are such that an effective time delay term in the pilot model is an adequate approximation to the high frequency effects.

Simple tracking task pilot model forms and associated pilot-vehicle system properties begin with the ideal crossover model[10] of Fig. 3. In this model the pilot adjusts his dynamic characteristics so that the open-loop pilot-vehicle dynamics are approximately K/s over the frequency band immediately above and below the gain crossover. The model also indicates that in

15.4

C -3

$$\omega_C \doteq \text{constant}$$

$$\neq f(K_C)$$

$$\omega_C \sigma_e = \text{constant}$$

$$\left|Y_p Y_c\right| = \left|\frac{\omega_C}{j\omega}\right|$$

$$Y_p Y_c = \frac{K_c K_p \, e^{-\tau j \omega}}{j\omega}$$

Isotonic
Stick
(free moving)

Isometric
Stick
(force)
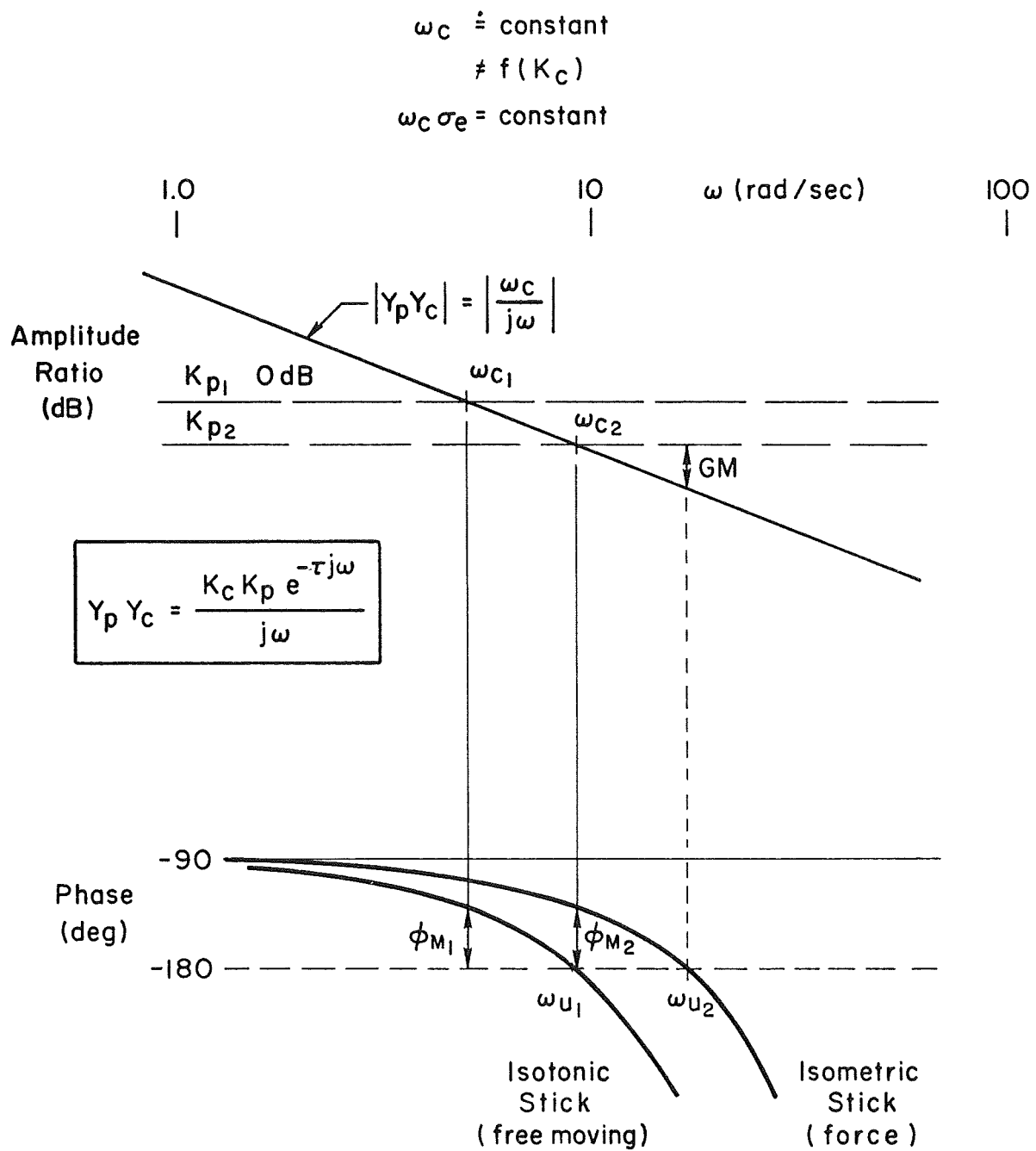
$$\Delta\tau = \tau_{\text{isotonic}} - \tau_{\text{isometric}} \doteq 0.1\,\text{sec}$$

Figure 3. Ideal Crossover Model

15.5

full attention tracking operations the pilot will adjust his gain to offset any variation in controlled element gain in order to maintain a nearly fixed control system bandwidth. Thus the full-attention closed-loop bandwidth ω_c (identified as the crossover of the 0 dB gain line with the K/s amplitude ratio plot) is independent of the controlled element gain. Furthermore, the pilot tends to keep the product of the crossover frequency and the task RMS error, $\omega_c \sigma_e$, constant.

In the crossover model the exponential term with time delay τ approximates all the lag contributions due to pilot and vehicle high frequency dynamic modes. The effective time delay is a function of, among other things, the force/displacement characteristics of the manipulator. As shown in Fig. 3, an isometric (force) stick results in less lag than does an isotonic (free moving) stick. Past experimentation[13] has identified the difference to be approximately 0.1 sec.

In Fig. 3 if the pilot gain were set at the value represented by K_{p2} with an isometric stick, the bandwidth would be indicated by ω_{c2} and would result in a system stability phase margin, ϕ_{m2}, and gain margin, GM. If this same gain were employed with the isotonic stick, the phase margin would be 0, and a low frequency continuous oscillation (PIO) would result. This oscillation can then be alleviated by pilot gain reduction to the value represented by K_{p1}, thereby accepting a reduced bandwidth. Thus Fig. 3 can be used to demonstrate the common low frequency PIO problem which generally occurs in the vicinity of 0.5 Hz and which is relieved by reducing pilot gain. (In the crossover model an ω_u of 4 rad/sec corresponds to $\tau = \pi/2\omega_u = 0.4$ sec for the total pilot, control system, aircraft, etc., latency).

Limb-Sidestick Neuromuscular Model (and Its Implications). As previously noted, early studies on the neuromuscular system noted the presence of a neuromuscular system or limb-manipulator peak at 14-19 rad/sec well past the usual "crossover region."[7] The effects of various restraints on the limb/neuromuscular system include closed-loop neuromuscular system model fits to pilot/controlled-element describing function measurements for pressure and free moving manipulators.[8] An important part of the neuromuscular dynamics in each case is a quadratic mode with damping and natural frequency of

MANIPULATOR	NM/L DYNAMICS
Free Moving	[0.07, 17]
Isometric or Pressure	[0.138, 18.6]

There is also a neuromuscular system mode which is approximated by a first-order lag break at about 10 rad/sec. This mode is also somewhat dependent on the nature of the manipulator restraints.[13,14]

The reason that the neuromuscular actuation system dynamics differ when the manipulator restraints are changed is physiological -- the neuromuscular

apparatus involved depends on the restraints and limb movements. While greatly oversimplified, the neuromuscular actuation elements of the human may be viewed as a two loop system. The inner loop principally involves Golgi, muscle spindle, and other receptors with short pathways directly to spinal level and back to the musculature. Viewed from the output end this loop is primarily sensitive to forces, and because of the short neutral pathways the time lags of information flow are small. The effective bandwidth of this loop can, therefore, be quite high. The second or outer loop includes joint receptors as major feedback elements. Their neural pathways, and associated delays, are longer, leading to a lower outer loop bandwidth. In isometric (force-stick) manipulator conditions, there is little or no joint movement, so the inner loop elements should be dominant. With isotonic (free-moving stick) conditions, on the other hand, the joint receptors are major elements. As already indicated in connection with Fig. 3 the net difference, in terms of an effective latency, is approximated at low frequencies by a difference in effective τ of about 0.1 sec.

If we now employ the detailed model of the neuromuscular system (instead of only approximating its phase lag contribution as in Fig. 3) and superimpose it on the controlled element K/s as in Fig. 4, we see an open-loop resonant peak in the 2 to 3 Hz frequency range due to the neuromuscular system. The correspondence of the neuromuscular/limb quadratic mode numerical values and observed roll ratchet frequencies is very unlikely to be a coincidence. So, at observed roll ratchet frequencies the neuromuscular/limb mode clearly should be taken into account. Since their primary effect is a resonant peak from which a "Gain Margin" might be measured,[*] these properties may be of central importance for high gain pilot situations.

EXPERIMENT GOALS AND SETUP

The experimental goals were to investigate and quantify limb/manipulator dynamics and interactions between the neuromuscular subsystem, force sensing side-stick configuration, high gain command augmentation, and command filtering; and to investigate possible relationships between these interactions and the roll ratchet phenomenon. A longer range goal is to provide and enhance guidelines for manipulator-system design.

The experimental setup is depicted in Fig. 5. A roll tracking task was selected in which the pilot matches the bank angle of his controlled element with that of a "target" having pseudo random rolling motions. The random motions are obtained via a computer generated sum of sine waves. The error

---

[*]While the "Gain Margin" shown in Fig. 4 indicates the magnitude difference between the $|Y_p Y_c|_{dB}$ peak and the zero dB line, the phase at or near this frequency may differ appreciably from that required for instability. Thus when the "Gain Margin" shown is zero only one of the two conditions for instability may be satisfied. Consequently this is not necessarily a true gain margin in the conventional sense. It does, however, indicate a resonant tendency contributed by the pilot.

$$Y_pY_c = \frac{K_c}{(0)} \times \frac{K_p e^{-\tau s}}{(T_N)[\zeta_N, \omega_N]}$$

$$(\omega_c = K_c K_p)$$
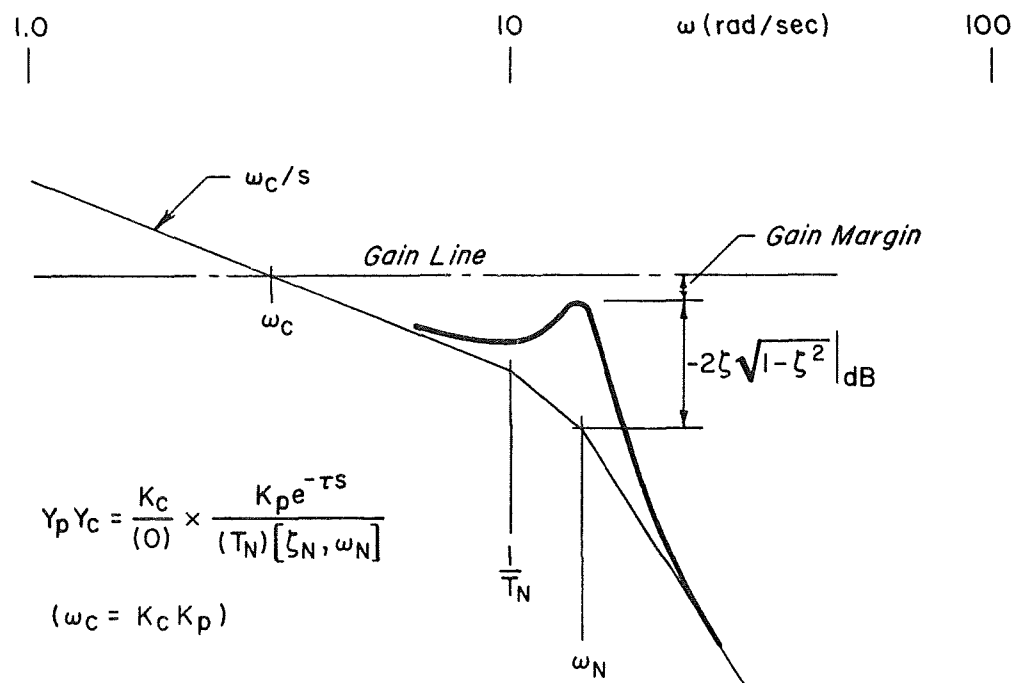
Figure 4. Bode Amplitude Ratio Plot for Neuromuscular System
Contribution to Roll Ratchet Potential

$$Y_c = \frac{K_c e^{-\tau s}}{s(Ts+1)} \quad ; \quad \begin{array}{l} \tau = 0 \to 0.1 \\ T = 0 \to 0.4 \end{array}$$
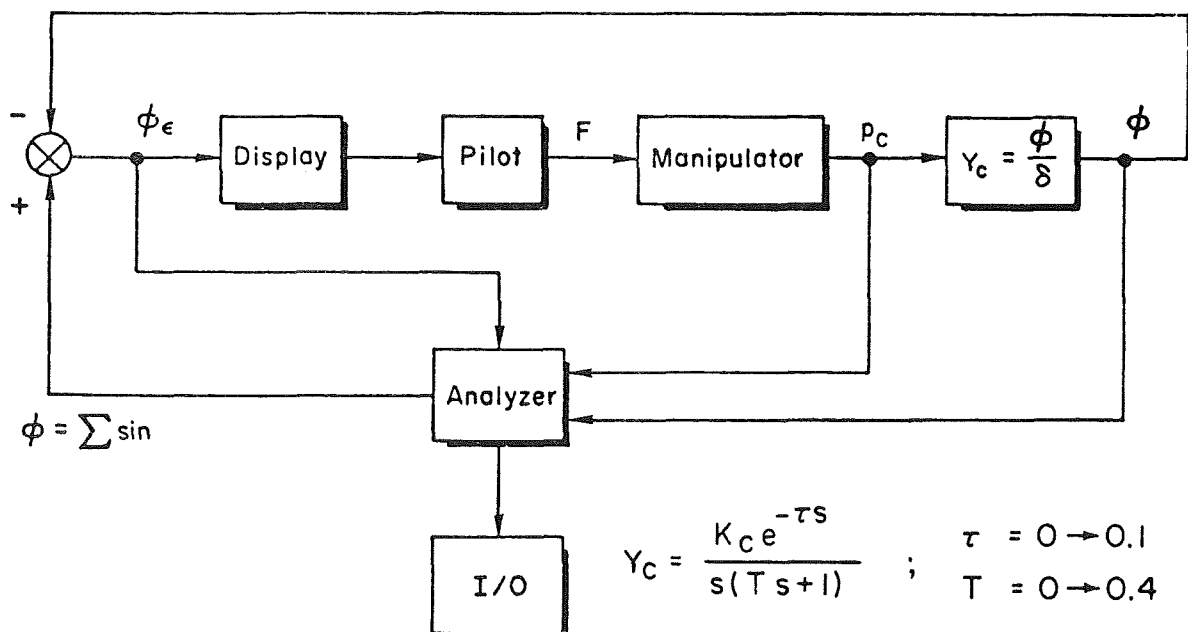
Figure 5. Experimental Setup

is displayed on a CRT and the pilot attempts to null the error by applying force to the manipulator, the output of which becomes the command to the controlled element, Y_c. The form of the controlled element is identified in Fig. 5 along with the range of lag time constants and time delays utilized in the experiment. This controlled element approximates a high gain roll rate command system. The time lag parameter, T, may be considered to be the effective roll subsidence time constant or a flight control system prefilter (between the pilot's stick command and the flight control system), whichever is larger. For very small values of τ the pure time delay may be a realistic approximation to digital flight control system sample and hold dynamics. More generally it is a low frequency approximation for all the high frequency lags in the system which are not covered by the time lag T. Because we are interested primarily in modern flight control systems, the parameter values for T and τ used in the experiment are generally consistent with values that would be present in a system designed to be Level 1 on the basis of flying qualities specifications. Thus, the parameter values used, in the main, should produce excellent effective controlled elements providing the gain is appropriately adjusted.

The manipulator was a McFadden force loader system used in many aircraft research and development simulations. Three stick displacement configurations were employed. One was a fixed (no displacement) stick as in the F-16.[11] The second had 0.77 deg/lb (small) stick motion. The third had 1.43 deg/lb (large) stick motion. The latter two matched the displacement/force characteristics employed in an NT-33 flight test.[12] Analog signals from the manipulator force sensor (p_c) and the resulting controlled element roll response ϕ were passed through an A $\rightarrow$ D converter to a digital computer where $Y_p Y_c$ describing functions and various performance measures were computed using STI's Frequency Domain Analysis (FREDA) program. The computations were essentially on-line and printed out at the conclusion of each run. Some 530 data runs were accomplished which provided a tremendous data base from which to determine or identify the various interactions of interest.

No accounts have been found where roll ratchet has been observed or recognized in fixed- or moving-base simulations. It apparently has only occurred in actual flight and then on a more or less random basis. The first objective of this experimental setup therefore was to tune the controlled element, manipulator, and command/force gradients to try to achieve roll ratchet, or at least maximize roll ratchet tendencies, in the fixed-base simulation. A key factor was that describing function measurements must cover the limb neuromuscular peaking frequency region, and forcing functions should be adjusted to emphasize good data in the neuromuscular subsystem region. The experimental runs were accomplished using the summation of sine waves presented in Table 1.

15.9

TABLE 1. ROLL TRACKING FORCING FUNCTION

Sine Wave (i)	1	2	3	4	5	6	7	8	9
Frequency (ω_i)	0.467	0.701	1.17	1.87	3.51	7.01	11.2	14.0	18.7
Amplitude (A_i)	15.2	15.2	15.2	7.6	3.04	0.76	0.38	0.228	0.152
Relative Amplitude	1	1	1	0.5	0.2	0.05	0.025	0.015	0.01

$$\phi_I = \Sigma A_i \cos \omega_i t \quad (\text{deg})$$

EXPERIMENTAL RESULTS

Human Pilot Dynamics

Consistency of Crossover Frequency. It will be recalled that in the ideal crossover model the crossover frequency remains constant even though the controlled element gain may vary. Figure 6 shows results obtained using the fixed side-stick manipulator configuration and a wide range of command/force gradients (controlled element gains). The initial command/force gradients for the F-16[11] and the NT-33[12] experimental flight programs are identified for comparison. The controlled element forms range from K/s to $Ke^{-0.07s}/s(0.1\ s + 1)$. The data for various time delay or time lags are indicated by the symbols. The data points of Fig. 6 indicate two aspects.

Sym	τ	T
X	0	0
⊙	.067	0
⊡	.067	.1
△	.1	0

$$Y_c = \frac{K_c e^{-\tau s}}{s(Ts + 1)}$$
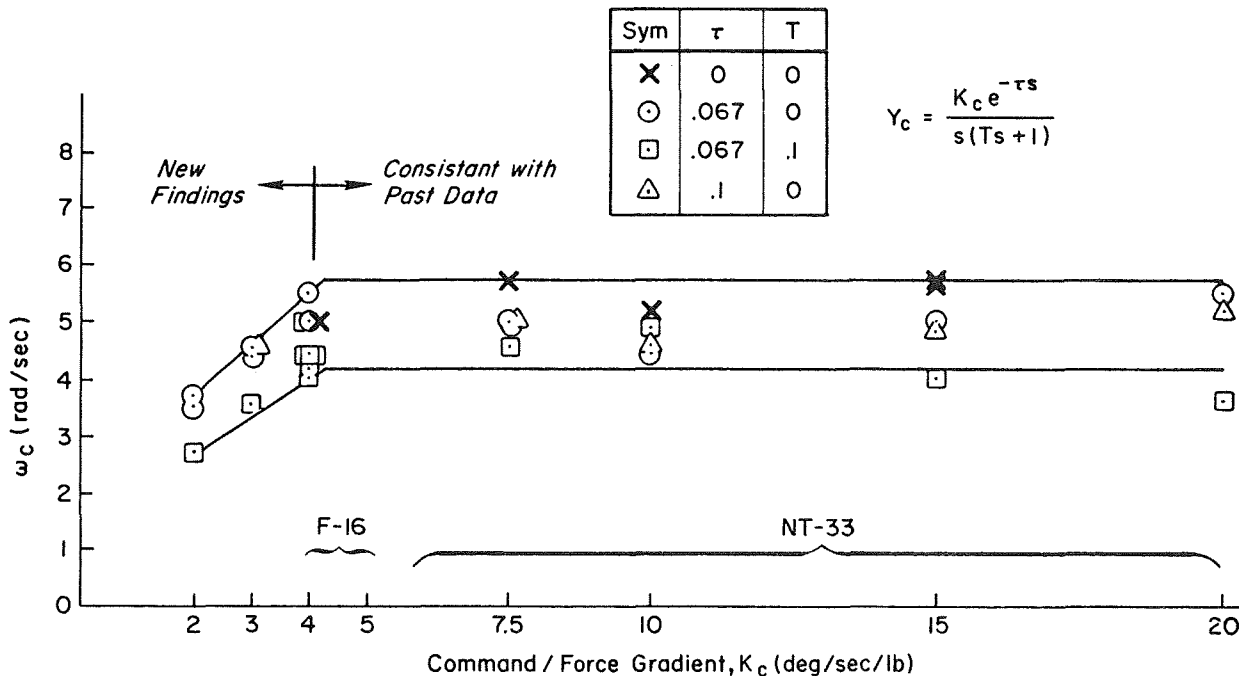
Figure 6. Influence of Command/Force Gradient on Crossover (Fixed Stick)

First they reflect a general decrease in ω_c as controlled element lags increase. Second they show that crossover frequency, as expected, is essentially independent of controlled element gain over a very broad region. But, as the controlled element gain becomes quite low and the manipulator forces required to achieve the desired rolling response become very large, a point is reached where the pilot can no longer accommodate and a rapid drop off in bandwidth results. Interestingly, the F-16 initial command/force gradients lie right at the break in ω_c and therefore represent the lowest values which might be considered acceptable to pilots.

Similar results were obtained with the small and large displacement sidestick configurations except that the crossover frequencies decreased slightly as the displacement was increased.

Neuromuscular System Peaking Tendencies

Turning attention now to the neuromuscular system, Fig. 7 presents the describing function measurements for 3 runs using the fixed force stick and a controlled element having a command/force gradient of 4 deg/sec/lb, no time lag, and a time delay of about 70 ms. The straight line reflects the resulting ω_c/s crossover characteristics. Amplitude departures from this asymptote are the contributions of the pilot's neuromuscular system at high frequency and his trim lag-lead at low frequency. In the region of crossover Y_pY_c is almost exactly ω_c/s as suggested by the ideal crossover model. The amplitude ratio departures from the asymptote at the highest 3 frequencies shows a peaking in the vicinity of the 14 rad/sec forcing function for 2 of the 3 runs. It also might be noted that there is remarkable consistency in both the amplitude and phase measurements across all frequencies for all 3 runs. In Fig. 7, two of the amplitude data points at 14 rad/sec lie slightly above the 0 dB line. We would therefore expect this to represent a neutral or slightly unstable dynamic mode if the phase angle were near -180 deg at this frequency. This then could be interpreted as affecting roll ratchet.

The two data points at 14 rad/sec are 10 dB above the asymptote and may or may not be exactly the actual neuromuscular system peak, i.e., the peak itself may occur at a slightly higher or lower frequency. The peaking tendency shown in Fig. 7 is representative of a large amount of the data obtained. This frequency is consistent with the roll ratchet frequencies observed in the flight traces.

Influence of Effective Controlled Element Characteristics

The sensitivity of the 14 rad/sec peaking tendency to time delay is shown in Fig. 8. The circles reflect the average values at each frequency and the bars indicate ± 1 σ ranges. The controlled element is $K_c e^{-\tau s}/s$. The manipulator is the fixed stick configuration. Results show that a time delay of approximately 0.065 to 0.07 tends to maximize the neuromuscular system peaking. At time delays either below or above these values, the peaking tendency decreases. Of all the controlled elements examined, K_c/s shows the minimum tendency for a peak. Interestingly, the time delay values which maximize the neuromuscular peaking would be considered good from the
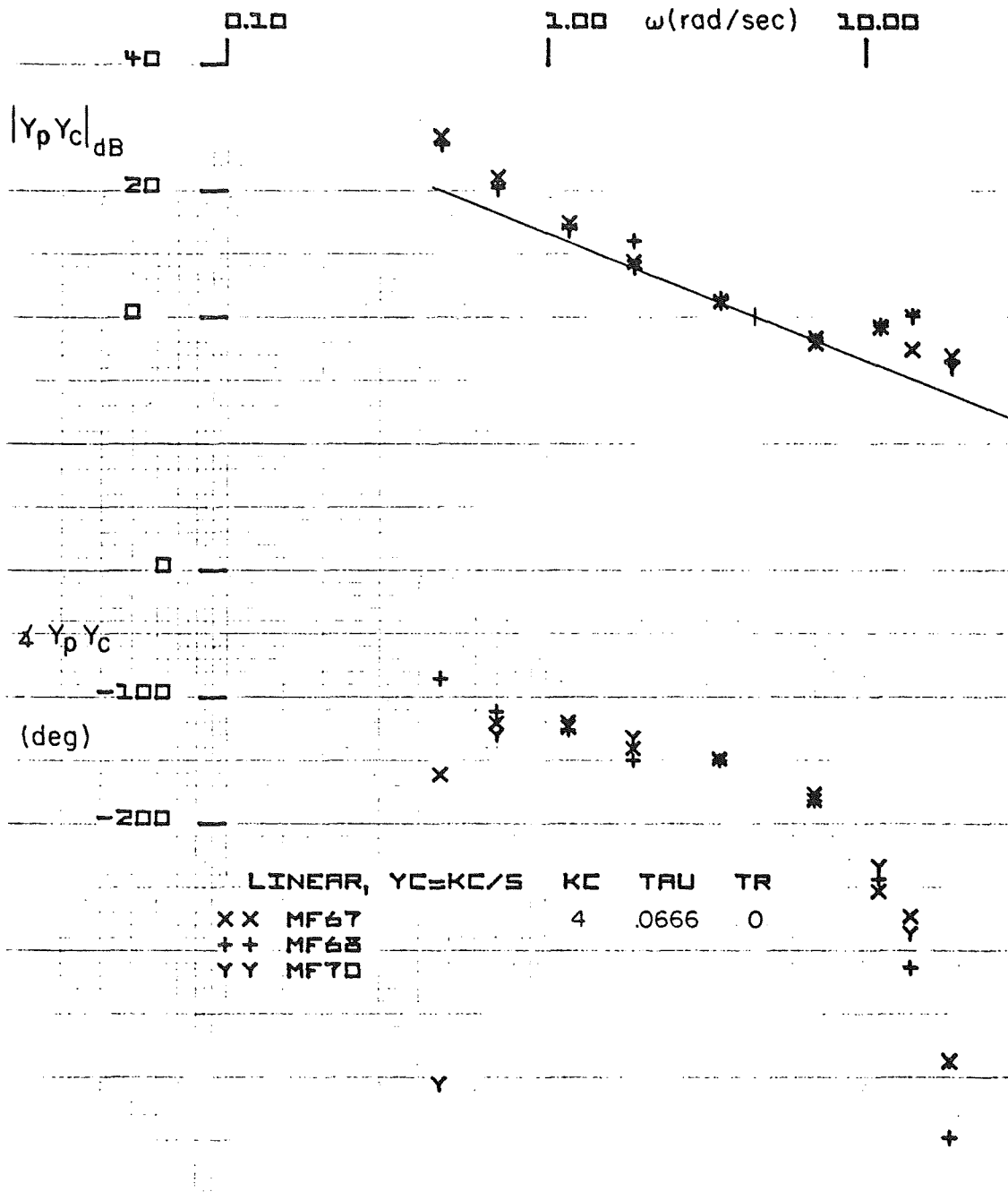
Figure 7. Y_pY_c Describing Function Amplitude and Phase Plot for $Y_c = 4/s \; e^{-0.067s}$

15.12
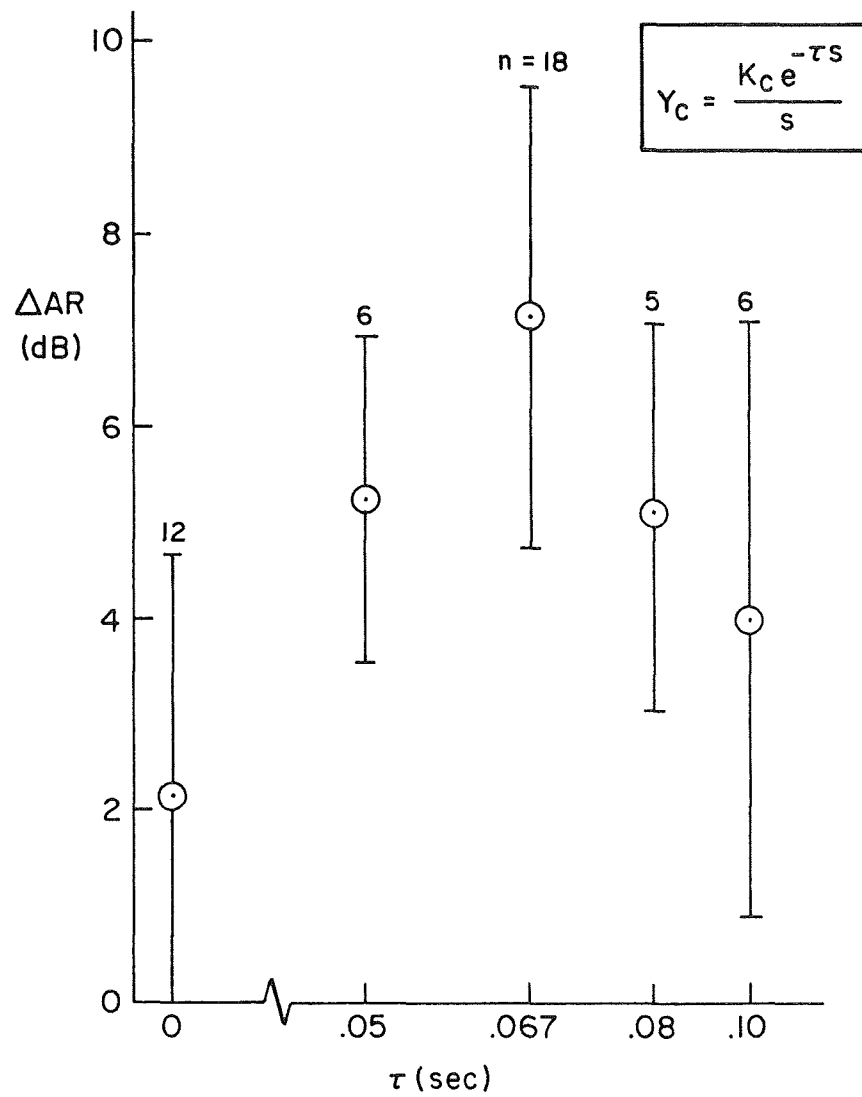
Figure 8. Neuromuscular System Amplitude Ratio Peaking
With Controlled Element Time Delay (Fixed Stick)

15.13

MIL-8785 flying quality specification standpoint. In essence, these data show that the tendency to peaking can be "tuned" by the adjustment of the controlled element effective lag, with a maximum effect near 0.07 sec.

The neuromuscular system peaking sensitivity to controlled element command/force gradient is shown in Fig. 9. Here the command/force gradient ranges from 3 deg/sec/lb (which is slightly lower than that employed on the F-16) up through 15 deg/sec/lb which was utilized in the NT-33. The data were obtained using the fixed stick and a time delay of 0.067 sec. Data for time lags of 0 and 0.1 have been combined. These data show a slight increase in peaking tendency in the vicinity of 7.5 deg/sec/lb command/force gradient. This is about the same value as the response/force ratio for the Fig. 2 flight traces of ratchet. This may or may not be coincidental. However, it is significant that there is appreciable peaking of the neuromuscular system across the entire gain range investigated in these experiments.

Influence of Stick Characteristics

The influence of stick motion is summarized in Fig. 10. These plots reflect the amplitude ratio peaking at the 3 higher frequencies (11, 14, and 19 rad/sec) for the fixed, the small deflection, and the large deflection stick configurations at 3 different values of the controlled element time delay: 0.0, 0.067, and 0.1 secs. All of these data were taken with the command/force gradient of 10 deg/sec/lb. The results show that there is relatively little difference between the fixed and small deflection force stick. Both show an increase in neuromuscular peaking tendency for the
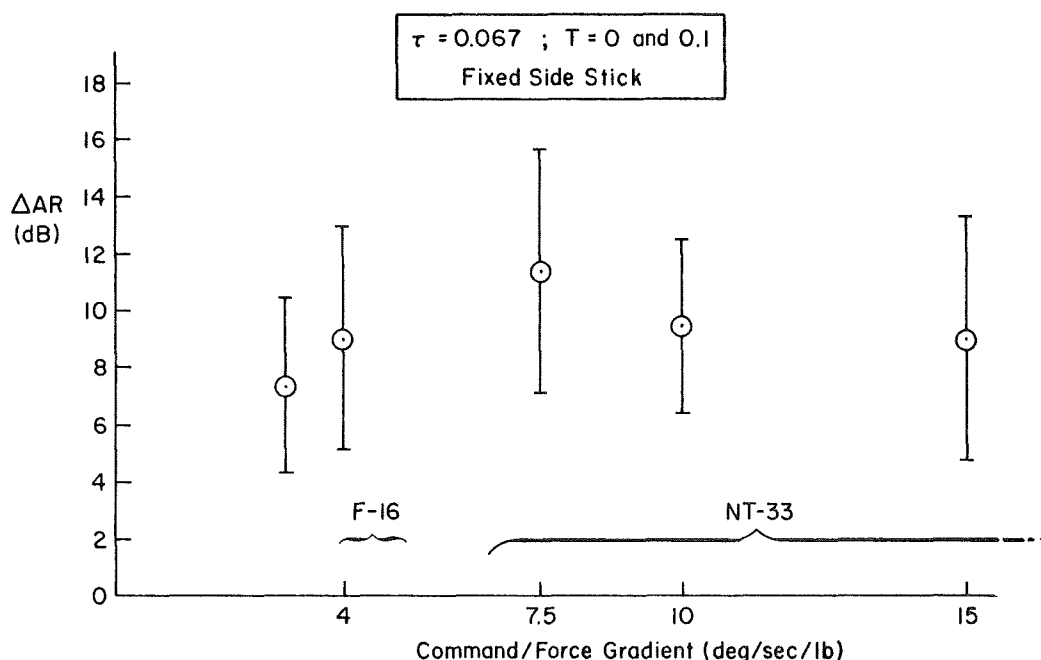
Figure 9. Neuromuscular Peaking Sensitivity to Controlled Element Command/Force Gradient

15.14
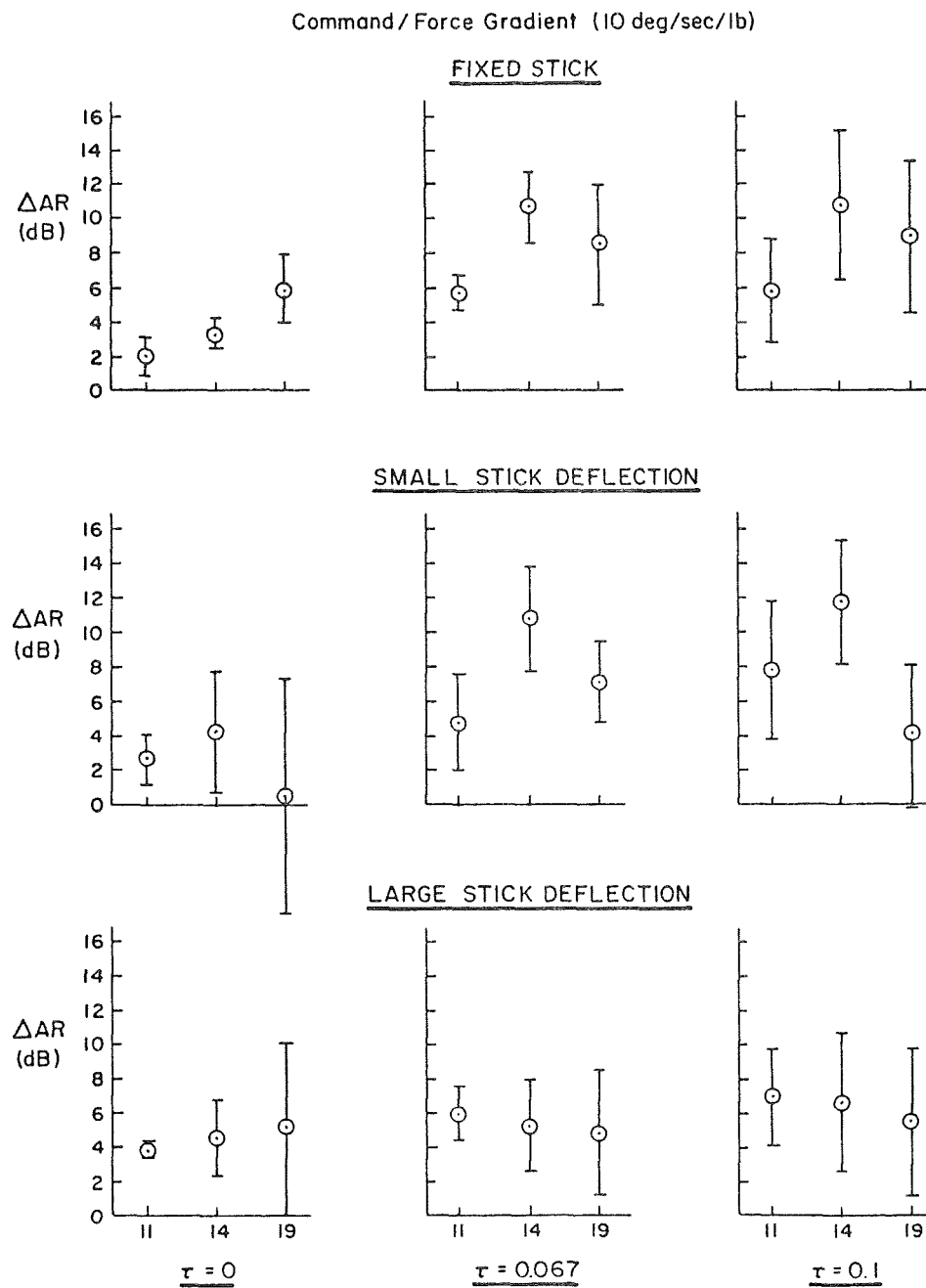
Figure 10. Influence of Stick Displacement on Neuromuscular
Peaking Tendency

15.15

0.067 and 0.1 sec time delays. They both show a tendency to maximum peaking in the vicinity of 14 rad/sec and in both cases there is considerably less peaking for the zero time delay cases. The large deflection stick, on the other hand, shows a relatively constant amplitude departure from the controlled element asymptote across the 11 to 19 rad/sec frequency band and a lack of sensitivity to the controlled element time delay.

Adjustment of Pilot Lead

The influence of the lag time constant on the neuromuscular system peaking and the possible adoption of lead by the pilot is reflected in Figs. 7 and 11 through 13. Figure 7 shows the neuromuscular peaking obtained with the controlled element command/force gradient of 4 deg/sec/lb, a time delay of 0.067 secs, and no lag. The maximum peaking was noted to be approximately 10 dB and occurred at 14 rad/sec. The addition of a first-order lag time constant of 0.1 sec is shown in Fig. 11. Here the solid line represents the controlled element (Y_c) Bode asymptote adjusted to go through ω_c. The crossover occurs in a region that is K/s in appearance, and the amplitude peaking again is approximately 10 dB, and occurs near the 14 rad/sec data point. The peaks are quite close to the 0 dB gain line, which indicates a likely tendency to roll ratchet. Comparison of the phase plots between Figs. 7 and 11 indicate that the pilot is generating little if any lead to offset the time lag. (Detailed analyses[15] indicate that there is a pilot lead near 8 rad/sec for these cases and for Y_c = K/s which tends to compensate for the high frequency lags in general, but cancels none of them.)

In Fig. 12 the time lag has been moved to 0.2 secs. Comparison of the phase angle data points in Figs. 7 and 12, or Figs. 11 and 12, indicates that the pilot has introduced lead in the Fig. 12 case which essentially cancels the time lag at 0.2 secs. The asymptote for the Y_p/Y_c open-loop system is thus represented by the solid line below the time break point and the dashed line above that break point. Again the amplitude ratio is ω_c/s-like in the vicinity of the crossover. However, there is now considerable scatter in the data points in the region of the neuromuscular system peaking dynamics. In only one of the three runs shown in Fig. 12 was there a peaking tendency for the neuromuscular system and this appears to be concentrated in the vicinity of 11 rad/sec rather than the 14 as noted previously. In the other two runs, the amplitude data points lie quite closely to the Y_pY_c asymptote.

In Fig. 13 the lag time constant has been moved down to 0.4 sec. Again comparison of the phase plots shows that the pilot has now moved his lead down to precisely cancel the controlled element time lag contribution so that the resulting Y_pY_c has the appearance of an ω_c/s throughout the frequency region of interest. The peaking tendency of the neuromuscular system is no longer evident and there should be little chance of roll ratchet. However, the roll control bandwidth has now been reduced to approximately 2.5 rad/sec whereas it was approximately 4.5 rad/sec with the time constant of 0.1 sec. If the pilot were to attempt to achieve a 4.5 rad/sec bandwidth in the presence of the lag characteristics shown in Fig. 13, a PIO would
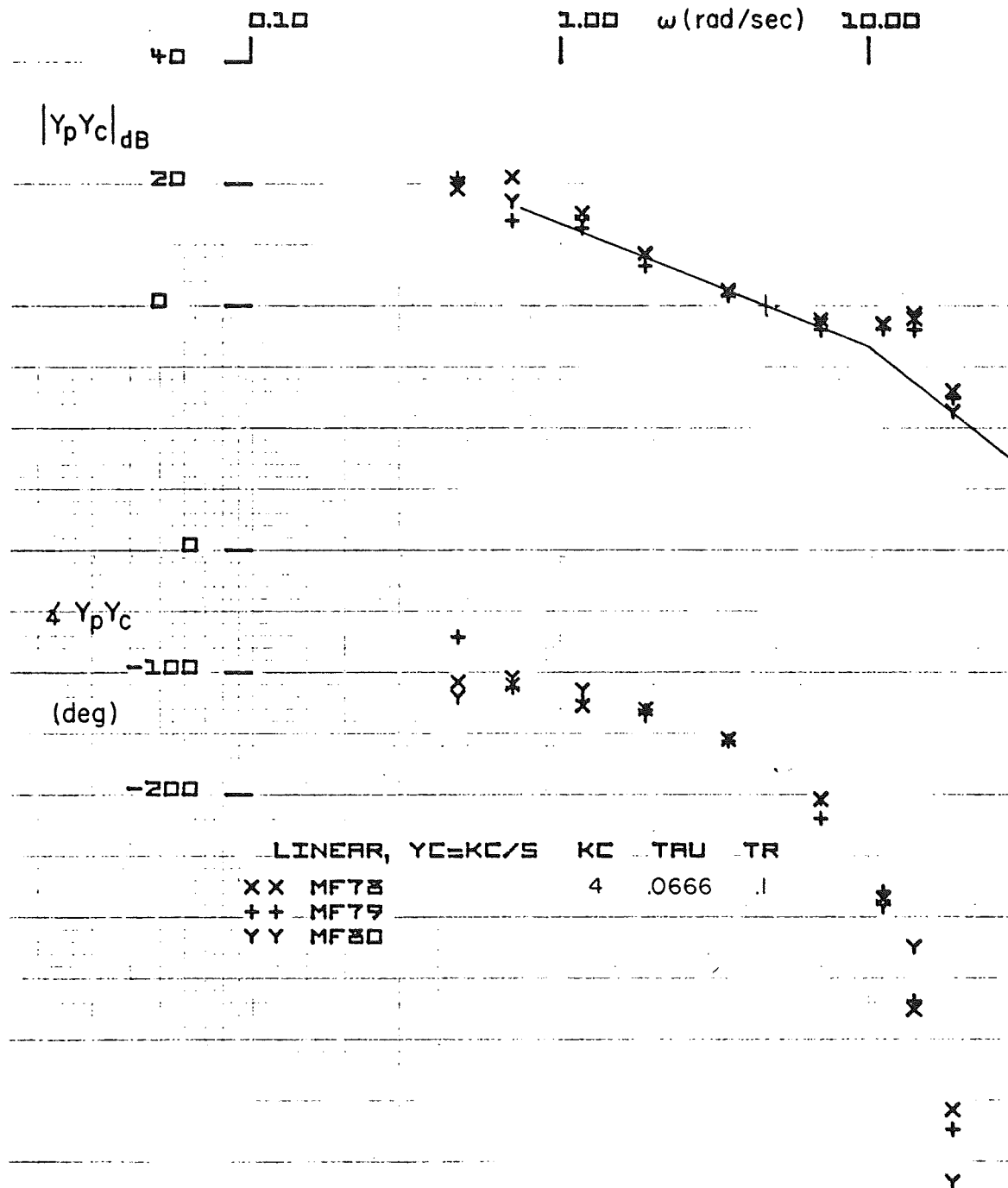
Figure 11. Y_pY_c Describing Function Amplitude and Phase Plot
For $Y_c = \dfrac{4e^{-0.067s}}{s(0.1s+1)}$

15.17
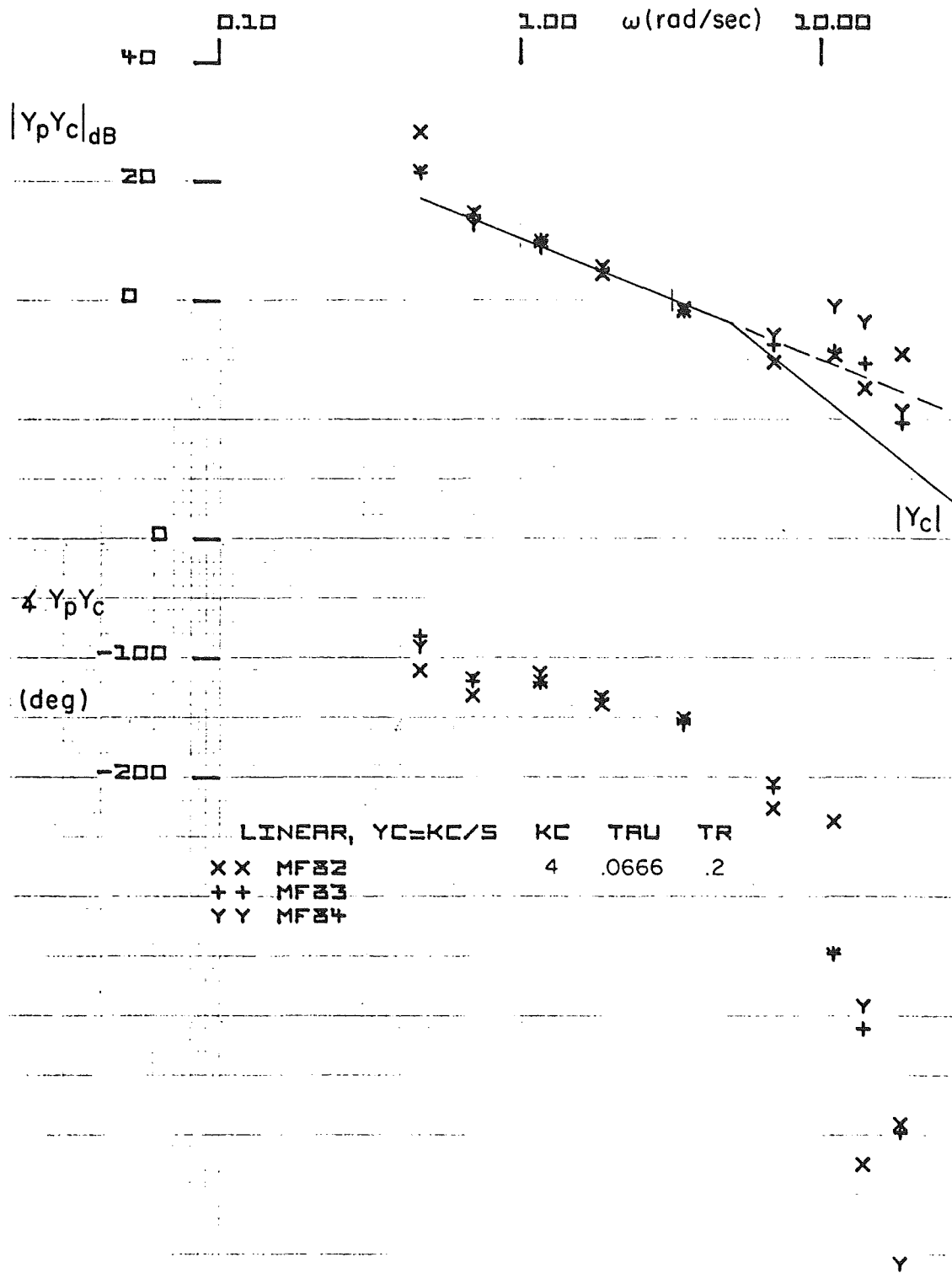
Figure 12. $Y_p Y_c$ Describing Function Amplitude and Phase Plot

For $Y_c = \dfrac{4e^{-0.067s}}{s(0.2s+1)}$

15.18
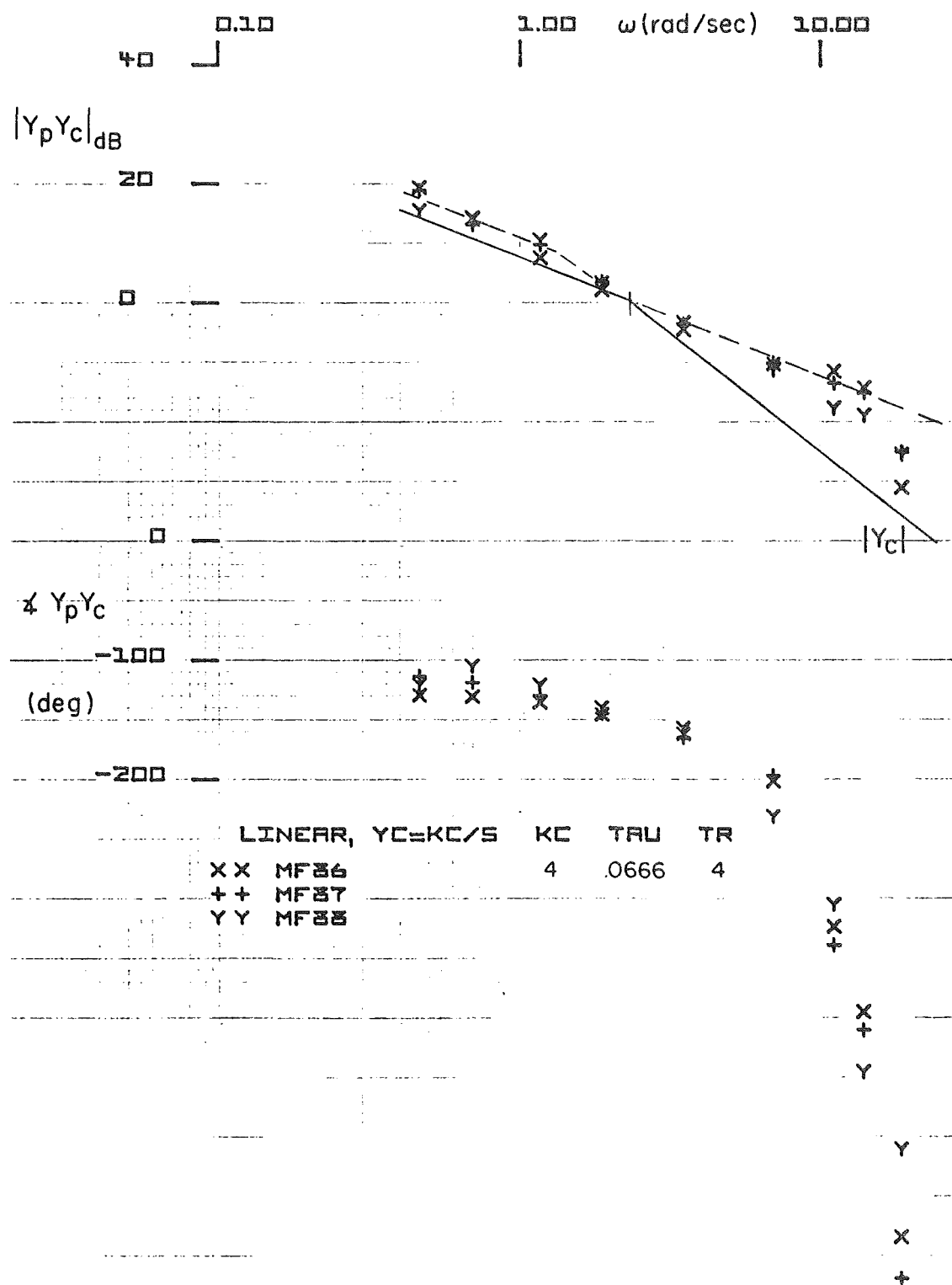
Figure 13. $Y_p Y_c$ Describing Function Amplitude and Phase Plot

For $Y_c = \dfrac{4e^{-0.067s}}{s(0.4s+1)}$

15.19

occur at roughly that frequency (4 rad/sec). Thus in reducing or eliminating the roll ratchet tendency, we may have substituted a tendency for the lower frequency PIO.

Closed-Loop Pilot-Vehicle System Characteristics

Observed Fixed-Base Roll Ratchet. The previous sections have emphasized the neuromuscular peaking tendency as a harbinger of the roll ratchet phenomenon. Yet, in the data presented, the open-loop system phase angle has generally been greater in magnitude than -180 degrees. This means that the gain differences between the peak and the 0 dB line are not necessarily true gain margins. The closed-loop pilot-vehicle systems will, therefore, not necessarily show an oscillation at the neuromuscular peaking frequency although the resonant peak will ordinarily be indicated in the closed-loop system. The pilot remnant, being relatively broadband in character, will therefore act as a driving mechanism to excite the resonant peak.

In some cases the experimental data actually indicated a roll ratchet-like oscillation under conditions similar to those where the phenomenon was found in flight. Most commonly these were stretches in the time histories which involved nearly steady-state rolling velocity commands. An example is given in Fig. 14. Here a short segment of the roll attitude command input is nearly triangular, and the pilot's stick force trace indicates a 2-3 Hz oscillation. Because the forcing function is a random appearing time signal, with only very occasional segments akin to the triangular or steady rolling commands shown, this type of ratchet-like pilot output trace is atypical in the context of a total experimental run. The pilot subjects, in fact, did not report that they had encountered the condition since it was so transitory. Yet it appeared quite commonly once the conditions were favorable -- i.e., neuromuscular peaking tendency present and momentarily steady rolling velocity command. Consequently the fixed base simulation can be said to have successfully demonstrated roll ratchet-like phenomena.

Fixed Force Stick Tracking Task

$$K_c = 3 , \tau = 0.067 , T = 0.1$$
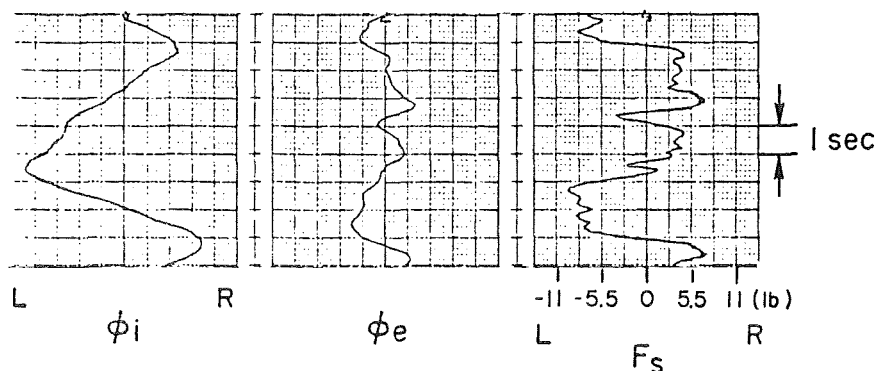
Figure 14. Example of Roll Ratchet-Like Oscillation
in Stick Force Trace

15.20

It is also useful to re-examine the open-loop describing function data when a first-order correction is made to the data to account for the effect of rapid rolling motion on the pilot during flight. The pilot's angular motion sensing neurological apparatus acts very much like a rate gyro inner loop in the frequency range near and slightly above crossover.[10] This inner loop, present when superthreshold rolling velocities are imposed on the pilot, has the effect of reducing the effective time lags in the pilot's visual-input/manipulator output response. The reduction can be as much as 0.1 second from the fixed-base data. When changes of phase lag of the magnitude 0.1ω are made on typical describing function data showing major neuromuscular peaking, the net phase shift in the frequency region about the peak is very often near -180 degrees. Figure 15 shows a typical example for the fixed force stick configuration with $T = 0$, $\tau = 0.067$, and $K_c = 10$ deg/sec/lb. Therefore one can conclude that the fixed-base neuromuscular peaking examples which show negative gain margins of the amplitude ratio peak relative to 0 dB are quite likely to result in oscillations in the flight situation. The roll ratchet phenomenon in these cases would therefore be high-frequency PIO's which intimately involve the pilot's limb-manipulator neuromuscular system dynamics.

Comparisons with Flight Data

The controlled elements in Figs. 11-15 essentially duplicate the F-16 configurations tested[11] and the qualitative results and trends are the same. The compromise selection for the prefilter in the F-16 was a time constant of 0.2 rad/sec which is shown in Fig. 12 to allow a comfortable bandwidth slightly above 3 rad/sec and having 30 to 35 deg of phase margin and a much reduced neuromuscular peaking tendency. Thus there should be minimum tendency for either low or high frequency PIO although the data scatter in the higher frequency range of Fig. 12 show that conditions favorable to roll ratchet could pop up from time to time.

Yet another comparison between simulation results and flight data can be drawn from the investigation of roll ratchet and various prefilter configurations flown in the NT-33.[3] In this case one set of effective controlled elements are a close match to this simulation. A major difference, however, was the use of a center-stick in the NT-33. The roll ratchet encountered in this flight test was described as "response which was objectionably abrupt, resulting in a very high frequency, pilot-induced-oscillation (wing rocking) or having 'square corners' or being very 'jerky.'" The frequency was approximately 16 rad/sec.

Figure 16 is a replot of data from Ref. 6 with command/force gradient plotted versus the roll time constant, T_R. The circles identify configurations flown; the open symbols reflect no ratchet obtained, the shaded symbols reflect roll ratchet observed by one or more of the evaluation pilots over the range of time delays investigated. (It should be noted in passing that in almost every case, the ratchet only occurred with non-zero τ as was the case in the lab simulation.) The triangular symbol at $T_R = 0.2$, $K_c = 12.5$ is another NT-33 data point obtained from a flight program in which the roll time constant was selected at 0.2 sec for up-and-away tasks and 0.5 sec for landing tasks.[12] In addition, two 20 rad/sec first-order filters were
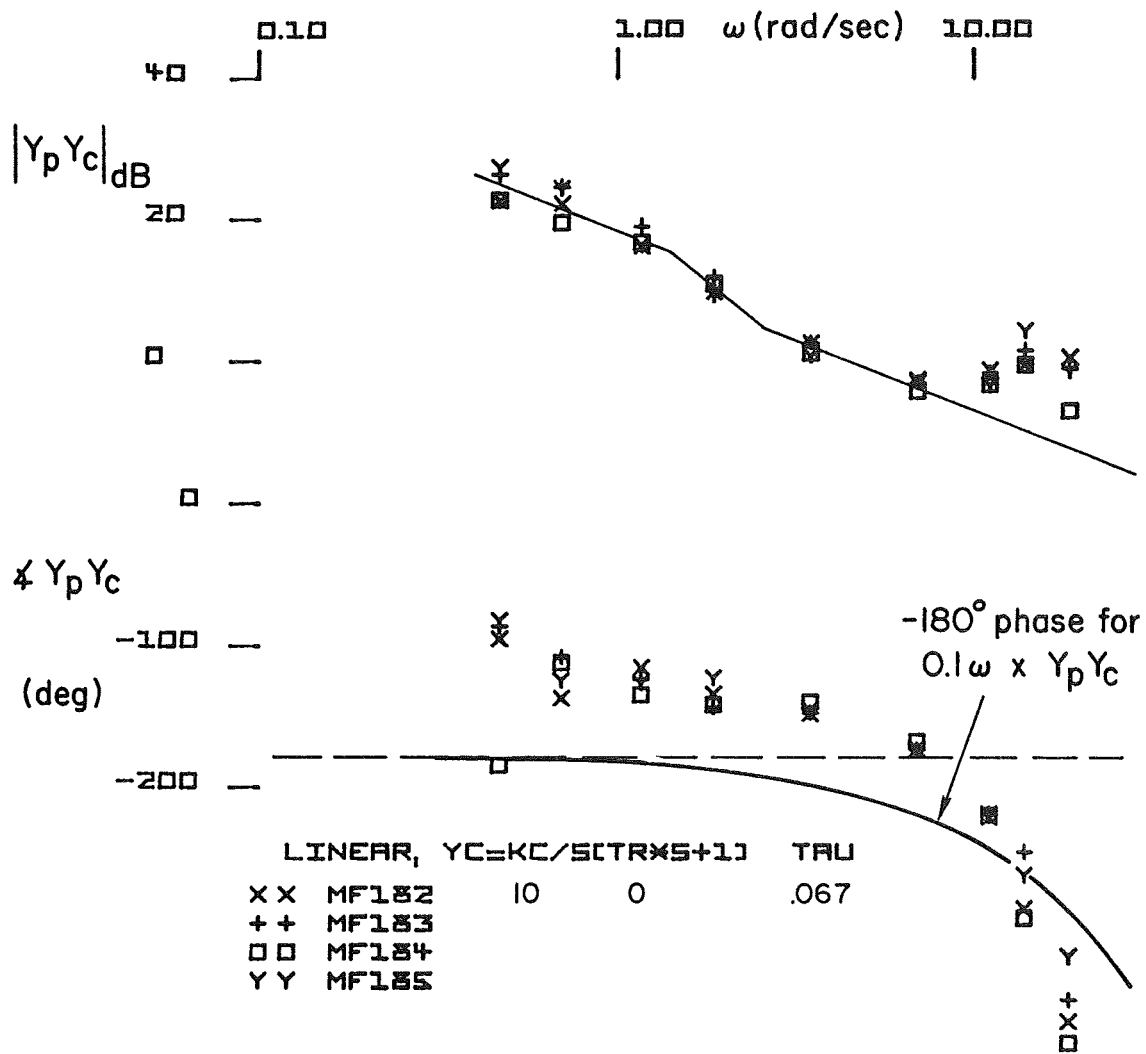
Figure 15.   Expected Phase Shift Due to Motion Effects
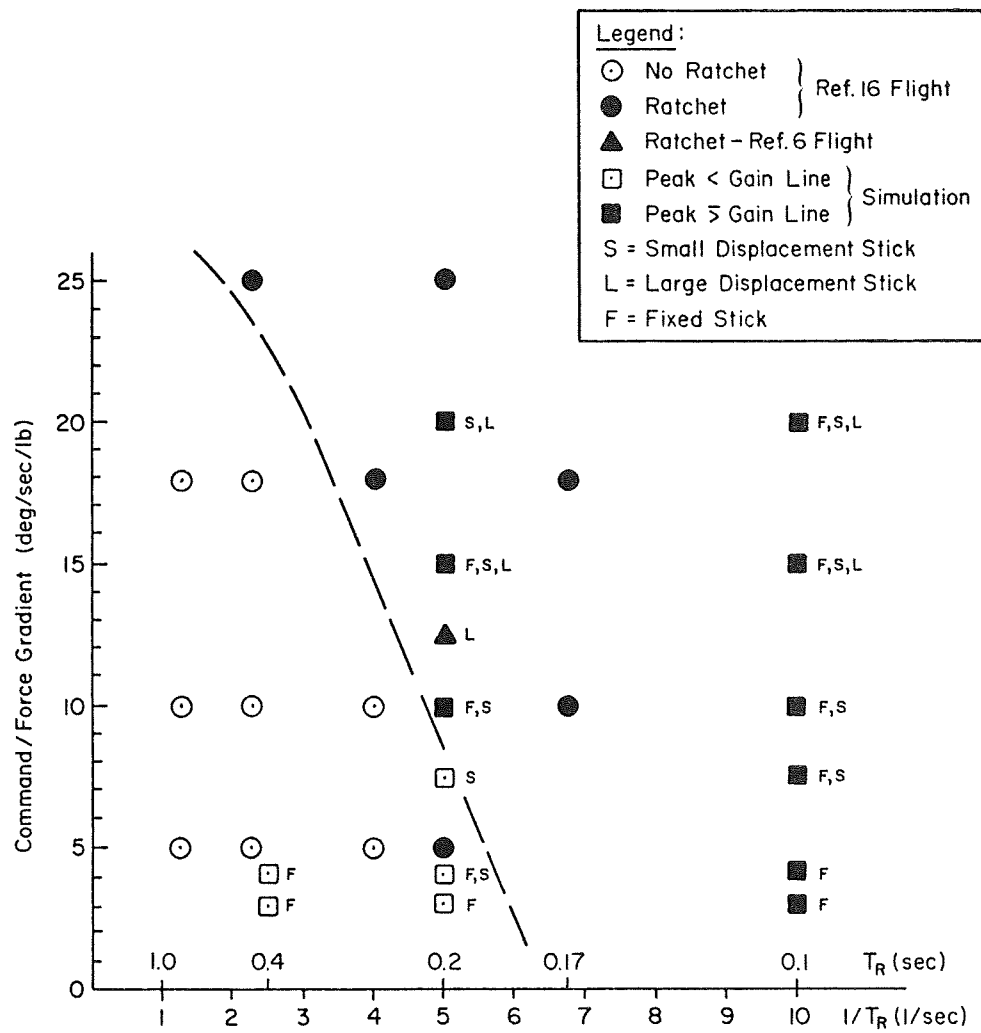
Figure 16. Roll Ratchet Comparison, Flight and Simulator

15.23

included in the roll rate command prefilter to "eliminate high frequency noise." Even so, this one case of ratchet tendency was observed.

The square symbols in Fig. 16 are configurations investigated in the fixed-base simulation. The open symbols identify configurations for which the Y_pY_c zero dB line did not pass through the neuromuscular peak (no ratchet possibility). The shaded squares identify configurations for which the zero dB line passed through the peak (ratchet possibility). The letters F, S, L reflect the displacement of the simulator side-stick. It is likely that the L side-stick most closely matched the NT-33 center-stick characteristics.

There is very good correlation between the flight and lab simulation ratchet tendencies shown in Fig. 16. The dashed line appears to separate the non-ratchet from the ratchet configurations except for the two or three lowest command/force gradient configurations at $T_R = 0.2$ sec. It is possible that this difference may be related to wrist (simulation side-stick) versus arm (flight center-stick) neuromuscular subsystem contributions at the lower command (higher force) configurations. The good agreement between flight and simulator results is interpreted as an encouraging validation of the simulator definition of ratchet potential -- i.e., neuromuscular peaking cut by the Y_pY_c zero dB line.

Pilot-Manipulator System Asymmetries

It was noted in the discussion of the influence of the command/force gradient on crossover in Fig. 6, that the control bandwidth ω_c decreased markedly as the command/force gradient decreased below 4 deg/sec/lb. The reason for this can be observed in the time traces of Fig. 17. The trace on the left is the random rolling motion of the target. The trace in the middle is the roll error between the target and the controlled element, the trace on the right is the stick force input to the controlled element. It will be noted on the force trace that in roll to the right the stick force rarely exceeds 5.5 lbs, but in rolls to the left the force frequently is as high as 8 lbs and shows a maximum peak at 11 lbs. This is consistent with the commentary[3,11] where the pilots indicate difficulty in generating rolls to the right using the thumb, but have little difficulty in rolls to the left where they can use the entire palm of their hand to generate the force. Thus we see bi-modal control in the traces of Fig. 17 with larger magnitude, shorter duration forces in rolls to the left and lower magnitude, longer duration forces being used in rolls to the right. Notice that the roll error average is approximately zero in the middle trace. Thus the area under the force traces for left vs. right maneuvers must be approximately the same. For right rolls, lower forces are held for longer periods of time. This results in a lower crossover or bandwidth for right rolls as compared to left rolls and hence a lower average bandwidth for the run. This bi-modal control characteristic was most evident for the 3 deg/sec and 4 deg/sec/lb controlled element or command force gradients, but was also evident up as high as the 7.5 deg/sec/lb. Thus the reduced bandwidth shown in the Fig. 6 plots for the low gain systems. For higher command/force gradients, the forces employed in the tracking task were sufficently low that there was little difference between left and right maneuvers.
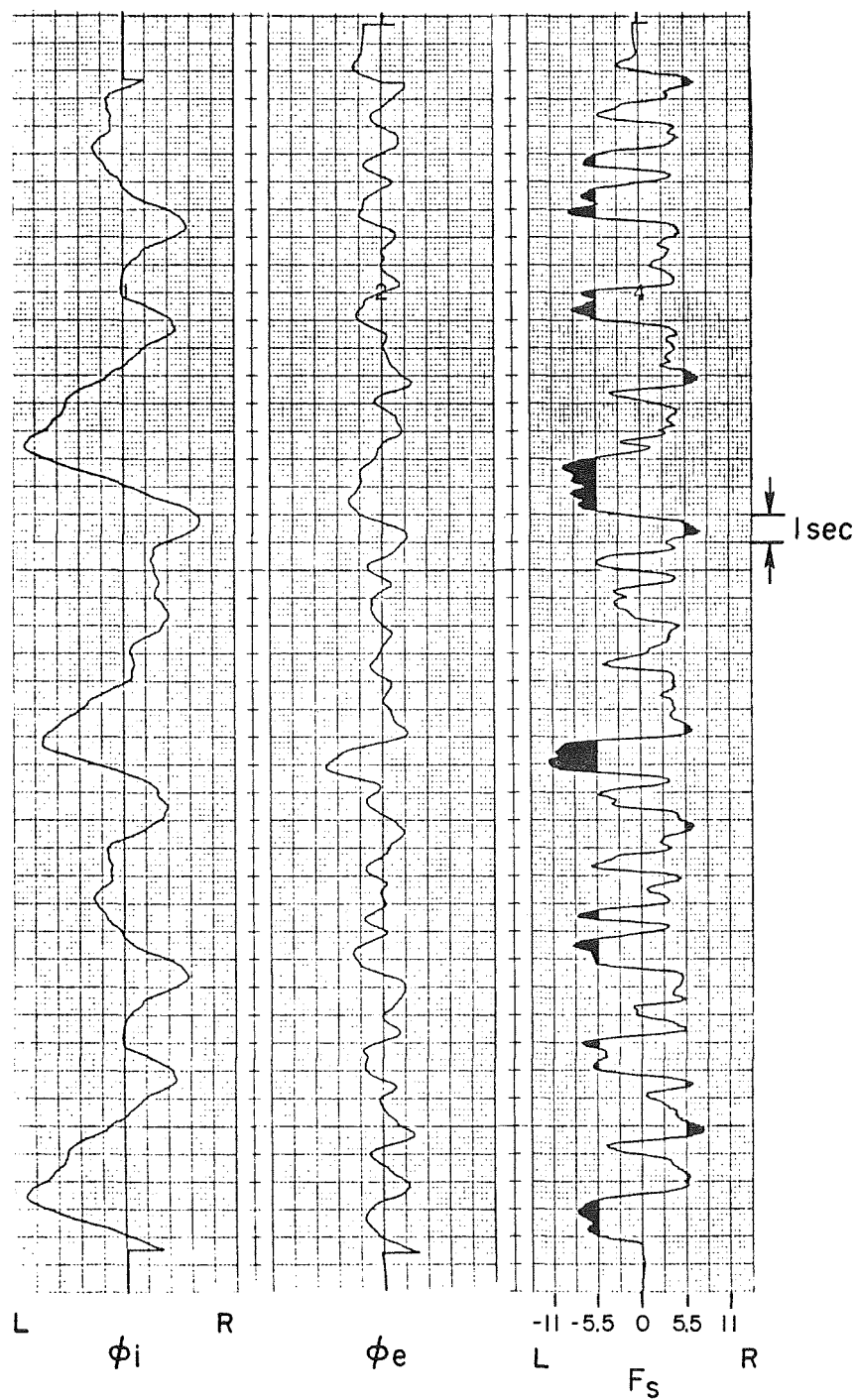
Figure 17. Time Traces of Fixed Force Stick Tracking Task
Run 115, $K_c = 3$, $\tau = 0.067$, $T = 0.1$

15.25

CONCLUSIONS

This fixed-base experimental investigation has identified and quantified interactions between the pilot's neuromuscular subsystem and such aspects of typical modern, high response, roll rate command control system mechanizations as:

- side-stick type manipulator force/displacement configuration

- command augmentation forward loop gain

- controlled element effective lag time constant

- flight control system effective time delay

The simulation results provide insight to high frequency roll ratchet oscillations, low frequency PIO, and roll-to-right control and handling problems previously reported in the production F-16, NT-33 side-stick, and NT-33 roll rate command augmentation investigations. The experimental configurations encompass and/or duplicate a number of actual flight situations and have reproduced control problems observed in flight.

Specific conclusions relating to human pilot dynamic characteristics and possible connection to roll ratchet are summarized in the following.

Human Pilot Dynamic Characteristics

1. Crossover Model Refinements

- The property $\omega_c(Y_c)$ = constant extends over an order of magnitude variation in K_c changes in force gradient. ω_c begins to fall off as very small K_c demand great pilot effort (large K_p) to keep ω_c constant.

- Controller element lags for $Y_c = K_c/(Ts + 1)$ are:

  -- almost exactly cancelled by pilot lead when T $\geqslant$ 0.2 second (lag breakpoint of 5 rad/sec);

  -- partly offset by pilot lead of approximately 1/8 second when T $\leqslant$ 0.2 second.

  Thus the adjustment rule indicating that pilot lead will offset controlled element lags by nearly exact cancellation now has a lower limit at about 1/8 second.

2. Human Pilot Limb-Manipulator Dynamics

- The classical third-order system approximation for the limb-manipulator portion of the human neuro-muscular system is both adequate and an essential minimum form needed to consider pilot-aircraft system dynamic interactions in the frequency range from 8-20+ rad/sec.

- The peaking tendency (damping ratio, ζ_N) of the quadratic component of the third-order approximation is a very strong function of the controlled element dynamics -- in essence this feature can be "tuned" by adjusting controlled element properties.

- For all stick force/displacement characteristics investigated the highest ζ_N (smallest peaking tendency) occured for $Y_c = K_c/s$ controlled elements.

- Pure time delay induces a greater peaking tendency than an equivalent time lag.

- Distinct peaking tendencies occured for fixed and small stick deflections for $\tau = 0.07$ and 0.1 second.

- The controlled element form which exhibited the maximum peaking tendency ($\Delta AR = 7$ dB) was $Y_c = K_c e^{-\tau s}/s$, for $\tau = 0.07$ sec. Higher and lower values of τ resulted in less peaking.

- For large stick deflections the peaking tendency is minimized or non-existent.

Roll Ratchet Connections

- The data strongly support the suggestion that the roll ratchet phenomenon is a closed-loop pilot-vehicle system interaction in which the pilot's neuromuscular dynamics play a central role.

- Ratchet tendencies can be detected in fixed-base simulations by careful tailoring of the forcing function and examination of particular stretches of data. Unlike the case in flight, the pilot may not be aware of the occasional ratchet.

15.27

- The ratchet potential of a given configuration is associated with the degree of neuromuscular system peaking. This peaking tendency can be "tuned" or "detuned" by controlled adjustments in the effective vehicle dynamics.

- This is readily assessed in a fixed-base simulation by describing function measurements in tracking tasks conducted with an appropriate forcing function. Such procedures are recommended as pre-flight development tests with modern fly-by-wire command augmentation systems.

- Ratchet tendencies are most severe on force sensing sidestick manipulators with small stick deflections.

ACKNOWLEDGEMENTS

REFERENCES

[1]Mitchell, D. G., and R. H. Hoh, "Flying Qualities Requirements for Roll CAS Systems," AIAA Paper 82-1356, presented at the AIAA 9th Atmospheric Flight Mechanics Conference, San Diego, CA, 9-11 Aug. 1982.

[2]Harper, Robert P., Jr., In-Flight Simulation of the Lateral-Directional Handling Qualities of Entry Vehicles, Calspan Report No. TE-1243-F-2, Feb. 1961.

[3]Monagan, Stephen J., Rogers E. Smith, and Randall E. Bailey, Lateral Flying Qualities of Highly Augmented Fighter Aircraft, AFWAL-TR-81-3171, Mar. 1982.

[4]Smith, R. E., Evaluation of F-18A Approach and Landing Flying Qualities Using an In-Flight Simulator, Calspan Report No. 6241-F-1, Feb. 1979.

[5]Chalk, C. R., "Excessive Roll Damping Can Cause Roll Ratchet," J. Guidance, Control, and Dynamics, Vol. 6, No. 3, May-June 1983, pp. 218-219.

[6]Allen, R. Wade, Henry R. Jex, and Raymond E. Magdaleno, Manual Control Performance and Dynamic Response During Sinusoidal Vibration, AMRL-TR-73-78, Oct. 1973.

[7]Magdaleno, Raymond E., Duane T. McRuer, and George P. Moore, Small Perturbation Dynamics of the Neuromuscular System in Tracking Tasks, NASA CR-1212, Dec. 1968.

[8]Magdaleno, R. E., and D. T. McRuer, Experimental Validation and Analytical Elaboration for Models of the Pilot's Neuromuscular Subsystem in Tracking Tasks, NASA CR-1757, Apr. 1971.

[9]McRuer, D. T., L. G. Hofmann, H. R. Jex, et al., New Approaches to Human-Pilot/Vehicle Dynamic Analysis, AFFDL-TR-67-150, Feb. 1968.

[10]McRuer, D. T., and E. S. Krendel, Mathematical Models of Human Pilot Behavior, AGARDograph No. 188, Jan. 1974.

[11]Garland, Michael P., Michael K. Nelson, and Richard C. Patterson, F-16 Flying Qualities with External Stores, AFFTC-TR-80-29, Feb. 1981.

[12]Hall, G. Warren, and Rogers E. Smith, Flight Investigation of Fighter Side-Stick Force-Deflection Characteristics, AFFDL-TR-75-39, May 1975.

[13]Magdaleno, R. E., and D. T. McRuer, Effects of Manipulator Restraints on Human Operator Performance, AFFDL-TR-66-72, Dec. 1966.

[14]McRuer, D. T., and R. E. Magdaleno, Human Pilot Dynamics with Various Manipulators, AFFDL-TR-66-138, Dec. 1966.

[15]Johnston, D. E., and D. T. McRuer, Investigation of Interactions Between Limb-Manipulator Dynamics and Effective Vehicle Roll Control Characteristics, Systems Technology, Inc., TR-1212-1, June 1985 (forthcoming NASA CR).

FLIGHT TEST OF A DISPLACEMENT SIDEARM CONTROLLER

Annual Conference on Manual Control 17 June 1985

Andrew L. Lippay, Ronald Kruk, Michael King
CAE Electronics Ltd, 8585 Cote de Liesse,
St Laurent, Quebec, Canada, H4L 4X4

and

Murray Morgan,
National Aeronautical Establishment,
Flight Research Laboratory,
Uplands Airport, Bldg U-61, Ottawa, Canada

ABSTRACT

A six-axis displacement-stick sidearm controller was developed to enable single-handed control of remote manipulator operations in space. Application of such a device to vehicular flight control has been a prime objective ever since CAE Electronics was involved in the TAGS program. With a working model available, piloted evaluation became possible in a fly-by-computer variable-stability research aircraft, originally a Bell 205 helicopter.

Following preliminary trials, the original mechanization was limited to three rotational axes and a linear one, analogous to the collective stick. A newly designed short stickgrip was mounted and the spring force pattern adjusted to suit the helicopter flight control environment.

A standard set of test maneuvers was flown by four experimental pilots with conventional helicopter flight controls and with sidearm controllers equipped with two different handgrips. Existing data from flight tests with an isometric-stick controller were added to complete the comparison. The displacement controller consistently achieved a rating of 3.0 to 3.5 on the Cooper-Harper scale, on par with the conventional controls. The learning period was generally short, with the controller becoming "transparent" to the pilot, giving the subjective impression of direct control of the helicopter lift vector.

The same basic controller design has been tested in spacecraft and remote manipulator simulations with very promising results. In each application operator/system integration was rapid and positive. The results demonstrate

feasibility and support the design philosophy of using deflection as well as force to generate proprioceptive feedback.

Preliminary evaluations in space systems simulations generally showed good operator/astronaut acceptance, reduced training/familiarization requirements and - in some cases - significant improvement in time-to-target control performance. A second-generation engineering effort is currently in progress to produce high-quality units for formal testing and eventual flight qualification.

1.0   INTRODUCTION

The appearance of on-board computers and advanced flight control systems has greatly increased the scope of aircraft performance and mission complexity that could be handled by human pilots and has caused radical changes in the nature of the piloting task. It has, therefore, become necessary to re-examine the physical interface which puts the pilot in direct contact with the flying task, namely the manual flight controls.

Conventional helicopter controls occupy all limbs of the pilot most of the time. This leaves no further capability for command tasks (e.g. forward speed control in future helicopters with auxiliary thrust). The controls occupy much prime cockpit space and are seldom operable by either hand to enable a wounded pilot to fly home. In precision maneuvers the collective-cyclic stick configuration may force the pilot into a "helicopter crouch" with resulting fatigue and spinal ailments due to the combination of poor posture and the high vibration environment. A multi-axis sidearm controller would leave one hand free and could relieve most of the other problems as well.

In some space applications currently under development there are scenarios where a vehicle and a dextrous manipulator may have to be operated concurrently. No one expects human operators to control 12 or more individual parameters simultaneously, continuously and accurately. However, a device whose dynamic characteristics and geometry correspond directly to the outer loop parameters may become "transparent" to the operator and promotes an intuitive mode of manual control. A pair of such transparent, function-oriented command devices may be manageable, with some sequential limitations, even in a proportional control system.

The principal difficulty in this proposition lies not in the derivation of electrical or mechanical command signals, nor in their processing, but rather in the packaging and cascading of the command axes in such a way that the controller movements remain compatible with the articulations of the human arm and hand, while matching the desired end results and system responses. Ideally, any related displays should also be harmonized with controller movements.

The objective of the present effort is to achieve a basic flightworthy controller design that satisfies the principal human-machine interface requirements and which could be optimized for a wide range of flight and remote manipulator applications with a minimum of modifications. The basic rationale for controller design, if correctly stated, should hold for a long time and for many control system variations.

This paper is intended as a progress report rather than as a comprehensive study of the state of the art. A brief summary of principal considerations and development drivers is offered by way of rationale.

2.0 SIX-DEGREE-OF-FREEDOM CONCEPT

2.1 Single Point Command Input

A single-point input device was envisaged, capable of commanding all vehicle responses, operable by either hand, in rate or position control modes. Leaving one hand free for such tasks as display management or communications selections was considered important. Auxiliary controls operable by the same hand were also to be accommodated.

2.2 Command Harmony

Spatial command harmony was considered essential, that is, the inputs (controller movements) would be followed by a vehicle or system response in the same sense and direction as the controller has moved, enabling the normative or inner model developed by the pilot to serve as a predictor in terms of the desired end results. Agreement between the predicted and actual responses largely determines the pilot's assessment of the task difficulty and the handling characteristics of the vehicle, and greatly influences overall success and performance.

2.3 Force Feedback

Manual controls also fulfill the role of a tactile display. The human hand can interpret loading forces appearing on the handgrip in terms of demands imposed on the system and its expectable response, enabling the pilot to develop a beneficial phase lead. This method of limiting accelerations or demand is preferable to that of derating vehicle responses in the control system; the latter may appear as sluggishness and invite poor pilot acceptance or even pilot-induced oscillations. (PIO)

Active force feedback raises a very severe packaging problem in integrated controllers, especially if redundancy is required. It appears, however, that passive forces generated within the controller and optimized for the command task may be adequate for most purposes.

2.4 Displacement vs Force Stick

On the basis of physiological characteristics and experimental results it may be said that the human operator is able to control motion or displacement with much greater ease and accuracy than he can control force. It was concluded that the intrinsic and near-instantaneous proprioceptive feedback on the command inputs developed by controller movements combined with a harmonious force pattern was essential. A deflection-stick concept was adopted despite the many obvious engineering advantages of the rigid stick.

2.5 Spring Return and Damping

Traditionally, spring return forces have ben regarded as necessary to restore zero command or trim outputs for the hands-off condition. During informal simulation trials it was found that pilots could not differentiate between spring and damping forces in the short term, and that a heavily spring-loaded stick will cause drift with or against the force gradient because of accommodation to constant pressure which develops quite quickly. It is proposed that for many rate control applications, rate dependent damping and good null identification may be sufficient.

3.0 RELATED WORK

During 1968-72 a flight demonstration of the Tactical Aircraft Guidance System (TAGS) was conducted as a joint Canada-US Army project involving a CH-47 helicopter equipped with a digital triplex redundant fly-by-computer system. One principal objective was to increase flight safety and mission capability with the prospect of using marginally trained pilots in Viet Nam.

A Canadian contribution was a four-axis sidearm controller with linear fore-aft movement controlling forward speed, roll movement giving lateral speed at hover or flight path direction over 35 kts forward speed. A stick twist input controlled spot turn at hover or aircraft heading at speed. A pivoting armrest controlled vertical speed. This was later relocated to the conventional collective stick.

The mechanical design left much to be desired due to a highly constrained installation, which also prevented the armrest to be correctly adjusted to the individual pilot. Hence the failure of the vertical control in which the pilot lost contact with the arm support and hand reference. Nevertheless, 103 test flights were conducted successfully, including sling loads, precision and cross-country flights, and much valuable experience was gained.

In 1974-77 the Remote Manipulator System of the Space Shuttle required a command device. A six-axis controller was recommended but was later considered a high schedule risk and two three-axis controllers were used instead. One controls translations of the end effector, its near-linear movements are coordinated with the prime display means associated with the operation in a fly-to fashion. The rotational controller has three angular freedoms and controls the attitudes of the end effector.

In response to a NASA request, CAE Electronics performed a study to show the feasibility of a six-axis controller for spacecraft flight and remote manipulator systems. A state-of-the-art survey and literature search revealed many attempts but no mature designs with six degrees of fredom, and precious few with more than three. (1979) As a follow-on effort to this study, CAE developed controller models which were used in the Manipulator Development Facility of the NASA Johnson Space Center, in the Manned Maneuvering Unit (MMU) simulation at Martin-Marietta Denver. The original demonstrator model is currently installed at NASA-Marshall Space Center in a dextrous manipulator system being developed for spacecraft servicing.

In early 1984 , CAE approached the National Aeronautical Establishment of the National Research Council of Canada to test the device as the primary flight controller of a highly maneuverable helicopter. A four-axis version was configured and preliminary flight tests were conducted. An improved engineering model was built and is undergoing flight testing. The results of flight testing this unit are presented later in this report.

4.0 DESCRIPTION

The basic controller design has three rotational and three linear motions. A universal ball-shaped handgrip contains the gimbal for two of the rotational axes (pitch and roll), the third is centered on the shaft supporting the ball. The three linear axes (X, Y, Z) are contained in the enclosure below the handgrip, together with the base-mounted electronics which pre-process the transducer outputs. Figure 1 shows the basic configuration.

This geometry allows all hand forces to pass through the same point, i.e. the center of the ball; the linear (translational) axes are constrained against torques developing due to their offset from this center. Thus any tendency to cross-coupling between axes is minimized and the ball is largely insensitive to hand position providing that the controller is located correctly with respect to the forearm and armrest.

The rotational displacements are approximately +/- 15 degrees, the linear excursions +/- 3/8 inch. The total vertical movement as configured for the helicopter collective is approximately 1.1 inches.

Spring breakouts and gradients are adjustable by replacing the spring sets, and can be made non-symmetrical. The vertical axis has damping which is rate dependent and pilot-adjustable over a vernier scale of its total force range.
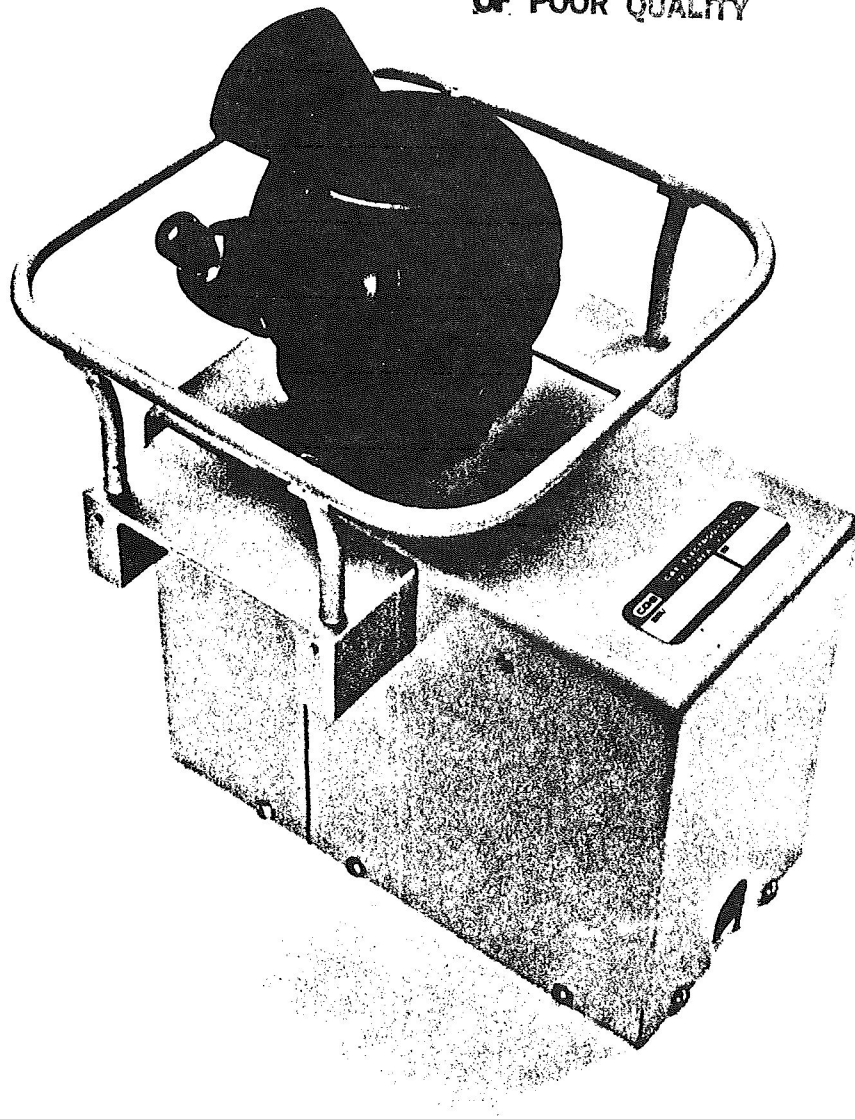
FIGURE 1   BASIC CONTROLLER CONFIGURATION

16.7

5.0 MANNED TESTING AND DEMONSTRATION

The following is based on flight tests conducted at the National Aeronautical Establishment (NAE) Ottawa, Canada. Additional information derived from NASA simulations and related tests is included as appropriate to the topics being discussed.

5.1 Initial Investigations

An early model was installed in the NASA-Johnson MDF (Manipulator Development Facility), in a position corresponding to the rotational controller of the CANADARM remote manipulator system, and the MDF arm was used for capturing and positioning moving targets. The Manned Mobility Unit (MMU) simulation at Martin-Marietta` Denver was temporarily equipped with the model, replacing two three-axis controllers. This unit is currently installed at the NASA-Marshall Space Center where it is used to operate a dextrous manipulator in a development project for satellite servicing and Orbital Maneuvering Vehicle (OMV) operations.

5.1.1 Helicopter configuration

Before conducting the first helicopter experiment, two informal flight development periods were held to adapt the controller characteristics to the helicopter flying task and investigate different handgrip shapes. Two of the generic model's translational axes were disabled (immobilized) and the third was modified as described below. The current version used for helicopter trials is an improved engineering model with helicopter-specific features.

5.1.2 Vertical Axis Modifications

The initial version had a center null position on the vertical axis with spring centering and breakout. For an open-loop collective drive in the helicopter the available range was objectionably short. The null was moved close to the bottom of this linear (vertical) stroke. The light friction levels in the axis resulted in a tendency to PIO. As a quick fix, friction damping was installed but this predictably produced lumpiness in the control due to its stick-slip properties. The final version had fluid damping and no spring return on the vertical motion.

For manipulator control and Manned Maneuvering Unit (MMU) flight the center-null vertical axis was found acceptable.

5.1.3 Ball Handgrip vs Stick Grip

The first models were equipped with a universal spherical
handgrip of approx 3.5 inches in diameter. Helicopter pilots
expressed a marked dislike of the ball handgrip, especially
for large-amplitude maneuvers.   For a quick trial, an
existing   handgrip   developed   for a   semi-rigid   stick
configuration was installed with an adapter ring attached to
the top part of the ball. Figure 2 shows this configuration
The   resulting offset in the hand pressure point introduces
some cross-coupling between the vertical and the pitch axes,
but   the pilots seem to accept this additional workload as
long as they can have a stick grip.   Figure 3 shows the
current helicopter version with a combination ball-grip which
minimizes the offset and combines the ball concept with
special advantages of a vertical stick grip.

For space operations, the ball was found quite suitable even
with an inflated spacesuit glove and was insensitive to hand
positions with remote manipulators.   A thin fin was later
added for fore-aft hand reference, slipping between the index
and middle finger. Pilots and astronauts alike recommended a
smaller, baseball-sized grip.   The diameter of the ball was
eventually reduced to 2.9 inches.
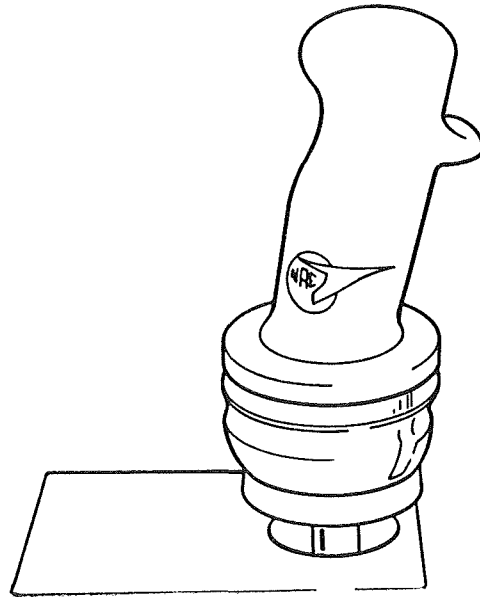
FIGURE 2   EXISTING STICK GRIP ADAPTATION
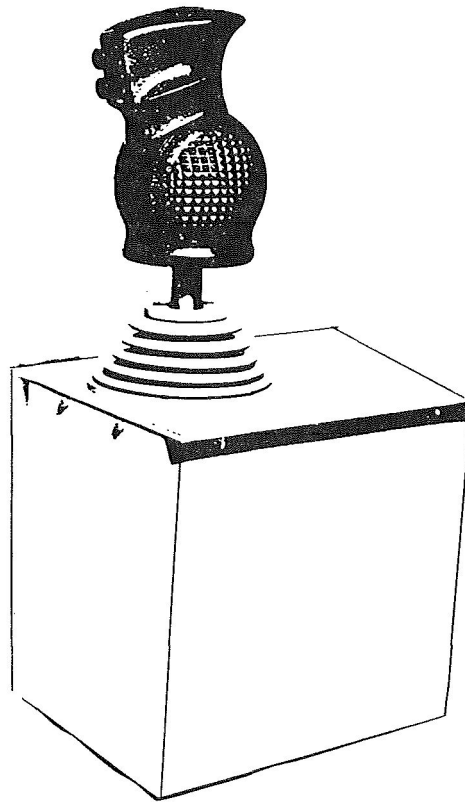
FIGURE 3   HELICOPTER STICK-GRIP CONFIGURATION

5.1.4   Installation Ergonomics

Due to schedule and manpower limitations, rigorous ergonomic
investigations have not yet been carried out to optimize the
hand pressure point with respect to the armrest (where there
is one present) or to the relaxed or preferred hand position.
In most cases, the controller was simply placed to have the
ball center fall where previous devices had their hand
pressure points or where a suited astronaut said he could see
and reach the controller within the framework of existing
vehicle or cockpit design.

In the helicopter cockpit the pilot's armrest acts as an
essential reference surface but may also become an obstacle
to wrist movement in dynamic maneuvers such as autorotation
or quick stop.   As a first step to resolve the ergonomic
problem of arm support and wrist freedom, an adjustable
armrest now replaces the standard unit on the helicopter

seat. Experience thus far indicates that orientation and positioning of the controller are critical factors in control performance and pilot acceptance. Therefore these issues will be addressed in greater detail as the development program continues.

5.1.5 Rotational Command Harmony

The order in which the rotational axes in the manipulator were cascaded resulted in the pitch and roll sensing axes rotating with a yaw input; this meant that there was no fixed relationship of pitch and roll inputs to airframe movements. Helicopter pilots had difficulty compensating for this effect. The problem was temporarily corrected by software transformation as a function of controller yaw angle. This aligned the command axes with the airframe but introduced variations in the effective spring rates in pitch and roll with respect to the transformed sensing axes. While this effect was noticeable under laboratory conditions, it was not reported by any of the evaluation pilots as a difficulty. Nevertheless, it might have had an influence on the overall handling qualities assigned.

The problem was removed by altering the cascading of axes in the next model such that the pitch and roll movements remained aligned to the aircraft pitch and roll axes.

No equivalent problem was reported by manipulator operators and MMU simulation pilots.

5.2 The NAE Airborne Simulator Facility

The Flight Research Laboratory (FRL) of the National Aeronautical Establishment of Canada (NAE) has been actively engaged in research into the use of integrated side-arm controllers in an airborne flight simulator for the last four years.

5.2.1 The Airborne Simulator

The FRL has extensively modified a Bell 205-A single-engine single main rotor helicopter to generate a variable-stability test bed with full-authority fly-by-computer command capability. An on-board digital system senses many environmental and aircraft state parameters, processes them in a variable-configuration flight control system and has a 64-channel digital recording capability. The facility is fully described in Reference (5).

5.2.2 Control Signal Conditioning

In addition to the sense axis transformation described above, inputs from the controller were subject to the following processing:

    o   Normalising gain
    o   Filtering (16 rad/sec first order low pass)
    o   Deadband
    o   Sensitivity setting gain

The two gains in series, while redundant, were useful because of ease of comparative documentation. A typical input conditioning chain is shown in Figure 4., the values used for the various conditioning parameters are given in Table 1.

Normalised stick input → $\dfrac{16}{S + 16}$ → Filter → Deadband → Sensitivity G →
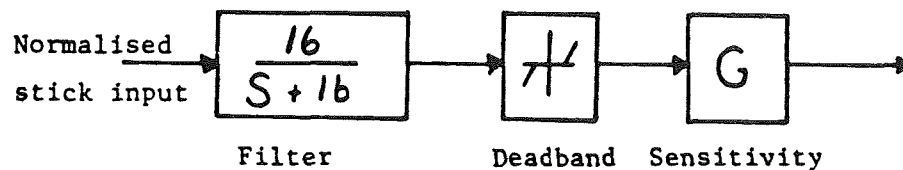
FIGURE 4   TYPICAL INPUT CONDITIONING

16.12

Axis	Filter B/point	Deadband	Sensitivity
Roll	16.0 r/sec	0.5%	0.5
Pitch	16.0	0.5%	0.5* 1.0**
Yaw	4.0 r/sec	0.5%	1.0
Collective	NIL	NIL	0.8

* Grip configured
** Ball configured

TABLE 1   SIGNAL CONDITIONING PARAMETERS

5.3   Experiment Design

Seven representative tasks were flown over a course with position markings alid out on the ground. The tasks included off-level landings and takeoffs, lateral flight, rearward flight, quickstop, spot turn and spot turn with hesitations. The course itself and maneuver standards are described in Reference 3 and are routinely used by the FRL.

Instructions to pilots for the off-level landing and takeoff maneuver were as follows:

> Establish a 10 foot hover, land within the marked box with a continuous downward motion of the aircraft, no hesitations and no vertical velocity reversals. Desired performance: complete task safely.

Raise the aircraft to a level attitude with the up
hill skid in contact with the ground, hesitate for 5
seconds in that condition then make a clean transition
to a 10 foot hover. Desired performance: Safe
completion with no return to both skids and no
premature lift-off from partial contact hover.

Each of the four FRL research pilots flew the full set of
tasks using conventional centre mounted controllers, the CAE
controller with ball grip, and the same device with the NAE
grip. Cooper Harper ratings were requested for each task and
verbal comments and written debriefs were taken also.
Previous results on the same tasks flown with a force-stick
sidearm controller were included in the comparative
statistics, as recorded in Reference 3.

5.3.1   Control System Configuration

The aircraft control configuration for the primary experiment
was a primitive system permitting comparison with
conventional controls. This configuration had rate damping,
augmentation in pitch, roll and yaw, a model of the 205
stabiliser bar, and collective inputs were de-coupled from
the yaw axis. The rate damping augmentation was scheduled
with airspeed to provide a vehicle with approximately -2 deg
per second damping throughout the envelope in all three
rotational axes. Collective control was simple direct drive.
A slow follow-up trim system was installed which summed a low
gain [0.25] integral of the controller output with that
output. A typical control system channel is shown in Figure
5, while the gains used are tabulated in Table 2.
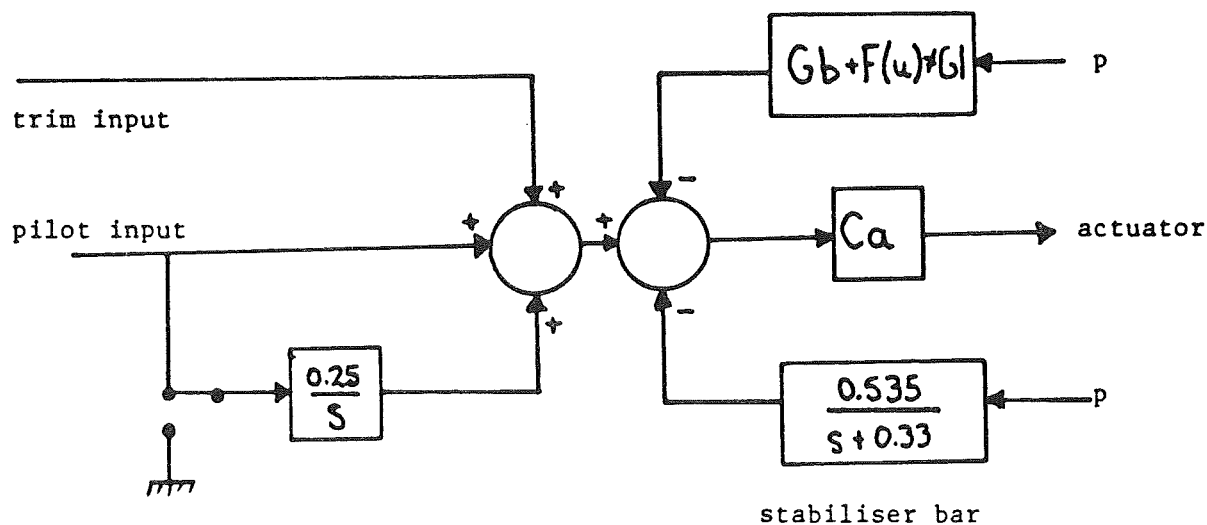
FIGURE 5   TYPICAL CONTROL CHANNEL [ROLL]

Axis	Delta Gain	Basic Gain	Actuator Coefficient
Roll	0.276	0.22	0.61
Pitch	0.39	0.61	0.66
Yaw	0.41	0.20	0.85
Collective	N/A	N/A	0.81

TABLE 2   CONTROL SYSTEM GAINS

5.4   Results and Discussion

5.4.1   Familiarization and Training Time

As a general characteristic, both versions of the controller
required very little time to become familiar to pilots,
astronauts and operators.   This is attributed to the spatial
command harmony achieved and the absence of mode switching
and other activities which normally result in breaking of
contact between the hand and the controller.

The NAE pilots had extensive helicopter experience, some
including sidearm controllers.   They all became sufficiently
familiar with the controller during the first hour of flight
to perform to the required standards.   At least one other
pilot with no previous sidearm experience was able to fly
nap-of-the-earth after approximately 20 minutes; his
comments during debriefing indicated that he was able to
treat the controller as if it were transparent, and fly the
aircraft intuitively.

The manipulator and spacecraft simulations showed that operators need only rudimentary instructions and a self-paced training period which is extremely short in comparison with other control mechanizations. The MDF arm was repeatedly operated with surprising proficiency by personnel of various backgrounds without the benefits of even a basic introduction. Preliminary trials with the dextrous manipulator at the Marshall Space Center showed a tendency of significantly reduced task times as well as training requirements even with novice operators.

5.4.2 Pilot Ratings

Figure 6 shows means of Cooper-Harper ratings and standard deviations for all maneuvers and test periods executed to date, for a global comparison between conventional controls, an isometric vertical grip, the CAE controller with the vertical grip and with the ball grip.
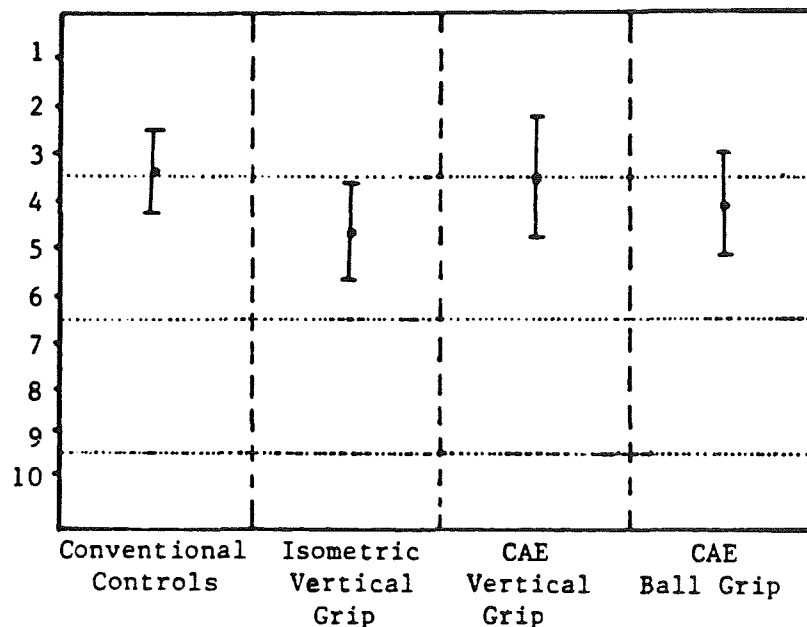
FIGURE 6 COOPER-HARPER RATINGS; COMPARATIVE SUMMARY

As the Bell 205-A with controls configured for the present study is a marginal Level One vehicle in handling qualities, there were few occasions when pilot compensation for handling deficiencies was not a factor. Within this overall constraint, however, there is a consistent hierarchy of handling quality ratings among the controller types evaluated.

Generally, the conventional controls are rated highest, with a mean of 3.3, satisfactory but with some mildly unpleasant characteristics. The CAE controller with the FRL stick grip is rated next, with a mean of 3.6, acceptable but with unpleasant characteristics. The CAE unit with the ball grip is rated 4.1, still acceptable but unpleasant. Last is the force stick at 4.7, tending towards unacceptable for normal operation.

The data from the individual tasks shows the same trend with one variation (See Figure 7a to 7g). The force stick provides unequivocal Level Two handling qualities for landing on flat surfaces. Off-level landings and takeoffs were not conducted systematically with this device. The performance of the CAE controller with the ball grip is rated much poorer than the conventional controls or the same unit with a vertical grip. In lateral and rearward flight the conventional controls are rated much better than any of the others and the force stick is again last. In the quick stop maneuver the conventional controls and the CAE controller with the FRL grip are similar, but the ball grip is worse than the force stick. In spot turns with and without hesitation the CAE controller is rated slightly ahead of the conventional controls but with a greater spread in ratings.
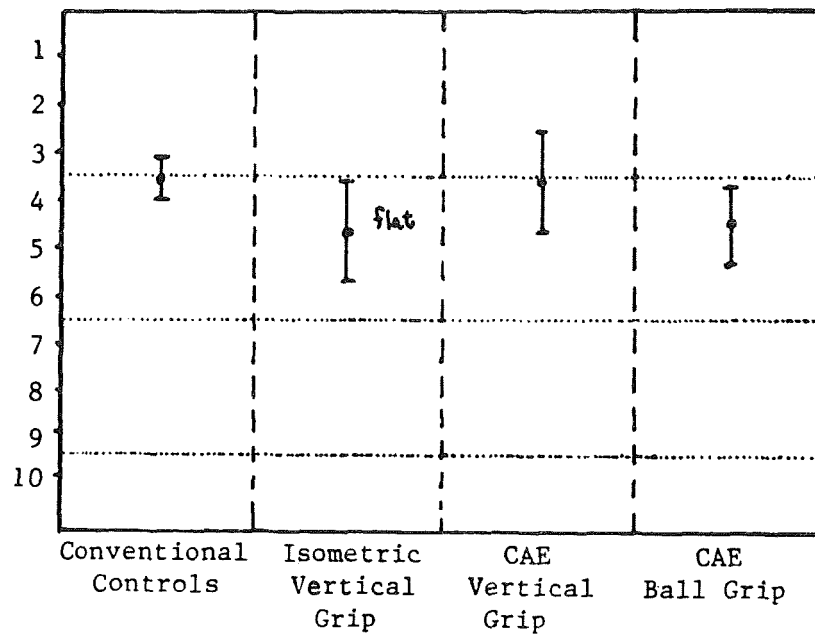
FIGURE 7a   OFF-LEVEL LANDING MANEUVER RATINGS
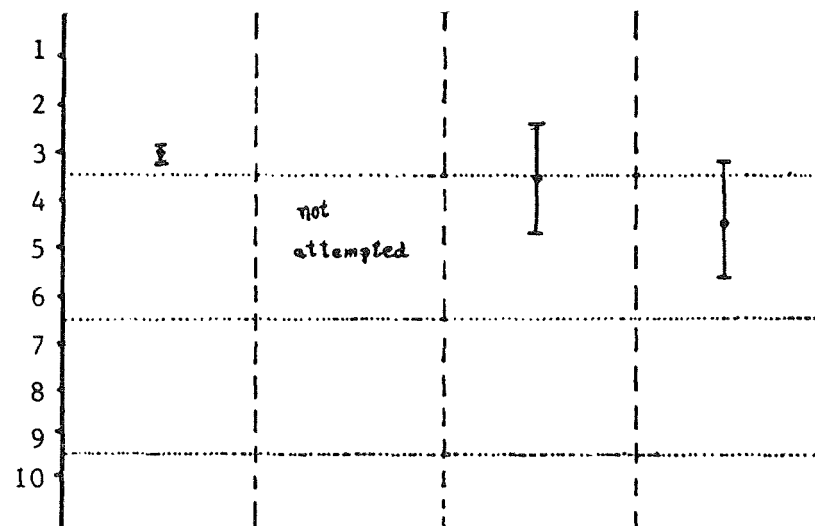
FIGURE 7b   OFF-LEVEL TAKEOFF MANEUVER RATINGS

16.18
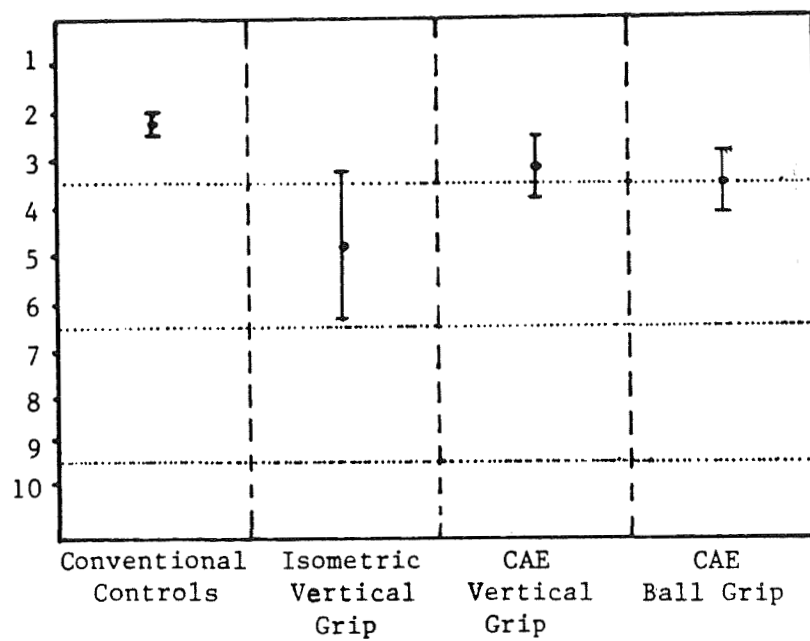
FIGURE 7c   LATERAL FLIGHT MANEUVER RATINGS
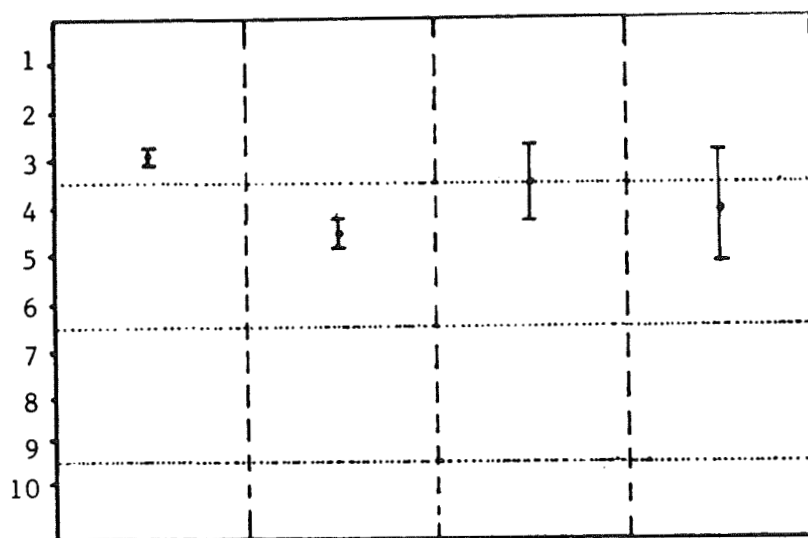
FIGURE 7d   REARWARD FLIGHT MANEUVER RATINGS
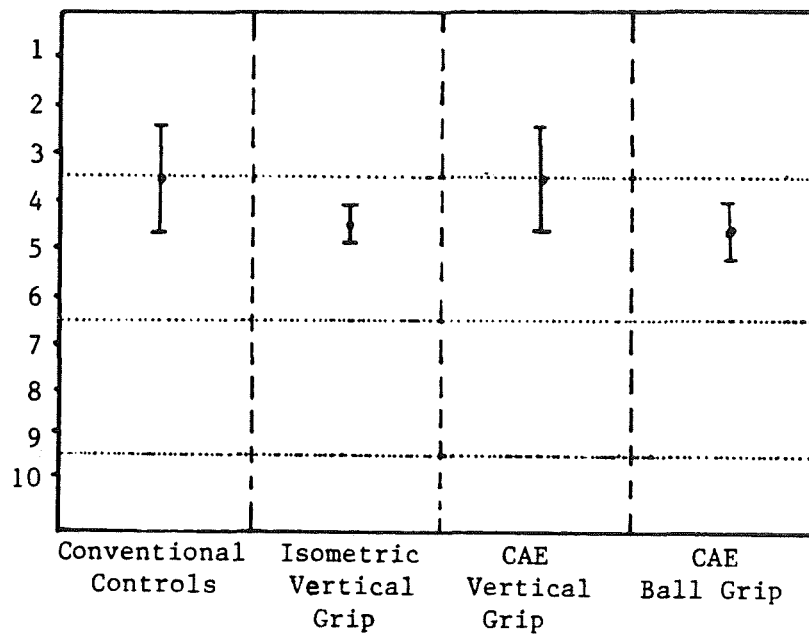
FIGURE 7e   QUICK STOP MANEUVER RATINGS
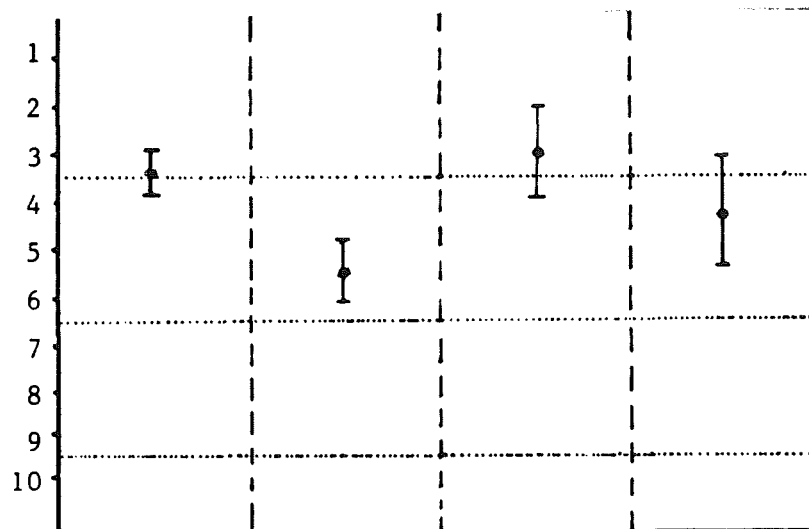
FIGURE 7f   SPOT TURN MANEUVER RATINGS

16.20
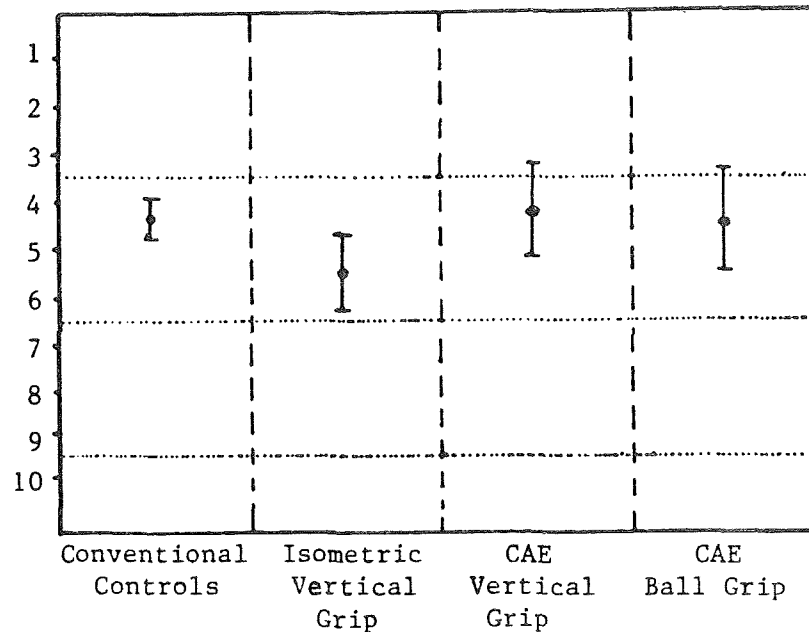
FIGURE 7g   INTERRUPTED SPOT TURN RATINGS

6.0 DISCUSSION

In nearly all cases, performance with sidearm controllers is degraded as compared to conventional controls. This is most severe in cases where finely coordinated multi-axis inputs are required such as in off-level landing and takeoff. Degradation is least severe or is even reversed where vehicle characteristics are the limiting factor such as in spot turns; the aircraft demonstrates a powerful yaw/roll coupling when rapid yawing motions are abruptly terminated. This results in significant lateral instability and the vehicle is a definite Level Two machine in these conditions even with conventional controls.

Identifying the cause for poor performance with the force stick is relatively easy. Lack of immediate feedback from the controller itself on control inputs means that the pilot must wait until the vehicle responds to assess whether the input was appropriate. This introduces a lag which raises pilot workload substantially and even so, stability may be inadequate to permit off-level landings to be conducted as a routine maneuver.

16.21

With displacement controllers the case is somewhat more
complex. Immediate feedback on control input is certainly
available. However, ensuring that this feedback is
appropriate in terms of rate and direction proves to be a
distinctly non-trivial task. With the ball grip, the
"natural" hand position seemed to rest the palm over the top.
This did not generate inherent corrrelation between the
sensed hand position and the lift vector. As well, in
dynamic maneuvers involving the collective a tendency to
cross couple needs to be actively neutralized. With upward
collective inputs the ball does not provide a supporting grip
surface, producing a subjective impression that the aircraft
will fall out of the sky unless the ball is held in a death
grip. This leads to white knuckles and fatigue, combined
with excessive tension in the hand and forearm muscles which
further reduces the precision of inputs and hence the ability
to compensate for cross-coupling.

These problems are most noticeable when large-amplitude up-
collective inputs must be combined with precision in other
axes, such as in quickstop maneuvers and off-level takeoffs.
Shifting the handgrip to the side of the ball reduces these
problems somewhat, but the lack of adequate grip surface and
tendency to cross couple remains. (The ball was left bare
and smooth in order not to force a given hand position on the
evaluation pilots.)

As the data shows, assessment of the controller with a
vertical grip improved dramatically over the ball, coming
close to conventional controls, and this by pilots with up to
2000 hours of conventional helicopter experience. That this
performance could be achieved despite the fact that the
damping characteristics and spring forces were still not
optimal, the rotational inputs were off-axis and no
systematic ergonometric work has been done to verify the
installation, was a good indication that the concept of using
displacement for control feedback in a sidearm controller is
valid.

Interestingly, the ball grip was preferred by astronauts and
operators in manipulator and MMU simulations. No significant
cross-coupling problems were reported even with inflated
space gloves. There may be several factors here, one being
the strong familiarity of the vertical stick grip to
helicopter pilots and its obvious analogy to the lift vector.
Furthermore, the effects of command inputs in low-altitude
precision hover and the resulting whole-body feedback cueing
have a much greater effect than in a slow-moving manipulator
where the operator is much more loosely coupled.

REFERENCES AND BIBLIOGRAPHY

(1)   Sinclair, M, and Morgan, M. : AN INVESTIGATION OF
MULTI-AXIS ISOMETRIC SIDE-ARM CONTROLLERS IN A VARIABLE-
STABILITY HELICOPTER. NAE LR-606, NRC No. L9629, August
1981

(2)   Morgan, M.: SOME PILOTING EXPERIENCES WITH MULTI-
FUNCTION ISOMETRIC SIDE-ARM CONTROLLERS IN A HELICOPTER. Proc
joint NASA/AHS Helicopter Handling Qualities Specialists
Meeting, NASA CP 2219, April 1982

(3)   Morgan, M.: A PILOT'S EXPERIMENT IN THE USE OF MULTI-
FUNCTION SIDEARM CONTROLLERS IN A VARIABLE STABILITY
HELICOPTER. Ninth European Rotorcraft Forum, Stresa, Italy,
September 1983

(4)   Heffley, R.K.A: COMPILATION AND ANLYSIS OF HELICOPTER
HANDLING QUALITIES DATA. VOL 1, Data Compilation NASA CR-
3144, August 1979

(5)   Sattler, D: THE NATIONAL AERONAUTICAL ESTABLISHMENT
AIRBORNE SIMULATION FACILITY. NAE Misc. 58, 31st Annual CASI
General Meeting, Ottawa, Canada, May 1984

(6)   Morgan, M: A COMPARATIVE STUDY OF AN ARTICULATED
INTEGRATED SIDE-ARM CONTROLLER IN A VARIABLE STABILITY
HELICOPTER Proc 41st Annual Forum of the American Helicopter
Society, May 1985

(B)   Sherbert, A.T: AN INVESTIGATION OF ADVANCED FLIGHT
CONTROLLERS FOR AEROSPACE VEHICLES. The Boeing Company,
Vertol Division D8-2424-1, July 1969 Proprietary Document.

(B)   Pitrella, F.D: PROBLEMS AND METHODS IN EVALUATING
FORCE STEERING AND OTHER FLIGHT CONTROL MODES.
Forschungsinstitut fuer Anthropotechnik, Meckenheim, West
Germany, Bericht Nr 20, A 3124, December 1974

(B)   Lippay, A.L: MULTI-AXIS HAND CONTROLLER FOR THE SPACE
SHUTTLE REMOTE MANIPULATOR. Proceedings, 13th Annual
Conference on Manual Control, 1977

(B)   CAE Technical Report SRMS-036, TRANSLATION HAND
CONTROLLER EVALUATION AND DESIGN RECOMMENDATIONS.

(B)   Lyman, Prof John, UCLA Los Angeles, Personal
Communications on the NORMATIVE MODEL.

(b)   Bejczy, A., Jet Propulsion Laboratory, Pasadena:
Personal Communications on MULTI-AXIS CONTROLLER PROPERTIES
AND RESEARCH.

ANTHROPOMETRIC CONSIDERATIONS FOR A
FOUR-AXIS SIDE-ARM FLIGHT CONTROLLER

William B. DeBellis
U.S. Army Human Engineering Laboratory
Aberdeen Proving Ground, Maryland

INTRODUCTION

This investigation is the first in a series of studies to generate a data base on multiaxis side-arm flight controls. The rapid advances in fly-by-light technology, automatic stability systems, and onboard computers have combined to create flexible flight control systems which could reduce the workload imposed on the operator by complex new equipment. This side-arm flight controller combines four controls into one unit and should simplify the pilot's task. However, the use of a multiaxis side-arm flight controller without complete cockpit integration may tend to increase the pilot's workload.

Background

One of the purposes of developing a multiaxis side-arm flight controller is to eliminate the three flight controls (cyclic stick, collective lever, and yaw pedals) required to control a helicopter and combine their functions into a single control. The new flight controller should reduce the piloting task by freeing the pilot's left hand for other tasks.

Fly-by-light technology is being developed through a combined effort of the Army's Aeromechanics Laboratory and Boeing Aircraft Corporation and through the advanced digital/optical control system (ADOCS) program. This technology uses encoded signals which are transmitted over fiber optic cables. The main purpose of the ADOCS program is to demonstrate that an Army helicopter can be flown with a multiaxis side-arm controller and fly-by-light technology. The impact on the pilot's workload has not been addressed.

Because of rapid technological advances in flight controls, there is not yet a data base for crew station designers and evaluators to work with. We believe that many positive benefits may be realized through the use of the multiaxis side-arm flight controller in Army aircraft. The controller will have a strong influence on aircrew station design. There will be more flexibility in seating posture and airframe design, and fabrication will be simplified. A greater range of male and female personnel may be able to fly; and control inputs can be "tuned" to each pilot, airframe, aircraft, flight phase, and mission phase for optimum effectiveness.

Two possible drawbacks to this new technology are that the piloting task may be increased and current operational capabilities may not be fully realized. The standard cyclic and collective control heads contain a significant number of switches which are used to operate various subsystems onboard the helicopter; the ADOCS programs have not addressed the issue of where to locate these switches if a single flight controller is used.

In addition, normal mission and piloting tasks have not been imposed on the simulation studies.

The U.S. Army Human Engineering Laboratory (HEL), through the use of its simulation and computational facilities, has designed a series of investigations to develop the data base and to determine if the side-arm flight control concept is operationally beneficial.

In a following investigation, pilots will fly the HEL simulator with the controller adjusted either orthogonal to the airframe or for the comfort of the pilot. If it can be shown that a position based on comfort is suitable, fatigue may be reduced and the piloting task simplified.

OBJECTIVES

The main objectives of this investigation were to: (a) determine the physical location of the multiaxis side-arm flight controller and armrest which is the most comfortable in a static situation and (b) determine the effects of CB protective gear on those location parameters.

METHOD

Description of Multiaxis Controller

Figure 1 shows the multiaxis controller used during this investigation. It is a small deflection force controller with characteristics as shown in Table 1. The design is not based on any specific Army requirement and was purchased off the shelf.

Figure 2 shows the test setup. Both the armrest and multiaxis controller could be adjusted in rotation and position with respect to each other and with respect to the seat reference point (SRP) as defined by MIL-STD-1333. A nonform-fitting armrest provided consistency within the investigation by not forcing the forearm into a particular position.

Figure 3 shows a pilot in partial mission-oriented protection posture (MOPP). The pilots were fully covered except for their faces. Masks were carried to their left side.

Subjects

Seventy nonpilots and seven Army helicopter pilots were picked from available personnel. Ten percent of the subjects were left-handed and twenty-three percent were female. Included in the subject sample were military personnel assigned to the HEL. All pilots were male. Anthropometric measurements indicate that the subjects were representative of the population as a whole. All subjects were cooperative and did not appear to introduce any artifacts into the data.

17.2

Figure 1. Multiaxis controller.

TABLE 1

CONTROLLER CHARACTERISTICS

MODEL 404-G717 MEASUREMENT SYSTEMS, INC			
PARAMETER	X & Y AXES	Z AXIS	TORQUE AROUND Z
FORCE OVER LINEAR RANGE	+/- 20 lbs	+/- 40 lbs	+/- 60 in-lb
MAXIMUM ALLOWED FORCE	+/- 160 lbs	+/- 528 lbs	+/- 1056 in-lb
SENSITIVITY +/-10%	0.5 volts/lb.	0.25 volts/lb.	0.17 volts/in-lb
DEFLECTION AT MAX OPERATING FORCE	+/- 0.4 in	+/- 0.1 in	+/- 4.0 degs/in-lb

Figure 2. Test setup.

Figure 3. Pilot in partial MOPP gear.

17.4

Procedure

The investigation was conducted in two phases which separated the pilot personnel from the nonpilot personnel. We anticipated that data generated from pilots would be influenced by flight experience and any experience with side-arm tracking controls which would have biased the perception of comfort.

The purpose of the investigation was explained and a series of anthropometric upper body measurements were taken of each subject. The subjects then sat in an AH-64 helicopter seat mock-up with the adjustable controller and armrest at their immediate right side. The subjects were told to sit squarely with their backs in contact with the back of the seat. They were then asked to relax but not to slouch forward. If the seated subjects lowered their right shoulder as if to anticipate contact with the armrest, they were asked to reassume a squared position. The experimenter adjusted the controller and armrest to where the subjects felt them to be comfortable. Once each subject was satisfied with the position of the controller and armrest, a film record was taken of the subject holding the control. Pilots would then wear MOPP and a second film record was taken.

The film record was obtained through the use of three orthogonal data cameras located at the subject's right side, top, and front. The cameras were started simultaneously and ran for approximately 3 seconds. Film records were read on a film analyzer and individual point coordinates were fed directly to the computer, where the data were reduced and analyzed.

RESULTS

Tables 2 through 9 summarize the data obtained in this investigation. Angular data are presented in degrees, while position data are presented in centimeters and referenced to the seat reference point (SRP). Figures 4 through 7 display the sign convention for measurements.

The statistical program used to generate the results was SAS, a statistical and data handling package from SAS Institute, Incorporated. The distributions presented in the summary tables were generated by the SAS univariate program, and the Q_1 and Q_3 values are the first and third quantiles using definition 4. For small sample sizes, the maximum and minimum values replace the quantiles. Selected individual comparisons were accomplished by t test using a pooled variance and assuming a normal distribution.

$$T = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sum X_1^2 + \sum X_2^2}{N_1 + N_2 - 2} \left(\frac{N_1 + N_2}{N_1 * N_2} \right)}}$$
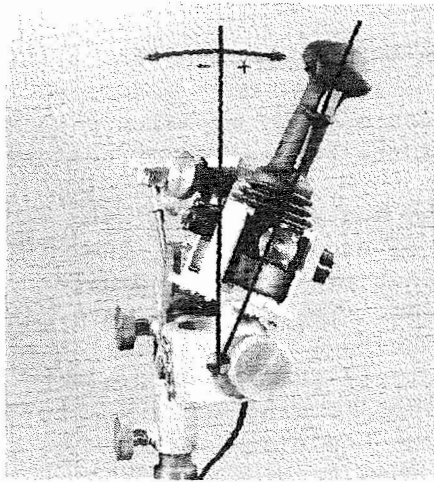
Figure 4. Angular conventions
as viewed form the front.

Figure 5. Angular conventions as
viewed from the right side.
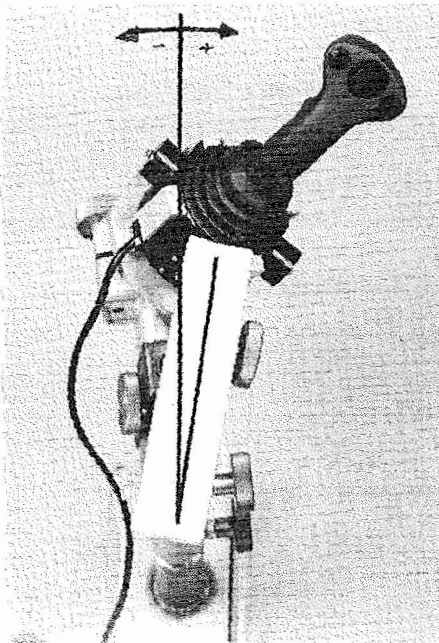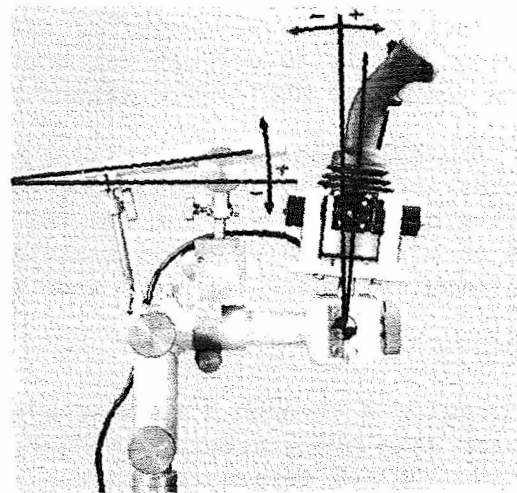
Figure 6. Angular conventions as
viewed from the top.

17.6
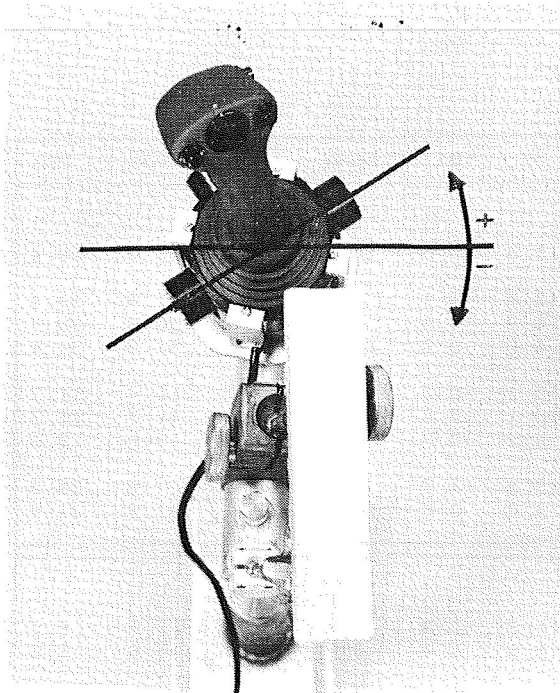
Figure 7. Controller rotation
convention as viewed
from the top.

Figure 8. Hand attack angle
showing a typical
10 deg. offset from
the controller rotation.
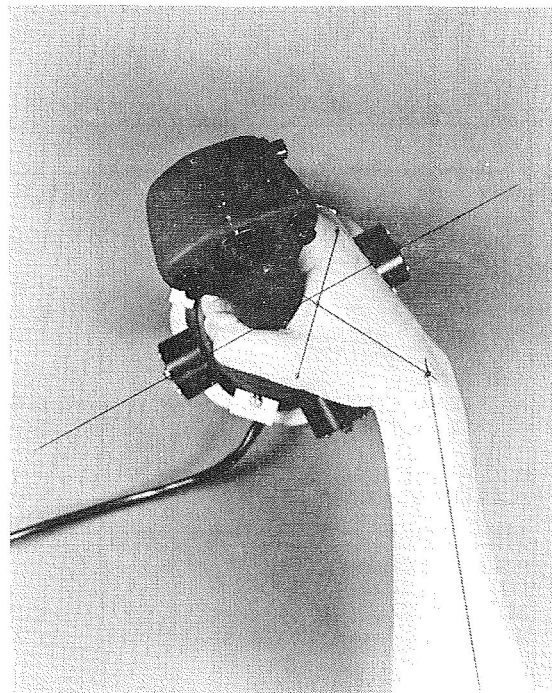
TABLE 2

CONTROLLER ROTATION
(degrees)

	N	MIN	5%	Q1	MEAN	Q3	95%	MAX
NONPILOT PERSONNEL								
————————————————————								
ALL	70	-23.6	-15.4	-4.0	4.4	11.4	30.0	38.4
ALL MALE	52	-23.6	-15.8	-3.1	5.8	14.4	31.9	38.4
ALL MALE RIGHT-HANDED	46	-23.6	-15.9	-4.0	4.9	13.9	30.8	38.4
ALL MALE LEFT-HANDED	6	0.8	0.8	0.8	12.8	24.5	31.9	31.9
ALL FEMALE	18	-15.1	-15.1	-5.2	0.1	7.7	11.7	11.7
ALL FEMALE RIGHT-HANDED	16	-15.1	-15.1	-5.6	-0.1	7.2	11.7	11.7
ALL FEMALE LEFT-HANDED	2	-4.7	-4.7	-4.7	2.2	9.0	9.0	9.0
ALL RIGHT-HANDED	62	-23.6	-15.6	-4.3	3.6	11.4	27.8	38.4
ALL LEFT-HANDED	8	-4.7	-4.7	0.8	10.1	19.4	31.9	31.9
PILOT PERSONNEL								
——————————————————								
ALL NOT WEARING CB GEAR	7	-15.8	-15.8	-2.3	0.2	6.7	7.7	7.7
ALL WHILE WEARING CB GEAR	7	-6.6	-6.6	-6.4	0.8	6.7	6.8	6.8

When viewed from the top a counterclockwise rotation is positive.

TABLE 3

CONTROLLER ANGLE FORE/AFT
(degrees)

	N	MIN	5%	Q1	MEAN	Q3	95%	MAX
NONPILOT PERSONNEL								
————————————————————								
ALL	70	-11.9	-3.3	3.6	8.6	13.8	24.2	30.0
ALL MALE	52	-3.2	-2.3	4.1	10.0	16.1	16.1	30.0
ALL MALE RIGHT-HANDED	46	-2.6	-1.7	4.1	10.8	17.9	27.8	30.0
ALL MALE LEFT-HANDED	6	-3.2	-3.2	-0.7	4.0	7.6	8.6	8.6
ALL FEMALE	18	-11.9	-11.9	1.7	4.5	9.3	16.7	16.7
ALL FEMALE RIGHT-HANDED	16	-11.9	-11.9	1.3	4.6	9.5	16.7	16.7
ALL FEMALE LEFT-HANDED	2	2.1	2.1	2.1	4.1	6.2	6.2	6.2
ALL RIGHT-HANDED	62	-11.9	-3.3	3.8	9.2	15.0	25.2	30.0
ALL LEFT-HANDED	8	-3.2	-3.2	-3.2	4.0	7.6	8.6	8.6
PILOT PERSONNEL								
——————————————————								
ALL NOT WEARING CB GEAR	7	-3.3	-3.3	-0.6	8.1	19.5	22.7	22.7
ALL WHILE WEARING CB GEAR	7	2.6	2.6	3.6	11.1	20.1	21.4	21.4

When viewed from the right side, a clock-wise rotation is positive.

17.8

TABLE 4

CONTROLLER ANGLE LEFT/RIGHT
(degrees)

	N	MIN	5%	Q1	MEAN	Q3	95%	MAX
NONPILOT PERSONNEL								
———————————————————								
ALL	70	-1.9	2.9	9.9	16.6	24.0	35.6	39.6
ALL MALE	52	-1.9	2.9	9.8	15.7	20.6	38.3	39.6
ALL MALE RIGHT-HANDED	46	-1.9	2.7	7.8	15.2	19.7	35.4	39.6
ALL MALE LEFT-HANDED	6	9.7	9.7	10.0	19.5	28.3	38.8	38.8
ALL FEMALE	18	0.1	0.1	13.3	19.2	26.4	33.5	33.5
ALL FEMALE RIGHT-HANDED	16	0.1	0.1	14.8	19.2	25.5	30.2	30.2
ALL FEMALE LEFT-HANDED	2	5.7	5.7	5.7	19.6	33.5	33.5	33.5
ALL RIGHT-HANDED	62	-1.9	2.5	9.6	16.2	23.7	30.3	39.6
ALL LEFT-HANDED	8	5.7	5.7	9.8	19.5	31.3	38.8	38.8
PILOT PERSONNEL								
———————————————————								
ALL NOT WEARING CB GEAR	7	-7.2	-7.2	2.3	6.2	0.7	25.7	25.7
ALL WHILE WEARING CB GEAR	7	-8.2	-8.2	-8.0	4.2	12.3	18.3	18.3

When viewed from the front, a clockwise rotation is positive.

TABLE 5

CONTROLLER POSITION FORWARD OF SRP
(centimeters)

	N	MIN	5%	Q1	MEAN	Q3	95%	MAX
NONPILOT PERSONNEL								
———————————————————								
ALL	70	33.2	35.4	39.8	42.8	46.1	51.3	55.8
ALL MALE	52	33.2	34.1	39.6	43.3	46.7	52.3	55.8
ALL MALE RIGHT-HANDED	46	33.2	33.8	40.3	43.8	46.9	52.6	55.8
ALL MALE LEFT-HANDED	6	37.2	37.2	37.8	40.2	42.1	42.3	42.3
ALL FEMALE	18	36.1	36.1	39.5	41.3	42.9	47.9	47.9
ALL FEMALE RIGHT-HANDED	16	36.1	36.1	39.9	41.5	43.1	47.9	47.9
ALL FEMALE LEFT-HANDED	2	38.4	38.4	38.4	40.4	42.3	42.3	42.3
ALL RIGHT-HANDED	62	33.2	34.7	40.1	43.2	46.5	51.8	55.8
ALL LEFT-HANDED	8	37.2	37.2	38.1	40.2	42.2	42.3	42.3
PILOT PERSONNEL								
———————————————————								
ALL NOT WEARING CB GEAR	7	35.1	35.1	35.5	40.1	44.8	48.1	48.1
ALL WHILE WEARING CB GEAR	7	39.4	39.4	39.8	42.6	46.4	48.1	48.1

When viewed from the right side, a position to the right of the SRP is
positive.

TABLE 6

CONTROLLER POSITION ABOVE THE SRP
(centimeters)

	N	MIN	5%	Q1	MEAN	Q3	95%	MAX
NONPILOT PERSONNEL								
ALL	70	20.1	26.0	30.8	32.6	35.0	37.4	38.1
ALL MALE	52	20.1	24.9	30.3	32.1	34.5	37.3	38.1
ALL MALE RIGHT-HANDED	46	20.1	24.2	30.1	31.9	34.0	37.4	38.1
ALL MALE LEFT-HANDED	6	30.6	30.6	30.8	33.3	35.0	35.6	35.6
ALL FEMALE	18	29.8	29.8	32.2	33.9	35.8	38.1	38.1
ALL FEMALE RIGHT-HANDED	16	29.8	29.8	32.4	34.0	36.0	38.1	38.1
ALL FEMALE LEFT-HANDED	2	30.5	30.5	30.5	32.9	35.4	35.4	35.4
ALL RIGHT-HANDED	62	20.1	25.6	30.8	32.5	35.0	37.5	38.1
ALL LEFT-HANDED	8	30.5	30.5	30.7	33.2	35.2	35.6	35.6
PILOT PERSONNEL								
ALL NOT WEARING CB GEAR	7	26.3	26.3	27.8	29.7	32.3	32.3	32.3
ALL WHILE WEARING CB GEAR	7	28.7	28.7	29.3	31.0	33.0	34.1	34.1

When viewed from the right side, a position above the SRP is positive.

TABLE 7

ARMREST ANGLE UPWARD
(degrees)

	N	MIN	5%	Q1	MEAN	Q3	95%	MAX
NONPILOT PERSONNEL								
ALL	70	-3.7	0.4	3.9	7.5	11.2	15.8	16.5
ALL MALE	52	-3.7	0.9	3.9	7.6	11.5	15.7	16.5
ALL MALE RIGHT-HANDED	46	0.5	1.3	4.2	7.7	11.6	15.9	16.5
ALL MALE LEFT-HANDED	6	-3.7	-3.7	1.9	6.3	10.5	11.7	11.7
ALL FEMALE	18	0.1	0.1	3.9	7.5	10.2	16.3	16.3
ALL FEMALE RIGHT-HANDED	16	0.2	0.2	4.3	8.1	10.8	16.3	16.3
ALL FEMALE LEFT-HANDED	2	0.1	0.1	0.1	2.4	4.8	4.8	4.8
ALL RIGHT-HANDED	62	0.2	1.2	4.3	7.8	11.5	16.0	16.5
ALL LEFT-HANDED	8	-3.7	-3.7	1.0	5.3	9.7	11.7	11.7
PILOT PERSONNEL								
ALL NOT WEARING CB GEAR	7	1.4	1.4	3.5	6.5	12.1	12.7	12.7
ALL WHILE WEARING CB GEAR	7	-1.2	-1.2	-0.2	3.4	7.8	8.6	8.6

When viewed from the right side, a counter-clockwise rotation is
positive.

17.10

TABLE 8

ARMREST ANGLE OUTBOARD
(degrees)

	N	MIN	5%	Q1	MEAN	Q3	95%	MAX
NONPILOT PERSONNEL								
————————————								
ALL	70	-17.3	-8.2	-2.5	-1.8	5.4	14.4	18.5
ALL MALE	52	-13.3	-7.9	-1.6	2.7	6.7	15.6	18.5
ALL MALE RIGHT-HANDED	46	-13.1	-8.7	-1.3	3.0	6.9	16.4	18.5
ALL MALE LEFT-HANDED	6	-4.5	-4.5	-3.9	0.6	5.0	7.1	7.1
ALL FEMALE	18	-17.3	-17.3	-4.4	-0.8	3.4	9.4	9.4
ALL FEMALE RIGHT-HANDED	16	-17.3	-17.3	-4.2	-0.4	3.9	9.4	9.4
ALL FEMALE LEFT-HANDED	2	-4.4	-4.4	-4.4	-3.6	-2.7	-2.7	-2.7
ALL RIGHT-HANDED	62	-17.3	-9.2	-2.2	2.1	5.7	14.6	18.5
ALL LEFT-HANDED	8	-4.5	-4.5	-4.2	-0.4	3.7	7.1	7.1
PILOT PERSONNEL								
————————————								
ALL NOT WEARING CB GEAR	7	-5.1	-5.1	-0.4	0.7	2.6	4.5	4.5
ALL WHILE WEARING CB GEAR	7	-8.6	-8.6	-3.3	1.3	6.7	14.9	14.9

When viewed from the top a clockwise rotation is positive.

TABLE 9

HAND ATTACK ANGLE
(degrees)

	N	MIN	5%	Q1	MEAN	Q3	95%	MAX
NONPILOT PERSONNEL								
————————————								
ALL	70	-10.5	-5.7	7.6	14.7	22.6	30.5	37.1
ALL MALE	52	-10.5	-7.7	6.5	14.5	23.9	30.8	37.1
ALL MALE RIGHT-HANDED	46	-10.5	-7.9	6.0	14.4	23.7	31.1	37.1
ALL MALE LEFT-HANDED	6	5.9	5.9	7.3	15.0	25.1	28.3	28.3
ALL FEMALE	18	2.2	2.2	12.0	15.1	19.7	27.2	27.2
ALL FEMALE RIGHT-HANDED	16	2.2	2.2	11.2	14.9	19.1	27.2	27.2
ALL FEMALE LEFT-HANDED	2	14.1	14.1	14.1	17.5	20.9	20.9	20.9
ALL RIGHT-HANDED	62	-10.5	-7.0	7.2	14.5	22.6	30.5	37.1
ALL LEFT-HANDED	8	5.9	5.9	8.4	15.7	23.3	28.3	28.3

When viewed from the top a counterclockwise rotation is positive.

17.11

DISCUSSION

In general, there are noticeable differences between the means of left-versus right-handed and male versus female personnel.

Controller Rotation (Table 2 and Figure 7)

The rotation data were obtained from the camera located over the subject's head. Cosine corrections were applied to adjust for both the forward and inward cant angles of the controller.

The range of adjustment required by pilot personnel with and without MOPP was within the range required by nonpilot personnel. An adjustment from about 16 degrees clockwise to 32 degrees counterclockwise rotation satisfied 90 percent of the males and females in our sample. Within this range, pilots tended to select a comfort position which was more orthogonal to the airframe axes because they were perhaps influenced by the current grip design and the need to operate switches on the control head itself. The difference between the mean rotational angle selected by males and females was significantly different at the 0.05 level with a t of 1.73 and a df of 68. The difference between left- and right-handed nonpilot personnel was not significant.

The most comfortable position for the hand when grasping the controller was to position the hand with 10 degrees more rotation than the rotation of the grip itself. This is depicted in Figure 8. The difference was greater for left-handed personnel than right-handed personnel while female personnel selected 15 degrees as the most comfortable position.

Fore/Aft Controller Angle (Table 3 and Figure 5)

The range selected by nonpilot males was not sufficient to include the range selected by pilot personnel. No physical differences were noted during data collection other than the flight clothing worn by the aviators. Therefore, the required range should be from 12 degrees rearward cant to 28 degrees forward cant. Within this range, there was a shift in means of almost 7 degrees between left- and right-handed male personnel. The effect of wearing MOPP narrowed the range of comfort selected by personnel without MOPP rather than to significantly shift it. The difference between male right and male left means was statistically significant at the 0.05 level with a t of 1.90 and a df of 50. The difference between the male and female means was also significant at the 0.05 level with a t of 2.43 and a df of 68.

Left/Right Controller Angle (Table 4 and Figure 4)

The range selected by pilots was from 7 degrees outboard to 26 degrees inboard. The range selected by nonpilots was from 0 degrees outboard to 39 degrees inboard. The mean position of 5.2 degrees selected by pilot personnel was not significantly different than the mean position of 15.7 degrees selected by nonpilot male personnel. The 9.5-degree shift toward a more upright position was tested at the 5-percent level using a two-tailed test.

Controller Position (Tables 5 and 6)

The controller position was based on a selected point centrally located within the grip. When personnel grasped the controller, this point remained relatively stable when compared to the angle of the controller within the hand. The range of adjustment was from 38 to 53 centimeters forward and 31 to 38 centimeters above the seat reference point as selected by nonpilot male personnel. The position selected by pilots was between 39 to 48 centimeters forward and 29 to 34 centimeters above the seat reference point.

Armrest Angle (Tables 7 and 8 and Figure 5)

Both the upward and outboard armrest angles selected as being comfortable tended not to follow the upward and outboard angles of the subject's forearm. Personnel seemed to want the armrest adjusted so that the muscular portion of the forearm was the only area in contact with the armrest. The perception of comfort seemed to be influenced by the need to have some flexibility in upper body movement which was observed as subjects shifted their upper torsos and shoulders while selecting a comfortable position. Normally, if one rests one's forearm along the arm of a chair when seated and attempts to shift the body, the arm of the chair restricts the motion of the body. Even though a fully supported forearm is better for control input, it is not always the most comfortable.

CONCLUSIONS

The data suggest that the classical approach of providing a side-arm controller which is orthogonal to the axes of the helicopter is not the most comfortable position. The controller must be significantly angled forward and inboard with a counterclockwise rotation. We realize that controller design has an impact on how a pilot selects a position of comfort and should be looked into with more detail. Of equal importance is that, even though the controller is comfortable to hold, the position may not allow the pilot to control the helicopter without noticeable cross coupling. The concern is, for example, if a control input to pitch forward were made by initiating a motion along the axis of the helicopter, a roll to the left would also occur.

An orthogonal position of the controller to the axes of the helicopter was within the range of comfort selected by the subjects.

MOPP gear did not expand or shift the comfort range selected by the subjects.

Studies are being planned to investigate the effects of controller attitude on simulator flight performance. In addition, the effects of operating switches on the control head will be examined with reference to flight performance.

BIBLIOGRAPHY

Aeromechanics Laboratory (1982). An experimental investigation of the interference between a right-hand sidearm controller tracking task (with a

four-axis isometric controller and with a three-axis displacement
controller) and a left-hand switch operation task (Contractor Report, NASA
Contract NA2-10980). Moffett Field, CA: NASA Ames Research Center.

Aiken, E. W. (1982). Simulator investigations of various side-stick
controller/stability and control augmentation systems for helicopter terrain
flight (AIAA Guidance and Control Conference Paper). Moffett Field, CA:
NASA Ames Research Center.

Aiken, E. W., Glusman, S. I., Hilbert, K. B., & Landis, K. H. (1984). An
investigation of side-stick controller/stability and control augmentation
system requirements for helicopter terrain flight under reduced visibility
conditions (AIAA-84-0235). Moffett Field, CA: NASA Ames Research Center.

Black, G. T., & Moorhouse, D. J. (1979). Flying qualities design requirements
for sidestick controllers (AFFDL-TR-79-3126). Wright-Patterson Air Force
Base, OH: Air Force Flight Dynamics Laboratory, Air Force Wright
Aeronautical Laboratories, and Air Force Systems Command.

Boeing Vertol (1983). Advanced digital/optical control system (ADOCS) flight
demonstrator control/display system design (Contractor Report, NASA Contract
NA2-10880). Moffett Field, CA: NASA Ames Research Center.

Clark, A. P., Jr. (1981). F-16 moveable side-stick evaluation (Letter Report,
TAC Project 81A-013F). Nellis Air Force Base, NV: USAFTFWC/CC.

Fry, E. B., Gerdes, R. M., & Schroers, L. (1973). A piloted simulation study
of the effects of controller force gradient in VTOL hovering flight (NASA
TM X-62230). Moffett Field, CA: Ames Research Center and US Army Air
Mobility Research and Development Laboratory.

Active Controllers and The Time Duration To Learn a Task

D. W. Repperger[*],   C. Goodyear[**]

[*] Air Force Aerospace Medical Research Laboratory
Wright Patterson Air Force Base, Ohio 45433
[**] Systems Research Laboratory
2800 Indian Ripple Road
Dayton, Ohio 45418

Abstract

An active controller was used to help train naive subjects involved in a compensatory tracking task. The controller is called active in this context because it moves the subject's hand in a direction to improve tracking. It is of interest here to question whether the active controller helps the subject to learn a task more rapidly than the passive controller.

At The Air Force Aerospace Medical Research Laboratory six subjects, inexperienced to compensatory tracking, were run to asymptote root mean square error tracking levels with an active controller or a passive controller. The time required to learn the task was defined several different ways. The results of the different measures of learning were examined across pools of subjects and across controllers using statistical tests. The comparison between the active controller and the passive controller as to their ability to accelerate the learning process as well as reduce levels of asymptotic tracking error is reported here.

Introduction

With the advent of microprocessor computer technology, one would like to use this new technology to help improve the interaction of humans with machines. One method to achieve this result is to use controllers or displays which exhibit the ability to adapt or change with time. An example of this type of application occurs with quickened displays where visual information is used to improve the man-machine interaction. In this case the display is "quickened" if it provides the operator with immediate knowledge of the effects of his own responses. Thus the human operator is able to more efficiently process information with this type of display.

Another way to use computers to improve man-machine interaction occurs if the hand controller the human interacts with is computer controlled to move the human arm and assist in the tracking. Intuitively this makes sense because it is known that golf or tennis teachers [1] physically force the limbs of a student through the appropriate movements for a specific stimulus. This appears to give rise to the quickest initial learning, however, the retention of this learning may be poor.

In this paper we consider a side stick controller which moves in one dimension laterally. The stick controller actually puts a force on the human subject's arm as a function of a smart stick algorithm and

physically moves the subject's hand. The subject can override this force depending on the commands he wishes to make.

The idea of using adaptive controllers has been considered previously in the manual control area. For example, in 1968, Herzog [2] investigated a manipulator that had mechanical characteristics matching the plant's characteristics in such a way that the control task of the operator is reduced to the problem of positioning the control stick. This was shown [2] to significantly improve tracking performance.

One must, however, separate the effects of practice from the effect of the subject interacting with the smart stick. In reaction time experiments one school of thought [5] views performance changing at all levels of practice. In fact in reference [5] the authors refer to a study in which performance of a simple manual operation involving a decision by operators in an industrial plant was found to be still improving after a million repetitions. Clearly, such investigations are beyond all pragmatic efforts within a laboratory.

The objective in this paper is to use the active (force producing) controller to observe the effect of this controller to help train subjects rapidly. It is desired to see if the use of an active controller may either reduce the time required to learn a task or possibly to help learning in some other manner.

The Experimental Apparatus

Figure (1) illustrates a block diagram description [3] of how the "smart stick" or active controller is presumed to work. The human body is modelled as a mass-spring-dashpot system. Within the dotted box is the "smart stick" controller which, for this paper, consists of a variable mass, spring, and dashpot, or possibly a programmed biomechanical force. The computer algorithm may possibly produce a programmed biomechanical force which will move the stick in a lateral direction to interact with the hand movements of the subject.

Figure (2) illustrates the mechanical components of this stick. A rack and pinion assembly is coupled to a gear and transmits force to the stick. A piston of area A within an airtight cylinder is moved to the right and left as a function of the pressure on each side of the piston. The pressures P_1 and P_2 are controlled by the two current-pressure transducers which regulate P_1 and P_2 via electrical currents I_1 and I_2. The algorithm from the computer determines the currents I_1 and I_2 which produces the desired force on the stick. Figure (3) illustrates the actual device.

Experimental Design

It is desired in this study to examine how this device may help or hinder the ability to learn a tracking task. Six young, healthy, male active duty Air Force personnel participated in this experiment. They were required to be "naive" trackers which, in this experiment, meant they had not previously participated in a tracking experiment at our laboratory involving compensatory tracking. All runs were conducted in a static (1Gz) environment on four days of a normal work week. Three of the subjects were the control group. The other three subjects were the experimental group. Each day a subject tracked nine trials of 85

seconds duration each with a 120 second rest between each trial. This required approximately 31 minutes daily of the subject's time. At the end of each trial the subject was given a display of his score on the screen of the CRT. The score number displayed was proportional to the root mean square tracking error level during the run. This score was illustrated to provide feedback to the subject on his performance level.

The three subjects in the control group tracked the nine trials each day for 4 days using a passive stick. The passive stick is defined as a simple displacement stick [4] with a relatively low spring constant. The remaining three subjects in the experimental group had the first two days of tracking with the passive stick, similar to the control group. On the third day, however, the experimental group tracked with the smart stick. On the fourth day the experimental group tracked again with the passive stick. It was initially hoped that a comparison of performance on the last day between the two groups may easily demonstrate the difference between the two training schemes. If, like the example from golf or tennis, the smart stick can demonstrate to the subject an improved method of tracking, then on the fourth day the subjects in the experimental group will presumeably track better with the passive stick.

Results

Figure (4) illustrates data from subject 3-PA (the third subject in the experimental group who tracked with both the passive and active stick). It is observed from this plot that the RMS error scores were lower on the third day (the active stick day) as compared to the previous two days involving the passive stick. On day 4, the subject now seems to perform slightly better with the passive stick as compared to days 1 and 2. It is necessary, however, to take out the effect of learning that would normally occur in the absence of an exposure to the smart stick.

Figure (5) illustrates the data from subject 1P (the first subject in the control group). The scores seem to asymptote on the second day with little change thereafter. These results were particular to these individuals but across subjects there existed other types of variation. Figure (6) illustrates data from a pilot (flight instructor). His reaction to the smart stick was of great interest because he was an experienced pilot as well as a flight instructor. On his first exposure to the smart stick he tried different strategies and by the eight trial he had settled down to his best performance level. On the fourth day he did show a small improvement in his error scores. It is necessary, however, to average these effects across subjects to see what can be said in a statistical sense.

Table I illustrates the RMS scores for each day and subject. The entries in the table are the minimum e_{RMS} score each day, the mean and standard deviation e_{RMS} score each day, and the coefficient of variation (ratio of s.d./mean). It is important to consider that learning data are exponential in nature [6] and the mean and standard deviation across all the trials that particular day does not have a great deal of meaning. It provides, at best, a crude estimate of

performance that particular day.

Table I- min e_{RMS}, mean, s.d., and C.V.(Coefficient of Variation)

Subject	Day	Minimum	Mean	Standard Deviation	Coefficient of Variation
1-P	1	11.4	17.1	10.3	.60
	2	9.3	10.2	1.6	.16
	3	9.0	9.9	0.8	.08
	4	9.1	9.9	0.6	.06
2-P	1	13.9	19.9	9.9	.50
	2	11.0	13.1	1.8	.13
	3	10.5	16.7	15.3	.92
	4	10.8	11.7	0.6	.05
3-P	1	9.0	11.7	5.9	.50
	2	8.6	9.3	0.6	.06
	3	9.3	11.7	5.1	.43
	4	9.8	10.7	0.5	.04
1-PA	1	7.4	10.2	5.8	.57
	2	7.7	8.6	0.9	.10
	3	4.9	10.0	4.8	.48
	4	6.9	7.7	0.5	.07
2-PA	1	10.7	19.4	14.4	.74
	2	11.2	13.2	2.0	.15
	3	5.6	6.9	0.8	.12
	4	11.6	12.3	0.4	.03
3-PA	1	10.1	12.8	4.2	.33
	2	8.8	11.7	3.4	.29
	3	6.6	11.1	9.6	.87
	4	8.1	9.0	1.2	.13

The coefficient of variation appears to be related to learning because one would expect (as a definition of learning) little variation from trial to trial (small values of s.d./mean). In a laboratory setting, we normally accept data as being consistent if the CV is .2 or less. This appears to occur on the second day for both the passive and active stick data.

To analyze these data, the minimum error RMS was determined for each subject on day 2 and day 4, and the percent change from day 2 to day 4 was calculated. These percent changes were used in a 2-sample T-test which found no significant difference between the PA group (mean=-5.0, s.d.=7.5) and the P group (mean=3.5, s.d.=9.4), T(4)=-1.2, p=.2876. Thus, using the active stick on day 3 did not result in significantly lowering the minimum error RMS scores for day 4 as compared with the P group. The following table contains the minimum error RMS scores used in the analysis:

Table II

Subject	E_{RMS} Min Day 2	E_{RMS} Min Day 4	% Change Day 2 to Day 4
1-P	9.3	9.1	-2.4
2-P	11.0	10.8	-1.5
3-P	8.6	9.8	14.3
1-PA	7.7	6.9	-10.2
2-PA	11.2	11.6	3.6
3-PA	8.8	8.1	-8.3

The coefficient of variation for error RMS was determined for each subject on day 2 and day 4, and the percent change from day 2 to day 4 was calculated. These percent changes were used in a 2-sample T-test which found no significant difference between the PA group (mean=-56, s.d.=21) and the P group (mean=-51,s.d.=19), T(4)=-0.3, p=.75238. Thus using the active stick on day 3 did not result in significantly lowering the variability of the error RMS scores for day 4 as compared with the P group. Table III contains the coefficients of variation obtained from these data.

Table III - Coefficient of Variation * 100

Subject	CV Min Day 2	CV Min Day 4	% Change Day 2 to Day 4
1-P	15.6	5.7	-63.5
2-P	13.4	5.4	-59.7
3-P	6.2	4.4	-29.0
1-PA	10.4	6.7	-35.2
2-PA	15.0	3.3	-78.1
3-PA	29.4	13.0	-55.7

The minimum error RMS was determined for each subject on day 2 and day 3, and the percent change from day 2 to day 3 then calculated. These percent changes were used in a 2 sample T-test which found a significant difference between the PA group (mean=-37.1, s.d.=12.5) and the P group (mean=0.3, s.d=7.2), T(4)=-4.5, p=.0109. Thus, there was a greater decrease in the minimum error RMS from day 2 to day 3 for the PA group than for the P group. The following table contains the minimum error RMS scores used in the analysis:

Table IV - Minimum Error Scores (RMS Values)

Sub.	Error RMS Min Day 2	Error RMS Min Day 3	% Change Day 2 to Day 3
1-P	9.3	9.0	-3.0
2-P	11.0	10.5	-4.7
3-P	8.6	9.3	8.5
1-PA	7.7	4.9	-36.0
2-PA	11.2	5.6	-50.2
3-PA	8.8	6.6	-25.2

Discussion

It was initially hoped that a comparison of performance results on the fourth day between the control group and the experimental group would demonstrate the advantage of the use of the smart stick to reduce the time to learn a task. Three questions were answered from this study. First, the question of whether the experimental group performed better on the fourth day as compared to the control group? It was demonstrated that the exposure to the smart stick did not produce any additional improvement in the passive stick scores from day 2 to day 4.

The second question of whether overall variability decreased was answered by studing the coefficient of variation. One could use as a definition of learning a measure of consistent and repeatable score levels. Perhaps the exposure to the smart stick would make the scores on day 4 more consistent which could be detected by a smaller value of the coefficient of variation. The results of the analysis of Table III indicated that subjects were no more consistent on day 4 following the

smart stick as the control group had following the passive stick on day 3.

The third question as to whether the smart stick actually improved tracking performance was obtained from analysis of Table IV. A significant difference was found across subjects and controllers in comparing Day 2 to Day 3 between the control group and the experimental group. The percent change reduction in e_{RMS} due to the smart stick exceeded 50% of the passive stick value for one subject.

Conclusions

An active controller was used to train naive subjects in a compensatory tracking task. The subjects apparently did not improve their passive stick scores after being exposed to the active stick anymore than a subject that had just tracked with the passive stick. The amount of variability across replications did not decrease after exposure to the smart stick. Finally, it was demonstrated that tracking with the active controller will significantly reduce error scores to levels sometimes 50% below the asymptotic levels for a passive stick.

References

[1] Sheridan, T. B., and Ferrell, W. R., "Man-Machine Systems", The MIT Press, 1981.

[2] Herzog, J. H., 1968, "Manual Control Using The Matched Manipulator Control Technique", IEEE Transactions on Man-Machine Systems, MMS-9, No. 3, pp. 56-60.

[3] Jex, H. R., and Magdaleno, R.E., "Biomechanical Models For Vibration Feedthrough To Hands and Head For a Semisupine Pilot", Aviation, Space, and Environmental Medicine, January, 1978, pp. 304-316.

[4] Levison, W. H., "Effects of Control Stick Parameters on Human Controller Response", Report No. 5510, Bolt, Beranek, and Newman, Inc., January, 1984.

[5] Mowbray, G. H. and Rhoades, M. V., 1959, "On The Reduction of Choice Reaction Times With Practice", Quarterly Journal of Experimental Psychology, 11: pp. 16-23.

[6] Gershoni, H., "An Investigation of Behavior Changes of Subjects Learning Manual Tasks", Ergonomics, 1979, vol. 22, No. 1, pp. 1195-1206.

"SMART STICK"

ACCELEROMETER SENSOR

A

COMPUTER WITH ALORITHM TO ADJUST STICK PARAMETERS

Ks

Bs

Ms

PROGRAMMED BIOMECHANICAL FORCE

STICK

G FIELD MEASUREMENT INPUT

BIOMECHANICAL REPRESENTATION OF THE PILOT
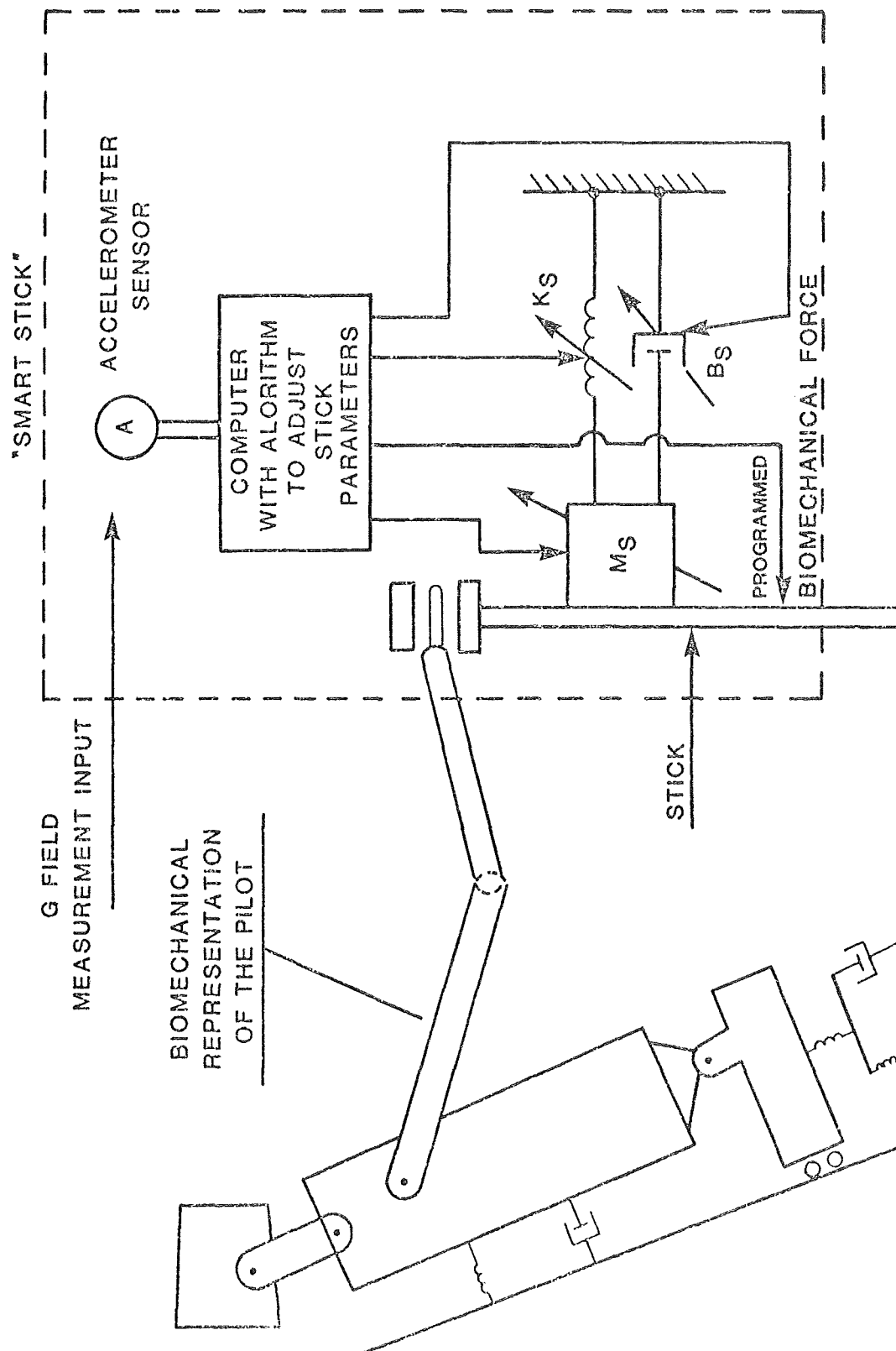
Figure (1) – THE "SMART STICK"
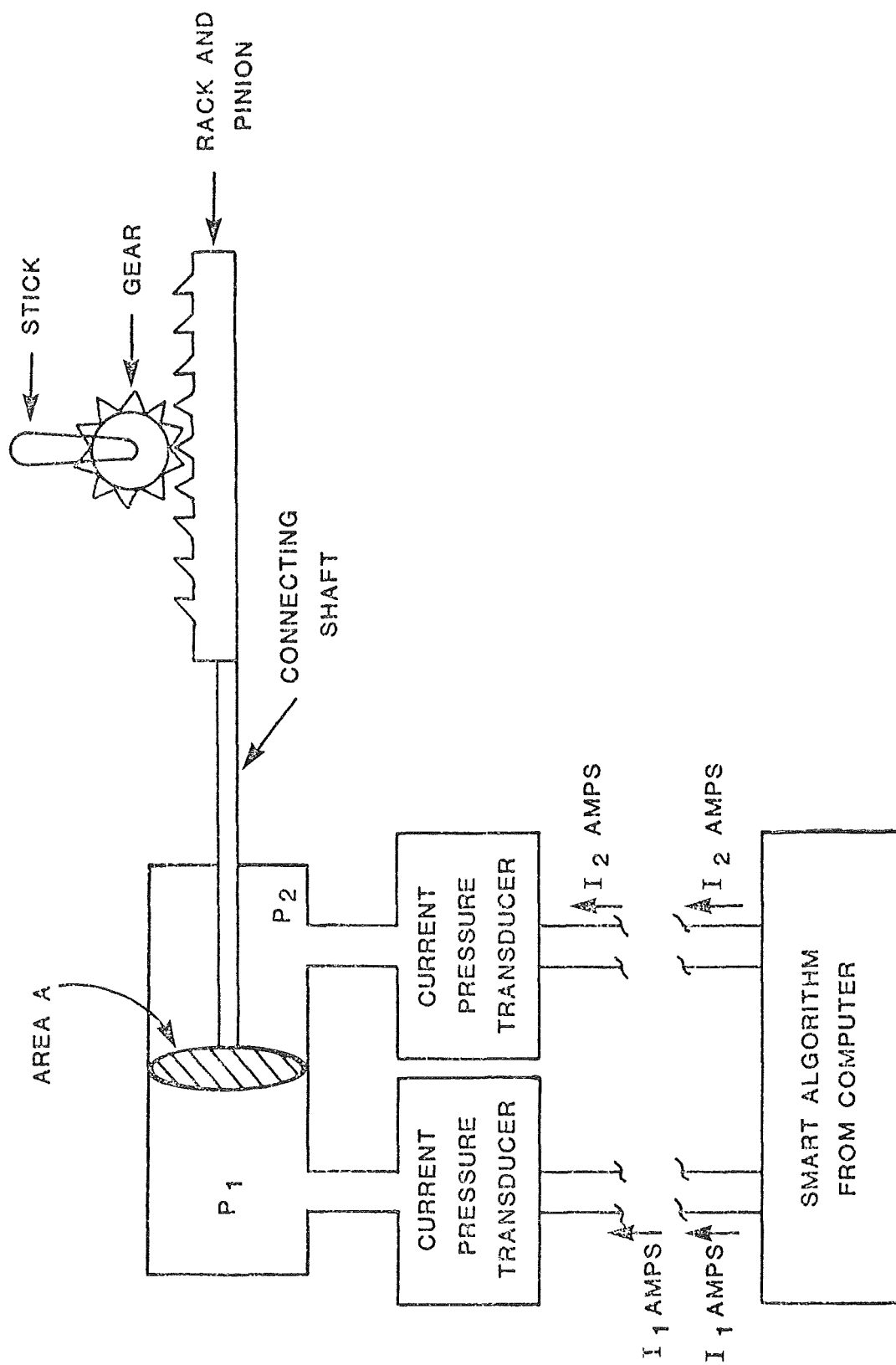
Figure (2) - THE ELECTROMECHANICAL DEVICE
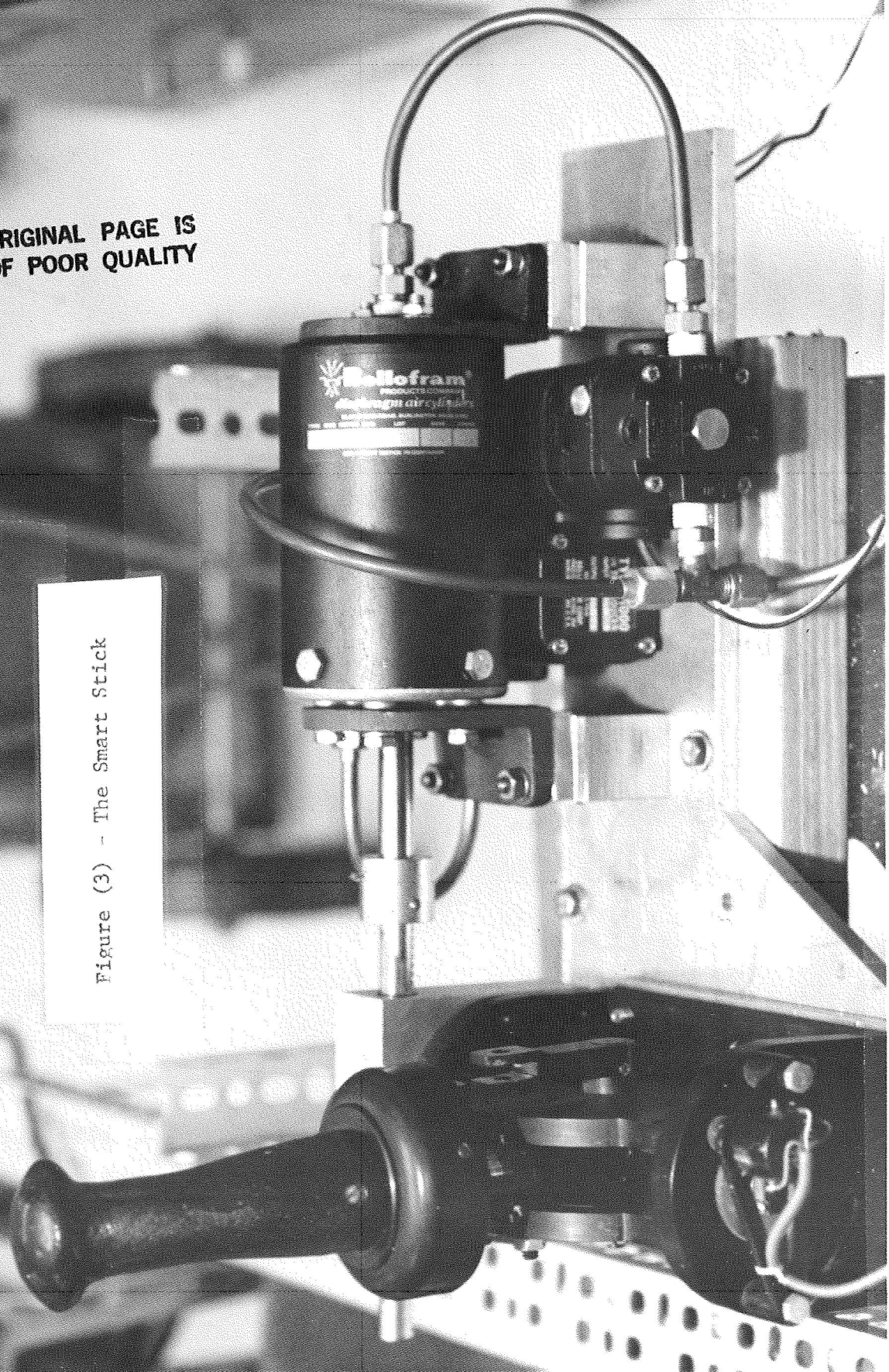
18.8

Figure (3) - The Smart Stick

PLOT OF ERROR RMS VS. DAY AND TRIAL

SUBJECT=3-PA
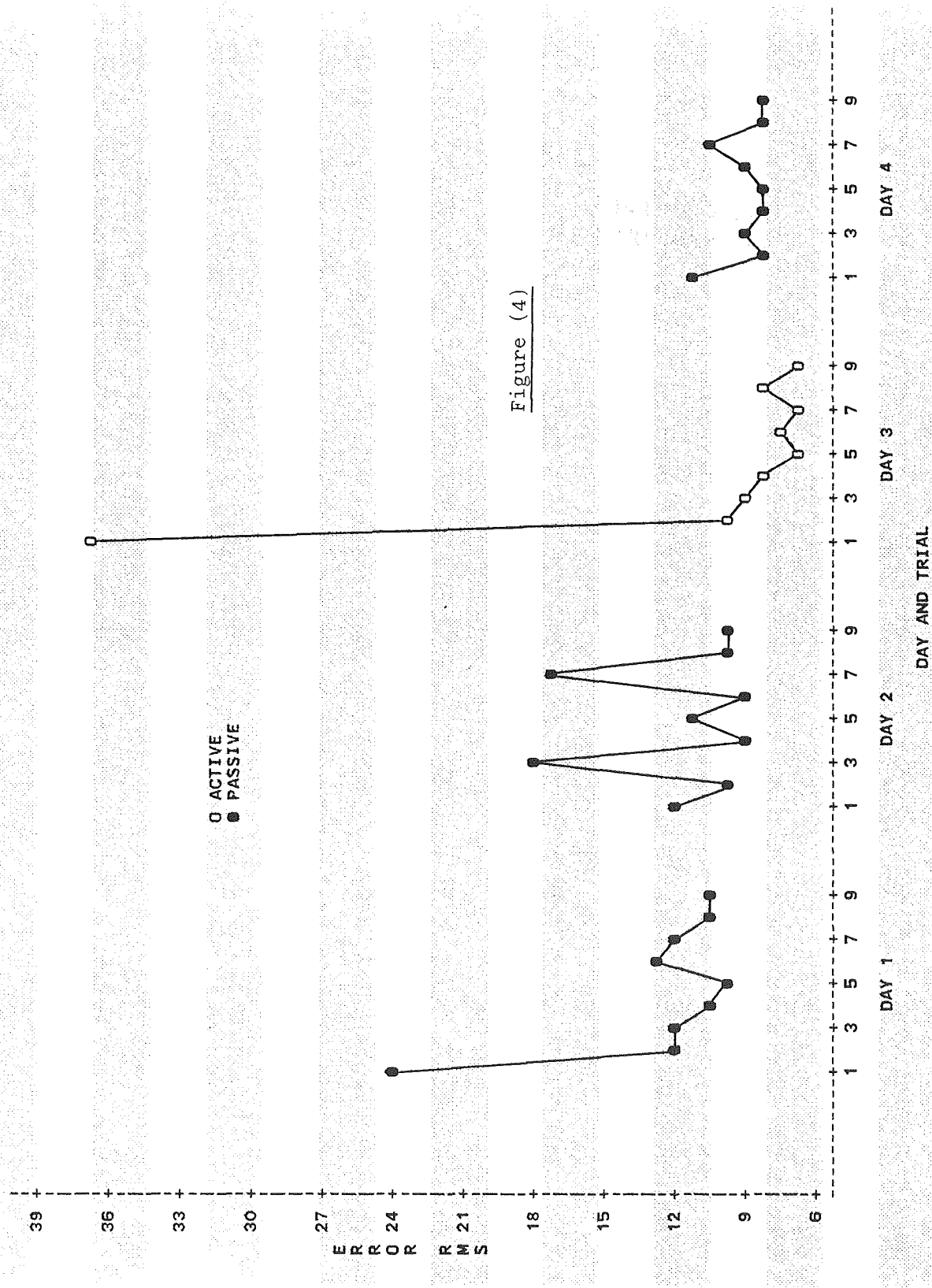
Figure (4)

18.10

PLOT OF ERROR RMS VS. DAY AND TRIAL

SUBJECT=1-P

Figure (5)

PLOT OF ERROR RMS VS. DAY AND TRIAL

SUBJECT=1-PA

O ACTIVE
■ PASSIVE

Figure (6)

ORIGINAL PAGE IS
OF POOR QUALITY

18.12
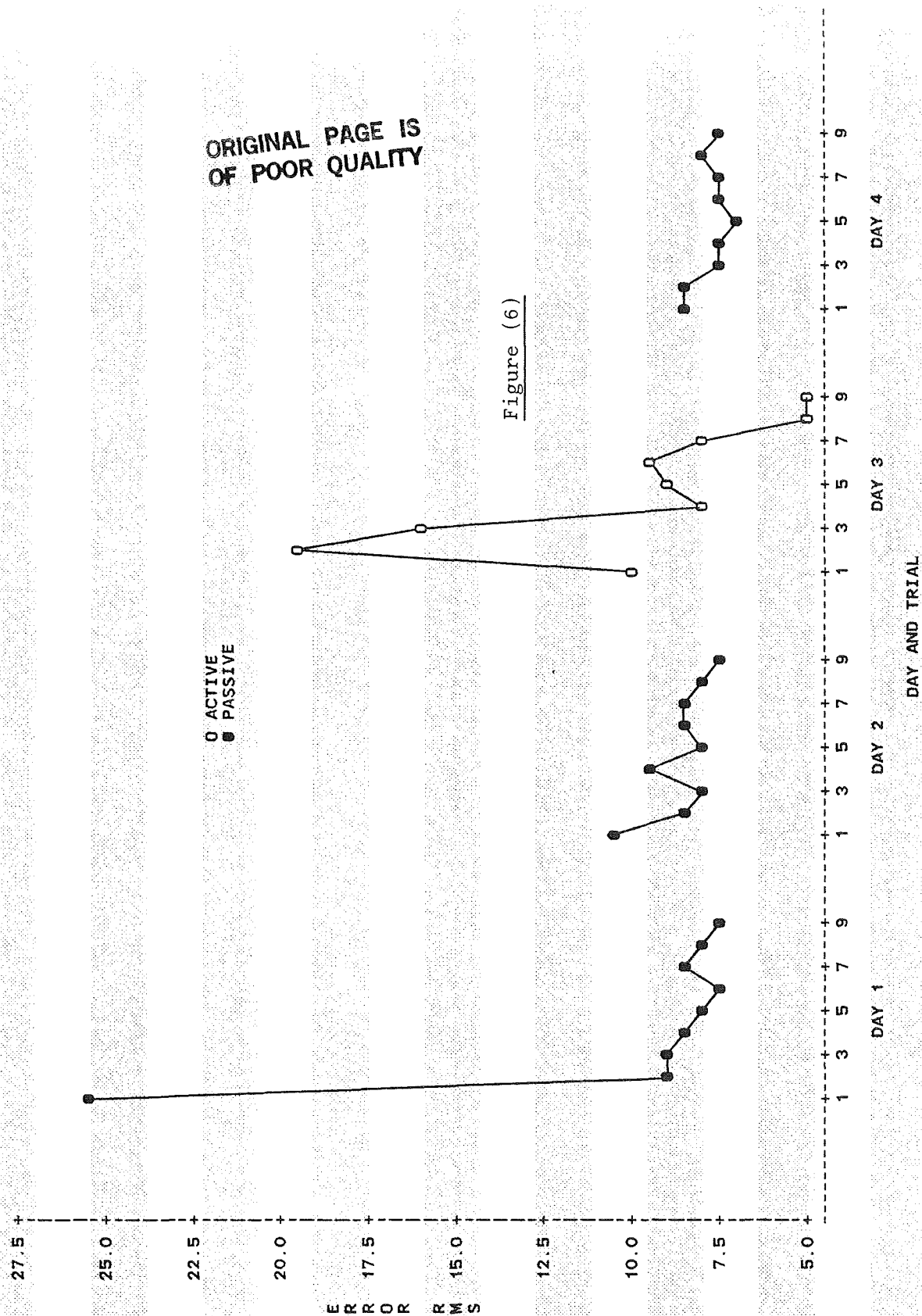
HESITATION IN TRACKING
INDUCED BY A CONCURRENT MANUAL TASK

Patricia A. Kelly and Stuart T. Klapp
California State University, Hayward

When people are required to track with one hand and perform occasional discrete responses with the other hand, there is a strong possibility that errors will be induced in tracking attributable to the simultaneous action by the other hand. We have been investigating this problem by pairing pursuit tracking (right hand) with a handle movement response (left hand) guided by an auditory stimulus. Tracking is assumed to represent flight control and the left hand response to represent other aspects of aircraft system management. The general goal of this research is to identify the types of errors induced into tracking by the requirement of a secondary response with the other hand.

In the previous Annual Manual conference (Klapp, Kelly, Battiste, & Dunbar, 1984) we reported that hesitations frequently occur in this situation. We defined a hesitation as holding the joy stick motionless for at least 1/3 sec. while the cursor was beyond the assigned tolerance. Overall, hesitations occurred on 48% of the instances of tracking sequences when accompanied by left hand secondary response, but only 6.5% of equivalent control instances with no secondary response. However, when tracking was emphasized by instruction and auditory alarm for out of tolerance cursor position, the rate of hesitations was reduced to 29% with left hand secondary response, and 4.5% on the control. This reduction in right hand hesitations was at the expense of increased left hand response simple reaction time (RT).

Now we report an attempt to determine if hesitations can be reduced further by combining tracking emphasis with a higher degree of practice. In addition a different type of joy stick controller was employed to determine whether the occurrence of hesitations generalizes beyond the particular joy stick and muscle groups involved in the earlier report. This new joy stick utilized finger muscles instead of those of the wrist and arm. Under these conditions hesitations occurred on 13.6% of the opportunities when the left hand response was present (but only on 0.78% of the control instances of tracking

19.1

unaccompanied by the left hand response). Although hesitations occurred each day, the frequency of hesitations decreased over days of practice (Table 1), F(3,21) = 3.6, p < .05.

The reduction of hesitations with practice was accompanied by a decrease in the RT of the secondary task, F (3,21) = 10.5, p < .001 (Table 1). Thus, the improvement of right hand tracking with practice cannot be attributed to developing a strategy of emphasis on tracking at the expense of left hand performance. By contrast to this effect of practice, emphasis on tracking improved tracking at the expense of left hand RT (Klapp, et al., 1984).

	Tracking Hesitations		Left Hand RT (msec.)
	Probe	Control	
Day 1	26.6%	0	480
Day 2	13.1%	0	416
Day 5	9.8%	0	380
Day 6	4.7%	3.1%	362
Mean	13.6%	.78%	409

Table 1. Hesitation rate and left hand reaction time.

An additional experiment is in progress which uses a third type of joy stick. Unlike the joy stick used in the experiment just reported, this one was spring loaded to bias movement in one direction. We assumed that subjects might release the stick rather than hesitate, so that the joy stick would move in the direction of the spring bias. Apparently this is not the case, because hesitations occur even with this joy stick. Four subjects have completed this experiment and hesitations occur on 18.6% of the instances in which the left hand must respond (and on 3.1% of the control responses). Apparently our subjects tend to "freeze" their right hand rather than to "let go."

We conclude that there is a tendency to freeze the tracking response when a discrete simultaneous response is required of the other hand. This type of error might be dangerous in flight control. Emphasis on tracking reduces hesitations at the expense of longer RT for the left hand response (Klapp, et al., 1984). By contrast, practice seems to reduce hesitations while also improving left hand

19.2

RT. Thus there appears to be a mode of control which permits tracking and discrete simultaneous responses to occur together. It would be desirable to understand how this is possible and how it might be facilitated.

Reference

Klapp, S. T., Kelly, P. A., Battiste, V., & Dunbar, S. (1984). Types of tracking errors induced by concurrent secondary manual task. In E. J. Hartzell and S. Hart (Eds.) <u>Proceedings of the Twentieth Annual Conference on Manual Control.</u>, Vol. 2, pp. 299-304.

Acknowledgement

Progression-Regression Effects in Tracking Repeated Patterns

Richard J. Jagacinski and Sehchang Hah

Ohio State University
Department of Psychology
Human Performance Center
404-B West 17th Avenue
Columbus, Ohio 43210

Subjects used a position control system to perform compensatory tracking
of a repeated input pattern. The input pattern was 20 seconds in duration
and was either an arctangent function or the sum of two sine waves.
Tracking error decreased with practice and increased with the addition of
a concurrent memory task. The shape of the ensemble-averaged tracking
error resembled the shape of the input velocity signal throughout these
changes in performance. Regression analyses were used to parameterize
these effects and compare these results with the predictions of several
conceptualizations of perceptual-motor learning.

THE ROLE OF IMPULSE PARAMETERS

IN FORCE VARIABILITY

Les. G. Carlton   K. M. Newell
Biomechanics Laboratory   Institute for Child
Freer Hall   Behavior and Development
University of Illinois   University of Illinois
906 So. Goodwin Avenue   51 Gerty Drive
Urbana, IL   61801   Champaign, IL   61820

One of the principle limitations of the human motor system is the
ability to produce consistent motor responses. When asked to
repeatedly make the same movement, performance outcomes are
characterized by a considerable amount of variability. This is
especially true for rapid actions or when salient feedback cues are
not available, requiring the performer or operator to function in an
open-loop manner. This occurs whether variability is expressed in
terms of kinetics or kinematics. Variability in performance is of
considerable importance because for tasks requiring accuracy it is a
critical variable in determining the skill of the performer. In
addition, understanding the factors affecting response variability
will provide important insights necessary for explaining Fitt's Law
(Fitts, 1954) and speed accuracy tradeoffs in general.

What has long been sought is a description of the parameter or
parameters that determine the degree of variability. Two general
experimental protocals have been used. One protocal is to use dynamic
actions and record variability in kinematic parameters such as spatial
or temporal error. A second strategy has been to use isometric
actions and record kinetic variables such as peak force produced.
While a number of hypotheses have been put forward, there are two
models which suggest that force parameters determine the amount of
variability in a variety of tasks.

Most recently, Schmidt, Zelaznik, Hawkins, Frank & Quinn (1979)
presented an impulse variability model which predicts a linear and
proportional relationship between the impulse produced and impulse
variability. As the level of force required to complete a response
increases, the variability in producing that force also increases.
Based upon this relationship, Schmidt et al. demonstrated that speed-
accuracy tradeoffs could be accounted for by variability in force
production. This work provided important advancements for providing
the link between variability at kinetic levels and variability in
kinematic variables consistent with speed-accuracy relationships.

A second model, which we label an impulse-ratio model, is an extrapolation
of the work by Bahrick, Bennett, and Fitts in 1955. They were interested
in the control of a spring loaded control stick and how changes of force
characteristics affected tracking performance. The model proposed that
amplitude, terminal torque and the change of torque from initial to final
torque levels influenced accuracy. Extrapolating to isometric tasks, the
impulse-ratio model would predict that force variability is proportional to
the ratio of the change in force from initial force to peak force, divided
by peak force.

Unfortunately, there has been little empirical support for either of these
models. For example, there is a large body of evidence which supports a
non-proportional relationship between force and force variability in both
isometric (Fullerton & Cattell, 1892; Jenkins, 1947; Newell & Carlton, in
press; Noble & Bahrick, 1956) and for dynamic movements (Newell, Carlton, &
Carlton, 1982). In addition, previous examinations of force variability
have confounded a number of force variables. For example, variations in
isometric peak force have co-varied with changes in impulse and rate of
force production.

The major purpose of this paper is to examine what might be the important
force related factors affecting variability and to provide an experimental
approach to examine the influence of each of these variables. The models
previously presented have implicated peak force, impulse, and change of
force. But when we consider that a motor response requires the generation
of force over time, it is noted that peak force is a function of the rate
of force production and the amount of time that the rate is generated.
Thus, the rate of force production and its time of application may be more
fundamental than consideration of peak force or impulse alone. Each of
these variables are depicted on a typical force-time curve generated in an
isometric force production task (Figure 1).
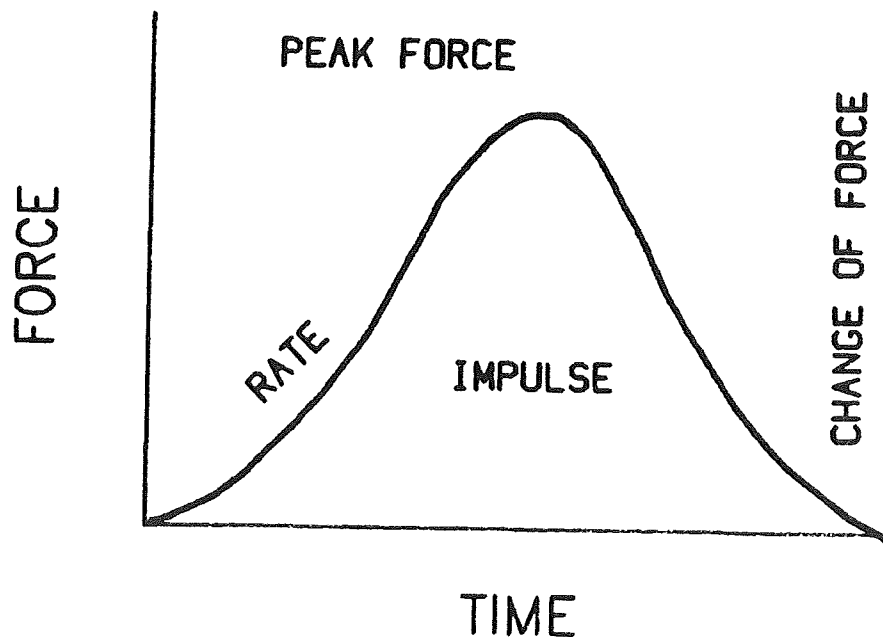
Figure 1. Typical isometric force-time curve.

Research Strategy

We suggest that a reasonable strategy would be to conduct a series of
experiments where each of the force parameters would be held constant while
allowing others to vary systematically. It is anticipated that synthesis
of the experimental findings would lead to an understanding of the
contribution of each impulse parameter to response variability. A priori,
it was reasoned that the impulse variability and impulse-ratio models had
focused on the non-essential variables of force production rather than the
essential variables.

Six experiments examining isometric force production are suggested. In each
study subjects are required to produce multiple discrete trials in order to
evaluate response variability. The subjects are provided a force-time
template which should be matched, and feedback after each trial regarding
the discrepancy between the template and actual response. The first three
experiments (Figure 2) manipulate the initial preload or steady force
exerted before each trial.

The Experiments

Figure 1A represents four conditions which have equal peak force but allow
for changes in the rate of force production as well as impulse size and
change of force. The triangulated force-time curves provide approximations
to the force-time manipulations for each experiment. Thus, as preload
increases the rate of force production and the change of force decreases.

The experiment outlined in Figure 1B keeps the change of force constant
across 4 conditions but allows the impulse size and peak force to vary
systematically. The rate of force production also remains constant. A test
of the impulse-ratio model is provided in Figure 1C. In each of the four
conditions the ratio descibed by the change of force divided by peak force
remains constant. The impulse size, rate of force production, and peak
force varies with conditions.

The second set of experiments (Figure 3) vary the time to peak force in
order to manipulate the desired force parameters. A test of the impulse
variability model is provided in Figure 3A. The size of the impulse
remains constant by increasing the time to peak force and reducing the
peak force attained. As a result, the rate of force production changes for
each condition. As far as we know this is the first strong test of the
impulse variability model. Figure 3B represents conditions with equal peak
force and different rates of force production as well as different impulse.
In Figure 3C the rate of force production is held constant while peak force
and impulse vary.

21.3
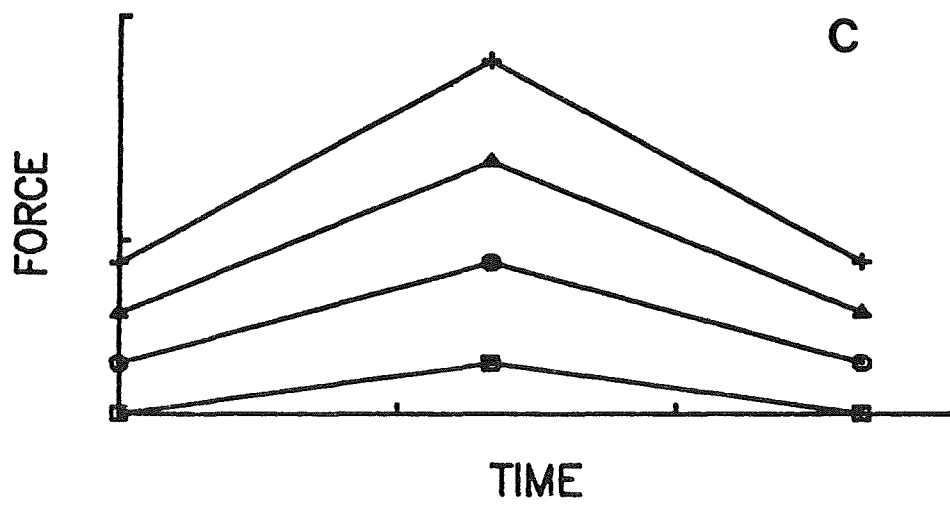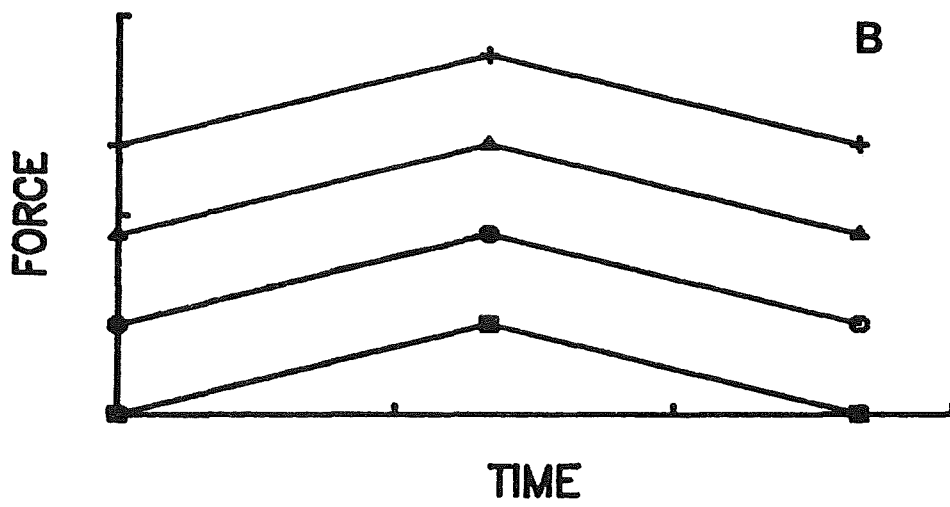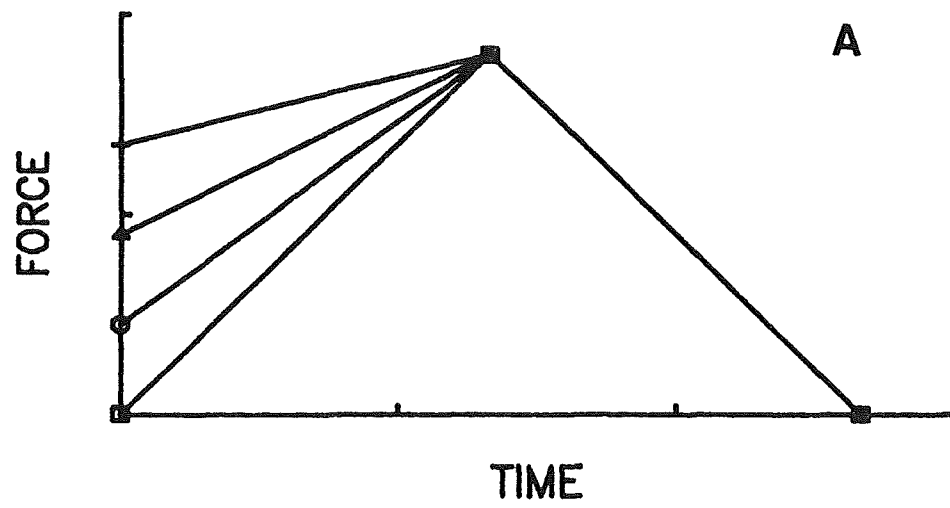
Figure 2. Triangulated force-time curves for experiments 1-3.

21.4
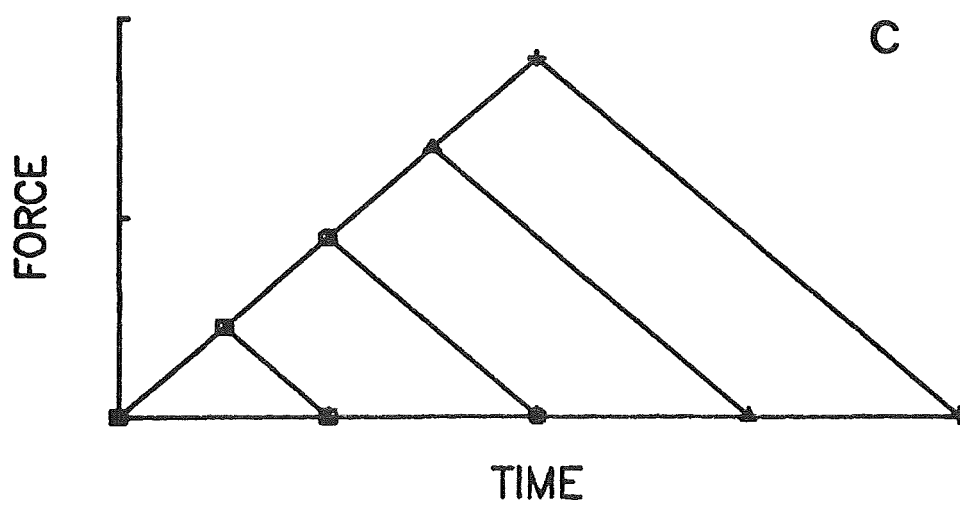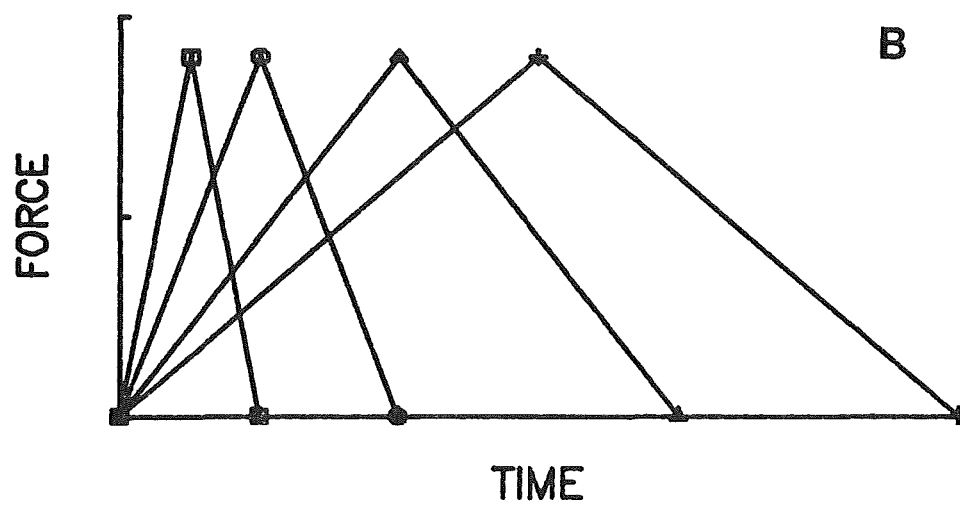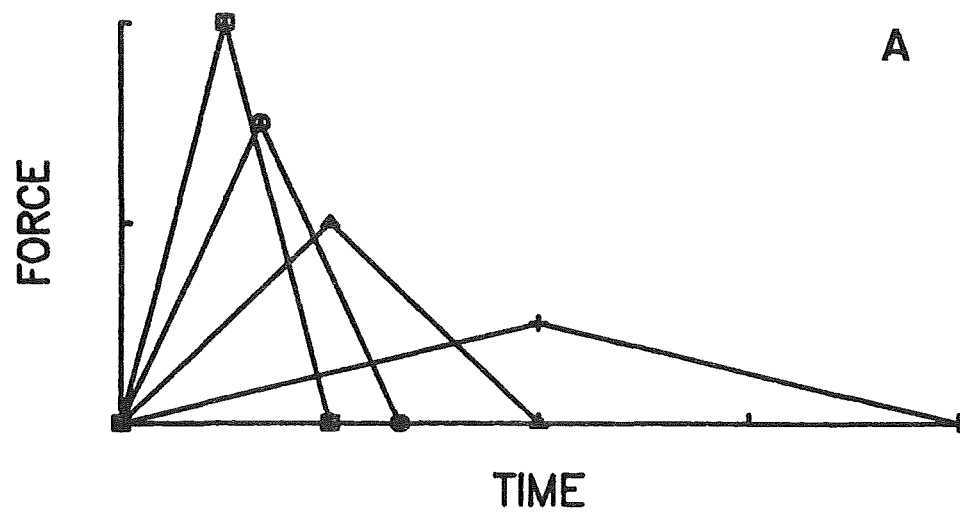
Figure 3. Triangulated force-time curves for experiments 4-6.

DISCUSSION

The pattern of results from the six experiments should provide an indication of the relative importance of each of the force related parameters to force variability. The simplest solution would be provided if variability remained constant as a function of one of the manipulations outlined. For example, if impulse variability remained constant across the four conditions oulined in Figure 3A, evidence would support the contention that impulse size determines variability. Changes in rate of force production and peak force would have no significant effect on variability. Such a finding would provide support for the impulse variability model.

We speculate, based on pilot data and the nature of force production, that no single factor will provide an accurate accounting of the force variability function. However, we believe a physical description is possible when multiple factors are considered. Rate of force production and the time for which that rate is developed would seem to be important features with other factors such as the change of force from initial to final force levels playing some role.

While these experiments have been outlined employing an isometric task, the same manipulations can be produced in dynamic actions. Although these tasks have differing control problems, both require the performer to functionally exert force over time, and hence, generate an impulse (time intergral of force). Newtonian principles of mechanics suggest that kinematic and kinetic approaches to response variability should be congruent and there have been recent attempts at mapping this relationship (Hancock & Newell, in press; Schmidt et al., 1979).

In summary, the results of the experiments should lead to an understanding of the contribution of each impulse parameter to response variability. More important than the relative contribution of these factors is the development of a physical description linking impulse parameters to response variability. The outlined experiments provide a direct test of the impulse-ratio and impulse variability model, but initial indications are that neither model accurately accounts for variability in performance. A model taking into consideration more fundamental properties of the force production mechanisms may provide a better description of response variability and associated phenomena such as Fitt's Law and other speed-accuracy tradeoffs.

REFERENCES

Bahrick, H. P., Bennett, W. F., & Fitts, P. M. (1955). Accuracy of positioning responses as a function of spring loading in a control. Journal of Experimental Psychology, 49, 437-444.

Fitts, P. M. (1954). The information capacity of the human motor system in controling the amplitude of movement. Journal of Experimental Psychology, 47, 381-391.

Fullerton, G. S., & Cattell, J. (1892). On the perception of small
    differences. University of Pennsylvania Philosophical Series, 2.

Hancock, P. A., & Newell, K. M. (in press). The movement speed-accuracy
    relationship in space-time. In H. Heuer, U. Kleinbeck, & K. J. Schmidt
    (Eds.), Motor behavior: Programming, control and acquisition. Berlin,
    West Germany: Springer.

Jenkins, W. O. (1947). The discrimination and reproduction of motor
    adjustments with various types of aircraft controls. American Journal
    of Psychology, 60, 397-406.

Newell, K. M., & Carlton, L. G. (in press). On the relationship between
    force and force variability in isometric tasks. Journal of Motor Behavior.

Newell, K. M., Carlton, L. G., & Carlton, M. J. (1982). The relationship
    of impulse to timing error. Journal of Motor Behavior, 14, 24-45.

Noble, M. E., & Bahrick, H. P. (1956). Response generalization as a function
    of intratask response similarity. Journal of Experimental Psychology,
    51, 405-412.

Schmidt, R. A., Zelaznik, H. N., Hawkins, B., Frank, J. S., & Quinn, J. T.
    (1979). Motor-output variability: A theory for the accuracy of rapid
    motor acts. Psychological Review, 86, 415- 441.

HItts' Law? A Test of the Relationship Between Information Load and Movement Precision

Mathew Zaleski
SWS Systems

Neville Moray
Department of Industrial Engineering University of Toronto

Recent Technological developements have made viable a man-machine interface heavily dependent on graphics and pointing devices. This has led to new interest in classical reaction and movement time work by Human Factors specialists.

Two experiments were designed and run to test the dependence of target capture time on information load (Hitt's Law) and movement precision (Fitts' Law). The proposed model linearly combines Hick's and Fitts' results into a combination law which then might be called Hitts' Law. Subjects were required to react to stimuli by manipulating a joystick so as to cause a cursor to capture a target on a CRT screen. Response entropy and the relative precision of the capture movement were crossed in a factorial design and data obtained that were found to support the model.

Introduction

Software engineers have always been under pressure from software users to provide friendly, easy to use interfaces to software systems of all kinds. The most effective seem to be those which combine with hardware to enable the user to point out places on the computer screen. Several systems are on the market which use a mouse for this purpose.

Briefly, the software systems of which we wish to make an example of are those that have come to be called "icon driven". A typical example is the Finder of Apple Macintosh. It is relevant to note that the users of these systems make commands to the computer by pointing to small pictures on the screen. The user's progress is sometimes limited by the speed with which these "icons" can be selected .

Human factors engineers have undertaken to study the properties of several of the pointing devices. Card, English and Burr [1978] demonstrated that the mouse and joystick are limited by the classical psychological result of Fitts [1954]. Further work made clear that using a joystick to control the motion of a cursor on a CRT is subject to the same fundamental limitations as are manual aimed motions, say with a stylus. Studies in this field typically ask the subject to manipulate a mouse or joystick so that a computer controlled cursor moves within a target area of the computer's CRT. The time required for the subject to make movements of varying length towards targets of varying size is measured (MT), and usually found to follow the well known result:

$$MT = a\ ID + b, \tag{1}$$

where ID is called the Index of Difficulty and is usually:

$$ID = log_2\ (2A/W). \tag{2}$$

Relation 1 is called Fitts' Law, and equation 2 is only one definition of ID. Many others have been proposed, for instance in Welford [1968].

Fitts Law is closely related to information theory. There is no derivation of Fitts Law in the rigorous sense, but fairly convincing analogies can be made which compare movements made by the human to transmitting information down a noisy channel. Consider a user about to make a cursor motion. It is intuitive that he is able to transmit more information with a precise movement than a crude one. If we further suppose man's motor system has a finite capacity to transmit information, then we expect that the time required to execute a motion ought to be proportional to the amount of information transmitted. Given that ID measures the information content of a motion, equation 1 follows.

There is another important element of the user's task, namely that he must often choose between discrete alternatives that are clearly presented on the screen before him. In many cases he is performing a similar task to that performed by the subject of a choice reaction time experiment. (Hick [1952], Hyman [1953])

The information content of a discrete target capture is quantified by a measure called response entropy (H). If we assume that man has a limited capacity to transmit information then we conclude that the length of time to make a choice (RT) will depend on the entropy of the required response.

$$RT = c + d\,H \tag{4}$$

Equation 4 (Hick-Hyman Law) is often written for equiprobable stimuli (so the probability of each is 1/n, where n is the number of stimuli) as:

$$RT = c + d\,log_2(n). \tag{5}$$

In light of the above discussion one might remark that there are aspects of the icon driven software interface which correspond to the view of both the Hick-Hyman and Fitts' Laws. A natural question to ask is whether a combination of the two laws might not be a useful way of modelling the behaviour of the user of such software. One might suggest that time taken to capture one of several targets would be described by:

$$CT = \alpha + \beta\,H + \gamma\,ID \tag{6}$$

where CT is the time to capture the target, H is the average response entropy and ID is the index of difficulty of the movement.

This combination of Hick's and Fitts' laws was proposed by Beggs, Graham, Monk, Shaw and Howarth [1972]. They performed an experiment which had inconclusive results and so it would appear that such a combination law has never been proved. If the combination law were found to hold it would offer a more complete model of the operator of icon or menu driven software systems in that it would incorporate two aspects of performance, namely the effects of both movement precision and response entropy on the average capture time.

In suggesting a additive combination of two fundamentally different psychological processes one enters a Great Debate in modern psychology. If a combination law such as equation (6) is found to hold does this imply that the underlying internal processes are serial and additive? Sternberg [1969] performed an elegant series of experiments in which certain memory searching processes appeared to be carried out in a highly serial way. His work gave rise to what has come to be called the additive factors methodology, which once was viewed as a way of detecting serial vs parallel processing. Taylor [1976] amongst others, suggested parallel processing schemes to explain the same data, and hence introduced a more conservative experimental approach which, unfortunately, is much more complex. This is mentioned in the context of this study because the data in the present study were analysed in a way similar to that used in additive factors and thus the results may be interpreted accordingly.

Goals of the experiments

The immediate goal of the study was to test whether the combination law holds for the task of manipulating a joystick so that in response to a visual stimulus a computer

SCREEN LAYOUT

fig 1

22.4
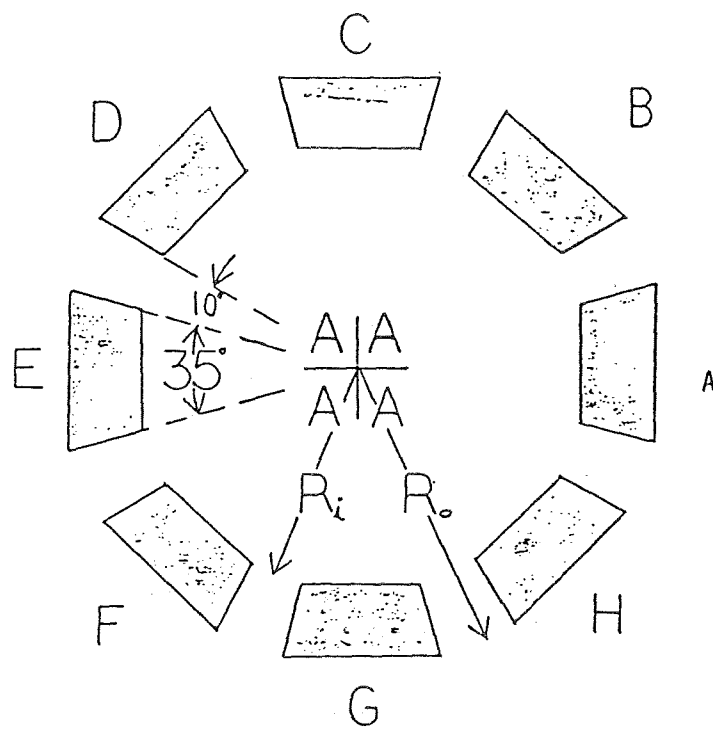
controlled cursor moves into a target area of a CRT display. A positive result would open the door to a predictive model which might be useful as part of the design process.

Description of pilot experiment

The pilot experiment set out to test equation 6 in a very typical choice reaction time study. The experiment was carried out on a micro computer with subjects manipulating a joystick and the computer arranging that a cursor move correspondingly. Reactions were to letters of the alphabet plotted in the centre of the screen. There was a one to one correspondence between targets and letters. In a given trial the number of targets and their dimension was manipulated. Stimuli were chosen such that they were equiprobable, so the average response entropy of a sequence of trials with n targets was $log_2(n)$. All the targets in a given trial were of the same relative size, and given the geometry of the situation the index of difficulty or ID was given by:

$$ID = log_2\left[(R_o + R_i)/(R_o - R_i) \right] \tag{6}$$

where R_o and R_i are defined in figure 1.

The pace of the experiment was sedate. First the targets and cross-hairs for a particular trial were plotted on the CRT. The subject was verbally instructed to centre the cursor on the cross-hairs during a delay of about 3.4 sec. An auditory warning followed the delay and 4 copies of the stimulus were presented (as in figure 1). Subjects then captured the target with the instructions to be as time efficient as possible without sacrificing accuracy. The cursor remained under joystick control for a further 3 seconds following the onset of the stimulus, at which time the screen cleared and there was a one second delay while data was written to disk. There was no quantitative feedback at the end of the trial or session of the subjects' performance. No attempt was made to instill competition between the subjects.

Experimental Design

There were three levels of H (see table 1) and three levels of ID (table 2). Subjects performed four pairs of sessions, each pair constituting one pass through the design. In most cases both halves of the design were performed one shortly after the other, with a rest in between. Each session was composed of six "blocks" of trials. All trials within each block had the same number of targets and hence the same response entropy. Each block was in turn divided into three "groups" of six trials each. All trials within a group had constant ID. Thus each session was made up of 12 trials in each of the nine cells of the design. Since it took two sessions for one pass through the entire design, each pass required 24 trials in each cell for 216 trials in all. Seven subjects, who were graduate students at the University of Toronto, each completed four passes through the experiment . Subjects were paid $5 per hour for their participation in the experiment.

22.5

C - 4

Table 1 - H manipulation - pilot

response entropy (H)	number of targets	targets sets	special features
3	8	ABCDEFGH	all targets used
2	4	ACEG BDFH	vert/Horiz quadrants diagonal quadrants
1	2	AE BF CG DH	right and left diagonal 180 ° apart top and bottom like BF, but rotated 90. °

Table 2 - ID manipulation - pilot

ID	R_I	R_o
3	60	77
4	75	85
5	80	85

In the blocks which contained fewer than all the targets there was the problem of choosing which subset of targets to use. The subsets chosen are listed in table 1. Note that what has been done is to restrict the screens to either vertical and horizontal or diagonal symmetry but no mixture of the two.

Implementation

The experiment was run on an Apple IIe 6502 based micro computer. All software was written in UCSD pascal except the clock and ADC drivers, which were written in assembler. A real time clock and ADC device handler was designed which collected data while the pascal mainline controlled the screen.

Subjects responded using a Measurement Systems joystick with no spring return to centre. The maximum possible deflection of the joystick was about 30°. Subjects were not located exactly with respect to the screen and joystick, but for the typical subject there was a gain of about 0.25° of visual angle for each 1° of joystick deflection.

With this apparatus the duration of each target capture (CT) was defined to be the interval between the onset of the stimulus and the beginning of a 350 millisecond capture of the target. Reaction Time (RT) was defined as the period from the onset of the stimulus until the joystick was deflected 0.3°. Movement time (MT) was the difference between CT and RT.

Results of Pilot Experiment

Statistical analysis was carried out in two main ways using analysis of variance (ANOVA) and regression analysis. This reflects the three main topics of interest, which are:

  i. How well do the independent factors ID and H predict the time required by subjects to select and execute a target capture response?

ii. In terms of the information hypothesis, what are the information capacities of the subjects to H and ID? Do they change with practice?

iii. Do the independent factors interact?

Results of ANOVA analysis for pilot experiment

An ANOVA was carried out for RT, MT and CT for each of the four runs the subject made through the experimental design. The ANOVAs assumed a three factor completely randomized mixed model. Only the results of the last run will be described here. Note however that examination of the training data showed that subjects were not stable at the end of the pilot experiment, as they were still improving significantly from the third to the fourth and last session. In the following ANOVA data the standard statistics are presented, namely the F score of the null hypothesis test (F), its Mean Square Error (MSE), the probability of the null hypothesis being true (p) and finally the fraction of the variance attributable to each factor, ω^2.

Reaction Time

The ANOVA shows that RT is influenced significantly only by H. Interaction terms and ID have no significant effect. Between subjects variation accounts for much of the total variance. The difference between subjects is highly significant.

We emphasize that RT has been defined operationally to be the time from the onset of the stimulus until the first small deflection of the joystick. Thus all factors which cause the subject to delay are grouped under RT. Nevertheless, as shown by the ANOVAs described below, only H has a significant effect on RT. This would imply that the particulars of the movement about to be made do not affect the duration of the delay before the movement.

Table 3 - Reaction Time Anova - pilot

factor	F		MSE	p	ω^2
Subject	$F(1,6)=$	79	2830	< 0.001	-
ID	$F(2,12)=$	1.2	22	0.30	-
H	$F(2,12)=$	32	99.6	< 0.001	0.24
interaction	$F(4,24)=$	.98	13	0.44	-

Movement Time

Movement time was found to depend significantly only on ID. H and interaction terms were found to be non-significant. As before subjects differed significantly.

Table 4 - Movement Time Anova - pilot

factor	F		MSE	p	ω^2
Subject	F(1,6)=	251	760	<0.001	-
ID	F(2,12)=	82	73	<0.001	0.60
H	F(2,12)=	0.062	31	0.94	-
interact	F(4,24)=	.53	66	0.71	-

Capture Time

Capture time showed significant effects of H and ID, but the interaction component of the model was not significant. The ω^2 column shows how ID accounts for about 20% of the variance and H for slightly more than 10%. The rest is due to between subject differences. It is clear from this data that there is no interaction taking place between the factor which affects the period of time until the beginning of the subjects' overt response (H) and the factor which affects the duration of the movement (ID).

Table 5 - Capture Time Anova - Pilot

factor	F		MSE	p	ω^2
Subject	F(1,6)=	173	4542	<0.001	-
ID	F(2,12)=	63	76	<0.001	0.20
H	F(2,12)=	20	166	<0.001	0.12
interact	F(4,24)=	3	73	0.53	-

Results of multiple regression analysis - pilot study

The result to be presented in this section is the multiple regression of average (across all subjects) CT vs ID and H. This regression is an experimental test of the combination of Fitts' Law and the Hick-Hyman Law given in equation 6. Simple regressions of RT vs H and MT vs ID were also performed to continue our examination of how these stages depend on H and ID. In this section we will concentrate on two main statistics:

i. The statistic r^2 is quoted to describe what proportion of the total variance can be explained using the simple linear model of equation 6. It is generally not safe to judge the quality of fit from r^2 alone.

ii. Residual Standard Error (RSE) is equal to the average square residual. This gives an indication of how far the average data point is from the fitted line.

The coefficients of the regression have the physical dimension of seconds per bit. Thus their reciprocal has dimensions of information capacity, or bits per second. The coefficients can be used to calculate the information capacity of the subjects with respect to H and to ID. Comparisons of the resulting information capacities to those of earlier studies is discussed in later sections.

RT vs H – points marked with ID
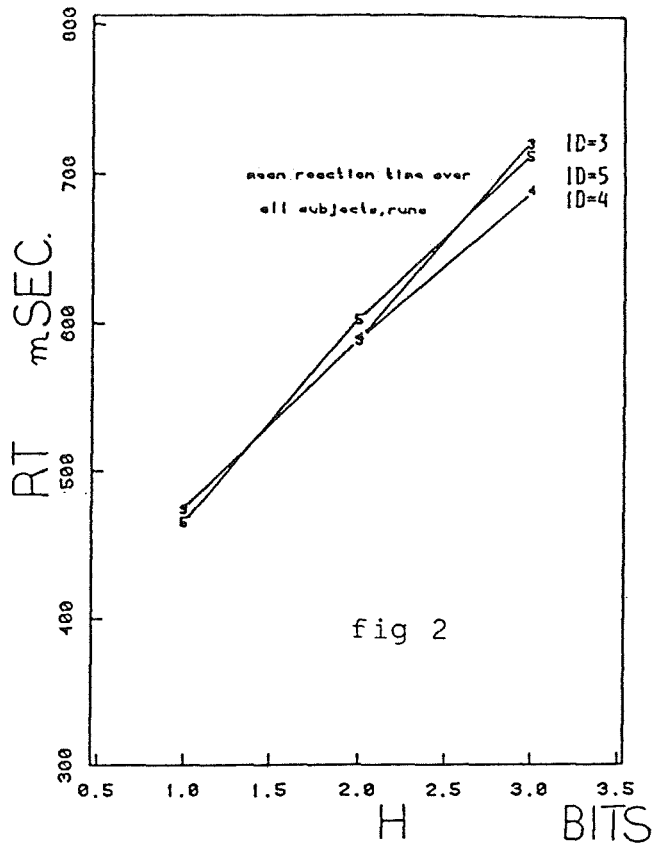
RT mSEC.

800

700

ID=3
ID=5
ID=4

mean reaction time over
all subjects, runs

600

500

fig 2

400

300

0.5 1.0 1.5 2.0 2.5 3.0 3.5

H   BITS

MT vs ID – points marked with H

800

H=3
H=1
H=2

averaged over all subjects

and runs

mean movement time (mSec)

700

MT

600

500

fig 3

400

300

2.5 3.0 3.5 4.0 4.5 5.0 5.5

ID   BITS

Capture Time vs ID – points marked with H

1600

average across all subjects and runs

H=3

1400

H=2

H=1

Average Capture Time (mSec)

1200

CT

1000

fig 4

800

600

2.5 3.0 3.5 4.0 4.5 5.0 5.5

ID   BITS

22.9
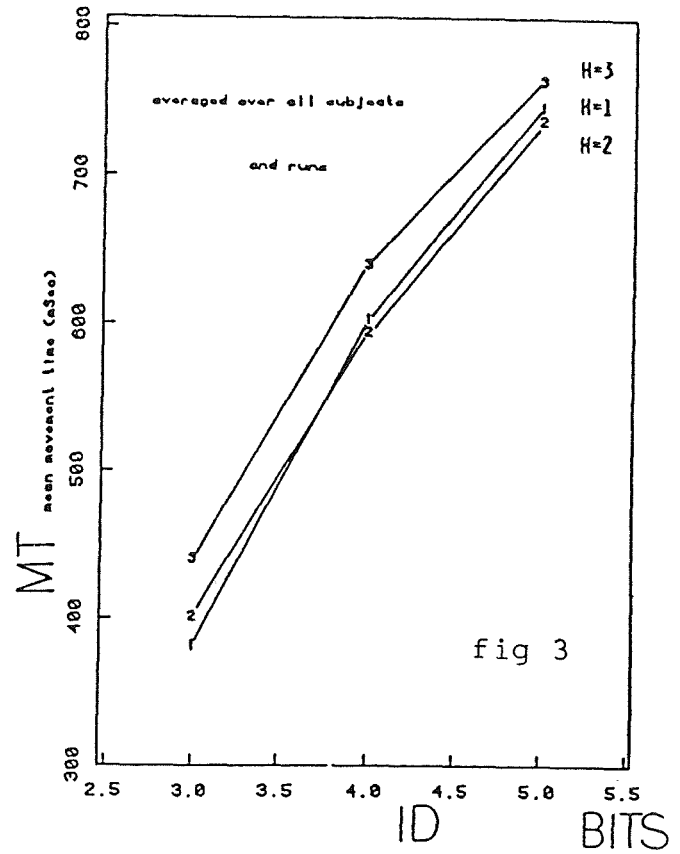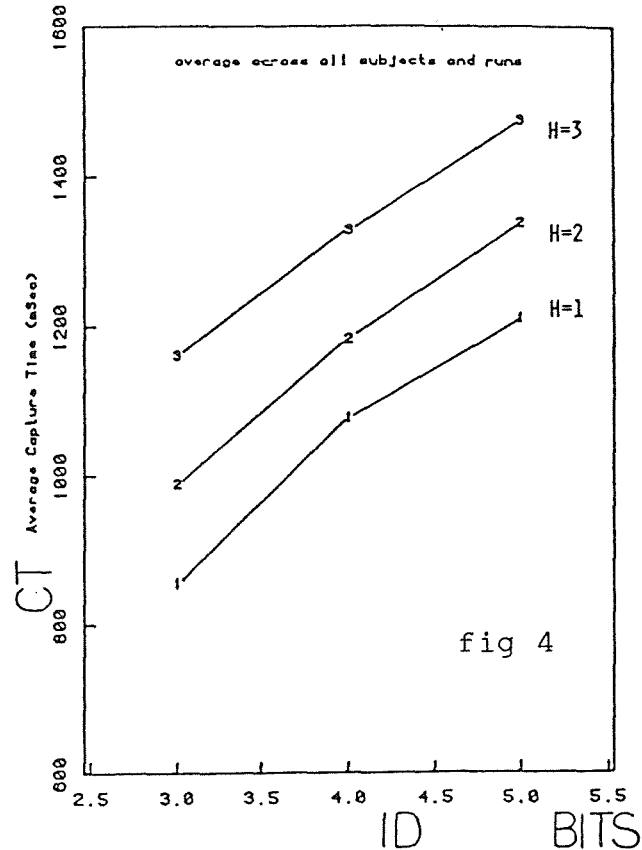
Reaction time

This section presents the extent to which the data follows the Hick-Hyman Law.See figure 2 for a graph of RT vs H for data pooled across all subjects' last run.

Table 6 - Reaction Time regression analysis - pilot

for model: RT=α+βH
(Hick's Law)

α	β	r^2	RSE	F(1,7)
349	123	0.98	14.7	421
mSec	mSec/bit			

Movement time

The Fitts' Law component of the data is described in Table 7. Presumably the value of the constant term would be different if the arbitrary boundary between RT and MT were changed.

Table 7 - Movement Time regression analysis - pilot

for model: MT=α+γID
(Fitts' Law)

α	γ	r^2	RSE	F(1,7)
-122	168	0.98	19.9	428
mSec	mSec/bit			

Capture time

The test of equation 6 with respect to the data of the pilot experiment is presented here. The r^2 of the regression is 0.99, so the model is explaining almost all of the variance in the pooled data. It is fair to say that the combination law describes the CT data just as well as the two classical laws describe RT and MT. Table 8 summarizes the regression results for the average across all subjects fourth run. Each subjects has made 648 responses previously.

Table 8 - Capture Time regression alalysis - pilot

α	β	γ	r^2	RSE
264	127	150	0.99	21
mSec	mSec/bit			

One might ask how such high linearity is present given the large between subjects variation measured by the ANOVAs. Table 9 presents the same regression

carried out in Table 8 except for each subject individually. Since there is less data there is *more noise*. Initially it was anticipated that the subjects would differ mostly in intercept with roughly similar coefficients. This, as is shown in table 9, is not borne out by the data at all. The per subject regressions illustrate again how simple models can predict mean behaviour well and yet cast little light on individual performance.

Contest - study 2

In the task of the pilot study the additive combination of Hick's and Fitts' laws was a very appropriate way of mathematically describing average subject performance. The results indicated that the subjects could be thought of a reacting in two sequential

Table 9
Per Subject Capture Time regression alalysis - pilot

model: CT=α+βH+γID			
subject	α	β	γ .	r^2
gc	203	107	152	0.82
ir	279	52	154	0.84
jp	149	103	219	0.87
kh	253	170	130	0.85
kv	321	122	112	0.88
mc	94	78	160	0.81
mk	549	257	127	0.94
average	264	127	150	.99
dimension	mSec	mSec/bit	mSec/bit	-

phases: a response selection stage followed by a movement stage. The task carried out by subjects in the pilot experiment differed from otherwise similar tasks performed by operators of icon driven software systems in several important respects:

i.   The trials were highly discrete. There was a gap between trials which was of   considerably   longer duration then the trials themselves. The experiment of Beggs et al [1972] was a continuous one and H and ID were found to interact. Practical software systems often require the user to make a series of captures, and often with little or no externally imposed temporal uncertainty.

ii.   The symmetry of the target capture motions   made in the pilot experiment was highly radial. The direction of the required motion corresponded one to one with the stimuli. This is artificial in the sense that in practical situations the stimulus corresponds to a target, but the direction of motion depends upon the starting position as well.

iv.   No feedback was given to the subjects of the pilot study of when they had captured their target (other than the position of the cursor on the screen) or how their performance compared to other subjects. This is very unrealistic, for in a practical setting there is little point in capturing targets if nothing is going to happen when you do so.

SCREEN LAYOUT

fig 5

22.12
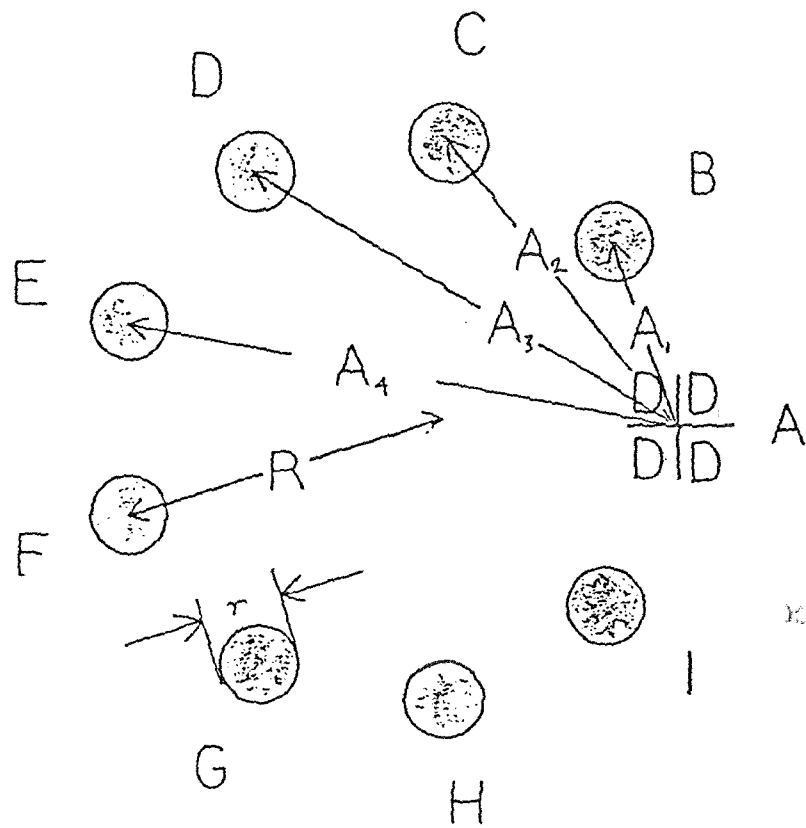
Description of Contest

The second study was designed in answer to these points. Enough feedback was built into the experiment that we refer to it as "Contest". Trial by trial feedback was provided by sounding a beep as soon as the capture was detected. At the end of each session a subject was told his total score, and a graph of these scores was clearly displayed on which each subjects progress was recorded in a different colour. Much effort was expended preparing software to make it possible to make the end of each session a competetive event, in which scores were compared (and excuses made). A prize of $10 was offered (above the hourly rate) for the best score.

Contest was continuous, which in the context of such experiments means only that delays between the trials have been minimized. The subject no longer waits for the stimulus, but initializes its onset himself by completing the previous trial. Subjects face no temporal uncertainty apart from that produced by short delays in the software. Hopefully this will allow more direct comparison to the continuous tapping style of Fitts Law experiment [Fitts 1954]. The response motions no longer correspond only to stimuli but also depend on the situation. In Contest the subjects do not return to cross-hairs at the centre of the screen but rather the cross-hairs are plotted over the last target captured, so that the next trial can start immediately. Figure 5 illustrates the screen layout of Contest.

It is desirable to use an Index of Difficulty as comparable to the one used in the pilot study as possible. Referring to figure 5 we note:

$$ID = log_2 (A_i/r).\qquad\qquad(7)$$

Experimental Design - Contest

As is visible from figure 5, up to four IDs are introduced by one choice of R and r. This has the effect of making it impossible to separate trials into groups of constant ID. For some "configurations" (figure 5 is an example of one configuration) there can be trials of different ID. The experimental design was difficult because it was convenient to have the same number of trials in each cell. However, we could not choose simple subsets of the targets as in the pilot study (see table 3.1.1) because there was no such set which had the same number of trials in each cell. The result was that a relatively large number of different configurations were chosen.

There were six subjects, three men and one woman graduate students, and two high-school age teenagers. They were paid $5 per hour and knew about the $10 first prize from the outset. Subjects participated in 11 or 12 sessions, until their behaviour had asymptoted as indicated by their score in each run.

There were 12 levels of ID and 4 levels of H; 48 cells in all. The design was made up of 32 blocks of three groups of 16 trials each for 3x16x32 = 1536 trials. Runs 1 through 10 used exactly the same sequence of trials, then for the last two trials the sequence was changed. The block and group structure was identical, but the stimuli were presented in a different order.

Results of Contest

The implementation of the experiment was essentially the same as for the pilot study, except for some rearrangement of the screen. The data processing had to be streamlined in order to detect target hits on line and to be ready at the end of a trial to feed performance back to the subject.

The statistical processing applied was also unchanged. The only difference was that it was found that the division of CT into RT and MT was not possible and so the statistics are quoted only for CT. In the pilot study there was a long forewarning period in which the subjects kept the cursor still on the cross-hairs. Thus the end of the Reaction Time period was reliably detected by the first deflection of the joystick. In Contest, continuous by design, subjects never held the joystick still for long enough to detect any transition between RT and MT. One attempt was to estimate the rate and acceleration of the cursor and make a decision based on them, but the results were not encouraging.

A session consisted of 1536 captures, the average trial requiring about 0.6 sec each, for an average session duration of about 50 minutes. Since the pace was set mostly by the subjects the percentage of the time on task actually spent in control of the cursor was about 60%. Subjects found the sessions quite tiring. In retrospect, a session of about 1000 trials would have been more appropriate.

Results of ANOVA analysis - Contest

ANOVA showed that variance in CT data pooled across all subjects' most highly trained session was almost entirely explained by the factors H and ID. H was responsible for 44% of the total variance, ID for 48%, leaving very little for between between subject differences and interaction. The interaction of H and ID was not significant at the 2.5% level, but it was at the 5% level. However, if the ω^2 of the interaction term is examined it becomes clear that the interaction has a negligible effect on CT. See table 10. It is fair to say that the interaction, even though statistically significant, is of no practical importance. It would appear that the steps taken to ensure that subjects are motivated and highly trained had a great affect upon between subject variance. Comparison of the ω^2 of tables 5 and 10 illustrates this clearly.

Capture time vs ID - points marked with H

fig 6

Capture time vs ID - points marked with H

fig 7

22.15
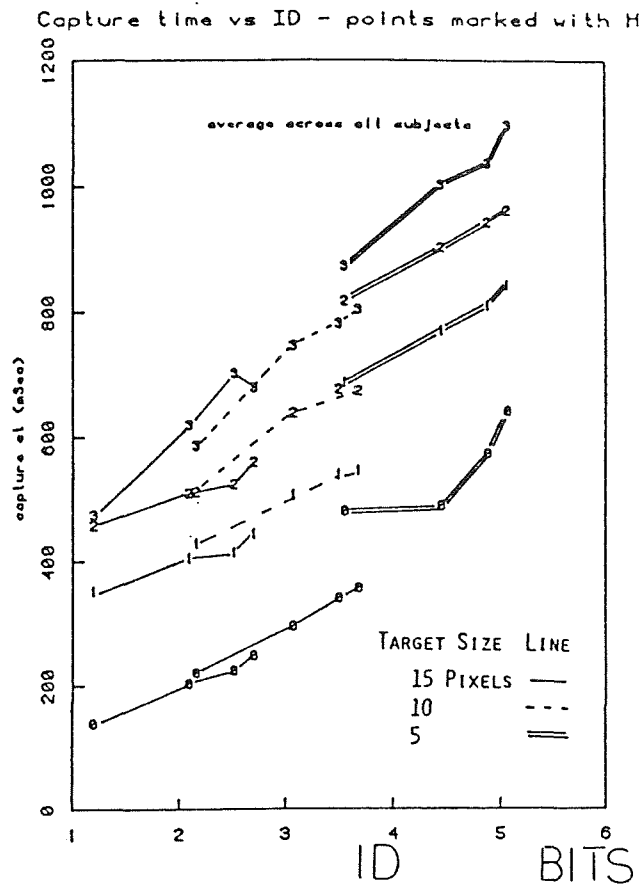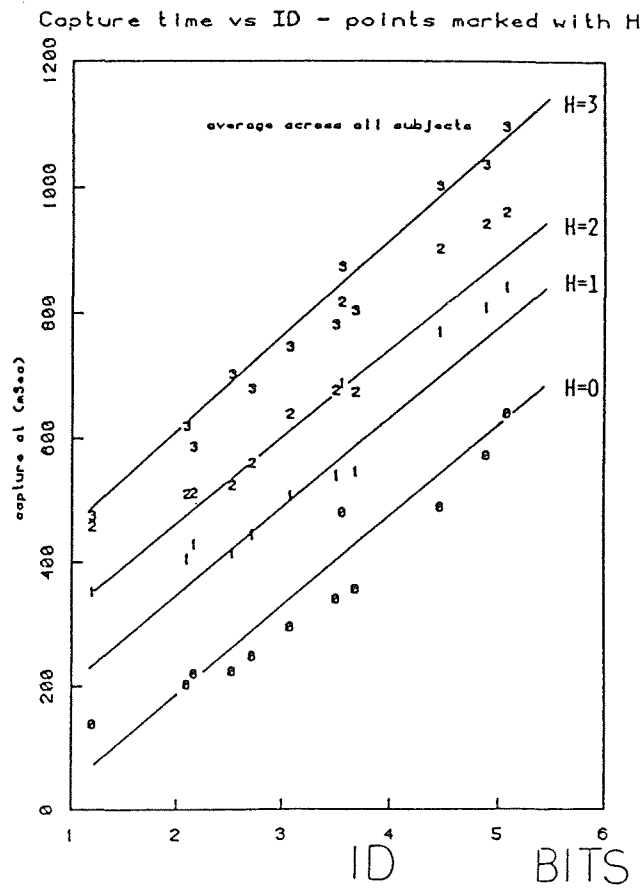
Table 10 - Capture Time Anova - Contest

factor	F		MSE	p	ω^2
Subject	F(1,4)=	353	4945	< 0.001	-
ID	F(11,44)=	200	31	< 0.001	0.48
H	F(3,12)=	226	92	< 0.001	0.44
interact	F(33,132)=	2.0	17	0.03	0.004

Results of multiple regression analysis - Contest

Due to the greater number of cells in the experimental design of Contest, one is more inclined to have confidence in the results of regression analysis. Comparing tables 8 and 11 we see that the linearity of Contest data is less than that of the pilot experiment. Nevertheless, both the r^2 and RSE indicate a very good linear fit to the model of equation 6. It would appear that the combination of Fitts' Law and the Hick-Hyman Law stands up to the more realistic task of Contest almost as well as to the task of the pilot experiment.

The capture time data for Contest is presented in figure 6 with the multiple regression line drawn in for several values of H, and in figure 7 with equal size targets connected by lines. Examination of figure 7 shows that although the model explains some 95% of the variance, clearly target size plays a role besides the one recognized by ID. Flgure 7 shows how the two larger target sizes (about $0.5°$ and $1.0°$ of visual arc) fall in line whereas the smallest targets (about $0.3°$ of visual arc) seem to take longer to capture. Jagacinski and Monk [In Press], have tested Fitts Law in two dimensions using similar apparatus and found Fitts' Law to hold for targets of this size. Their criterion for target capture was not quite as simplistic, in that they allowed the cursor to leave the target for very short periods of time during the capture in order to "avoid penalizing the subjects for slight amounts of jitter" [Jagacinski and Monk, In Press]. It is possible that the stringent operational definition of capture used in Contest lengthened CTs for small targets by accentuating the effects of muscular tremour.

Table 11 - Capture Time regression analysis - Contest
for model: CT=α+βH+γID

α	β	γ	r^2	RSE
-83 mSec	142 mSec/bit	144 mSec/bit	0.95	55

Practice effects

Practice effects were investigated by performing the analysis described above for each run of both the pilot study and Contest. Figure 8 is essentially the same graph that was displayed near the apparatus. It shows the anticipated flattening out of performance. Figure 9 shows how the form of the data does not change qualitatively from session to

score vs runs, Exp 2

fig 8

run 1   run 3

fig 9

run 6   run 9

22.17
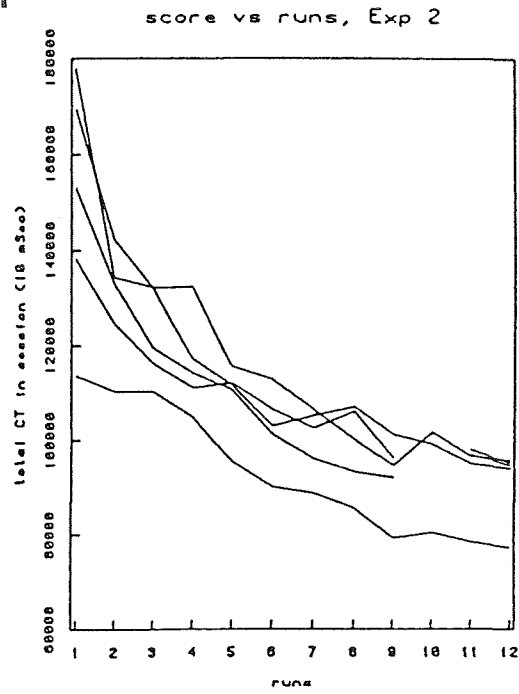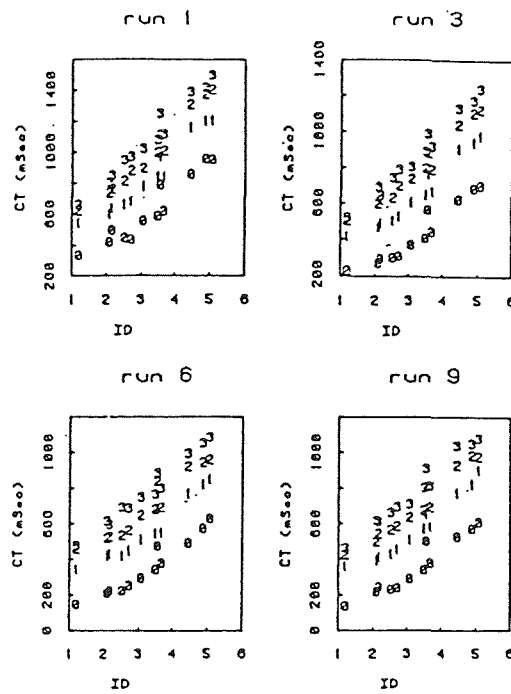
session even though quantitatively it can be seen to shift downwards.

Table 12
Regression analysis of Practice Effects - Contest

$$T = \alpha + \beta H + \gamma ID$$

run	α	β	γ	r^2	RSE
1	6.18	154	200	0.95	6.8
2	-47.7	149	184	0.95	6.5
3	-61.1	148	165	0.95	6.1
4	-76.5	140	158	0.95	5.8
5	-95.6	141	152	0.95	5.8
6	-75.9	141	143	0.95	5.6
7,8	NA	NA	NA	NA	NA
9	-90.6	141	149	0.95	5.6
10	-102	140	155	0.94	6.4
11	-68.8	141	143	0.94	5.8
12	-83.2	142	144	0.95	5.5
	mSec	mSec/bit	mSec/bit		

Perhaps the most interesting practice effect is evident in the CT data of the pilot experiment. Table 12 lists the regression results of each session of the pilot study whereas figure 10 shows the grand mean of RT, MT and CT for each run, pooled across subjects. We see that although MT decreases from the first session to the last session by about 13% RT changes little. One might have expected that with practice both RT and MT would decrease. Furthermore, the same trend is visible in the regression of CT. The intercept term increases even though the mean of CT gets smaller with practice. It would appear that as they learn the task subjects invest time at the beginning of each response which they can regain during the movement phase. Finally we observe that much improvement took place in the last two sessions of the pilot study, a clear indication that the subjects' performance had not stabilized.

In Contest, the intercept of the CT multiple regression started slightly greater than zero and steadily decreased. One assumes that the lack of temporal uncertainty in the task would chop a constant time out of the CT, but a negative intercept seems unrealistic at first glance. On closer examination one learns that several studies of discrete target capture behaviour found a negative Fitts' Law intercept [Fitts and Peterson, 1964]. The intercept of the regression in the extrapolation of the data to a point at which H and ID equal zero. Zero response entropy corresponds to the situation where the subject has no choice to make, and so is well defined. Zero Index of difficulty corresponds to an odd geometry in which the width of the target and the length of the motion are equal, an unrealistic scenario.

The regression coefficients of the CT regression decrease steadily with practice.

They vary inversely with information capacity and so we see the subjects' capacity to transmit information increasing with training.

Discussions and Conclusions

We have shown that CT is influenced by the degree of choice and by the required movement precision. By using a simple combination of Fitts' and the Hick-Hyman Laws, most of the variation in the data can be accounted for. These results hold both in a highly discrete and a continuous setting.

In the pilot experiment, RT was as well described by the Hick-Hyman Law as MT was by Fitts' Law. CT was as well described by the combination ("Hitts") law. The least squares multiple linear regression fit to the most (though still not fully) practiced data of the pilot experiment was:

$$CT = 260 + 130H + 150ID \quad (mSec)$$

whereas the most highly trained session of Contest yielded:

$$CT = -83 + 140H + 140ID \quad (mSec)$$

Following Fitts and Peterson [1964], the ID suggested by Welford, namely:

$$ID' = log_2(A/2r + 0.5) \tag{7}$$

was tried out to see what effect it would have. The quality of fit was unchanged and the information capacity with respect to ID was reduced by about 11%. There seems no reason with these data to favour equation (7) over equation (2).

At risk of becoming embroiled in controversy, we comment that the data of the pilot experiment support the hypothesis that response and movement are executed sequentially as two separate stages. The independent measurement of RT and MT suggests that the factors affecting RT do not affect MT and vice versa. It would appear that *sequential behaviour is a fact. Whether or not internal processing has the same structure* is another problem altogether.

In the second experiment, the ability to independently measure RT and MT has been lost and so no such claim can be made. All that can be said here is that the factors which affected RT and MT separately in the pilot study interact to a negligible extent. One could probably analyze the time series of joystick positions with more sophisticated analysis and divide CT in the more continuous task Our experience has shown that that this may be difficult to accomplish.

Table 13 compares several studies in the literature with our results. We point out that although the experimental methods differ greatly (for instance the studies shown differ widely in modality of stimulus and response) information capacity with respect to H

varys over all by only about 25%. Fitts Law values, however, differ widely between experiments.

Table 13 - Comparison of results with other studies

Study	with respect to	
	ID	H
Hick[1952]	-	6.4
Hyman[1953]	-	7.87
Fitts[1954]	11.5	-
Jacinski and Monk[In Press]	5.0	-
Fitts and Peterson[1964]	13.5	-
Experiment 1	6.7	7.9
Experiment 2	6.9	7.0
	bps	bps

This would suggest to us that there is much to be gained at the physical interface level to the user. There is a need to tune the dynamics of mouse driven systems. This implies some quantitative method is required to provide a criterion for optimising design. To illustrate the environment in which software engineers typically work we quote from the notes for software developers included with what is one of the world's leading mice:

> We strongly urge you to try 2X magnification. Most software engineers are reluctant to do so, but after trying it, they find the feeling of control and speed far outweigh the inability to choose single pixels.. [p 8, Mouse Systems Corporation, M-2 Optical Mouse Technical Reference Manual, Jan. 1984]

Information capacity with respect to ID is a good starting point for tuning an interface. Anyone who has used a mouse recognizes that there are tasks which require higher or lower gains depending on the average size of targets and lengths of motions. One hopes that eventually a body of knowledge and guide lines will appear for what dynamics to use in which typical situations.

BIBLIOGRAPHY

Beggs, W.D.A., J.C. Graham, T.H. Monk, M.R.W. Shaw, and C.I. Howarth, "Can Hick's Law and Fitts' Law be combined?,"*Acta Pschologica,* vol. 36, pp. 348-357, 1972.

Card, S. K., W. K. English, and B. J. Burr, "Evaluation of mouse, rate controlled joystick, step keys and Text keys for Text Selection on a CRT,"*Ergonomics* , vol.21, pp. 601-613, 1978.

Fitts, P.M., "The Information capacity of the human motor system in controlling the amplitude of movement," *Journal of Experimental Psychology,* vol. 47, pp. 381-391, 1954.

Fitts, P.M. and J.R. Peterson, "Information capacity of discrete motor responses," *Journal of E xperimental Psychology,* vol. 67, pp. 103-112, 1964.

Hick, W.E., "On the rate of gain of information," *Quarterly Journal of Experimental Psychology,* vol. 4, pp. 11-26, 1952.

Hyman, R, "Stimulus Information as a Determinant of Reac tion Time,"*Journal of Experimental Psychology,* vol. 45, pp. 188-199, 1953.

Jagacinski, R. and D. L. Monk, "Fitts Law in two dimensions with hand and head movements,"*Journal of Motor Behaviour,* (In Press).

Kornblum, S., "Choice Reaction Time for Repetitions and Non-repetitions: A Rexamination of the Information Hypothesis," in *Attention and Performance* , ed. A. F. Sanders, pp. 178-187, North Holland Publishing Co., Amsterdam, 1967.

Meyers, B. A., "The User Interface for sapphire: A Screen Allocation Package Providing Helpful Icons and Rectangular Environments," InPress.

Sternberg, S., "The Discovery of Processing Stages: Extensions of Donders' Method.," in *Attention and Performance II,* ed. W. G. Koster, pp. 276-315, North Holland Publishing Co., Amsterdam, 1969.

Taylor, D.A., "Stage Analysis of Reaction Time,"*Psychological Bulletin,* vol. 83, no. 2, 1976.

Welford, A.T., *Fundamentals Of Skill,* Methuen and Co.Ltd, London, 1968.

AIRCREW COORDINATION AND DECISIONMAKING:
Peer Ratings of Video Tapes made during a Full Mission Simulation

Miles R. Murphy, NASA-Ames Research Center, Moffett Field, CA
Cynthia A. Awe, San Jose State University, San Jose, CA

Abstract: Six professionally active, retired captains
rated the coordination and decisionmaking performances
of sixteen aircrews while viewing videotapes of a simu-
lated commercial air transport operation. The video-
tapes displayed a composite of four views of
crewmembers, and the cockpit, from cameras located
inside the simulator. The scenario featured a required
diversion and a probable minimum fuel situation. Seven
point Likert-type scales were used in rating variables
on the basis of a model of crew coordination and
decisionmaking. The variables were based on concepts
of, for example, decision difficulty, efficiency, and
outcome quality; and leader-subordinate concepts such
as person- and task-oriented leader behavior, and com-
petency motivation of subordinate crewmembers. Five-
front-end variables of the model were in turn dependent
variables for a hierarchical regression procedure. The
variance in safety performance was explained 46%, by
decision efficiency, command reversal, and decision
quality. The variance of decision quality, an alterna-
tive substantive dependent variable to safety perfor-
mance, was explained 60% by decision efficiency and the
captain's quality of within-crew communications. The
variance of decision efficiency, crew coordination, and
command reversal were in turn explained 78%, 80%, and
60% by small numbers of preceding independent vari-
ables. A principle component, varimax factor analysis
supported the model structure suggested by regression
analyses. Crewmembers for this study were diverse with
respect to airline of origin and recency, or currency
on the Boeing 707 - the aircraft simulated. Some
retired personnel were used. The results should be
interpreted accordingly.

INTRODUCTION

The aircrew interaction process has been implicated as contributing to numerous recent air transport accidents and incidents (Cooper, White, & Lauber, 1979; Murphy,1980; NTSB, 1976). And many interpersonal factors have been suggested as causes of ineffective crew performance {lack of decisive command, strained social relations, and pilot-copilot role relationships (Murphy, 1977)}. Problematic pilot-copilot role issues include the command responsibility of the captain when the first officer is flying, and the responsibility of the first officer when the captain deviates from safe or legal practices (Wiener, 1977).

Flightcrew communications patterns have been related to performance outcomes in a study of simulator data (Foushee and Manos, 1981). Mitigation level, a linguistic indication of tentativeness and indirectness in speech, has been identified as a factor in failures of crewmembers to get new topics discussed or suggestions ratified by the captain (Goguen, Linde, and Murphy, 1984). In their study of air transport accident transcripts they also showed mitigation level to vary with command and situation dimensions. Finally, a full mission simulator study of crew performance (Ruffell-Smith, 1979) related ineffective management of both human and material resources to increased decision times. Generally, however, suggested causal factors in air crew performance effectiveness have not been well defined through systematic study or research. One reason for this could be the lack of adequate methods for isolating and quantifying crew interaction factors and for relating these factors to flight task performance (Foushee, 1984; Murphy, 1977) – a situation comparable to that for small group performance generally (Hackman and Morris, 1975).

The major objective of this rating study was to initiate development of a hierarchical process model of aircrew coordination and decisionmaking. A secondary objective was to develop reliable measures of the crew interaction process that could be related to other substantive measures, such as flight task error measures, or to measures developed with coded communications data. This will be addressed in future reports.

This study used videotapes of aircrews performing a full mission simulation of a commercial air transport operation (Murphy, Randle, Tanner, Frankel, Goguen, and Linde, 1984). Such videotapes have been used in studying medical team-patient interactions (Frankel & Beckman, 1982). Leadership style and crewmember competency variables, included in the model of crew performance presented below, reflect findings from recent critical reviews of the leadership literature (House, 1984; House & Baetz, 1979; Kerr, 1984). Design

of the rating scales and procedures (including rater train-
ing), reflect findings from recent reviews of rating litera-
ture (Landy & Farr, 1983; Landy & Farr, 1980). The single
dimension , Likert-type scales and anchoring methods were
justified on the bases that the study is exploratory, and
research findings have shown little gain in performance when
more complex scales are used.

METHOD

The Model

Crewmember Behavioral Variables: Variables of major
interest were based on some focal concepts. Task-oriented,
person-oriented, and participatory leadership behaviors were
rated for both the captain and first officer. These vari-
ables were differentiated on the basis of specific
behaviors. Task-oriented leadership behaviors were those
concerned with establishing goals, clarifying responsibili-
ties, defining subordinate (others for the first officer)
roles and task requirements, coaching subordinates and pro-
viding task related feedback. Person-oriented leadership
behaviors included those evidencing concern with establish-
ing and maintaining positive crewmember relationships, pro-
viding psychological support, and enabling feelings of
satisfaction.

Participatory leadership was rated in regard to
supervision/resources management on the basis of behavior
that encouraged subordinates (or other crewmembers, for
first officer) to make suggestions regarding accomplishment
of tasks, independently analyze problems, give feedback, and
question the leader. Participatory leadership was also rated
in regard to decisionmaking. The criterion was behavior con-
cerned with ensuring that all crewmembers for whom a deci-
sion was relevant had a chance to influence that decision.
Relevance was indicated if a crewmember had significant and
pertinent information related to the decision, responsibil-
ity for implementing the decision, or significant ego
involvement for other reasons.

To address the question of interaction between a
captain's leadership effectiveness and a subordinate's capa-
city and willingness to participate, the first officer and
flight engineer were rated on a dimension of competency-
motivation. The criterion was evidence of a crewmember being
knowledgeable, skillful, and motivated with respect to ful-
filling the requirements of his position.

Within-crew communications quality was rated for all three crewmembers with respect to both specific decisionmaking processes and participation in the more general crew coordination process. Behavioral criteria were 1) hearable, understandable, appropriate (in style and content), accurate, and timely messages; 2) being a good listener who makes an effort to understand; and 3) achieving a reciprocal indication that understanding was reached.

Command Reversal – An Interactive Variable: Command reversal was also rated with respect to both specific decisionmaking processes and the more general crew process. If, for example, a first officer performed much of what would normally be the captain's general leadership function, a high rating for that crew on the variable "command reversal" would be expected. Similarly, if a first officer performed much of what would generally be the captain's decisionmaking function, a high rating for that crew on "decision command reversal" would be expected. No attempt was made to distinguish whether such command reversals were due to acquiescence of the captain or dominance of the first officer or whether the actions and decisions were or were not appropriate.

Intermediate Performance Measures: Command reversal was expected to negatively affect crew coordination and decision efficiency, two other focal variables of the hypothesized hierarchical process model (shown in part in Figure 1). A high rating on crew coordination would indicate strong rater agreement that, over a mission segment: individual crewmember knowledge and skills were allocated in an effective and timely manner to meet task and situation demands. A high rating on decision efficiency would indicate strong rater agreement that, for a particular decision process all significant information was acquired at an opportune time, adequately evaluated, and appropriately utilized.

Dependent Variables: Decision quality and safety performance, also shown in Figure 1, are alternative, primary dependent variables for this study. Like decision efficiency and some other variables mentioned above, decision quality was rated for eight decision processes that occured during the mission. A high rating on decision quality would indicate strong rater agreement that a choice made was "the best considering safety of flight and/or the attainment of all other mission goals". Unlike the 15 crew process and nine decisionmaking variables, safety performance was rated on the basis of relatively factual data after raters had observed all 16 crews. Safety performance ratings were essentially based on an assessment of risk in a crew's solution or attempted solution to the major scenario problem.

Data entering into safety performance included the airport where landing occured, fuel on board at landing, and altitudes reached during approaches below minimums when the runway could not be seen.

Other Variables: In addition to the focal variables discussed above, three variables of more peripheral interest were rated: crew cohesiveness, crew friendliness, and decision difficulty.

Identification and Definitions of Variables: All variables are identified with their concepts, referents, and instrument by which they were measured, in Table 1. Attaching the prefix "decision" to a concept such as participatory leadership distinguishes a variable referring to participatory behavior in decisionmaking as opposed to that in supervision/resources management. Attaching the suffix (P1), (P2), or (P3) to a concept distinguishes a behavioral variable, such as communications quality, as to whether reference is to captain, first officer, or flight engineer behavior, respectively. Definitions for all variables can be synthesized from the scale and criterion statements of Appendix A, presented so as to mirror concept presentation in Table 1.

The model will be further discussed below – including those assumptions leading to the partial formulation shown in Figure 1.

The Data

The primary data for the study were sixteen high quality, quad image tapes showing interaction and performance of sixteen three-man flight crews. The crews flew a full mission scenario in a Boeing 720B flight training simulator, a late version of the Boeing 707. Figure 2 shows a typical quad image videotape frame: captain and first officer (upper left and right quadrants, respectively); flight engineer (lower right quadrant); and a context image shot from the back of the simulator that preserved the same relative locations of crewmembers (lower left quadrant). This combined view was made from four small video cameras located in the simulator, out of sight of the crewmembers.

A current, professional air traffic controller was used in the simulation. The controller also participated with another member of the experimental team in simulating conversations with other aircraft, to provide background conversations on the Air Traffic Control (ATC) network.

Crewmembers: The crewmembers were paid volunteers. Their experience represented a wide range of airline of origin and recency, or currency on B-707 line operations. Some were current on the B-707. Many had recent B-707 line experience but were currently flying other jet aircraft in line operations. Some were retired from the line. Thus crew composition ranged from one in which all members were retired from the line to one currently flying the B-707 as an intact crew. The major objective of the overall simulation study was to develop methods for quantifying crew coordination and decisionmaking factors, and their relationships to flight task performance. Thus, this diversity in experience was considered of some importance as an aid in evaluating the sensitivity of candidate performance measures.

All crewmembers received six hours of classroom differences training and four to eight hours of simulator differences training. The number of hours of simulator differences training that a crewmember received was based on recency. Subjects were formed into crews prior to simulator training and were instructed in coordinated procedures during this training.

Scenario: Simply, the overall scenario represented a flight from Tuscon, continuing to Los Angeles (LAX) after a short stopover at Phoenix, with a forced diversion to an alternate upon reaching LAX. Each crew flew the scenario only once, without prior knowledge of the scenario problem. The intent of this procedure was to maximize a valid description of natural crew performance. The crew's enactment of the scenario began with a Captain's Briefing in the simulated operations room at Tuscon and ended upon stopping on the runway at the selected alternate {either Palmdale (PMD) or Ontario (ONT)}. This rating study used videotapes from the longer, problem-leg only - beginning as all three crewmembers entered the cockpit at Phoenix.

The scenario was designed to evoke a series of decisions about where to proceed following a missed approach at LAX due to nose gear not-down-and-locked indication. This situation was exacerbated because it occured at a time when the Los Angeles basin (which includes the planned alternates Ontario and Long Beach) was experiencing low and deteriorating ceiling and visibilities due to coastal fog. Following the missed approach and upon going through a complete gear check procedure that takes several minutes, the crews had to insure that the gear was down and pinned so they could assume that the panel light indication was faulty.

Eight Decisions: While on the ground at Phoenix the crew was given weather information indicating some

degradation at LAX. During the latter part of the cruise to LAX, they would be given direct information and other cues about further deterioration of ceiling and runway visual range (RVR) at LAX. Cues included being given an enroute hold due to traffic back-up and an ATC-net conversation regarding another aircraft's missed approach and return for a second attempt. If weather conditions at possible alternates were requested a crew could realize that conditions at other coastal airports such as Long Beach were similar to LAX; that ONT located inland from LAX, was lagging LAX in deterioration; and that PMD, located just over a mountain range out of the Los Angeles basin, was experiencing clear weather with good visibility. The decisionmaking behavior of the crews with respect to whether unusual contingency planning was required and what that planning should be was the first of eight decision processes to be rated in this Crews were cleared to approach LAX, from hold, when their fuel remaining was 14,000 lbs. This decision process was evaluated at the point of calling for gear down, a short time thereafter and near the outer marker at LAX.

The other decisions concerned: 2) whether to go around on the first approach to LAX - a decision that is essentially procedural in that attempts to recycle the gear did not extinguish the failure indication; 3) whether to reapproach LAX - some crews chose to proceed to an alternate and work the gear problem enroute 4) whether to go around (including early interruption) on the second approach to LAX - somewhat less procedurally based, depending on fuel remaining, for example; 5) the choice of ONT or PMD as an alternate; 6) whether to bring the nose gear up during cruise for fuel conservation; 7) whether to select another alternate after receiving company information on relative weather conditions at PMD and ONT and the companies' preference for ONT for passenger handling; and 8) what arrival status to declare.

The last decision, like the first, is a complex decision or planning process, and has components involving whether to declare an emergency or problem situation (due to the nose gear indication or for low fuel) and whether to request emergency equipment (if an emergency is not declared). The timing of information given by the company for decision process seven was designed to require crews to reconsider their alternate, but to maintain their original decision for a prudent outcome.

The Rating Procedure

Rating Scales and Administration: Figure 3 shows the eight videotape stop points at which decision processes were

rated. At each point, raters marked nine decisionmaking (DM) scales like that presented in figure 4 - each occupying one page of a booklet. A criterion statement shown below a scale defines the underlined modifier in the scale statement at the top. Thus a _very efficient_ decision process (see Figure 4) is one in which "all significant information was acquired at an opportune time, adequately evaluated and appropriately utilized."

The criterion statements tended to anchor scales at the top (7) scale value. During interactive rater training the kinds of outcomes that would merit rating at the other extreme (1) value and/or intermediate values, were discussed. A blackboard beside the video playback unit contained complete rater instructions and a large scale with anchor descriptions for scale numbers two through six: neutrality for four and incrementally equal interval tendencies toward strong agreement or disagreement for the others. Except for decision difficulty and decision quality scales, presented in that order at the beginning of each booklet, scales were presented in different random orders for each of the eight decisions.

Figure 3 also shows that stop points one, five, and eight and the point at which all three crewmembers had entered the cockpit, defined the boundaries of three mission segments: 1) pre-problem, 2) major problem, and 3) secondary problem. At these three stop points, following administration of the DM instrument, the 15-scale crew process (CP) instrument was administered. As contrasted to the DM instrument the CP instrument assessed qualities based on behavior of individual crewmembers, or the crew, throughout each segment. Examples are task- and person- oriented leadership qualities and crew coordination. The 15 scales were presented in different random orders for each segment.

As noted previously,the first videotape stop was made as "gear-down" was called at LAX. The eighth stop was made over the outer marker at the alternate. The other stops were keyed to completions of decision processes - usually signaled by the start of implementation. After stop point eight, the videotape was continued until the aircraft had stopped on the runway. At this time a fourth CP booklet was administered. These ratings on CP variables over the complete operation were made for comparison with average ratings over the three segments.

All 25 scales were identical to that shown in Figure 3 except for those that assessed first officer leadership styles. These contained a n/a position after number seven. N/a (not applicable) was to be circled only if _no_

opportunity arose for leader behavior.

As has been noted, the safety performance (SP) instrument was administered in a session after raters had completed all other ratings on the sixteen crews. The raters were not told that they would provide the SP ratings until the other rating sessions were completed.

Airport, airway, and simulator performance information was available to raters during the rating process. Calculated fuel requirements under emergency flight conditions, for both ideal and less-than-ideal aircraft configurations (e.g. gear down) were made available during rating of safety performance - for go-around at ONT and PMD and for flights between airports. During all rating sessions, the videotape would be stopped at a rater's request to clarify a crewmember utterance or other factual information. The raters took notes throughout the flights, and were particularly encouraged to do so during the cruise from Phoenix to LAX. A monitor was present during all rating sessions to insure independence of ratings.

Scenario conditions (or contextual events) were not entirely consistent over the 16 crews. These inconsistencies as well as how they were dealt with are discussed in (Appendix B).

Raters: The raters were six retired captains, all maintaining professional experience as analysts or researchers with the NASA-Aviation Safety Reporting System. All had experience on the Boeing 720B and/or 707. Their combined airline experience totaled 224 years. Four airlines were represented. Year of retirement ranged from 1978 to 1984.

Rating Design: The raters were formed into two balanced groups of three raters, an A- and B-group, based on ASRS research experience, airline of experience, and recency of retirement. Two raters who worked in proximity to each other at their ASRS position were assigned to different groups - all agreed not to discuss completed ratings with members of the other group.

The A-group rated the videotapes, and hence crews, in order one through 16. This crew identification order represented a randomization of the order that crews performed in the simulator. On each rating day, A-group rated two crews - one in a morning and one in the afternoon. The B-group rated crews in the general order of nine through 16 followed by one through eight - except that morning and afternoon videotapes were reversed. Their actual order was: 10,9,12,11.... The latin square type design provided control

for time of day effects, partial control for unanticipated sequential effects, and the possibility of examining data for such effects.

Rater Training: All raters received three 2-hour initial training sessions. In session one, the DM and CP rating booklets were presented and discussed. Feedback on these instruments was solicited and utilized when appropriate. A lecture was also given on rating theory – discussing, for example, assumptions of multidimensionality of jobs and situations; effectiveness levels, or degrees of qualities within dimensions; the rater as a measuring instrument; rating accuracy; need to reduce errors of halo, leniency, and midpoint cluster; rating skill components; and the desired end result of independent but reliable ratings. During sessions two and three accuracy and error reduction discussions were repeated. Also, DM and CP scales were utilized repeatedly on videotapes made during "shakedown" simulator runs. These were non-data runs made by crews that were not included in the study. Following each rating effort, ratings were posted and discussed. A-group received an added training session prior to starting ratings of crews nine through 16 due to an unplanned 1-week interruption of their rating activity.

RESULTS AND CONCLUSIONS

Inter-rater Reliabilities

Variable reliabilities were computed by use of the Spearman-Brown formula for multiple raters. These multiple r reliabilities are presented in Table 2. The average reliability over all variables is .92.

Variable Means and Standard Deviations

Crew Process Variables: Means and standard deviations were computed for the 15 crew process variables for each of the three segments. The means for each variable were compared with Fisher's "protected t" test. Differences between segments 2 and 3 means were not significant (p > .10). Differences were significant (p < .05) for all variables – except competency-motivation (for P2 and P3), command reversal, and task-oriented leadership (P2) – between the pre-problem segment and each of the problem segments (see Table 3). Thus, performance ratings declined on five of the six leadership behavior variables, on communications quality for all crewmembers, and on the three crew referenced variables

(cohesiveness, coordination, and friendliness) as the diffi-
culty of the scenario was increased. These effects,
apparently due to being in a less structured problem situa-
tion, could have implication for remedial training. Table 3
also presents the overall mean and SD for each of the 15
crew process variables. An identical "t" test procedure
also revealed no differences between an average rating on CP
variables over the three segments and the overall rating
made on the ground at the alternate.

The standard deviations for the crew process variables
across segments were fairly consistent within and between
variables. The range of standard deviations within each of
these variables across the three segments was approximately
equal to the average range of .13.

Decisionmaking Variables: Means for each of the nine
decision variables are presented for each of the eight deci-
sions in Table 4. The overall mean and standard deviation
for each variable is also presented in Table 4.

The standard deviations for the decisionmaking vari-
ables across decision points were fairly consistent within
and between variables. The range of standard deviations
across the eight decision points within each of these vari-
ables was approximately equal to the average range of .41.
Variables 16 and 17 had larger ranges (.72 and 1.01, respec-
tively). Inspection revealed that these larger ranges were
due to the low standard deviations at decision point two
(the first go-around) for these two variables.

Safety Performance: The overall mean for safety perfor-
mance (V25) was 3.53, and the overall standard deviation was
1.85.

Multiple Regression/Correlation Analyses

Correlation Matrix: An interpair correlation matrix was
obtained (n=96; 6 raters x 16 crews) (Table 5). The six
raters made independent ratings. However, in that each rater
rated each of the 16 crews, crews cannot be considered
truly independent. Rather, there is a relative independence
among the 96 points and some bias due to non-independence
had to be accepted. The correlations of Table 5 that are at
or above .267 are significant (p < .01) for 90 degrees of
freedom.

The correlations between variables representing con-
cepts assessed for both crew process and and decisionmaking
are of interest for methodological reasons. Command rever-
sal (V9) is seen to correlate .92 with decision command

reversal (V24). The possibility of combining these two variables for modeling purposes is suggested. The communications quality variables (V13-15) correlated .79, .73, and .78 with their decisionmaking counterparts (V21-23). The participatory leadership variables (V5,V6) correlated .83 and .77 with their decisionmaking counterparts (V19,V20).

Crew Cohesiveness (V10) and Crew Coordination (V11) are significantly correlated ($r=.93$). There was evidence that a few raters had some difficulty in distinguishing these two variables conceptually - and it is easy to conceive of difficulty in distinguishing them operationally. The expressed problem was in separating crew cohesiveness conceptually from crew coordination - not in rating crew coordination. For the prior reasons the high correlation may be, in part, an artifact.

Regression Analyses: The partial model in Figure 1 is based on a set of assumptions that would determine the order for entering variables into a hierarchical regression analysis. These assumptions include the usual assumptions for establishing causal priority. Also, based on the hierarchical command structure it is assumed that captain leadership and communications qualities would have larger effects than those of other crewmembers. This assumption accounts for captain quality variables, and not other crewmember variables, being included in the front-end, partial model. The rationale for including a task-oriented leadership variable rather than person-oriented or participatory leadership variables was derived from some evidence that a task-oriented leadership style is more effective in problem situations. The curved line between V13 and V21 of Figure 1 indicates correlation but implies no causal relationship, as the directed, signed lines do.

The analytic approach chosen was a hierarchical procedure initiated by a series of stepwise regressions, each subsequent procedure including decreasing numbers of the variables shown in Figure 1. That is, all of the variables except V13 {communications quality(P1)}, which was included only in the last of the five regressions in the series. This was done to reduce the ratio of the k (independent variables - IV's) to the n (96) for the first four regressions. Although these k/n ratios exceed what may be considered prudent for substantive findings, the exploratory nature and predominately predictive interest here, as well as use of an a priori hierarchical model for entry of initial variables is argued to justify the procedure. Through this procedure, k is restricted to small values relative to the large number of possible IVs.

The series of five stepwise regressions were performed in reverse order to that shown in Table 6. Table 6 indicates the dependent variables for these regressions as I) command reversal, II) crew coordination, III) decision efficiency, IV) decision quality, and V) safety performance.

Variables in the partial model that precede a particular dependent variable were the IVs for that regression. Entering order for the IVs were top-down from left to right. The stepwise regression program employed was BMDP2R (BMDP Statistical Software, 1983) with a minimum acceptable F value (to enter) of 4.00 ($p < .05$) and a maximum acceptable F value (to remove) of 3.90.

Following each of the five basic regression procedures, the partial correlation table for variables not in the equation was consulted to determine which F value to enter above 4, if any. If such a variable was present and logically prior to the dependent variable, another regression, adding this variable, was performed. This procedure was continued until no logically prior variables had an F value above 4 to enter.

Significant IVs were then entered into a final regression by a usual hierarchical procedure – if precedences were strictly established by the model. If not, some alternative paths were usually considered. Regression results are discussed in the order performed. Summary analyses are presented in Tables 6 and 7.

Safety Performance is seen (Table 7) to have 46% of the variance explained. Command reversal – a negatively correlated IV contributes 16%. Decision efficiency contributes 28%. Decision quality contributes 3%.

That decision quality only increments the variance explained by 3% could be due to two considerations. First, the definition of decision efficiency stops not too far short of including decision quality. The second and perhaps most important consideration is based on defining criteria for decision quality versus safety performance and is discussed below.

Decision Quality, considered an alternative substantive dependent variable to safety performance, is seen (Table 7) to have 60% of the variance explained by communications quality (P1) (12%) and decision efficiency (48%).

The suggested rationale for the relative percentages of variance explained for decision quality and safety performance is based on defining criteria for the two variables.

23.13

Defining criteria for decision quality were presented at the beginning of the rating effort and considered whether choices were most appropriate based on "safety of flight and/or the attainment of other mission goals." Safety performance ratings were based on the level of safety achieved (or risk avoided) in the end solution for the major scenario problem – or in unsafe attempts to solve the problem by landing at LAX – and criteria were presented only after all the other ratings had been made. Safety performance defining criteria thus excluded any consideration of attaining "other mission goals" – for example that ONT was preferable for passenger handling and was the designated alternate.

Decision Efficiency is seen (Table 7) to have 78% of the variance explained. The explanatory variables and suggested increments in variance explained are decision communications quality (P3) (17%), decision communications quality (P1) (33%), decision command reversal (8%), command reversal (5%), and crew coordination (14%).

This is the first stepwise procedure to be continued beyond the basic regression. Table 6 shows that, prior to these continuations 74% of the variance in decision efficiency was explained by the three variables of the formal partial model – decision communication quality (P1), crew coordination and decision command reversal. The added variables for the entering orders shown, contributed 2% increments of variance respectively. Table 7 however shows that if the flight engineer's decision communications quality were considered logically prior to the other significant variables, it contributes 17% of the variance. The flight engineer's communications concerning the nose gear and fuel is suggested to explain the significance of this variable for decision efficiency.

For most purposes, as will be further discussed below, V9 (command reversal) and V24 (decision command reversal) can be considered to be synonymous – or V24 can be used to assess the generic concept of command reversal, as the model of Figure 1 suggests. In explaining decision efficiency however, V9 increments the variance explained by 5%. Decision command reversal is shown logically prior to command reversal in Figure 7 on the basis of the substantive dependent variables being decision based.

Crew coordination is seen by Table 7 to have 81% of the variance explained. The order of explanatory variables and suggested increments in variance explained is essentially arbitrary. Perhaps the major implication from the regressions in Tables 6 and 7 is that crew coordination has most of its variance explained by a quality variable for each of

the three crewmembers: Communications quality (P1), competency-motivation (P2), and competency-motivation (P3). Comparing the regressions in Tables 6 and 7 also indicates that the effect of the communications quality variable essentially nullifies that of task-oriented leadership (P1). For reasons discussed previously, crew cohesiveness, highly correlated with crew coordination, was excluded from entering the stepwise regression equations.

Decision command reversal is shown by Table 6 to have 60% of the variance explained by five variables. In decreasing order of contribution, for the variable order shown, the IVs are task-oriented leadership (P2), decision difficulty, task-oriented leadership (P1), communications quality (P1), and decision participatory leadership (P1). The alternative order shown in the hierarchical procedure in Table 7 has no less an arbitrary variable order than the stepwise procedure of Table 6 and does not contradict the major suggestions from Table 6: The behavioral variables of most importance are the task-oriented leadership variables for the first officer and captain, correlated positively and negatively respectively with command reversal, and accounting for about 28% and 15% increments in variance explained respectively. Considering also the 10% increment of variance explained by decision difficulty, the suggestion is that a combination of low task-oriented captain leader behavior with high task-oriented first officer behavior fosters command reversal, and that this is particularly so as crews get themselves into difficult situations. The other two IVs together account for an 8% increment in variance explained; five percent of which is attributable to the communications quality of the captain, also negatively correlated with command reversal.

As mentioned above, decision command reversal (V24) correlated significantly (.92) with a more general leadership-associated measure of command reversal (V9). Some analyses including a case in which V24 was an independent and a dependent variable, were repeated with V9 substituted for V24. For example, repetition of the preceding stepwise regression, with V9 substituted for V24, produced similar results; All the variance explained was attributable to the first officers' and captains' task-oriented leadership behaviors and decision difficulty: about 35%, 11%, and 5% increments in multiple R squared respectively. A repetition of hierarchical Regression V of Table 7 also produced similar results with V9 substituted for V24. V24 in conjunction with decision efficiency and decision quality explained 46% of the variance in safety performance as opposed to 45% when V9 was substituted -- with 16% contributed by V24 as opposed to 9% contributed by V9. The overall evidence is that V9

and V24 differ little as currently defined and measured.

Although it may be appropriate in this study to consider these two variables as essentially synonymous, such is not necessarily recommended for future work - and improvements in definition may be appropriate.

A summary suggestion from the above series of regression analyses is that major differences for coordination versus decisionmaking pathways in the model are associated with the effect of command reversal for decisionmaking and competency-motivation of the first officer and flight engineer for coordination.

Factor Analysis

A factor analysis with principal components extraction and varimax rotation was performed using BMDP4M (BMDP Statistical Software, 1983). The number of factors was limited to the number of eigenvalues greater than one. Rotated factor loadings for the five orthogonal factors and variance explained by each factor are shown in Table 8.

The highest loadings for factor one (ranging from .844 to .923) clearly cluster captain participatory and person-oriented leadership variables, and captain communications quality variables. Crew cohesiveness, crew coordination, and decision efficiency loadings group separately and load on factor one (.590 to .641) but also load substantially on factors two, three, and four. Task-oriented leadership (P1) is clearly grouped with these three crew variables and also loads substantially on factor 4.

Factor two shows a similar pattern to factor one in that the highest loadings clearly cluster first officer participatory and person-oriented leadership variables, and first officer communications quality variables. Competency-motivation (P2) and task-oriented leadership (P2) group together somewhat separately from the above cluster.

Factor three shows a similar pattern, loading most heavily for flight engineer communications quality and the competency-motivation variables.

Factor four has its highest loadings on the command reversal variables (.842 and .727). The only other positive loading (.619) is for decision difficulty. The important negative loadings are for the substantive dependent measures, flight safety and decision quality; the intermediate performance measures, crew coordination and decision efficiency; and task-oriented leadership (P1).

Factor five loads most heavily on crew friendliness (.574) followed by task-oriented leader behavior of the first officer (.390) and captain (.369), and the negative loading for the first officer's participatory leader behavior in decisionmaking (-.312). It is also of some interest that crew friendliness loads (.522) on factor one – thus at about the same level as does task-oriented leadership (P1).

The outcome of the factor analysis tends to confirm the model structure suggested by regression analysis; adding knowledge concerning the clustering of communications quality variables and other-than-task-oriented leadership variables; and identifying a factor that loads most heavily on crew friendliness. Further factor analyses, and regression analyses using factors, may be appropriate.

Crew Differences

Figure 5, graphically displays the means and SDs for the 16 crews on each of the variables of the partial model – and also for task-oriented leadership (P2). The crew presentation order is on the basis of ratings on safety performance. Table 9 indicates where each crew landed and the amount of fuel on board at landing.

The following two conversations are presented only as examples of communications related to the crew coordination process and to decisionmaking. They are from the crews rated highest and lowest on safety performance (numbers 13 and eight), and exemplify successful coordination and unsuccessful decisionmaking, respectively.

> Captain: Don't forget to fly the airplane
>   First officer: Yeah, I am (sounding
>   slightly defensive).
>   - pause -
> First officer: Everything's under control.
> Captain: That's your main responsibility.
> First Officer: Yep.

An exchange like the above occured six times between the captain and first officer of crew 13 during this flight leg. This particular clarification of responsibility occured while both the captain and first officer were copying weather.

> Captain: (On approach to LAX) We're going to land
>   this way. Tell him to get the fire trucks,

we're probably going to smear the nose wheel.
    First officer: (To ATC) We-ah-could have a problem
                with our nose gear, so be aware of that
                please.
    Captain: (continues approach)
    First officer: We don't have the green light on our
                nose gear. Ah we'll continue the approach.
    Captain: If that 12K (referring to 12,000 lbs of
                fuel) is realistic, we don't need to smear
                this thing.

This conversation occurred within crew eight, a crew for
which members appeared to function very independently of
each other (see their low coordination rating on Figure 5).
In this conversation the last stated insight of the captain
received no support from other crewmembers and the aircraft
was landed unsafely at LAX.

Concluding Statement

        A start has been made on the development of a model of
crew coordination and decisionmaking. The inclusion of deci-
sion efficiency and command reversal as variables in the
model appear to be useful advances in conceptualizing the
crew performance process. Some insight was suggested on the
dynamics of command reversal: Relatively low and high task-
oriented leader behavior of the captain and first officer
respectively - especially as the situation becomes difficult
appears to be the leading impetus to the occurrence of com-
mand reversal. Task-oriented leadership thus appears to be
distinguished some from person-oriented or participatory
leadership. The latter leader behavior variables are shown
by factor analysis to cluster closely with communications
quality variables.

        Both regression and factor analyses suggest the impor-
tant effects of all three crewmember behavioral qualities on
crew coordination and, through different pathways, on deci-
sion efficiency. Both analyses also suggest the important
effects of command reversal in decision pathways - on deci-
sion efficiency and quality variables, and on safety perfor-
mance. As outcome variables decision quality is seen to have
more variance explained (60%) than did safety performance
(46%). The reason is suggested to be that the defining cri-
teria for decision quality ratings are more inclusive, con-
sidering not only safety of flight but also the attainment
of other mission goals - and were available to raters
throughout the rating effort.

        Significant decreases in the ratings of leader
behaviors and communications qualities occurred for both the

captain and first officer, for the major problem segment as opposed to the pre-problem (take off and cruise to LAX) segment. Similar decreases occured for crew coordination, cohesiveness, and friendliness - but not for command reversal and competency-motivation of the first officer and flight engineer. It is perhaps significant for command reversal that the only leader behavior not rated lower on the major problem (as opposed to the pre-problem) segment was the first officers' task-oriented leadership.

One of the five orthogonal factors loaded most heavily for crew friendliness. However, the validity of crew cohesiveness as a substantial measure as defined and used in this study was questioned.

The rating of videotapes appear to have considerable promise in developing crew performance models. A suggested improvement on this study is the inclusion of a display of systems information, such as airspeed, altitude, fuel remaining, etc., adjacent to the video display. Such a display should reduce error variance in crew process ratings, and permit addition to the model of a variable, paralleling decision quality, assessing flight task execution quality.

Considerable reduction in error variance should also be realized through refinement of variables and/or their defining criteria, through improvements in rater training made possible by these videotapes, and through improvements in data generation and rating procedures.

References

Cooper, G. E., & White, M. D., & Lauber, J. K. (Eds.).
(1980). Resource management on the flight deck.
Proceedings of a NASA/Industry Workshop. Moffett Field,
Ca.: NASA Ames Research Center.

Dixon, W.J., & Brown, M. B., & Engelman, L., & Frane, J. W.,
& Hill, M. A., & Jennrich. R. I, & Toporek, J. D. (Eds.).
(1983). BMDP Statistical Software (rev. ed.). Berkely:
University of California Press.

Foushee, H. C. & Manos, K. L. (1981). Information transfer
within the cockpit: Problems in intracockpit
communications. In C. E. Billings & E. S. Cheaney (Eds.)
Information transfer problems in the aviation system.
(Report No. NASA TP-1875). Washington, DC: National
Aeronautics and Space Adminstration.

Frankel, R. M., & Beckman, H. B. (1982). Impact: an
interaction-based method for preserving and analyzing
clinical transactions. In L.S. Pettegrew (Ed.),
Explorations in provider and patient interaction (pp. 71-
85). Humana Inc.

Goguen, J. Linde, C., & Murphy, M. (1984) Crew communication
as a factor in aviation accidents. In E J. Hartzell & S.
Hart (Eds.), Papers from the 20th Annual Conference on
Manual Control, 2, (pp. 217-248). Moffett Field, CA: NASA,
Ames Research Center.

Hackman, J. R. & Morris, C. G. (1975). Group tasks, group
interaction process, and group performance effectiveness:
A review and proposed integration. In L. Berkowitz (Ed.),
Advances in experimental social psychology (Vol. 8, pp
45-99). New York: Academic Press.

House, R. J. (1984). Commentary on management research. In
A. P. Brief (Ed.), Productivity research in the behavioral
and social sciences (pp. 268-282). New York: Praeger.

House, R. J., & Baetz, M. L. (1979). Leadership: Some
empirical generalizations and new research directions.
Research and organizational behavior (Vol. 1, pp. 341-
423). JAI Press Inc.

Kerr, S. (1984). Leadership and participation. In A. P.
Brief (Ed.), Productivity research in the behavioral and
social sciences (pp. 229-251). New York: Praeger.

Landy, F. J., & Farr, J. L. (1980). Performance rating.

    <u>Psychological Bulletin</u>, <u>87</u>, 72-107.

Landy, F. J., & Farr, J. L. (1983). <u>The measurement of work
    performance</u>. New York: Academic Press.

Murphy, M. R. (1977). Coordinated crew performance in
    commercial aircraft operations. <u>Proceedings of the 21st
    Human Factors Society Annual Meeting</u>, 416-420.

Murphy, M. R. (1980). Analysis of eighty-four commercial
    aviation incidents: implications for a resource management
    approach to crew training. <u>1980 Proceedings Annual
    Reliablility and Maintainability Symposium</u>, 298-306.

National Transportation Safety Board, Special Study. (1976).
    <u>Flight crew coordination procedures in air carrier
    instrument landing system approach accidents</u> (Report No.
    NTSB-AAS-76-5). Washington, DC: National Transportation
    and Safety Board.

Ruffell Smith, H.P. (1979). <u>A simulator study of the
    interaction of pilot workload with errors, vigilance, and
    decisions</u> (Report No. NASA TM-78482). Washington DC:
    National Aeronautics and Space Adminstration.

Table 1

Concept-Variable Relationships

	Concept	*Referent	Var. No.	**Instrument
a.	Task-Oriented Leadership	P1,P2	1,3	CP
b.	Person-Oriented Leadership	P1,P2	2,4	CP
c.	Participatory Leadership	P1,P2	5,6	CP
			19,20	DM
d.	Competency-Motivation	P2,P3	7,8	CP
e.	Command Reversal	P1-P2	9	CP
			24	DM
f.	Crew Cohesiveness	Crew	10	CP
g.	Crew Coordination	Crew-Task	11	CP
h.	Crew Friendliness	Crew	12	CP
i.	Communications Quality	P1,P2,P3	13,14,15	CP
			21,22,23	DM
j.	Decision Difficulty	Situation	16	DM
k.	Decision Quality	Crew Outcome	17	DM
l.	Decision Efficiency	Crew-Task	18	DM
m.	Safety Performance	Crew Outcome	25	SP

*P1 = Captain, P2 = First Officer, P3 = Flight Engineer
**CP = Crew Process, DM = Decision Making, SP = Safety Performance

Table 2

Variable Reliabilities Determined by Spearman-Brown Formula for Multiple r

No.	r	No.	r	No.	r	No.	r	No.	r
1	.91	6	.87	11	.96	16	.88	21	.89
2	.93	7	.93	12	.90	17	.93	22	.86
3	.78	8	.96	13	.94	18	.94	23	.92
4	.90	9	.96	14	.92	19	.88	24	.95
5	.92	10	.96	15	.96	20	.84	25	.99

Table 3

Changes in Crew Process Variable Means Across Segments

No.	Name	Means for Segments			‾Decrease		Overall	
		1	2	3	1 to 2	1 to 3	Mean	SD
1	Task-Oriented Leadership (P1)	5.08	4.71	4.56	*.37	**.52	4.79	1.28
2	Person-Oriented Leadership (P1)	4.48	3.93	3.91	**.55	**.57	4.11	1.32
3	Task-Oriented Leadership (P2)	4.97	4.91	4.86	ns	ns	4.92	1.01
4	Person-Oriented Leadership (P2)	4.72	4.34	4.30	*.38	**.42	4.46	1.09
5	Participatory Leadership (P1)	4.71	3.91	3.84	**.80	**.87	4.16	1.47
6	Participatory Leadership (P2)	4.91	4.13	4.38	**.78	**.53	4.48	1.18
7	Competency-Motivation (P2)	5.10	4.87	4.98	ns	ns	4.99	1.29
8	Competency-Motivation (P3)	4.50	4.15	4.42	ns	ns	4.36	1.34
9	Command Reversal	3.60	3.86	3.81	ns	ns	3.76	1.59
10	Crew Cohesiveness	4.56	3.89	4.02	**.67	*.54	4.16	1.56
11	Crew Coordination	4.49	3.86	3.83	**.63	**.66	4.07	1.57
12	Crew Friendliness	4.83	4.35	4.49	**.48	*.34	4.55	1.03
13	Communications Quality (P1)	4.82	4.17	4.06	**.65	**.76	4.36	1.51
14	Communications Quality (P2)	4.97	4.53	4.53	*.44	*.44	4.68	1.23
15	Communications Quality (P3)	4.53	4.14	4.44	ns	ns	4.37	1.19

*Difference significant (p<.05)
**Difference significant (p<.01)
‾Differences between segments 2 and 3 means were not significant (p>.10)

Table 4

Means of Decisionmaking Variables for the Eight Decisions

No.	Name	Decision Number								Overall	
		1 Mean	2 Mean	3 Mean	4 Mean	5 Mean	6 Mean	7 Mean	8 Mean	Mean	SD
16	Decision Difficulty	3.11	2.21	4.23	3.24	3.89	3.41	2.49	2.54	3.16	1.45
17	Decision Quality	4.54	6.11	3.97	5.10	4.71	4.77	5.65	4.29	4.86	1.65
18	Decision Efficiency	3.84	4.52	3.18	3.32	3.18	3.43	4.38	3.48	3.65	1.72
19	Decision Participatory Leadership(P1)	4.65	4.05	3.78	3.75	3.77	3.55	3.63	3.53	3.86	1.42
20	Decision Participatory Leadership(P2)	4.65	4.24	4.26	4.11	4.20	3.75	3.70	4.27	4.20	1.25
21	Decision Communications Quality(P1)	4.52	4.50	3.85	3.96	4.01	3.97	3.97	3.77	4.08	1.42
22	Decision Communications Quality(P2)	4.83	4.44	4.23	4.18	4.39	3.88	4.22	4.36	4.34	1.22
23	Decision Communications Quality(P3)	4.26	4.19	4.00	3.74	3.81	4.04	3.92	4.07	4.01	1.21
24	Decision Command Reversal	3.53	3.36	3.58	3.78	3.69	3.05	2.82	3.54	3.45	1.56

Table 5

Correlation Matrix

		1	2	3	4	5	6	7
1	Task-Oriented Leadership (P1)	1.0000	0.4768	0.2226	0.1267	0.4425	0.0883	0.2476
2	Person-Oriented Leadership (P1)	0.4768	1.0000	0.2164	0.1652	0.8170	0.1622	0.3990
3	Task-Oriented Leadership (P2)	0.2226	0.2164	1.0000	0.4983	0.1629	0.6470	0.6412
4	Person-Oriented Leadership (P2)	0.1267	0.1652	0.4983	1.0000	0.1857	0.6880	0.5766
5	Participatory Leadership (P1)	0.4425	0.8170	0.1629	0.1857	1.0000	0.2756	0.4206
6	Participatory Leadership (P2)	0.0883	0.1622	0.6470	0.6880	0.2756	1.0000	0.6969
7	Competency-Motivation (P2)	0.2476	0.3990	0.6412	0.5766	0.4206	0.6969	1.0000
8	Competency-Motivation (P3)	0.2311	0.2353	0.5283	0.4028	0.2783	0.4666	0.6182
9	Command Reversal	-0.3241	-0.1315	0.5005	0.2427	-0.0368	0.4061	0.2668
10	Crew Cohesiveness	0.5528	0.6368	0.4442	0.4076	0.6549	0.4730	0.7436
11	Crew Coordination	0.5453	0.6145	0.4031	0.3523	0.6693	0.4490	0.6016
12	Crew Friendliness	0.5144	0.6676	0.4290	0.5288	0.5224	0.3034	0.4589
13	Communications Quality (P1)	0.5785	0.7998	0.1434	0.1581	0.7992	0.2367	0.4499
14	Communications Quality (P2)	0.2790	0.3028	0.6137	0.6432	0.3012	0.7007	0.8098
15	Communications Quality (P3)	0.1310	0.1635	0.4760	0.3340	0.1988	0.3686	0.4803
16	Decision Difficulty	-0.2881	-0.0472	-0.1302	-0.0734	0.0171	-0.0074	-0.0581
17	Decision Quality	0.3888	0.2913	0.1722	0.1186	0.3210	0.2309	0.3343
18	Decision Efficiency	0.5639	0.5157	0.2518	0.2533	0.5957	0.3701	0.5638
19	Decision Participatory Leadership (P1)	0.3334	0.6917	0.1798	0.2087	0.8276	0.3105	0.3989
20	Decision Participatory Leadership (P2)	-0.0892	0.0716	0.5180	0.6101	0.2010	0.7667	0.5593
21	Decision Communications Quality (P1)	0.4759	0.7191	0.1516	0.1716	0.7559	0.3080	0.4683
22	Decision Communications Quality (P2)	0.0263	0.2107	0.5025	0.5399	0.2720	0.7214	0.6524
23	Decision Communications Quality (P3)	-0.0085	0.1907	0.3125	0.2250	0.2330	0.3393	0.4156
24	Decision Command Reversal	-0.3853	-0.0992	0.4257	0.1760	-0.0242	0.2978	0.2035
25	Safety Performance	0.4094	0.2673	0.0935	0.1517	0.2988	0.1998	0.2358

	8	9	10	11	12	13	14	15	16
1	0.2311	-0.3241	0.5528	0.5453	0.5144	0.5785	0.2790	0.1310	-0.2881
2	0.2353	-0.1315	0.6368	0.6145	0.6676	0.7998	0.3028	0.1635	-0.0472
3	0.5283	0.5005	0.4442	0.4031	0.4290	0.1434	0.6137	0.4760	-0.1302
4	0.4028	0.2427	0.4076	0.3523	0.5288	0.1581	0.6432	0.3340	-0.0734
5	0.2783	-0.0368	0.6549	0.6693	0.5224	0.7992	0.3012	0.1988	0.0171
6	0.4666	0.4061	0.4730	0.4490	0.3034	0.2367	0.7007	0.3686	-0.0074
7	0.6182	0.2668	0.7436	0.6816	0.4509	0.4499	0.8098	0.4803	-0.0581
8	1.0000	0.1852	0.6868	0.6674	0.3927	0.2925	0.5268	0.8686	-0.1803
9	0.1852	1.0000	-0.1056	-0.0744	-0.0973	-0.2549	0.1687	0.2174	0.2703
10	0.6868	-0.1056	1.0000	0.9337	0.6295	0.7539	0.6731	0.5342	-0.2410
11	0.6674	-0.0744	0.9337	1.0000	0.5603	0.7427	0.6035	0.5090	-0.2232
12	0.3927	-0.0973	0.6295	0.5603	1.0000	0.5777	0.5048	0.3409	-0.2322
13	0.2925	-0.2549	0.7539	0.7427	0.5777	1.0000	0.4069	0.1687	-0.0831
14	0.5268	0.1687	0.6731	0.6035	0.5048	0.4069	1.0000	0.4035	-0.1853
15	0.8686	0.2174	0.5342	0.5090	0.3409	0.1687	0.4035	1.0000	-0.0999
16	-0.1803	0.2703	-0.2410	-0.2232	-0.2322	-0.0831	-0.1853	-0.0999	1.0000
17	0.3028	-0.1437	0.4780	0.5461	0.2237	0.3488	0.2307	0.3937	-0.2500
18	0.5054	-0.1928	0.7727	0.8206	0.4375	0.6528	0.4606	0.4115	-0.2305
19	0.2867	0.0970	0.5673	0.5362	0.4385	0.6437	0.3062	0.2471	0.0185
20	0.4559	0.4838	0.3379	0.2861	0.1641	0.1188	0.5573	0.4000	0.0772
21	0.2915	-0.1446	0.6917	0.6836	0.4168	0.7938	0.4040	0.1928	-0.1462
22	0.4775	0.3584	0.4748	0.4160	0.2638	0.2916	0.7287	0.4394	-0.0189
23	0.6974	0.2541	0.4362	0.4025	0.1611	0.1368	0.2508	0.7797	0.0571
24	0.1534	0.9198	-0.1259	-0.1286	-0.1112	-0.2500	0.0934	0.1851	0.3729
25	0.2707	-0.3049	0.4166	0.4589	0.2828	0.2891	0.1591	0.2059	-0.3415

	17	18	19	20	21	22	23	24	25
1	0.3088	0.5639	0.3334	-0.0892	0.4759	0.0263	-0.0085	-0.3853	0.4094
2	0.2913	0.5157	0.6917	0.0716	0.7191	0.2107	0.1907	-0.0992	0.2673
3	0.1722	0.2518	0.1798	0.5180	0.1516	0.5025	0.3125	0.4257	0.0935
4	0.1186	0.2533	0.2087	0.6101	0.1716	0.5399	0.2250	0.1760	0.1517
5	0.3210	0.5957	0.8276	0.2010	0.7559	0.2720	0.2330	-0.0242	0.2988
6	0.2309	0.3701	0.3105	0.7667	0.3080	0.7214	0.3393	0.2978	0.1998
7	0.3343	0.5638	0.3989	0.5593	0.4683	0.6524	0.4156	0.2035	0.2358
8	0.3028	0.5054	0.2867	0.4559	0.2915	0.4775	0.6974	0.1534	0.2707
9	-0.1437	-0.1928	0.0970	0.4838	-0.1446	0.3584	0.2541	0.9198	-0.3049
10	0.4780	0.7727	0.5673	0.3379	0.6917	0.4748	0.4362	-0.1259	0.4166
11	0.5461	0.8206	0.5362	0.2861	0.6836	0.4160	0.4025	-0.1286	0.4589
12	0.2237	0.4375	0.4385	0.1641	0.4168	0.2638	0.1611	-0.1112	0.2828
13	0.3488	0.6528	0.6437	0.1188	0.7938	0.2916	0.1368	-0.2500	0.2891
14	0.2307	0.4606	0.3062	-0.5573	0.4040	0.7287	0.2508	0.0934	0.1591
15	0.3937	0.4115	0.2471	0.4000	0.1928	0.4394	0.7797	0.1851	0.2059
16	-0.2500	-0.2305	0.0185	0.0772	-0.1462	-0.0189	0.0571	0.3729	-0.3415
17	1.0000	0.7533	0.3305	0.1409	0.4453	0.2104	0.3852	-0.2444	0.5795
18	0.7533	1.0000	0.5490	0.2096	0.6876	0.3484	0.4180	-0.2963	0.6204
19	0.3305	0.5490	1.0000	0.4071	0.8041	0.4429	0.4047	0.1368	0.2742
20	0.1409	0.2096	0.4071	1.0000	0.3301	0.8384	0.4975	0.4817	0.1215
21	0.4453	0.6876	0.8041	0.3301	1.0000	0.4762	0.3554	-0.1437	0.4184
22	0.2104	0.3484	0.4429	0.8384	0.4762	1.0000	0.5020	0.3382	0.0959
23	0.3052	0.4180	0.4047	0.4975	0.3554	0.5020	1.0000	0.2871	0.2649
24	-0.2444	-0.2963	0.1368	0.4817	-0.1437	0.3382	0.2871	1.0000	-0.3979
25	0.5795	0.6204	0.2742	0.1215	0.4184	0.0959	0.2649	-0.3979	1.0000

Correlations at or above .267 are significant for n=90 df.

Table 6

Stepwise Regression Analysis Summaries

Step	Variable Entered	Variable Removed	Multiple R	RSQ	Change in RSQ
I DECISION COMMAND REVERSAL (V24)					
1	13.Communications Quality(P1)		.2500	.0625	.0625
2	1.Task-Oriented Leadership(P1)		.3867	.1496	.0871
3		13	.3853	.1485	-.0011
4	3.Task-Oriented Leadership(P2)		.6509	.4237	.2752
5	16.Decision Difficulty		.7216	.5207	.0970
6	19.Decision Participatory Leadership(P1)		.7441	.5537	.0330
7	13.Communications Quality(P1)		.7774	.6043	.0506
II CREW COORDINATION (V11)					
1	1.Task-Oriented Leadership(P1)		.5453	.2973	.2973
2	21.Decision Communications Quality(P1)		.7279	.5299	.2325
3	7.Competency-Motivation(P2)		.8313	.6910	.1611
4	8.Competency-Motivation(P3)		.8794	.7734	.0824
5	24.Decision Command Reversal		.8857	.7844	.0110
6	13.Communications Quality(P1)		.9053	.8196	.0352
7		24	.9024	.8144	-.0053
8		21	.8991	.8083	-.0060
III DECISION EFFICIENCY (V18)					
1	1.Task-Oriented Leadership(P1)		.5639	.3180	.3180
2	21.Decision Communications Quality(P1)		.7384	.5452	.2272
3	11.Crew Coordination		.8461	.7158	.1706
4	24.Decision Command Reversal		.8591	.7380	.0221
5		1	.8577	.7357	-.0023
6	23.Decision Communications Quality(P3)		.8720	.7604	.0247
7	9.Command Reversal		.8849	.7831	.0228
IV DECISION QUALITY (V17)					
1	1.Task-Oriented Leadership(P1)		.3888	.1512	.1512
2	21.Decision Communications Quality(P1)		.4886	.2387	.0875
3	11.Crew Coordination		.5627	.3167	.0779
4		21	.5568	.3101	-.0066
5		1	.5461	.2983	-.0118
6	24.Decision Command Reversal		.5737	.3291	.0308
7	18.Decision Efficiency		.7638	.5834	.2543
8		11	.7536	.5680	-.0154
9		24	.7533	.5675	-.0005
10	13.Communications Quality(P1)		.7766	.6031	.0356
V SAFETY PERFORMANCE (V25)					
1	1.Task-Oriented Leadership(P1)		.4094	.1676	.1676
2	21.Decision Communications Quality(P1)		.4819	.2323	.0646
3	24.Decision Command Reversal		.5539	.3068	.0745
4		1	.5400	.2915	-.0152
5	11.Crew Coordination		.5841	.3412	.0496
6		21	.5721	.3274	-.0138
7	18.Decision Efficiency		.6610	.4369	.1095
8		11	.6597	.4352	-.0017
9	17.Decision Quality		.6795	.4617	.0266

23.25

Table 7

Hierarchical Regression Analysis Summaries

		Multiple		Change
Variable Entered		R	RSQ	in RSQ
I DECISION COMMAND REVERSAL (V24)				
16. Decision Difficulty		.3729	.1391	.1391
19. Decision Participatory Leadership(P1)		.3949	.1559	.0169
3. Task-Oriented Leadership(P2)		.6080	.3697	.2137
13. Communications Quality(P1)		.7323	.5363	.1660
1 .Task-Oriented Leadership(P1)		.7774	.6043	.0680
II CREW COORDINATION (V11)				
13. Communications Quality(P1)		.7427	.5516	.5516
1. Task-Oriented Leadership(P1)		.7561	.5717	.0201
7. Competency-Motivation(P2)		.8515	.7251	.1534
8. Competency-Motivation(P3)		.8991	.8083	.0833
III DECISION EFFICIENCY (V18)				
23. Decision Communications Quality(P3)		.4180	.1747	.1747
21. Decision Communications Quality(P1)		.7122	.5073	.3326
24. Decision Command Reversal		.7680	.5898	.0825
9. Command Reversal		.8012	.6419	.0520
11. Crew Coordination		.8849	.7831	.1412
IV DECISION QUALITY (V17)				
13. Communications Quality(P1)		.3488	.1216	.1216
18. Decision Efficiency		.7766	.6031	.4814
V SAFETY PERFORMANCE (V25)				
24. Decision Command Reversal		.3979	.1583	.1583
18. Decision Efficiency		.6597	.4352	.2769
17. Decision Quality		.6795	.4617	.0266

Table 8

Factor Analysis Using Principal Components and Varimax Rotation

VARIABLES	FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4	FACTOR 5
5. Participatory Leadership (P1)	0.923	0.	0.	0.	0.
2. Person-Oriented Leadership (P1)	0.888	0.	0.	0.	0.
13. Communications Quality (P1)	0.882	0.	0.	0.	0.
21. Decision Communications Quality (P1)	0.846	0.	0.	0.	0.
19. Decision Participatory Leadership (P1)	0.844	0.	0.	0.	0.
10. Crew Cohesiveness	0.641	0.393	0.409	-0.306	0.252
11. Crew Coordination	0.639	0.333	0.430	-0.327	0.
18. Decision Efficiency	0.590	0.	0.406	-0.517	0.
1. Task-Oriented Leadership (P1)	0.505	0.	0.	-0.479	0.369
6. Participatory Leadership (P2)	0.	0.882	0.	0.	0.
20. Decision Participatory Leadership (P2)	0..	0.812	0.	0.	-0.312
14. Communications Quality (P2)	0.	0.812	0.	0.	0.
22. Decision Communications Quality (P2)	0.	0.801	0.	0.	0.
4. Person-Oriented Leadership (P2)	0.	0.794	0.	0.	0.
7. Competency-Motivation (P2)	0.356	0.711	0.324	0.	0.
3. Task-Oriented Leadership (P2)	0.	0.657	0.341	0.	0.390
15. Communications Quality (P3)	0.	0.255	0.889	0.	0.
23. Decision Communications Quality (P3)	0.	0.	0.836	0.	0.
8. Competency-Motivation (P3)	0.	0.368	0.814	0.	0.
24. Decision Command Reversal	0.	0.315	0.	0.842	0.
9. Command Reversal	0.	0.408	0.	0.727	0.
25. Safety Performance	0.254	0.	0.258	-0.671	0.
16. Decision Difficulty	0.	0.	0.	0.619	0.
17. Decision Quality	0.300	0.	0.482	-0.528	0.
12. Crew Friendliness	0.525	0.308	0.	0.	0.574
VARIANCE EXPLAINED	6.083	5.393	3.513	3.304	1.483

Loadings less than 0.2500 have been replaced by zero

Table 9

Airport of Landing and Fuel Remaining at Landing

Crew No.	Airport	*Fuel Klbs	Crew No.	Airport	*Fuel Klbs
1	PMD	6.2	9	PMD	2.9
2	PMD	7.9	10	ONT	2.0
3	ONT	3.5	11	PMD	7.5
4	ONT	4.6	12	PMD	8.9
5	ONT	3.4	13	PMD	7.7
6	PMD	6.0	14	PMD	2.8
7	ONT	4.0	15	PMD	9.8
8	LAX	8.8	16	PMD	5.4

* Accuracy of fuel remaining = +/- 10%
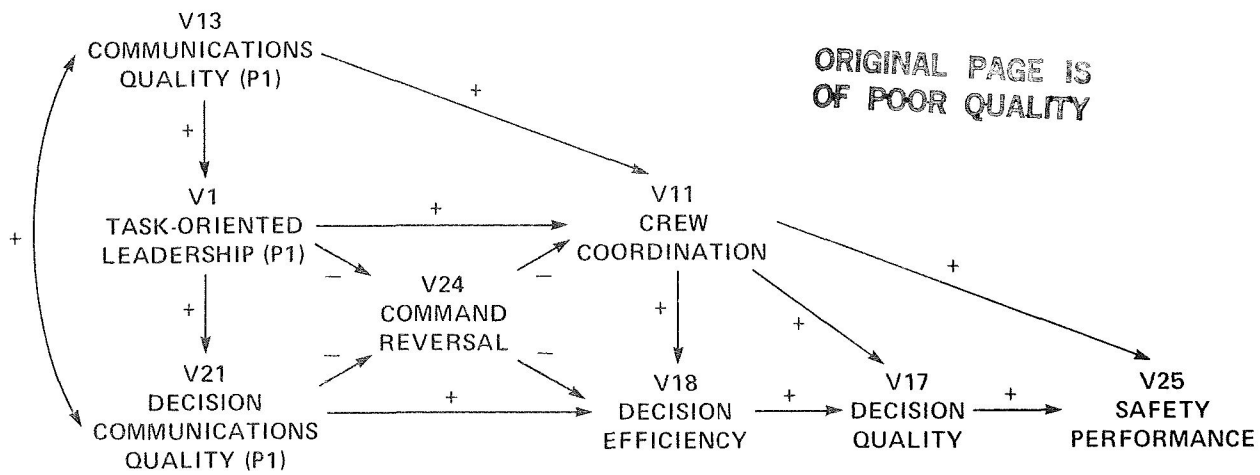
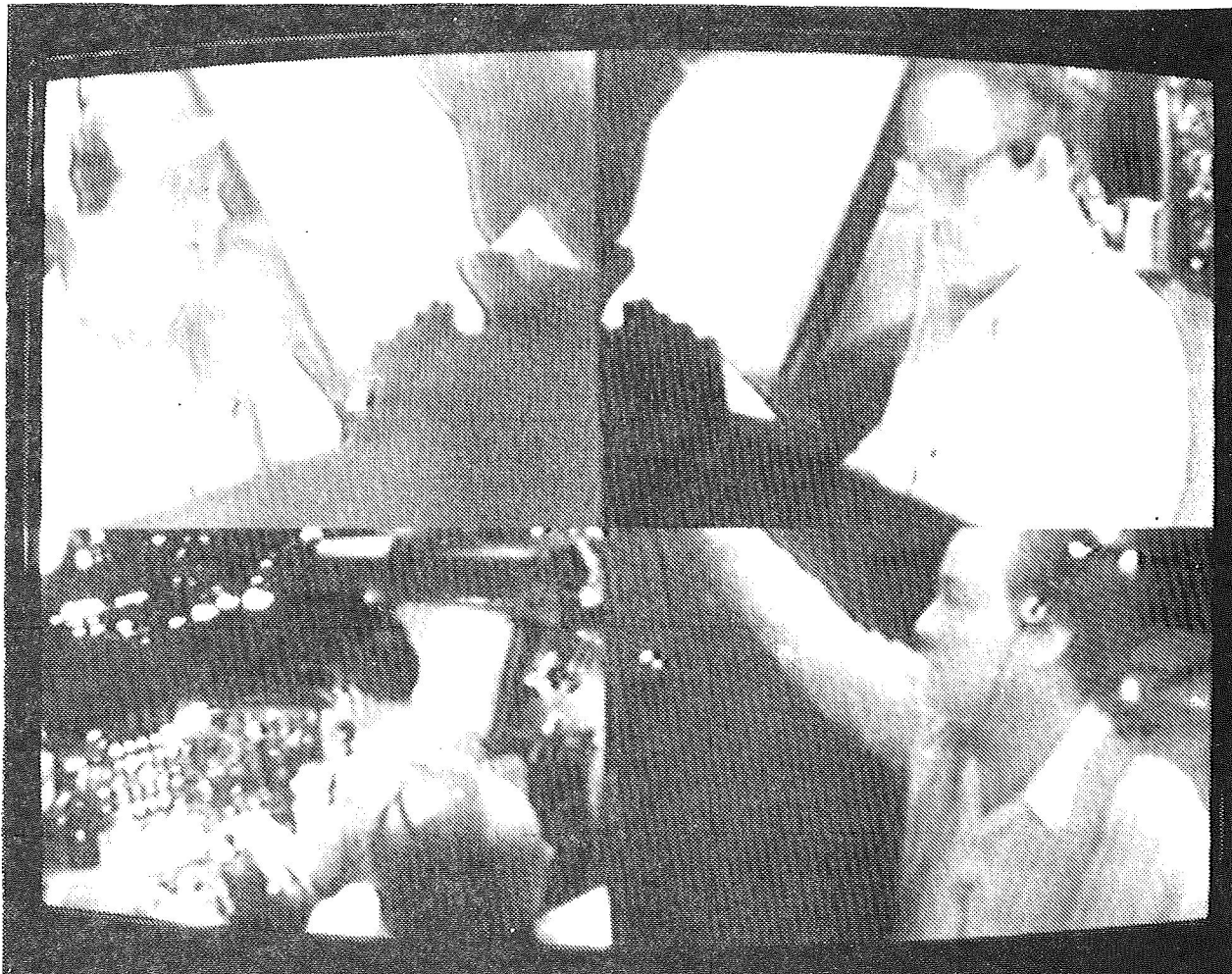Fig. 1 Partial Model of Crew Coordination and Decisionmaking
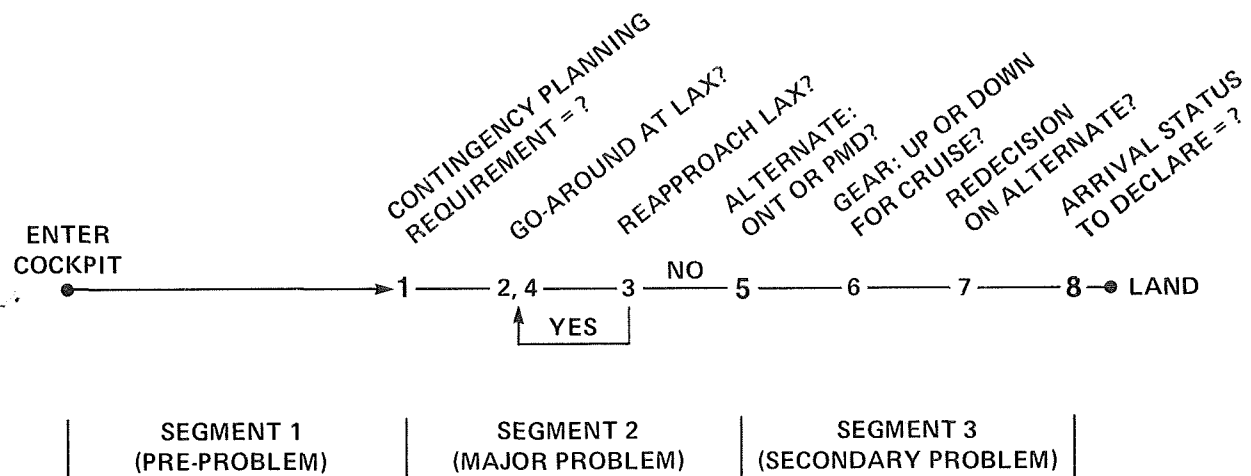

Fig. 2 Video Still Frame

ENTER
COCKPIT

CONTINGENCY PLANNING
REQUIREMENT = ?

GO-AROUND AT LAX?

REAPPROACH LAX?

ALTERNATE:
ONT OR PMD?

GEAR: UP OR DOWN
FOR CRUISE?

REDECISION
ON ALTERNATE?

ARRIVAL STATUS
TO DECLARE = ?

●———————————————►1 ——— 2, 4 ——— 3 —— NO —— 5 ——— 6 ——— 7 ——— 8 —● LAND

↑ YES |

| SEGMENT 1 | SEGMENT 2 | SEGMENT 3 |
| (PRE-PROBLEM) | (MAJOR PROBLEM) | (SECONDARY PROBLEM) |

Fig. 3 Videotape Stopping Points for Decisionmaking
and Crew Process Ratings

THE DECISION PROCESS WAS VERY EFFICIENT

STRONGLY DISAGREE   1   2   3   4   5   6   7   STRONGLY AGREE

(ALL SIGNIFICANT INFORMATION WAS ACQUIRED AT AN OPPORTUNE TIME,
ADEQUATELY EVALUATED, AND APPROPRIATELY UTILIZED)

COMMENT:

Fig. 4 Rating Scale Example

Crew No.	Task-Oriented Leadership (P2) V3	Comm. Quality (P1) V13	Task-Oriented Leadership (P1) V1	Decision Comm. Quality (P1) V21	Decision Command Reversal V24	Crew Coord. V11	Decision Efficiency V18	Decision Quality V17	Safety Performance V25
13									
2									
12									
15									
11									
8									
16									
4									
14									
9									
7									
3									
5									
1									
10									
6									

Fig.5 Means and Sds of Sixteen Crews on Variables of Model - Descending Order of Performance on V25 (Safety Performance)

23.30

Appendix A

Scale and Criterion Statements for Variables - Ordered by Concepts

a. Task-Oriented Leadership/Variables 1 and 3:
The CAPTAIN'S (FIRST OFFICER'S) leader behavior was highly task-oriented. (His behavior was highly concerned with establishing goals, clarifying responsibilities, defining *subordinate roles and task requirements, coaching subordinates and providing task related feed-back and evaluation)

b. Person-Oriented Leadership/Variables 2 and 4:
The CAPTAIN'S (FIRST OFFICER'S) leader behavior was highly sociomotively-oriented. (His behavior was highly concerned with establishing and maintaining positive crewmember relationships, pro-viding psychological support, enabling feelings of satisfaction, mak-ing tasks interesting or enjoyable)

c. Participatory Leadership/Variables 5 and 6:
The CAPTAIN'S (FIRST OFFICER'S) leader behavior was highly participa-tive in regard to supervision/resources management. (His behavior was highly concerned with encouraging subordinates to make sugges-tions regarding accomplishment of tasks, to independently analyze problems, to give feedback, and to question the leader)

Participatory Leadership/Variables 19 and 20:
The CAPTAIN'S (FIRST OFFICER'S) leader behavior was highly participa-tive in regard to decisionmaking. (His behavior was highly concerned with ensuring that all crewmembers for whom a decision was relevant had a chance to influence that decision -i.e. those crewmembers who had significant and pertinent information, responsibility to imple-ment the decision, or significant ego involvement for other reasons)

d. Competency-Motivation/Variables 7 and 8:
The FIRST OFFICER (FLIGHT ENGINEER) exhibited high competence and willingness to be responsible with respect to fulfilling the require-ments of his position. (appeared highly knowledgeable, skillful, and motivated in his behavior and task performance)

e. Command Reversal/Variable 9:
The FIRST OFFICER performed much of what would generally be the captain's leadership function. (whether due to acquiescence of the captain or dominance of the first officer)

Command Reversal/Variable 24:
The FIRST OFFICER performed much or what would generally be the captain's decisionmaking function. (same criterion statement)

f. Crew Cohesiveness/Variable 10:
The CREW functioned in a highly cohesive manner. (showed high crew solidarity or harmony; i.e. appeared well integrated into a unit)

g. Crew Coordination/Variable 11:
ACTIVITIES OF CREWMEMBERS were well coordinated to task and situation demands. (individual crewmember knowledge and skills were allocated in an effective and timely manner to meet task and situation demands)

h. Crew Friendliness/Variable 12:
The CREWMEMBERS related in a highly friendly manner. (interactions were most usually accompanied by verbal and/or non-verbal signs of friendliness, or warmth)

i. Communications Quality/Variables 13,14,15:
The CAPTAIN (FIRST OFFICER,FLIGHT ENGINEER) exhibited very good within-crew communications. (highly hearable, understandable, appropriate - in style and content, accurate and timely messages; good listener who made effort to understand; achieved reciprocal indication that understanding was reached)

Communications Quality/Variables 21, 22, 23:
The CAPTAIN (FIRST OFFICER,FLIGHT ENGINEER) exhibited very good within-crew communications in regard to decisionmaking. (same criterion statement)

j. Decision Difficulty/Variable 16:
The DECISION was a very difficult one. (involved complex, interacting operational factors - some of which may have been contingent on uncertain future events; or conflicting goals of safety, operational efficiency, company and/or ATC requirements - or preferences)

k. Decision Quality/Variable 17:
The CHOICE (or outcome of the decision process) was most appropriate. (the best choice considering safety of flight and/or the attainment of other mission goals)

l. Decision Efficiency/Variable 18:
The DECISION PROCESS was very efficient. (all significant information was acquired at an opportune time, adequately evaluated, and appropriately utilized)

m. Safety Performance/Variable 25:
The LEVEL OF SAFETY achieved through this crew's performance, considering the major scenario problem and any **special circumstances(s), was very high.(based on: safety of approach (es) to LAX; the airport of landing - considering differential risks of LAX,ONT,and PMD; fuel-on-board at touchdown - considering go-around and go-to-another-alternate fuel requirements, as well as additional fuel for a reasonable margin of safety)

*The word "subordinate" was used in referring to "other crewmembers" only in reference to the captain's leadership role. The latter phrase, or a derivative, was used throughout in reference to the first officer's leadership role.
**See Appendix B for a discussion of these special circumstances.

Appendix B

Data Quality
_____

For many reasons scenario conditions or (contextual events) were
not entirely consistent over the 16 crews for these long data runs.
Simulator failures, crew generated events, experimenter team errors,
planned interventions to reduce possibilities of crashes (for example,
from remaining too long in the LAX area) contributed to such incon-
sistencies. An advantage of ratings are that contextual effects can be
taken into consideration. During the rating procedure all such incon-
sistencies were identified and discussed with the raters. For the safety
performance ratings such inconsistencies were documented, by crew, on
fact sheets.

Simulator problems included stabilizer chatter that led crew one to
assume a runaway stabilizer and declare and emergency prior to arrival
at LAX. This necessitated clearing their flight to approach LAX with
17,500 lbs of fuel on board. Crews three, seven, and nine were cleared
for but did not execute the enroute hold prior to fuel levels reaching
14,000 lbs, and their subsequent clearance to LAX. Although varying
time-in-hold was planned (to compensate for usual crew differences in
fuel burn), simulator burn rates were determined to be high for these
flight segments. A few crews were given runway visual ranges that were
below the minimum 2400 ft. on their second approach to LAX. Although
the crews could and some did, legally continue their approach across the
outer marker by declaring an emergency, bias toward early interruption
of approach could be present for some crews. Finally, the flight
engineer of crew two had some prior knowledge of the major scenario
problem. All experimenters and raters agreed however that his role play
as a naive crewmember was successful and should result in little bias
for that crew's performance.

AN EXPERIMENTAL PARADIGM FOR TEAM DECISION PROCESSES

by

Daniel Serfaty

and

David L. Kleinman

CYBERLAB
University of Connecticut
Dept. of Electrical Engineering
and Computer Science
Storrs, CT   06288

ABSTRACT

The study of distributed information processing and decision making is presently hampered by two factors:

(i)   The inherent complexity of the mathematical formulation of decentralized problems (control, detection, data fusion, etc.) has prevented the development of efficient and practical theoretical models that could be used to predict actual performance in a distributed environment.

(ii) The lack of comprehensive scientific empirical data on human team decision making has hindered the development of significant descriptive models. Most of the organizational behavior and applied psychology research in the field focuses on centralized group decision making rather than on team decision making in which the element of decentralization is essential.

As a part of a comprehensive effort to find a new framework for multihuman decision making problems, we have developed a novel experimental research paradigm involving human terms in decision making tasks [1]. The paradigm focuses on the problems of distributed resource management and task processing in an uncertain dynamic environment. The task environment is an abstraction of a Naval Battle Group Command, Control and Communications (C^3) system in which a number of geographcally scattered commanders must make coherent decisions based on decentralized information on enemy actions.

This information is presented to each decision maker through graphical and alphanumerical displays providing data on tasks' status, attributes, resources, and communication messages.

The paradigm is flexible enough to be tested across a large range of experimental conditions in which the main independent variables are: the team configuration, the team information and communication structure, the uncertainty level in both inputs and consequences of action, the level of expertise and functional overlapping between the different decision makers. Our first baseline experiment involves a dyad - i.e. a symetric team of two decision makers with no hierarchical relationship. No communication is allowed and silent coordination is assumed with the main variable being the degree of overlapping in the decision makers' functional area of responsibility.

The flexibility of the paradigm will be used to study various cognitive factors which have found empirical evidence in the literature: need and use of communication in a well cooordinated and cohesive team; "risky" or "cautious" shift in team decision polarization; "selfish" behavior and misperception of team reward structure; conservatism and uncertainty avoidance in human organizations.

Attempts to construct parts of an integrated model with ideas from queueing networks, team theory, distributed estimation and decentralized resource management are described. Future development of these normative--descriptive models of human team behavior depends strongly on the availability of data to be provided by the experimental paradigm.

[1] Kleinman, D.L., D Serfaty and P.B. Luh, "A Research
     Paradigm for Multi-Human Decisionmaking", Proc.
     American Control Conference, San Diego, CA, 1984.

Investigation of Crew Performance
in a Multi-Vehicle Supervisory Control Task

Richard A. Miller, Brian D. Plamondon, Richard J. Jagacinski
and Alex C. Kirlik

Ohio State University
Department of Industrial and Systems Engineering
and
Department of Psychology
1971 Neil Ave.
Columbus, Ohio   43210

The primary objective of this research is to measure and represent crew information processing and decision making in a supervisory control task which is loosely based on the mission of future generation light helicopters. Subjects control the motion and activities of their own vehicle (the "scout") and direct the activities of four additional craft. The task involves searching an uncertain environment for "cargo" and "enemies," returning cargo to home base and destroying enemies while attempting to avoid destruction of the scout and the supervised vehicles.

A series of experiments with two-person crews and one-person crews will be performed. Resulting crew performance will be modeled with the objective of describing and understanding the information processing strategies utilized. Of particular interest are problem simplification strategies under time stress and high work load, simplification and compensation in the one-person cases, crew coordination in the two-person cases, and the relationship between strategy and errors in all cases. The results should provide some insight into the effective use of aids, particularly aids based on artificial intelligence, for similar tasks. In this informal paper we will describe the simulation which is used for the study and discuss some preliminary results from the first two-person crew study.

FORCE/TORQUE DISPLAY FOR SPACE TELEOPERATION CONTROL EXPERIMENTS AND EVALUATION

Kevin Corker
10 Moulton Street
Bolt, Beranek and Newman, Inc.
Cambridge, Massachusetts 02238

Antal Bejczy and Barry Rappaport
Jet Propulsion Laboratory
4800 Oak Grove Drive
Pasadena, California 91109

ABSTRACT

Experiments were performed at the Johnson Space Center (JSC), Manipulator Development Facility using the full scale Shuttle Remote Manipulator System (SRMS) to evaluate the effect of visual presentation through perspective display of the orthogonal forces and torques sensed at the manipulator end effector. The experiments investigated the effect of the display information on the management of forces and torques generated during payload berthing and deployment, as well as simulated satellite module change-out operations. The evaluation also addressed (i) issues of display format, including: force/torque scaling, point of resolution, and display mixing with video generated imagery, and (ii) task related variables of payload size, alternative sources of guidance information, and control mode.

This paper briefly presents the results of a first-pass informal analysis of the analog, strip chart-recorded data from these evaluation tests. The results provide a relative measure of improvement in force management through the use of such a display, as well as information regarding the impact of display variables and task demands on operator performance.

1.0 INTRODUCTION

Two experiments were performed at the JSC Manipulator Development Facility using the full-scale Shuttle RMS and the JPL two hundred pound range force/torque (F/T) sensor, four-claw end effector, and a perspective visual display of the forces and torques sensed at the end-effector. The equipment used in these tests, with the exception of the perspective display system, are described in a previous evaluation report by Bejczy and co-workers (1982).

The two evaluation sessions provided an assessment of the effect of the F/T sensor and display system on SRMS performance. The first session investigated operator handling in large payload berthing. The second session dealt with small tool handling and simulated module change-out performance. Figure 1 provides a plan view of the payloads, their size, and location for the tests, in relation to the Rockwell
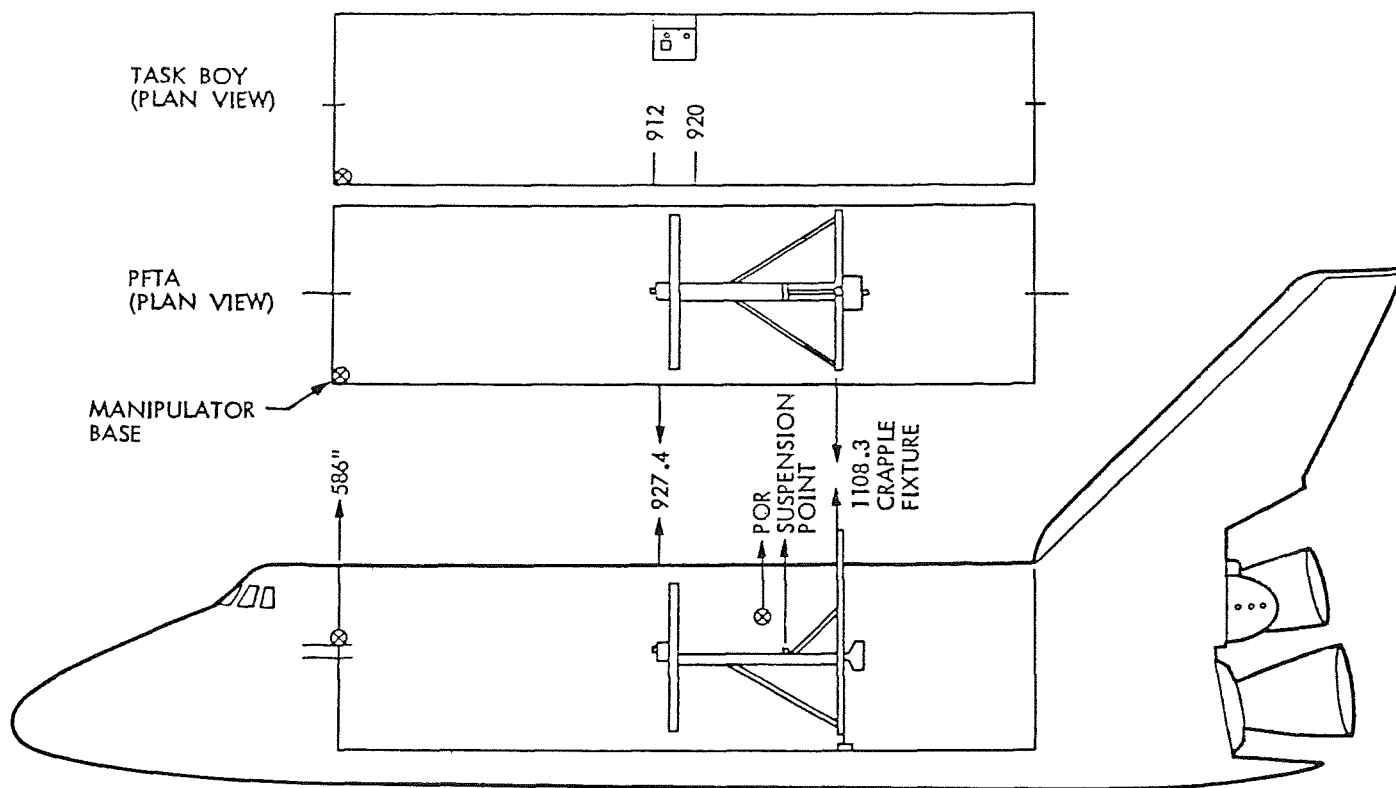
PAYLOAD LOCATIONS IN JSC MDF
FORCE/TORQUE DISPLAY

Figure 1.

26.2

SCALE: ANALOG CHART RECORDING
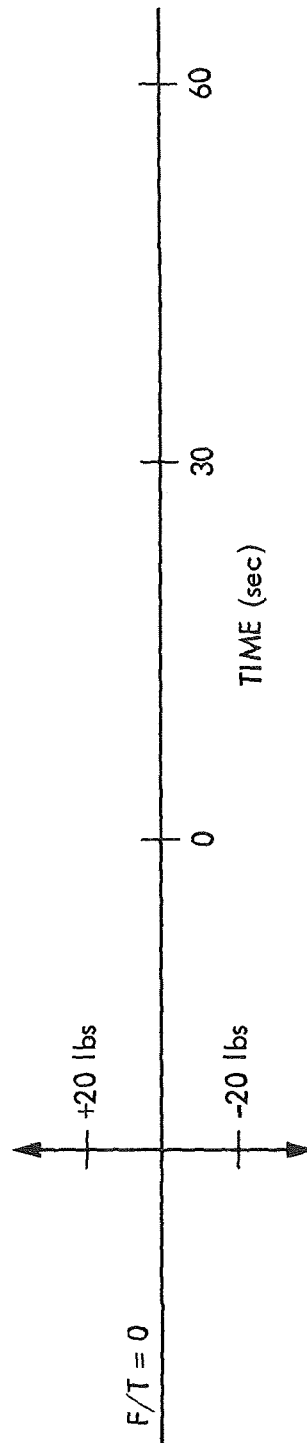FORCES AND TORQUES FOR TOOL USE
AND MODULE CHANGE-OUT

Figure 1b.

Shuttle point of reference (POR), i.e., 236 in. forward of and 400 in. below the orbiter nose point.

The evaluation tasks were performed by four JSC personnel who were trained and MDF qualified in the use of the shuttle RMS simulator. The tests were performed in two sessions each of one week duration and separated by a six month hiatus.

1.1 Display Characteristics

The characteristics of the display format used for these evaluation are presented here. Since the time of these tests, we have made substantial progess in creating a three dimensional perspective display. This display technique is described in the final section of this paper on future research efforts.

(i) The display, pictured in Figure Two, presents force and torque as filling from the center of the six axis perspective frame. The point of reference for the axes can be manipulated in software to correspond to the control reference frame of the operator, or any other reference frame deemed appropriate to the task. In the case of the PFTA payload, the X axis relates to the fore/aft axis of the orbiter, the Z axis refers to the elevation in and out of the payload bay, and the Y axis designates port/starboard across the payload bay. The torques about these axes are designated by filling of the pitch, roll, and yaw frames associated with each of the torques. In the case of the tool handling and module change out procedures, the display is referenced to the end effector and sensor reference frame as illustrated in Figure 3b.

(ii) The display provides force and torque readings to the operator referenced to the point of resolution (POR) of the PFTA payload, in the first evaluation, and referenced to the sensor reference frame in the second evaluation. (The POR can be varied through software manipulation of the data provided by the sensor system and can be calculated for the desired operator perspective, dependent on payload geometry.) The POR chosen for the large payload berthing was the center of geometry of the payload. This POR is forward of the center of mass of the payload to compensate for the small residual frictional forces associated with the payload counterweight system. The MDF counterweight system serves to simulate zero gravity operation for high mass payloads, such as the PFTA.

(iii) The "sense" of the displayed forces shows the effect of the operator's control input on the payload. For example, in the case of PFTA manipulation, a roll to port that generates contact forces with the \port trunions is displayed as an increased torque to port and an increased Z force. The corrective control action to reduce these forces and torques is to roll starboard, i.e., the operator acts as if to push the extending display bar to zero, the center point. Operators generally found this "fly to" arrangement intuitive. However, when the payload is viewed from the aft cameras the sense of the display in terms of required corrective action is reversed. This caused some confusion,
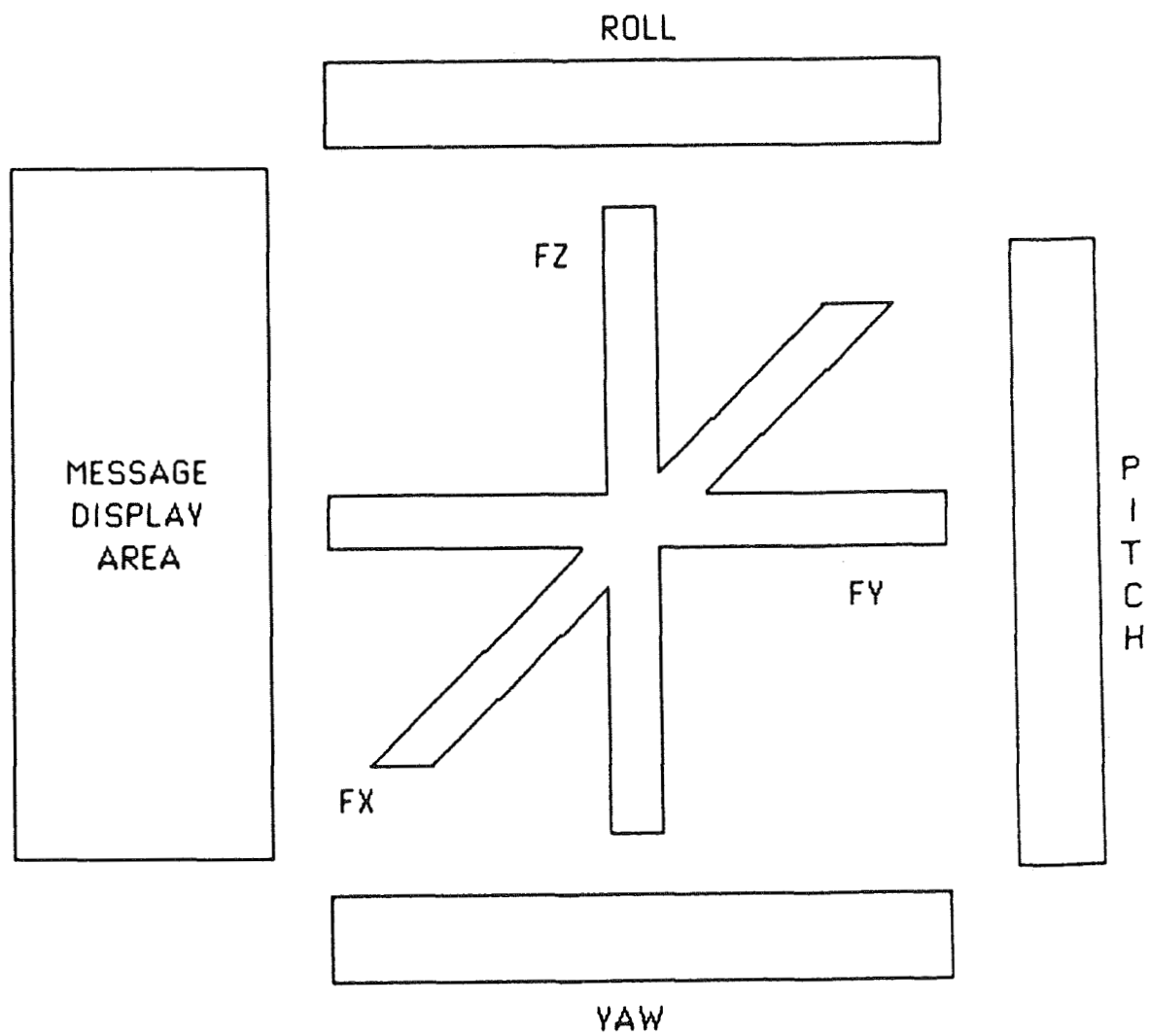
Figure 2. Display Format

and argues for a display reference that is dynamically referenced to the point of regard of the operator. Experiments and software requirements for such transformations are currently under consideration by the authors.

(iv) Force/Torque display scaling proved sensitive to the payload geometry. Because of the large moment arm of the PFTA payload, torques generated at the bay trunions saturated the torque scaling more quickly than forces about the POR. Software decoupling and rescaling of the torque display was accomplished, but there is some danger in this approach, in that sensor saturation may not bear a clear relation to display saturation. Future work will seek to provide both sensor and display saturation scales to the operator.

(v) The display size could be reduced to allow split screen mixing with an operator selected camera view of the payload.

1.2 Performance Data

The data collected were (i) total task time, defined as operator control initiation to payload berthed and latched condition, (ii) analog chart recording of the forces and torques sensed about three orthogonal force and three orthogonal torque axes of the sensor POR during the berthing operation, and (iii) digital recording of these forces and torques. In this preliminary evalution, statistical analysis is precluded by the large number of treatment conditions in relation to the number of data points gathered in the analysis. The evaluation was designed to survey the relative impact of the provision of and the format of visual F/T feedback, rather than to establish statistically robust parameterization of that effect.

2.0 EVALUATION PROTOCOL

2.1 Large Payload (PFTA) Berthing

Task:

The performance required for this evaluation involved berthing the PFTA payload after it was deployed to a random position above the paylad bay trunion guides. The task represents the precision placement portion of a payload berthing task. The berthing task was performed ten times by each subject after familiarization and briefing runs on the display characteristics. The ten test trials were performed under varied feedback and control conditions as illustrated in Table I. The control point of reference for these tests was the orbiter control mode, in which the operator controls the end effector of the RMS in relation to the shuttle body. Translation axes of the two-handed controller refer to for/aft, port/starboard, and elevation in/out of the bay. Rotational axes of pitch, roll and yaw are referenced to these trannslational axes. (The control mode for the majority of the tests was a resolved rate control. The exception to this was a joint by joint control mode which had its greatest impact in dramatically increasing required performance

Trial	Camera				Display		Control Mode	
	Port Aft	Starbrd Forward	Starbrd Aft	Elbow	Digital	F/T	Orbiter Loaded	Single Joint
1	*	*	*	*	*	*	*	
2	*		*		*	*	*	
3		*		*	*	s/s	*	
4	*				*		*	
5	*	*	*		*		*	
6						*	*	
7	*	*	*		*			*
8		*			*	*		*
9				*	*	*	*	
10	*					*		*

* Indicates sensor available or mode used.

S/S = Split Screen

Table 1. Feedback and Control Conditions
for PFTA Berthing Experiments

time for all operations.)

Results:

A very general discussion of results is presented for this task. Analysis of the digital data, as opposed to the analog chart recording, is being pursued with the intent to describe the effect of the varied feedback views in conjunction with the F/T display. At this point we will confine our discussion to the management of forces and torques with and without the visual display from the sensor.

(i) Force/Torque generation:

- Provision of force/torque information via the visual display reduced the loads on the PFTA payloads and payload guides during berthing by 30-50% of the values generated without the provision of the display.

- For those forces generated in excess of 50% of the dynamic range of the sensor, visual display of the force/torque values reduce the duration of the application of that excessive force by 60-80%.

(ii) Task completion time:

- Task completion time was most dependent on the individual operator's control strategy. The directions stressed both accuracy and speed in task completion; however, speed was consistently sacrificed to performance accuracy.

- Provision of F/T information slightly increased the usual task completion time for a given operator. This was probably due to the requirement for shared attention between visual displays of payload position and the force/torque display.

- Several operators noted that the provision of the F/T display expedited trajectory planning in the case of excessive force application. The F/T information could be used diagnostically to identify the cause of the problem and to provide a basis for replanning the maneuver. This was especially true in the case of keel trunion misalignment; because the source of such an error is not readily visual available.

- As noted, the effect of the varied feedback conditions will be examined through analysis of the digital force/torque data.

2.2 Tool use and Simulated module change out

Task:

The tool use and module change out task involved manipulation of the modules of the task board illustrated in Figure 2. The flat screw driver blade was used to unlatch the box module and replaced in the

appropriate receptical. The module was then grasped, removed and reinserted. The screw driver blade was then retrieved and used to latch the module back in place. The task was performed in the end effector control mode, in which the control and display was referenced to the end effector position, independent of its position in the shuttle bay. Figure 3b illustrates the coordinates of the end effector reference frame. Figure 1b illustrates the placement of the task box in relation to the shuttle bay.

Results:

It is significant to note that three of the four subjects were unable to compete the module extraction task without the provision of visual force and torque information.

Several representative figures have been abstracted from the analog performance record to illustrate typical performance profiles.

- Figure 4 shows the calibration scale for the data represented.

- Figures 5a-5b shows the basic extraction/insertion sequence. The generation of excessive forces and torques in the absence of the F/T display is illustrated in 5a. In fact, the trial was aborted when the forces were sufficient to damage the module during the test. Successful completion of the same task sequence is demonstrated in 5b.

- Figures 6a-6b provide a direct comparison of module insertion sequences with and without the F/T display. A comparison of 6a and 6b illustrates increased levels of force/torque generation and increased task completion time for the single subject who was able to complete the module change out in the absence of the F/T display.

- Figure 7a provides a demonstration of a jam in which module extraction is aborted due to excessive force in the X and Y axis and torque about the Z axis. The diagnostic capability of the display is illustrated in Figure 7b, in which, despite the occasional generation of high force and torque values, the subject is able to successfully complete the module extraction.

- Figure 8a-8b shows successful force management in the tool use sequence of the task as a function of the provision of the force/torque display.

3.0 FUTURE RESEARCH EFFORTS

One of the major concerns in the presentation of force/torque information is the speed vs. cognitive information transmission dilemma. In other words, it is the dilemma of trying to transfer to the operator as much information as fast as possible without having a degradation of performance. This information should be presented so that the operator can cognitively understand and utilize it.

ORBITER BODY AXIS SYSTEM

ORBITER ROTATION AXIS SYSTEM
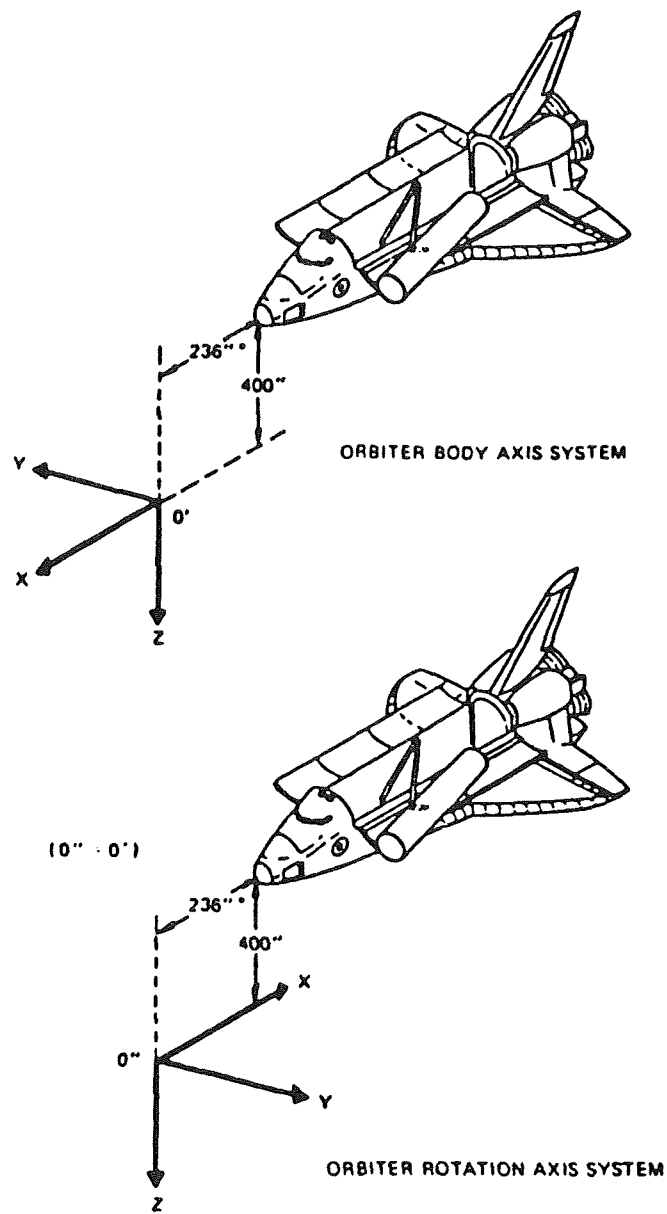
* For Orbiter 102 only

Figure 3a.   Relationship between Orbiter Axis System
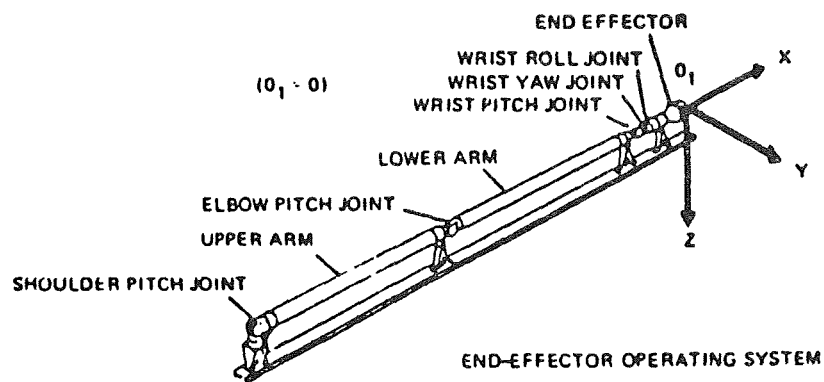   and Orbiter Rotation Axis System
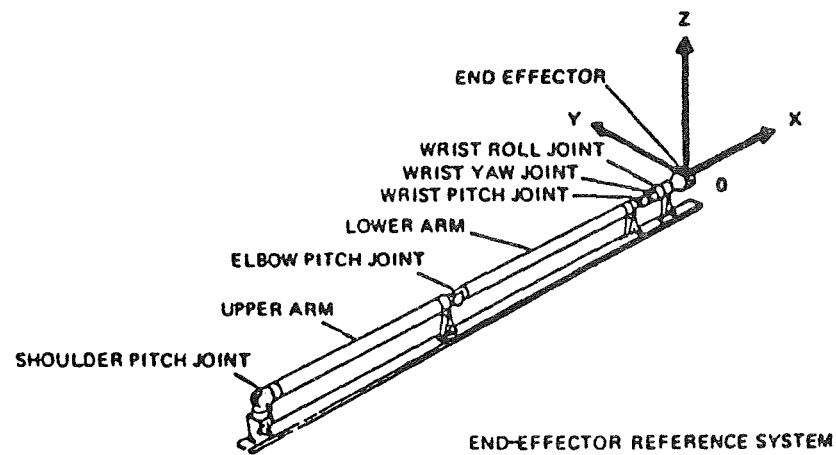
26.10

Figure 3b. Relationship between End-Effector Reference System
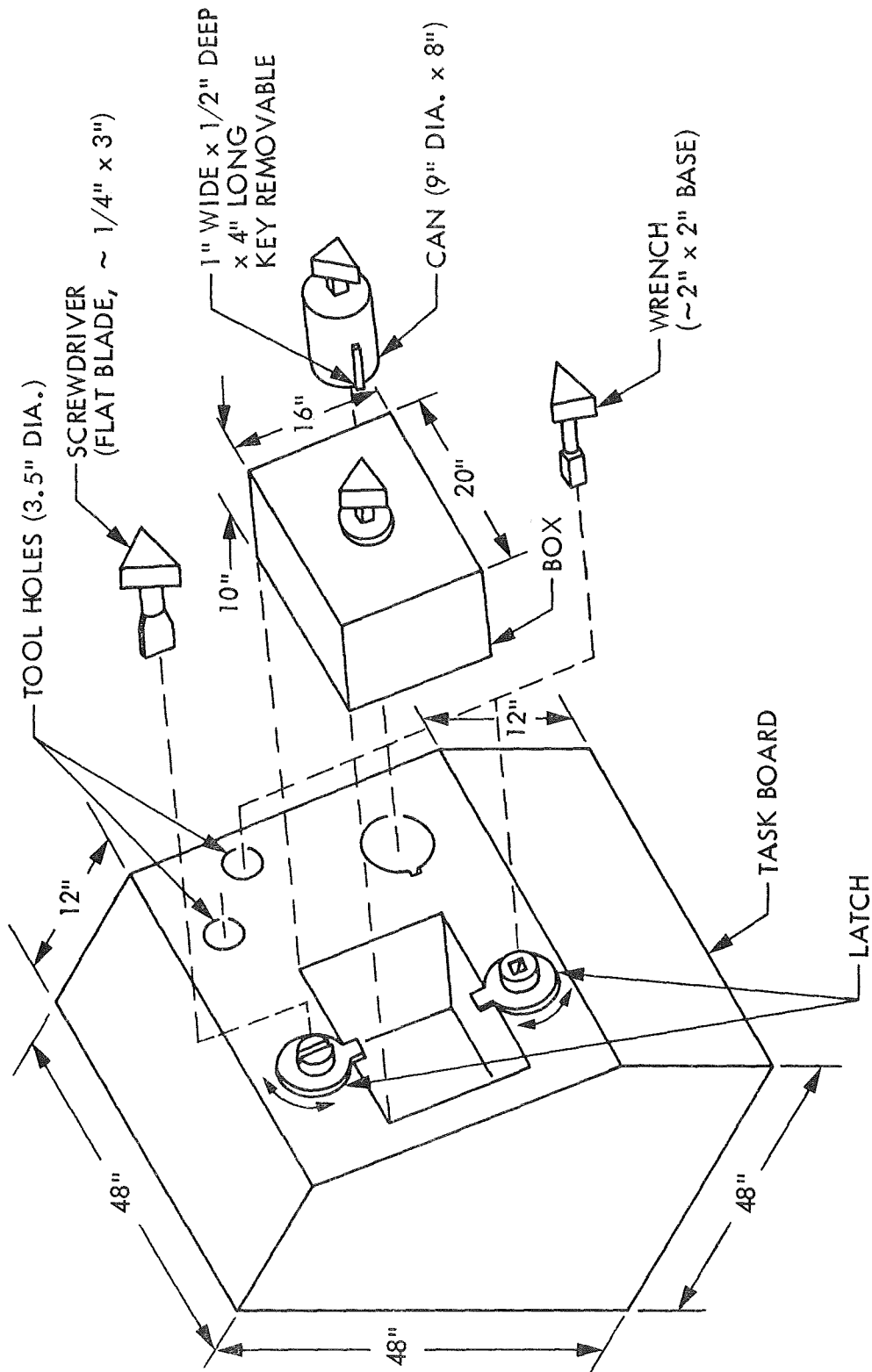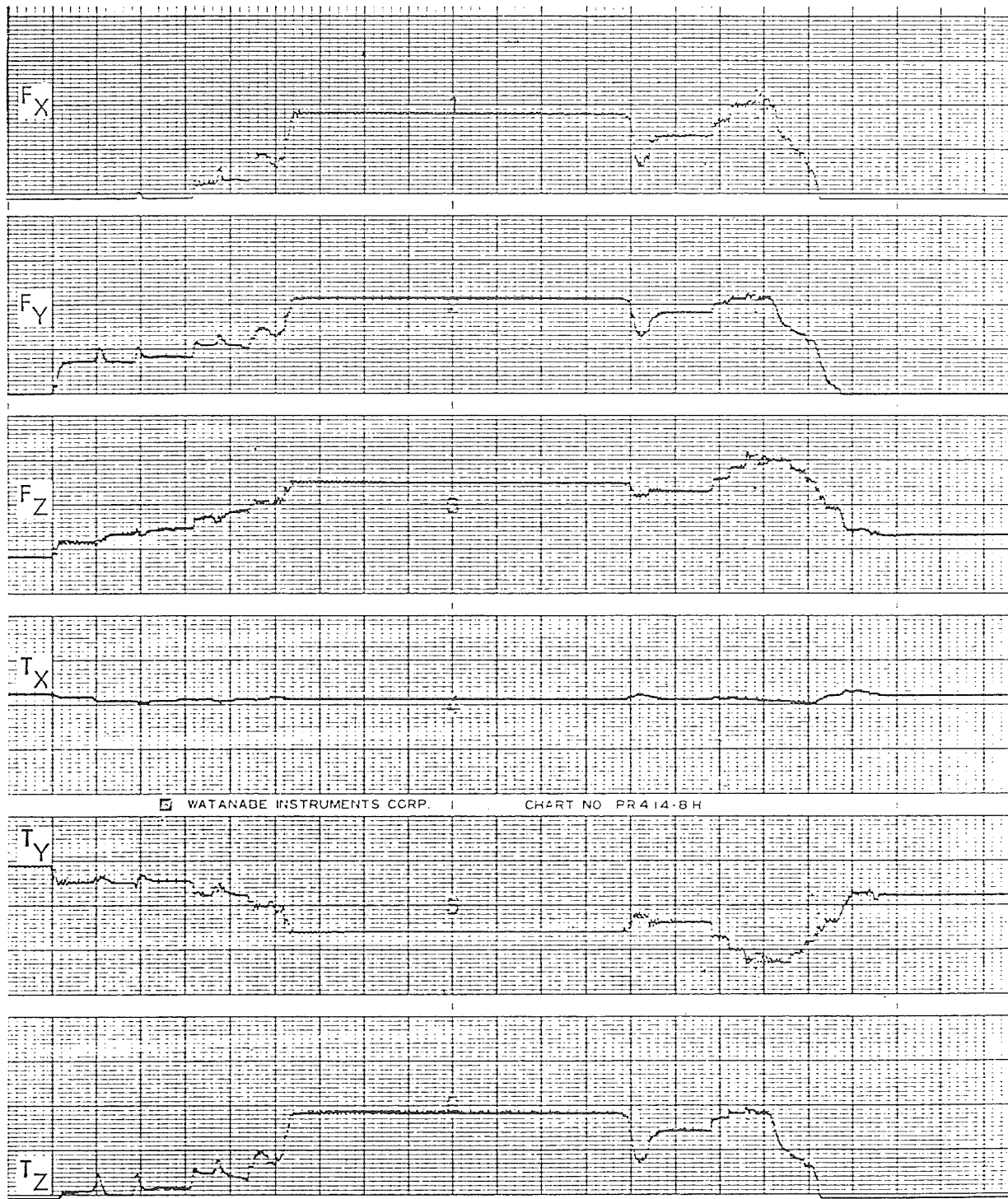and End-Effector Operating System

TOOL HOLES (3.5" DIA.)

SCREWDRIVER
(FLAT BLADE, ~ 1/4" × 3")

1" WIDE × 1/2" DEEP
× 4" LONG
KEY REMOVABLE

CAN (9" DIA. × 8")

WRENCH
(~2" × 2" BASE)

16"

20"

10"

BOX

12"

12"

48"

48"

48"

TASK BOARD

LATCH

Figure 4.

26.12

WATANABE INSTRUMENTS CORP.   CHART NO PR414-8H

SUBJECT :   DP
CONTROL: STANDARD
TASK   :   MODULE EXTRACTION
FEEDBACK: CAMERA
             DIRECT VISION

ABORT
TRIAL

Figure 5a.

26.13
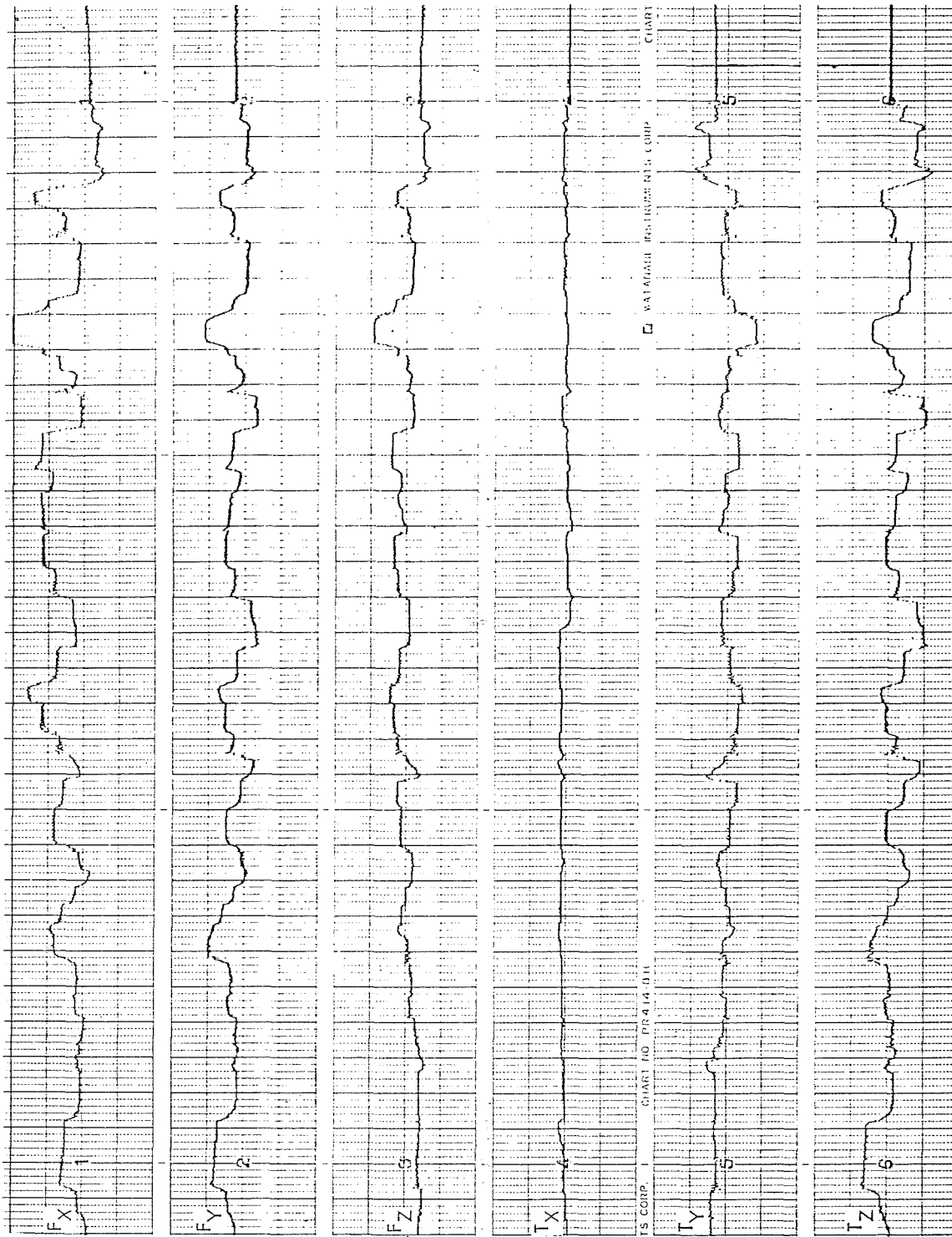
Figure 5b-1.

SUBJECT : DP
CONTROL: STANDARD
TASK : MODULE INSERTION
FEEDBACK: FORCE/TORQUE DISPLAY
          CAMERA
          DIRECT VIEW

26.14
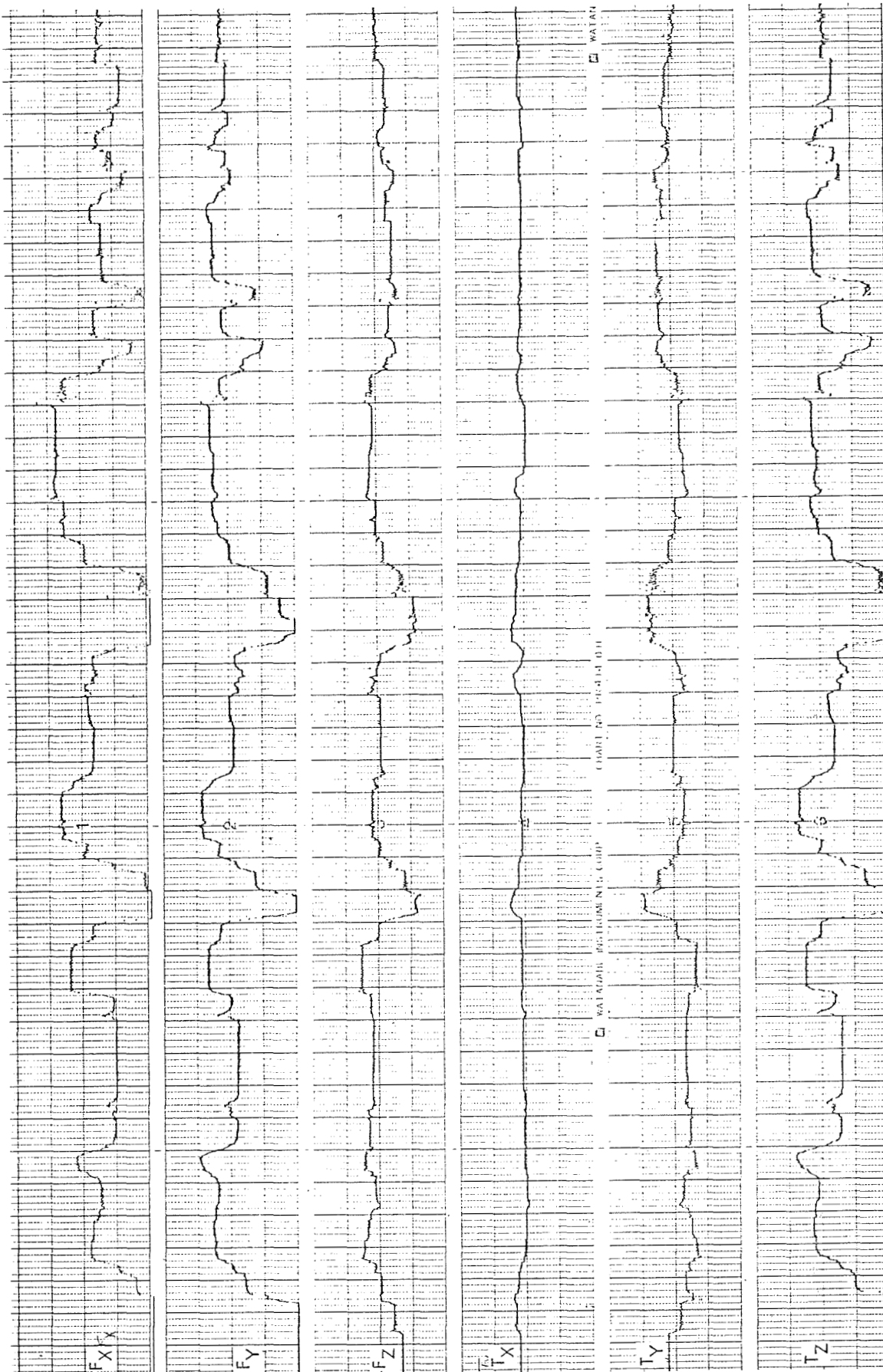
Figure 5b-2.

26.15

SUBJECT   : DP
CONTROL: STANDARD
TASK   : MODULE EXTRACTION
FEEDBACK: FORCE/TORQUE DISPLAY
           CAMERA
           DIRECT VIEW

ORIGINAL PAGE IS
OF POOR QUALITY

Figure 6a.

SUBJECT :   JB
CONTROL: STANDARD
TASK   :   MODULE INSERT
FEEDBACK: CAMERA
            DIRECT VISION

26.16

SUBJECT : JB   Figure 6b.
CONTROL: STANDARD
TASK   : MODULE INSERTION
FEEDBACK: FORCE/TORQUE DISPLAY
             CAMERA
             DIRECT VISION

26.17
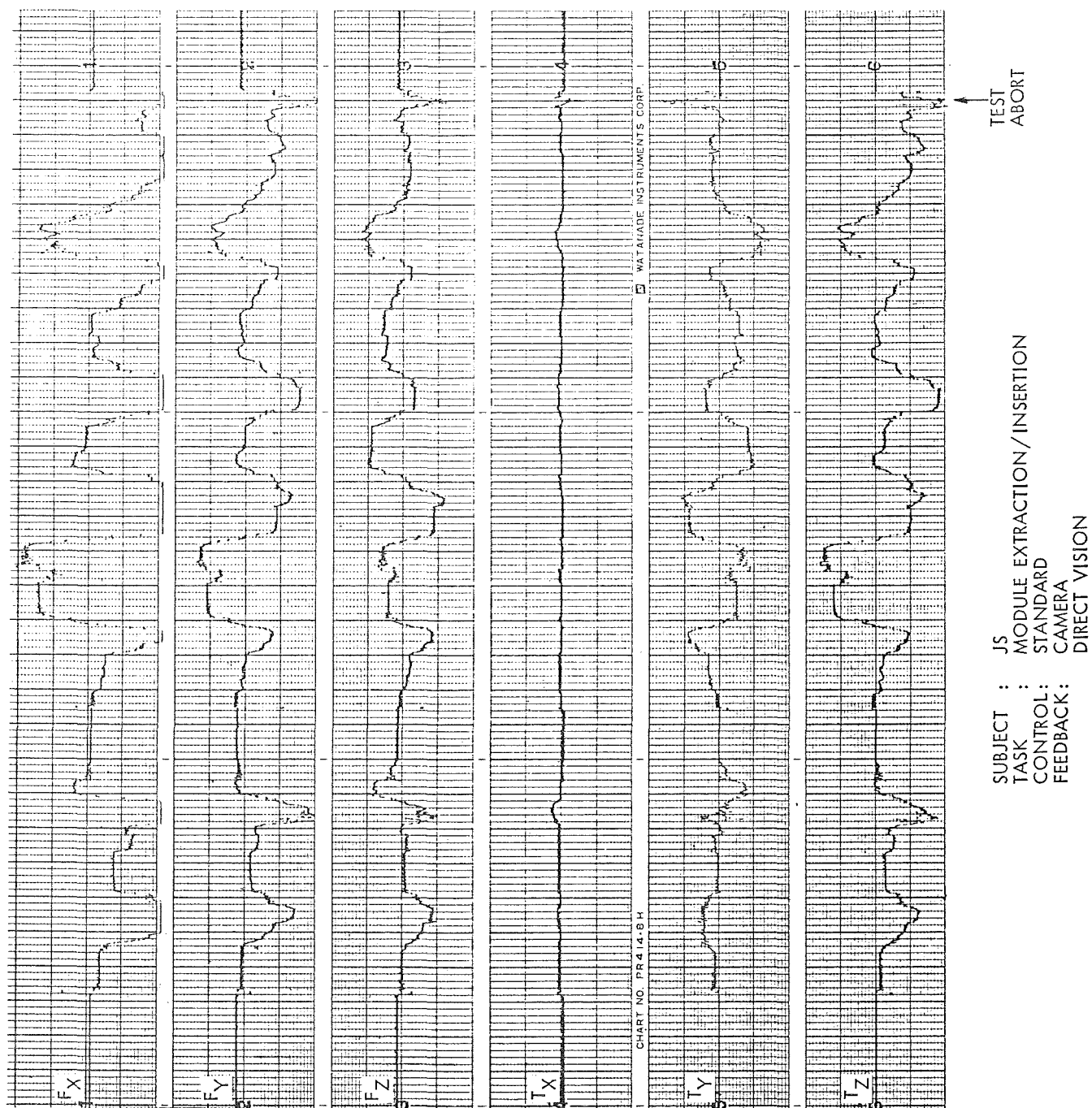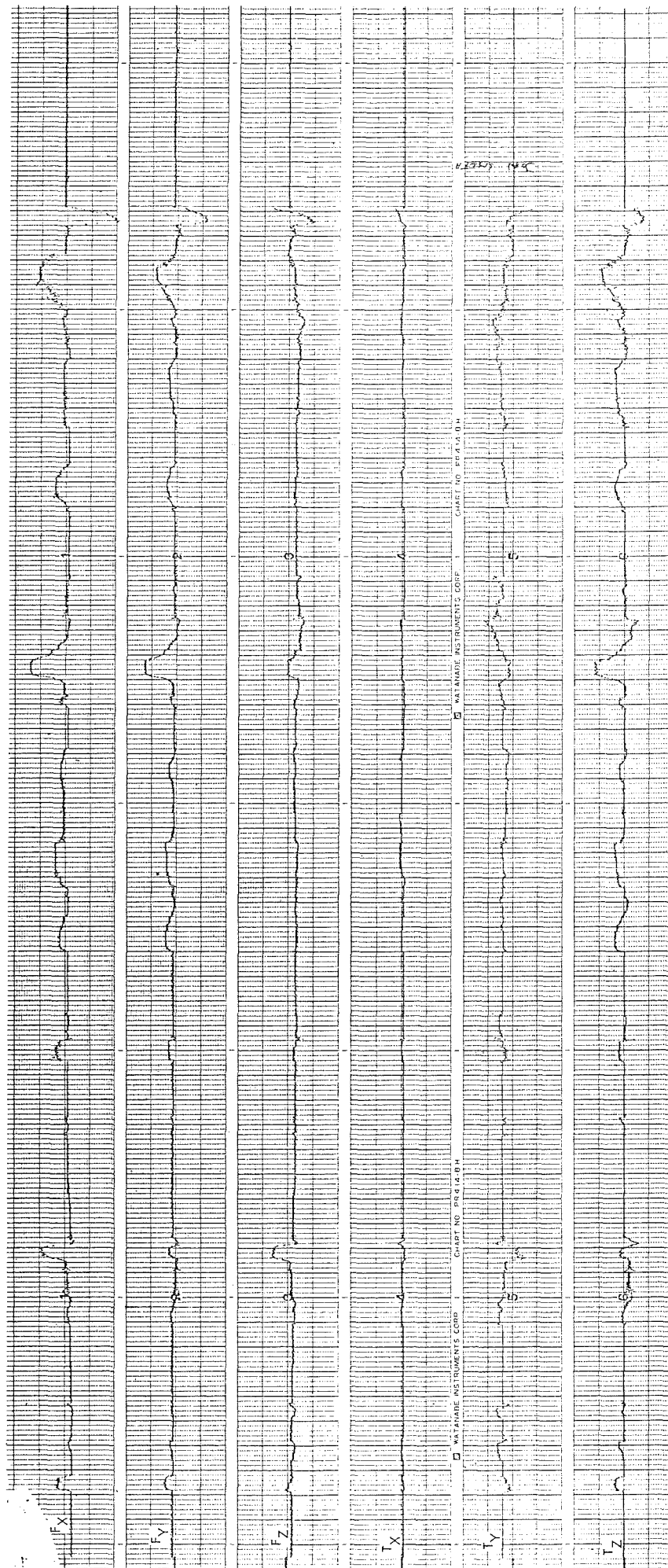
Figure 7a.

SUBJECT : JS
CONTROL: STANDARD
TASK : MODULE EXTRACTION/INSERTION
FEEDBACK: FORCE/TORQUE DISPLAY
CAMERA
DIRECT VISION

Figure 7b.

26.19
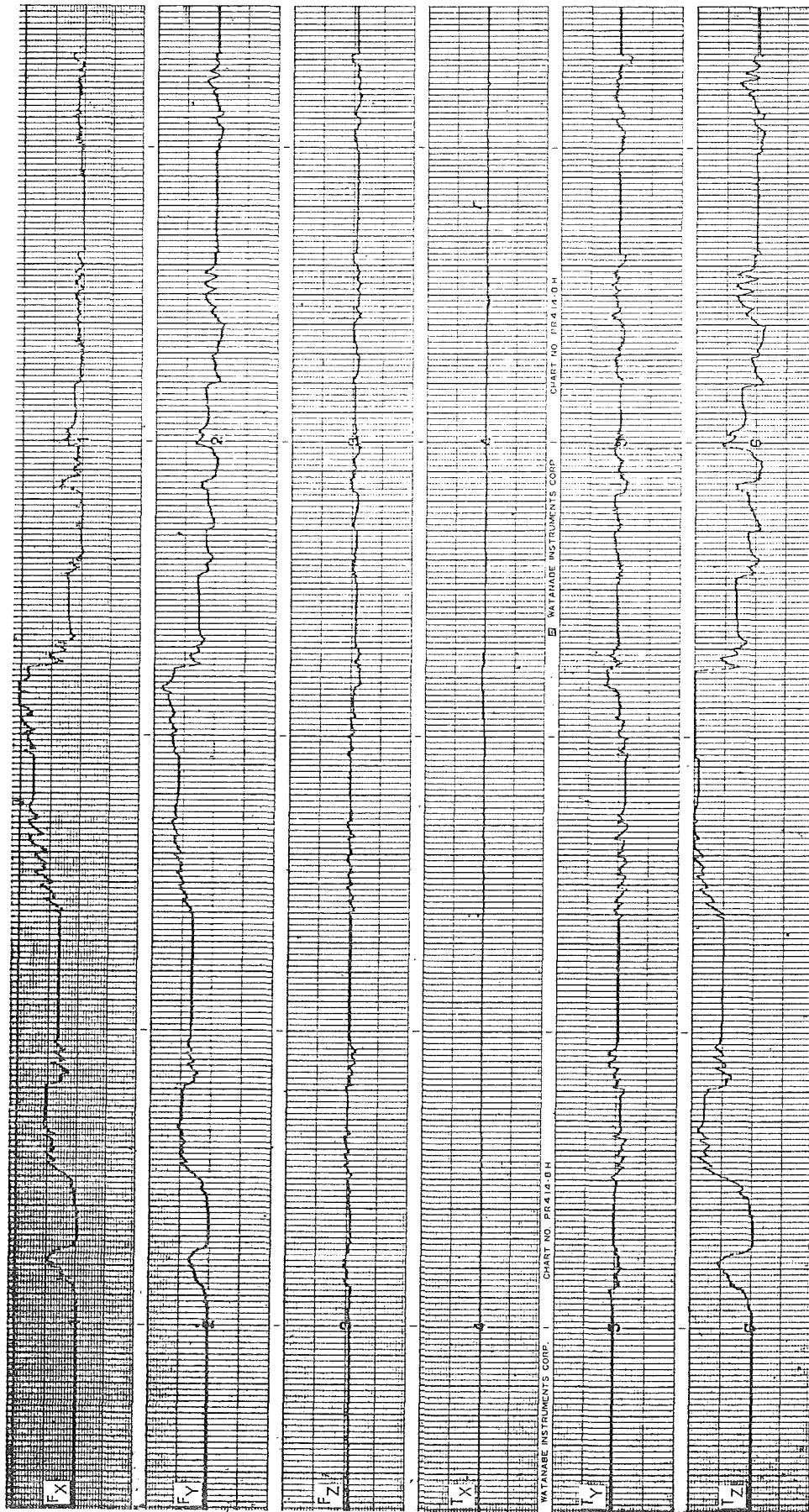
SUBJECT : PV
CONTROL: STANDARD
TASK : TOOL USE
FEEDBACK: CAMERA
              DIRECT VIEW

Figure 8a.

SUBJECT   :   PV
CONTROL: STANDARD
TASK   :   TOOL USE
FEEDBACK: FORCE/TORQUE
            CAMERA
            DIRECT VISION
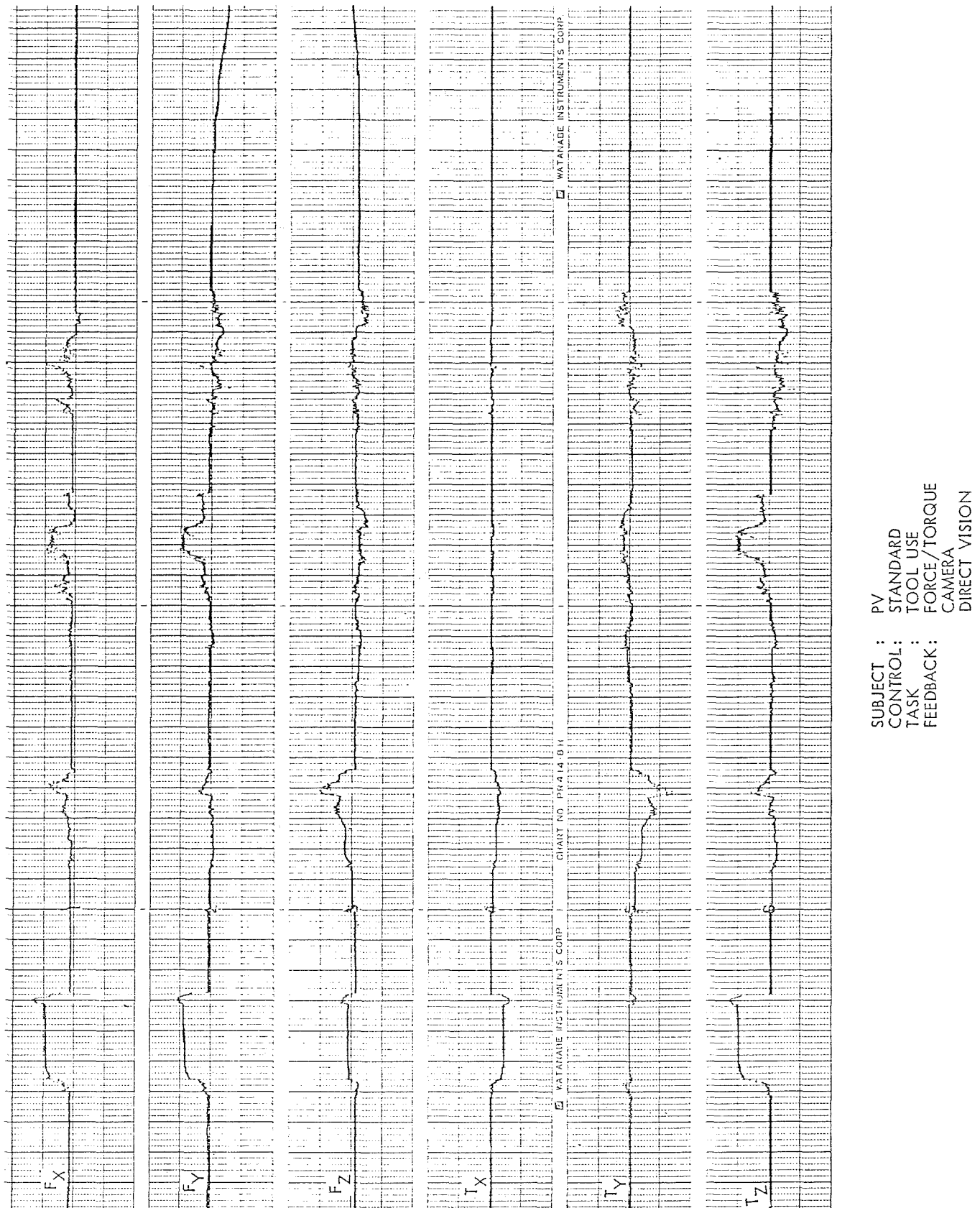
Figure 8b.

Items for display improvement:

i) The display should be as smooth as possible. The operator should be concentrating on the information present in the display, not on the display itself.

Most computer graphics display hardware is display-bound. The more pixels and polygons drawn on the screen, the slower the pixel write speed. Since most hardware internal graphics subroutines (draw rectangles, draw circles) are faster than software generated subroutines, it is optimal to use as many hardware oriented commands as possible.

ii) The display should present the information in a natural manner (i.e., true perspective view).

iii) Color should be used to enhance contrast between different display parts.

The true perspective 3-D Force/Torque display:

We have been able to make progress in the development of real-time 3-D displays because the substantial leap in the speed of current computer graphics hardware. The displays we used at JSC had a refresh rate of 4 to 5 hertz and there was a significant speed difference between the X/Y axis and the Z axis. With current display technology, a refresh rate of 30 hertz is easily achieved with much more true and complex display of forces and torques (Figure 9).

The torques and forces are color and directional coded. Red indicates a negative force or torque and blue indicates a positive force or torque. The torques follow the right-hand rule around the force axis. The display is projected in true perspective. The box around the display enhances the perspective image. The reticular marks divide the force bars into quarters. These marks help the operator gauge force on each axis. This is true especially in the case of the negative z force axis.

We thought about adding a grid on the bottom of the box to enhance the perspective image but it was decided that it would add too much clutter to the display.

4.0 CONCLUSION

In general, the operators considered the F/T display informative, and the data illustrate the fact that management of forces and torques improved when the display was used. In fact, the precision module extraction and tool use task was only able to be performed with the display aiding. There were a number of factors noted that could contribute to an improvement of the display format, and these have been the focus of our efforts in the development of the three dimensional perspective display. In particular, the following issues are being
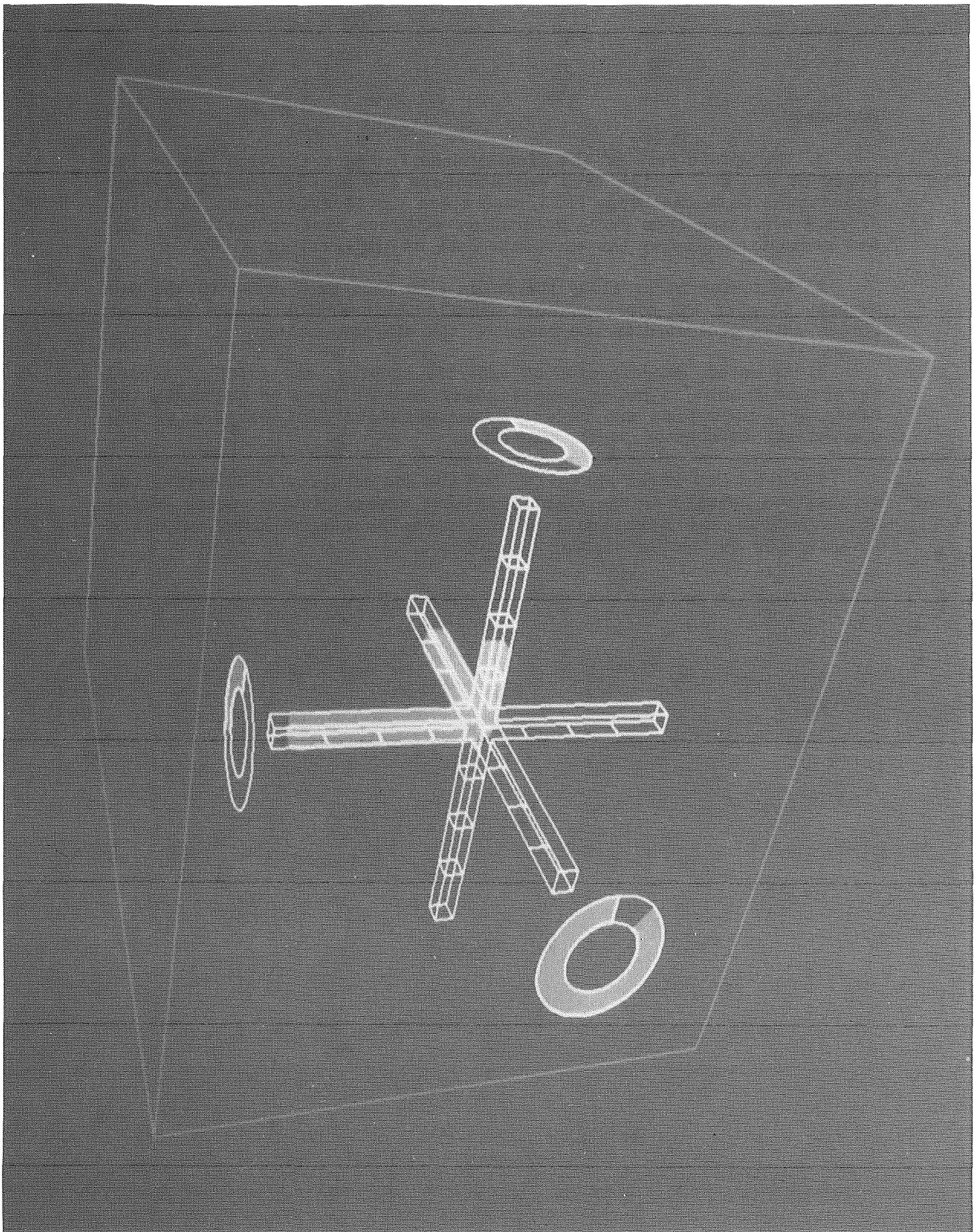
Figure 9. New Force-Torque Display

26.23

addressed:

1. The update rate of the display used in the evaluation was on the order of 4-5 Hz. While this was adequate for slowly moving payload operations with the large PFTA, there was a noticeable jumping in the display resulting from force generation with the smaller payloads. The new generation display has an update rate on the order of 30 Hz.

2. Reticular marks along the frame axes have been added in the new display to give the operator more detailed information on the level of forces being generated in the range of the display scale.

3. As noted, coordination of control, display, and point of regard reference frames is being investigated in an effort to maintain the operator's situation and reduce disorientation in interpreting the operational effects of force generation.

4. There is a great potential for the use of color to increase the information density of the display without adding clutter. Color coding of direction and magnitude of the force/torque vectors is being investigated in the new display development.

MODEL ANALYSIS OF REMOTELY CONTROLLED RENDEZVOUS AND DOCKING
WITH DISPLAY PREDICTION

Paul Milgram and Paul H. Wewerinke
National Aerospace Laboratory NLR
Anthony Fokkerweg 2, 1059 CM Amsterdam, The Netherlands.

ABSTRACT

Manual control of rendezvous and docking (RVD) of two spacecraft in low earth orbit by a 'remote' human operator is discussed. Experimental evidence has shown that control performance degradation for large transmission delays (between spacecraft and operations control centre) can be substantially improved by the introduction of predictor displays. An initial Optimal Control Model (OCM) analysis of RVD translational and rotational perturbation control has been performed, with emphasis placed on the predictive capabilities of the combined Kalman estimator/optimal predictor with respect to control performance, for a range of time delays, motor noise levels and tracking axes. OCM predictions are then used as a reference for comparing tracking performance with a simple predictor display, as well as with no display prediction at all. Use is made here of an 'imperfect internal model' formulation, whereby it is assumed that the human operator has no knowledge of the system transmission delay.

1. INTRODUCTION

In the course of early missions in space (eg. Gemini, Apollo, Skylab), humans played an important role, especially during launch, early orbital phases and spacecraft systems checkout during actual flight. That role was often managerial; system variables were compared with nominal values and, in the case of unacceptable deviations, the spacecraft subsystem would be commanded to a standby or safety mode. In other space operations to date, including shuttle arm manoeuvres, furthermore, the human operator's (HO's) activities have been scheduled and well-defined and in practically all cases the HO's role has been very well rehearsed. For future space operations, especially contingency operations, on the other hand, faster responses and more adaptiveness, flexibility and innovation are going to be required.

During the execution of any (tele)operation in space, the HO, whether on the ground or in space, may be considered in some way to be 'remote' --i.e. spatially, temporally and/or functionally-- with respect to the system being supervised or controlled. The combination of remoteness and the need for extending human (perceptual, decision making and problem solving) capabilities into space will necessitate further technological developments both towards increasing local autonomy through artificial intelligence and towards

augmenting the HO's ability to influence events at a remote worksite through 'telepresence' (Akin et al, 1983). In order systematically to define and optimise the distribution of machine and human intelligence within such remote teleoperator systems, designers of these systems will need to base their decisions, among others, upon analytical quantitative predictions of the human operator's performance as a system supervisor and controller.

One of the most important operations in space is rendezvous and docking (RVD), whose purpose is to bring together and achieve a physical union between two orbiting spacecraft. The capability of achieving this physical union opens up the possibility of execution of a large variety of space operations, such as transfer of spacecraft or spacecraft elements to new orbits, removal of debris in space, assembly of spacecraft in orbit, maintenance of spacecraft and exchange of spacecraft payloads.

When RVD operations are performed with unmanned spacecraft, operations are controlled from a (ground-based) Operations Control Centre (OCC). Contact between the OCC and the space segment (both spacecraft) involves activities such as periodic checkout of spacecraft systems, calibration, transmission of go/no-go commands, monitoring of manoeuvres and, in a number of cases, on-line, closed loop control by a human operator at the OCC.

Direct communication between an OCC on the ground and the space segment is possible only when 'coverage' exists; that is, when there exists a data transmission path between ground segment and space segment, and vice versa. Direct coverage exists when the spacecraft are within the optical field of view of the OCC. However, the times at which this occurs may be inappropriate, and also very brief. Such difficulties can be overcome by using a Data Relay Satellite (DRS) in geostationary orbit. In all cases the transmitted signals will be delayed to some extent, for both uplink and downlink transmission, and these delays will in turn tend to diminish the ease and efficiency of regulating RVD from the OCC. Sources of signal time delays include data synchronisation and limited data transmission capacity (in both space segment and ground segment), distance to be travelled by the signal, data sampling and processing, data routing via one or more DRS's and a non-colocated ground antenna and OCC.

In this paper we present a model analysis of performance during manually controlled RVD for a hypothetical 'chaser'-'target' system, as illustrated in Fig. 1. The central aspect addressed here is the effect on performance of a communication time delay between the HO's control station and the RVD worksite and the improvement in performance which can be achieved through the introduction of display prediction. We have allowed the delay to range as an independent parameter of the analysis, between zero (representing RVD directed by the HO from within the chaser for example) and several seconds (representing RVD directed from the ground, with communication established via one or more DRS's and ground stations). The other factors which are examined here are the effects of multi-axis controlling and the effect of HO-injected disturbances. The direct manual control case has been chosen specifically in order to investigate the feasibility and limits of performance for this fundamental operational mode, since, in light of current progress in telepresence technology, manual control need not necessarily be regarded solely as a

mode of 'last resort'. Complete details of this analysis may be found in the reference by Milgram et al. (1984).
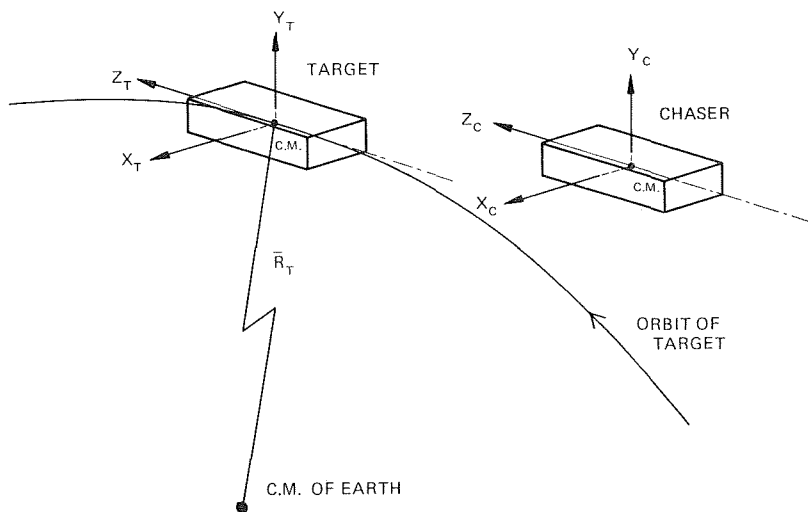

Fig. 1   Chaser-target reference frames.

2. OUTLINE OF ANALYSIS

A large number of early investigations into tracking performance in the presence of time delays have indicated that performance degrades rapidly as transmission delays increase, if continuous closed-loop tracking is attempted without some kind of mechanism for compensating for these delays. (Otherwise the HO will adopt an open-loop 'move-and-wait' control strategy.) Such mechanisms may be either extrinsic or intrinsic to the HO. Some common examples of extrinsic compensation devices include predictor displays, quickened displays, preview displays and 'flight director' displays. Even if no such external aids are supplied, the HO still possesses 'internal' information processing capabilities which act intrinsically to compensate for system delays, to an extent which depends upon the particular tracking situation (i.e. display characteristics, number of tracking axes, order and complexity of system dynamics, disturbance amplitude and bandwidth, etc.). This characteristic is modelled within the Optimal Estimator-Predictor part of the well known Optimal Control Model (OCM) (Kleinman, 1969; Kleinman et al, 1970).

In conventional applications of the OCM the HO is modelled specifically as being able, by means of the optimal predictor, to compensate for his/her own combined perceptual delays (along the order of 0.2 s). At what point the validity of such a delay compensation (sub)model breaks down for larger time delays, either intrinsic or extrinsic or combined, has not yet to our knowledge been carefully investigated. It will, as mentioned above, in any case depend on the characteristics of the task. In the analysis which follows, this aspect of the OCM has been extrapolated beyond its likely range of vali-

dity. In doing so we have not presumed that the HO is actually equipped with such inherent predictive capabilities. Rather, our first goal is to estimate an upper bound on performance, based on the usual assumed limitations of the human operator (observation noise, neuromotor noise) but excluding explicit perceptual delays (which have been neglected here relative to the much larger system transmission delays). By modelling a HO whose predictive capabilities are able to compensate optimally for extrinsic system transmission delays, what we obtain is an estimate of the best possible system performance, that is, the mean performance which might be expected when well-trained HO's are provided with an optimal predictor display.

Regarding such OCM results as forming a hypothetical upper bound on performance, given the constraints of the task and inherent limitations of the optimal predicting HO, it is convenient also to estimate a corresponding hypothetical lower bound on performance, based on exactly the same constraints, limitations and assumptions of optimality, but assuming that the HO performs no prediction. (The reason why this model is hypothetical is obvious: clearly the HO will always make some effort to compensate for system delay. The implication of not doing so is to presume that the HO zeroes system errors on the basis of currently displayed information, even though it is clear, on the basis of accumulated observations, that this is 'outdated' information.)

Finally, with respect to the above two cases, which collectively form a performance envelope for this analysis, we examine the case in which the HO is presented with a (simple) predictor display, which is designed to ameliorate tracking performance by performing the transmission time delay compensation extrinsically for the operator. By presenting the model results in this manner, i.e. in relation to the estimated performance envelope, it is clear i) what performance gains have been made by introducing the particular predictor display, and ii) what performance gains conceivably remain to be achieved with respect to optimal performance.

The optimal prediction modelling approach is outlined in the following section, and the no-prediction and predictor display analyses are described in section 4.

3. OPTIMAL PREDICTION MODEL

A schematic representation of the Optimal Control Model (OCM) as applied here is given in Fig. 2. In that figure both the uplink and downlink time delays, T_u and T_d, are indicated explicitly. In order to justify applying the OCM "as is" in the context of continuous tracking in the presence of communication time delays (and in the absence of extrinsic predictor aiding), we commence by postulating how such a "human optimal feedback controller" might conceivably behave under such circumstances. Assuming that, in addition to knowing the system dynamics and noise statistics, the HO also knows both the downlink and uplink delays, T_d and T_u, the essential elements of such a model are:

1. The HO receives noisy delayed display information, on the basis of which $\hat{\underline{x}}(t-T_d)$, an optimal estimate of the source of the T_d-delayed information from the remote system, is made.

2. The HO knows that if he/she were to generate a control command based upon an estimate of the present system state only, i.e. $\hat{\underline{x}}(t)$, such a command would arrive at the remote system at a time T_u too late. The HO must therefore generate a prediction of the future state of the remote system, i.e. $\hat{\underline{x}}(t+T_u)$, based upon past control inputs and past and present state estimates.

3. The HO generates a control signal, $\underline{u}_c(t)$, proportional to $\hat{\underline{x}}(t+T_u)$. The delayed input to the system in space, $\underline{u}_d(t)=\underline{u}_c(t-T_u)$, is the optimal control input.

On the basis of these hypotheses, and assuming stationarity, it can be shown that the 'conventional' approach to implementing the OCM can be used to analyse such optimal feedback regulation problems with up- and downlink delays simply by lumping together $T=T_u+T_d$ and substituting this delay into the standard OCM submodel of HO predictive compensation for internal perceptual time delays. In doing so we assume henceforth that the effects of the HO's own perceptual time delays are implicitly included within the total (lumped) system time delays.
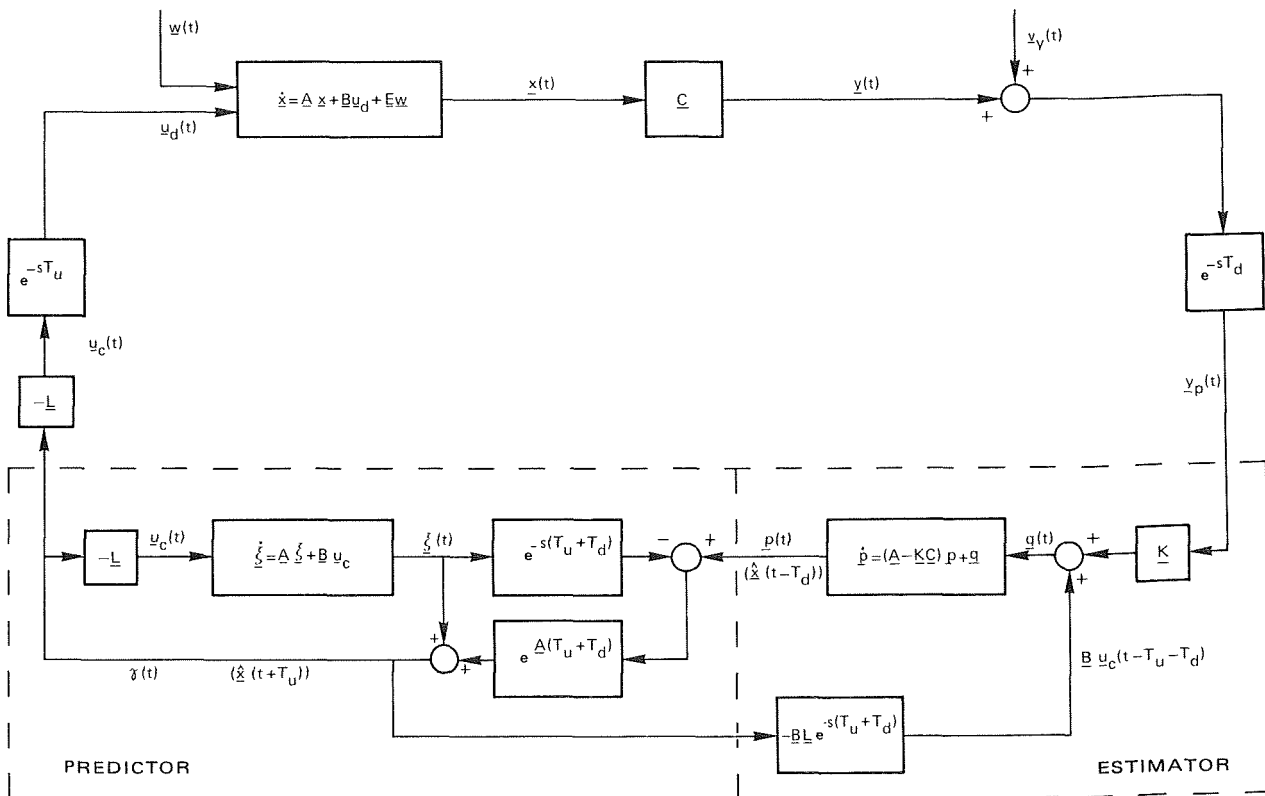

Fig. 2 Optimal feedback controller for system with up- and downlink time delay and observation noise.
(adapted from Kleinman, 1969)

27.5

4. NO-PREDICTION AND SIMPLE PREDICTOR DISPLAY MODELS

The essential difference between the above model and the model formulations used for the analysis of the no-prediction and the simple predictor display cases lies in the HO's knowledge of and response to the (lumped) system time delay. In the OCM analysis, where the optimal predictor is intrinsic to the HO, the HO is assumed to have perfect knowledge of the delay T. In the present two cases, however, prediction is either extrinsic (in the display) or completely absent. Since in these cases the HO is not required to have any knowledge of T, it is consequently assumed for the analysis that the HO has no knowledge of T. The reason for modelling these two cases in a similar fashion is that for both configurations the task of the operator is identical: to regulate out system disturbances on the basis of currently displayed information.

The absence of the HO's knowledge of T is a sufficient nonconformity from the conventional OCM structure to prevent us from employing the usual closed-form solution for ensemble average performance estimates. What is necessary is to formulate a model structure where the actual time delay is incorporated within the dynamic equations of the physical system, together with a model of the predictor display if there is one, but where the time delay is absent from the HO's internal model of that system. In other words, an analysis must be performed whereby the HO has an imperfect internal model of the physical system to be controlled.

As pointed out recently by Baron (1984), very little work has been done on modelling situations in which the system to be controlled is ill-defined for the HO. The approach taken here parallels that outlined in Baron & Berliner (1975) and the basic concepts are illustrated in Fig. 3. The formulation for the "real" system is expressed in the figure in the standard state space form as shown:

$$\dot{\underline{x}}(t) = \underline{A}\,\underline{x}(t) + \underline{B}\,\underline{u}_c(t) + \underline{E}\,\underline{w}(t) \tag{1}$$

where $\underline{u}_c(t)$ is the HO's command input and $\underline{w}(t)$ is the independent, gaussian white system disturbance. Since the "real" system includes all physical elements external to the HO, if there is any transmission delay in the system it will be included in the upper block in Fig. 3. The display matrix $(\underline{C})$, including any predictive display, is also part of that block. The display information corrupted by observation noise which is perceived by the HO is expressed by:

$$\underline{y}_p(t) = \underline{C}\,\underline{x}(t) + \underline{v}_y(t) \tag{2}$$

where $\underline{v}_y(t)$ is a gaussian, white noise. Note that for this analysis, as for the optimal prediction model above, we have neglected the human operator's own internal perceptual time delay.

Opposite the "real" system block in Fig. 3 is the HO's internal model of that system, which may or may not be the same, i.e. perfect. For the sake of generality the HO's internal model of the system is expressed in terms of a different state vector, $\underline{z}$, as shown:

$$\dot{\underline{z}} = \tilde{\underline{A}} \, \underline{z}(t) + \tilde{\underline{B}} \, \underline{u}_c(t) + \tilde{\underline{E}} \, \underline{w}(t) \tag{3}$$

where the symbol $\tilde{}$ is used to distinguish the internal HO model parameters from those corresponding to the "real" system. The dimension of the HO's $\underline{z}$ vector may or may not be the same as $\underline{x}$. Whereas the HO's internal representation, as defined by the HO's $\tilde{\underline{A}}$, $\tilde{\underline{B}}$, $\tilde{\underline{C}}$, $\tilde{\underline{E}}$ matrices, may differ from the real system, the conventional assumption that the HO has a perfect internal representation of the covariance of the independent disturbance, $\underline{w}(t)$, of the observation noise $\underline{v}_y(t)$ and of his own injected motor noise, $\underline{v}_u(t)$, is retained for this analysis.

Similar to the OCM description above for no transmission delay, the HO is assumed to estimate the current presumed system state, $\hat{\underline{z}}(t)$, on the basis of both observed and expected display information, according to:

$$\dot{\hat{\underline{z}}}(t) = \tilde{\underline{A}} \, \hat{\underline{z}}(t) + \underline{B} \, \underline{u}_c(t) + \tilde{\underline{K}} \, (\underline{C} \, \underline{x}(t) + \underline{v}_y(t) - \tilde{\underline{C}} \, \hat{\underline{z}}(t)) \tag{4}$$

where $\tilde{\underline{K}}$ is the HO's Kalman gain. Note that the bracketed expression on the right hand side of equation (4) is the difference between the current perceived information in equation (2) and the HO's expectation $\tilde{\underline{C}} \, \hat{\underline{z}}(t)$. Further-
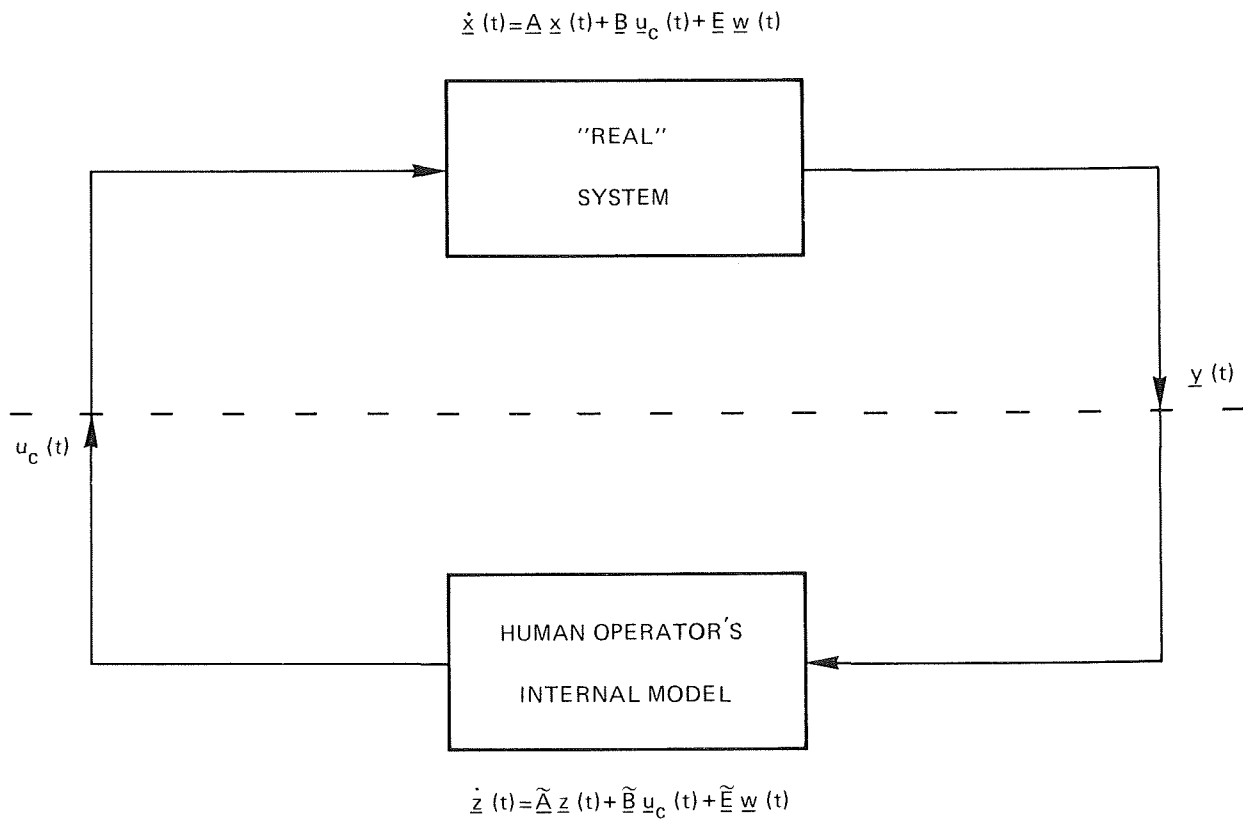
$$\dot{\underline{x}}(t) = \underline{A} \, \underline{x}(t) + \underline{B} \, \underline{u}_c(t) + \underline{E} \, \underline{w}(t)$$

$$\dot{\underline{z}}(t) = \tilde{\underline{A}} \, \underline{z}(t) + \tilde{\underline{B}} \, \underline{u}_c(t) + \tilde{\underline{E}} \, \underline{w}(t)$$

Fig. 3 Interface for imperfect internal model formulation.

27.7

more, as indicated also in Fig. 2, the HO is assumed to generate an optimal control command proportional to $\underline{\hat{z}}(t)$, given by:

$$\underline{u}_c(t) = -\underline{\tilde{L}} \, \underline{\hat{z}}(t) \tag{5}$$

which minimises a specified cost functional. The weighting factors which define this cost functional are assumed to be the same as for the optimal predictor case, since the goals of the task are the same for both cases. The $\underline{L}$ matrix must be computed on the basis of the $\underline{\tilde{A}}$ and $\underline{\tilde{B}}$ matrices, however, rather than on $\underline{A}$ and $\underline{B}$.

Substituting equation (5) into both equation (4) and equation (1), the results can be combined into a single linear system of matrix equations which describe the system in Fig. 3:

$$
\begin{bmatrix} \underline{\dot{x}}(t) \\ \\ \underline{\dot{\hat{z}}}(t) \end{bmatrix}
=
\begin{bmatrix} \underline{A} & -\underline{B}\,\underline{\tilde{L}} \\ \\ \underline{\tilde{K}}\,\underline{C} & \underline{\tilde{A}} -\underline{\tilde{B}}\,\underline{\tilde{L}} -\underline{\tilde{K}}\,\underline{\tilde{C}} \end{bmatrix}
\begin{bmatrix} \underline{x}(t) \\ \\ \underline{\hat{z}}(t) \end{bmatrix}
+
\begin{bmatrix} \underline{E} & \underline{0} \\ \\ \underline{0} & \underline{\tilde{K}} \end{bmatrix}
\begin{bmatrix} \underline{w}(t) \\ \\ \underline{v}_y(t) \end{bmatrix}
\tag{6}
$$

Assuming stationarity, the covariance of the combined $[\underline{x} \;\; \underline{\hat{z}}]'$ vector can be solved with conventional linear matrix operations.

The objective of incorporating the transmission time delay, T, within the actual system equations can easily be achieved by means of a linear Padé approximation. In the following analysis a second order Padé filter has been introduced at the output of the "real" system, that is:

$$\frac{o(s)}{i(s)} = \frac{1 - (1/2)Ts + (1/12)T^2 s^2}{1 + (1/2)Ts + (1/12)T^2 s^2} \tag{7}$$

which implies that, for an input $i(t)$ to the filter, $o(t) \sim i(t-T)$. This results in $\underline{\tilde{A}}$ and $\underline{\tilde{B}}$ matrices which are sub-matrices of $\underline{A}$ and $\underline{B}$.

Since the HO part is modelled identically for the predictor display, no-prediction and OCM cases, the HO's $\underline{\tilde{C}}$ matrix is identical for each. In this analysis no explicit display format has been examined, i.e. display vectors = observable state vectors. For the no-prediction case the $\underline{C}$ matrix merely defines the delayed outputs as displays. To define the $\underline{C}$ matrix for the predictor case, a simple second order truncated Taylor series has been used for generating a displayed prediction of the system output component $x(t)$:

$$y(t) = x(t) + T\,\dot{x}(t) + T^2/2\,\ddot{x}(t) \sim x(t+T) \tag{8}$$

Because third derivative information was unavailable, the observed rate of change of the predicted display is approximated by:

$$\dot{y}(t) = \dot{x}(t) + T\,\ddot{x}(t) \sim \dot{x}(t+T) \tag{9}$$

27.8

5. OUTLINE OF RVD FINAL APPROACH

The final approach phase of RVD is described in more detail in Milgram et al (1984). In summary, during station-keeping of the chaser at an aim point about 1000 m from the target, several activities are carried out on board both spacecraft, involving equipment checkout, readying of docking mechanisms, determination of relative position and attitude, etc. Upon receipt of a command from the OCC the chaser initiates the acquisition phase of RVD. The purpose of this phase is to bring the chaser from the aim point to a standoff point on the docking axis of the target, typically some 200 m from the target, upon which the chaser again engages in station-keeping and system checkout.

Upon receipt of another command from the OCC, the translation phase begins. The chaser now moves along the nominal docking axis of the target towards another standoff point some 20 m away from the target. Here further checks are carried out while the chaser is involved in station-keeping. The chaser then undergoes a series of controlled accelerations, decelerations and coasts, and finally achieves physical contact with the target, with carefully controlled relative translational and rotational errors and related rate errors.

In order to analyse this case it is clear that the various deceleration and acceleration manoeuvres from an initial to a final constant velocity constitute a terminal control problem. In the present RVD case, however, it has been specified that these manoeuvres are deterministically programmed for each flight profile. We therefore concentrate on the problem of regulating out disturbances, or perturbations, about the preprogrammed nominal flight profile and about the relative chaser-target orientation during constant velocity coasting. The problem of HO-mediated terminal controlling in RVD is nevertheless an important topic for future study.

The motion of each spacecraft (i.e. chaser and target) can be described in terms of translational motion of its centre of mass and rotational motion around its centre of mass. Roughly speaking, translation deals with position; rotation deals with orientation. In order to derive the equations of motion of relative position and relative orientation, certain assumptions have been introduced, specifically:
- the target moves in a near circular orbit around the Earth,
- the target is Earth-stabilised,
- the target docking axis lies along the principal axis of the target; in the nominal case this points in the direction of the orbital velocity vector (Fig. 1),
- the chaser reference frame is approximately aligned with the target reference frame; i.e. lateral position errors and their rates are small, orientation errors and their rates are small (Fig. 1),
- the chaser docking axis lies along the principal axis of the chaser.

These simplifications allow the relative translational perturbation dynamics to be expressed linearly as:

$$\ddot{x} = a_x - \omega_o^2 x$$

$$\ddot{y} = a_y + 3\omega_o^2 y + 2\omega_o \dot{z} \qquad\qquad (10)$$

$$\ddot{z} = a_z - 2\omega_o^2 \dot{y}$$

where ω_o is the orbital angular velocity and a_x, a_y, a_z are scaled thrust accelerations (in this case, maximum value = 0.01 m/s²) along the respective axes. For positional control the goal is to reduce x,y (lateral errors) and z (deviation from programmed axial relative closure profile) and their derivatives to zero. For a circular low earth orbit of 500 Km, ω_o = 1.1 rad/s. Substituting this into equation (10) it can be demonstrated that all terms involving ω_o in the perturbation equations (10) are negligibly small, given the maximum thrust acceleration magnitudes of 0.01 m/s², whence it may be shown that the translational dynamics are effectively uncoupled. In the following, therefore, analyses are performed for one representative generic tracking axis from the uncoupled system of translational perturbation dynamics.

Turning to the rotational dynamics, it is assumed that the target is stabilised with respect to the orbital reference frame. The attitude motion of the chaser relative to the target is therefore given by:

$$\ddot{\theta} = m_x \, , \qquad \ddot{\psi} = m_y \, , \qquad \ddot{\phi} = m_z \qquad\qquad (11)$$

where θ, Ψ, ϕ are the angles of orientation of chaser with respect to target and m_x, m_y, m_z are scaled rotation control accelerations (in this case, maximum value = 1°/s²). Since the goal of attitude control is to zero the three uncoupled orientation angles, which have been assumed to be small, equations (11) may clearly be regarded as perturbation dynamics. Also for the analysis of rotational control, therefore, one representative generic tracking axis has been chosen.

In Table 1 the nominal limits on state deviations for the generic translational and rotational tracking axes are given. For translational control these limits are range (R) dependent, as shown. The values selected for this analysis have been indicated by an asterisk. These limits, which emphasise rate of change as opposed to positional deviation, define the control laws in the ensuing model analyses. (No other range dependent parameters have been assumed here. In particular, display outputs have been assumed equal to system state outputs. Had visual display cues been modelled explicitly, then range dependence would necessarily have to have been taken into account in this context.)

The specifying of the magnitude and statistical properties of external disturbances to this dynamic vehicular system is less straightforward, since most common 'terrestrial' factors, such as turbulence in the air or bumps on the road, are not present in space. The principle sources of noise which were assumed are:
i) fluctuations in the thruster outputs and thruster control system,
ii) cross-coupling between rotational and translational control systems,
iii) fluctuations in target attitude due to limit cycling in the attitude control system.

MAXIMUM POSITION MISALIGNMENT	* 0.5	(m)	(R ∿ 5 m)
	0.1	(m)	(R ∿ 1 m)
	0.02	(m)	(R ∿ 0.2 m)
MAXIMUM VELOCITY DEVIATION	* 0.01	(m/s)	(R ∿ 5 m)
	0.002	(m/s)	(R ∿ 1 m)
	0.0004	(m/s)	(R ∿ 0.2 m)
MAXIMUM ATTITUDE MISALIGNMENT	* 1.0	(deg)	
MAXIMUM ANGULAR VELOCITY DEVIATION	* 0.05	(deg/s)	

Table 1   Nominal limits on translational and rotational state deviations
for generic chaser-target system.

Another, more unconventional, independent disturbance was assumed: (motor)
noise introduced to the control system by the HO and which, due to the large
time delays, propagates throughout the system and becomes effectively indepen-
dent of the other state variables. In the following all independent system
disturbances have been lumped and modelled collectively as a low-pass gaus-
sian noise with bandwidth 0.2 rad/s and covariance equal to 1.5% of the
related maximum thrust and maximum torque, for translation and rotation res-
pectively. (The effect of varying bandwidth has also been analysed, but is
not presented here.)

Another 'problem' associated with analysing such space propulsion sys-
tems is the bang-bang nature of control inputs, i.e. a thruster is either on
or off. Such systems do not particularly lend themselves to straightforward
linear, stationary analysis. However, if the thruster control logic is con-
structed such that command inputs are translated into trains of discrete
firing pulses whose frequency determines the net effective thrust output
(i.e. PFM, or pulse frequency modulation), it is possible to treat the HO's
control input to the thrusters as quasi-linear and quasi-continuous. Such a
PFM control logic was assumed in the following.

6. MODEL RESULTS

In Fig. 4 and 5 are shown the OCM results for translational and rota-
tional motions respectively. In both figures the standard deviation of the
positional component is shown on the left and of the velocity component on
the right. The second independent parameter in both figures is the HO's
motor noise-to-signal ratio, P_u, representing the relative amount of noise
(in dB) injected by the HO into the system via his/her control actions. The
reason for allowing P_u to vary in this fashion is due to uncertainty about
precise levels of external disturbance, $\underline{w}(t)$, to the system. Since, as men-

tioned above, we are assuming that the HO is a potential source of approximately independent noise, the effect of different disturbance levels has been investigated in this fashion.

Since the noise levels examined here are relatively low, the performance for 'nominal' levels of P_u = -20 dB is quite stable in Fig 4 and 5; the HO/optimal predictor-controller is able to regulate the system quite well, even up to 10s delay. As relative noise level increases, however, performance becomes rapidly more divergent. This effect is more pronounced for rotational control, where noise levels do not go beyond P_u = -8 dB.

Relating these results to Table 1, we note that the 3σ levels in Fig. 4a and b remain well below the specified limits of 0.5 m and 0.01 m/s respectively over the ranges shown, for P_u = -20 dB and -10 dB. Comparing these to the rotational results in Fig. 5, however, we see that the 3σ levels exceed the specified maximum attitude misalignment and angular velocity deviation at approximately T = 1s and T = 0s respectively, for a (noisy) P_u of -10 dB. Clearly, the relative state and control weightings and comparative independent disturbance noise level for rotational control are such that this is a more difficult control task than translational control.

In both Fig. 4 and 5 the performance results are for one representative axis out of the three which are being simultaneously tracked. A 'full' attention level (P_o) of -17 dB has been assumed for each task (Baron, 1984). In Fig. 4 attention is evenly allocated across positional and velocity components; in Fig. 5, on the other hand, an optimal distribution of attention has been used. A separate analysis has confirmed, however, that due to low sensitivity in this region, the effective difference between the two approaches here is very slight.

In Fig. 6 and 7 are shown the results of varying the number of axes of tracking, i.e. 1, 3 or 6 axes. This has been simulated by means of varying the relative fraction of 'full' attention allocated across the various display outputs (eg. see Baron, 1984). The results are qualitatively similar for both translational and rotational performance. The important conclusion to be drawn from these results is that, although performance decrements in the direction expected as the human optimal estimator-controller is required to divide attention across increasingly more task dimensions, this performance decrement is not very large, that is, for the particular independent and dependent noise conditions which have been assumed. On the other hand, it can be expected that, as noise levels increase, the effect of multi-axis tracking will become more dramatic. This is because for higher noise levels the HO's uncertainty about the state of the system will become relatively greater more quickly. The consequence of this is that new displayed information becomes more important as expectations based on past observations become more unreliable. If under such circumstances the HO is required to allocate attention over more axes, the updating of display information will fall behind, total uncertainty will increase and performance will deteriorate.

OCM results from Fig. 4 and 5, for the intermediate case P_u = -14 dB, have been plotted in Fig. 9 and 10, together with the model results for the

no-prediction and predictor display analyses described above in section 4. In order to arrive at these results, a rather lengthy empirical attention optimisation procedure was followed, which is illustrated here in Fig. 8, for translational control. In that figure, where the selected optimal attention allocation is indicated by arrowheads, we see that, for increasing system time delay, velocity display information (equation 9) becomes less reliable and the HO must pay increasingly more attention to positional information (equation 8). Further details of the iterative algorithm necessary to generate the results in Fig. 9 and 10 for constant P_u are given in Milgram et al (1984).

Referring to Fig. 9 and 10, it must be noted that the abscissae differ in scale from those of Fig. 4 and 5. Note as well in Fig. 9 that for T = 1s numerical inaccuracy in the predictor and no-prediction estimates is indicated by a separate symbol.

As expected, Fig. 9 and 10 indicate best performance for the optimal predictor, worst performance for the no-prediction case and intermediate performance for the Taylor predictor display. It is perhaps surprising, in contrast to what might otherwise be suggested from previous experimental evidence, that the no-prediction performance has been maintained at all within the 3s range before diverging. The explanation for this can be shown to derive from the specific optimal control laws which have been computed and which have the equivalent effect of a large HO lead compensation. Evaluation of the validity of such control laws must explicitly take into account, however, the HO's visual thresholds for the observation of velocity information, which is necessary for realising the prescribed feedback control.

The principle factor underlying the control performance here, therefore, is the proportionately large weight assigned to minimising velocity deviations relative to positional misalignments, as indicated in Table 1. Indeed, we note that on the right hand sides of Fig. 9 and 10, i.e. for velocity deviations, the curves shown much more closely the expected pattern of rapid divergence of the no-prediction case as T increases and stabler performance for the Taylor predictor case. Clearly, a 'better' predictor display than the simple display defined in equations (8) and (9) would generate less rapidly increasing system output errors and would thus be able to extend the controllable time delay range even further, the limit of course being an 'optimal' predictor display, whose performance is indicated by the OCM curves.

7. CONCLUDING REMARKS

In this paper some factors related to the control of rendezvous and docking of two spacecraft in low earth orbit by a 'remote' human operator have been dealt with. In general, the remote control of systems in space, especially in the presence of large transmission delays, has long been recognised as a task which is ill-suited for the unaided human controller, and thus as a task which should be as fully automated as possible. As the need for more flexibility during scheduled and unscheduled operations grows, however, so will the need for more onsite 'intelligence'. One potential way to

bring the HO 'closer' to the remote sight is to compensate for transmission time delays by means of predictor displays (of all relevant sensory information). The model results presented in this paper provide an initial indication of some of the improvements in performance which may be gained through the use of such displays.

This paper has also attempted to illustrate the usefulness of adapting and applying existing human performance models for the analysis of this relatively unexplored class of human operator control problems. Further analyses are necessary in order to investigate the effects on performance, for example, of different external disturbance characteristics, different system dynamics and various advanced display concepts, including other predictor displays and integrated display formats such as perspective displays, preview displays and director displays. In addition to the application of existing models, new modelling approaches must be developed, including improved 'imperfect internal model' formulations, terminal control applications and open-loop 'move-and-wait' control models. The ultimate goal of these developments is to combine the use of skill-based behaviour models with models of cognitively more complex rule-based, and eventually knowledge-based, supervisory control behaviour, in order to be able systematically to analyse and evaluate a large range of potential teleoperator design alternatives and operational procedures.

8. REFERENCES

Akin, D.L., M.L. Minsky, E.D. Thiel and C.R. Kurtzman (1983): Space applications of automation, robotics and machine intelligence systems (ARAMIS), Phase II. NASA CR-3736.

Baron, S. (1984): Adaptive behaviour in manual control and the optimal control model; in Adaptive Control of Ill-defined Systems, O.G. Selfridge, E.L. Rissland & M.A. Arbib (ed's), Plenum Press, 51-73.

Baron, S. and J.E. Berliner (1975): MANMOD 1975: Human internal models and scene-perception model. U.S. Army Missile Command, TR RD-CR-76-3.

Kleinman, D.L. (1969): Optimal control of linear systems with time-delay and observation noise. IEEE Transactions on Automatic Control, Vol. AC-14, 524-526.

Milgram, P., P.Th.L.M. van Woerkom and P.H. Wewerinke (1984): Control loops with human operators in space operations; Part III: Rendezvous and docking operations and model analysis of performance with human-in-the-loop. (Netherlands) National Aerospace Laboratory, NLR Report TR 84116 L, Part III.
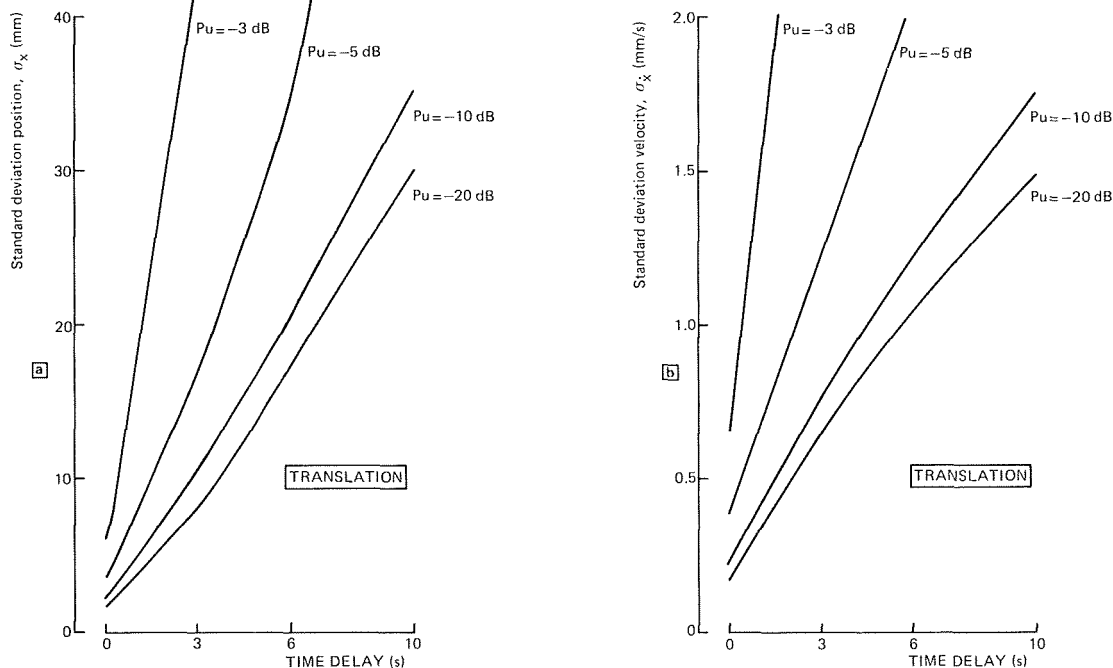
Fig. 4   Translational control: $\begin{cases}\text{(a) Position deviation}\\\text{(b) Velocity deviation}\end{cases}$ vs time delay

For various motor noise levels (P_u)
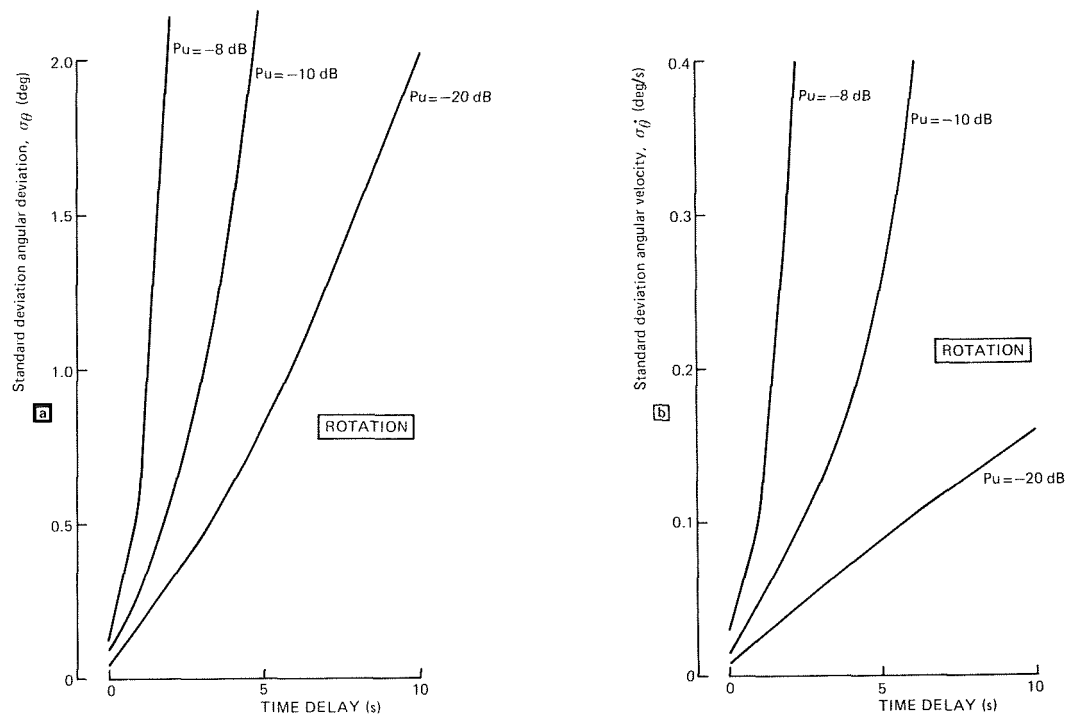(3-axis tracking: P_o = -17 dB, $f_1 = f_2$ = 1/6)

Fig. 5   Rotational control: $\begin{cases}\text{(a) Angular deviation}\\\text{(b) Angular velocity}\end{cases}$ vs time delay

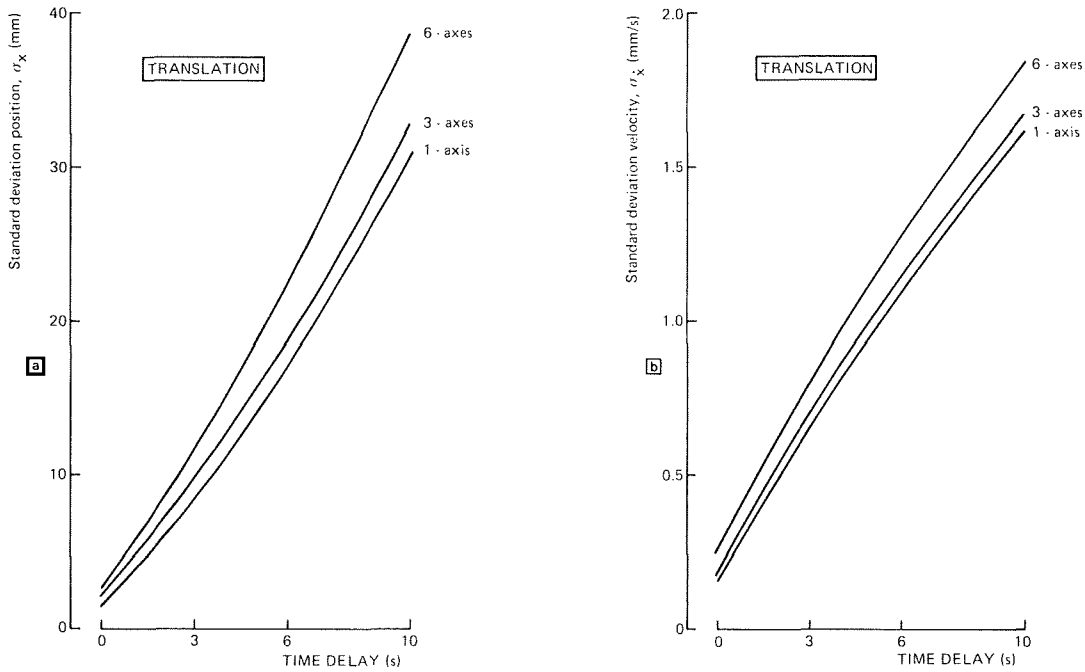For various motor noise levels (3-axis tracking:
P_o = -17 dB, optimal attention)

27.15

Fig. 6 Translational control: $\begin{cases}\text{(a)} & \text{Position deviation} \\ \text{(b)} & \text{Velocity deviation}\end{cases}$ vs time delay
For 1-, 3-, 6-axes of tracking.
(Optimal attention allocation, $P_u = -10$ dB)
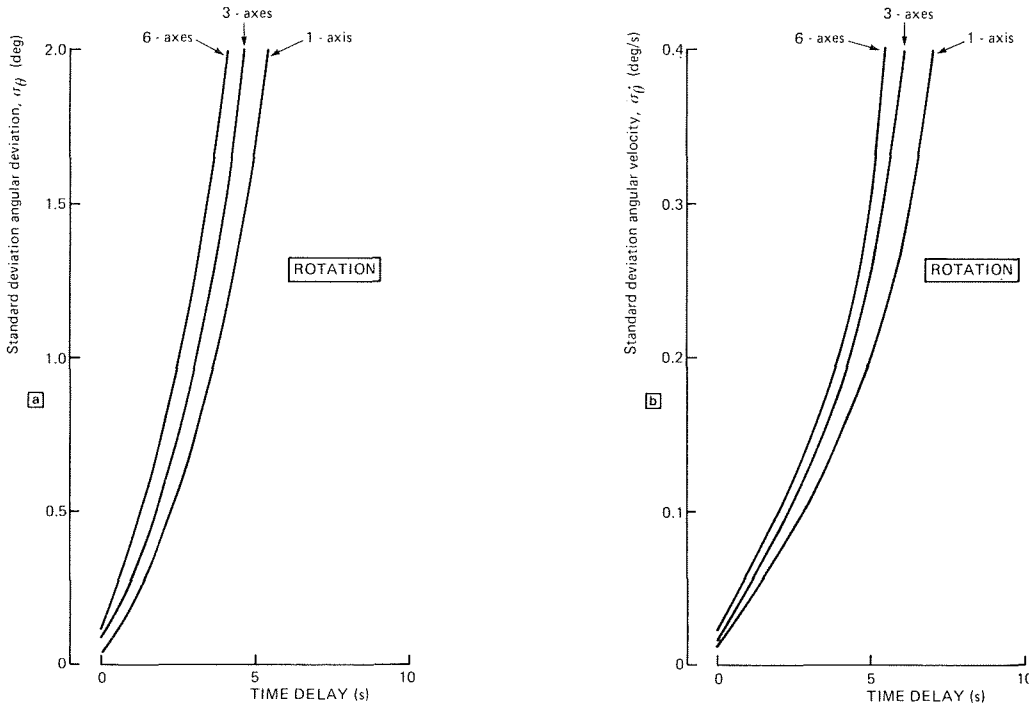
Fig. 7 Rotational control: $\begin{cases}\text{(a)} & \text{Angular deviation} \\ \text{(b)} & \text{Angular velocity}\end{cases}$ vs time delay
For 1-, 3-, 6-axes of tracking.
(Optimal attention allocation, $P_u = -10$ dB)

27.16

C - 5
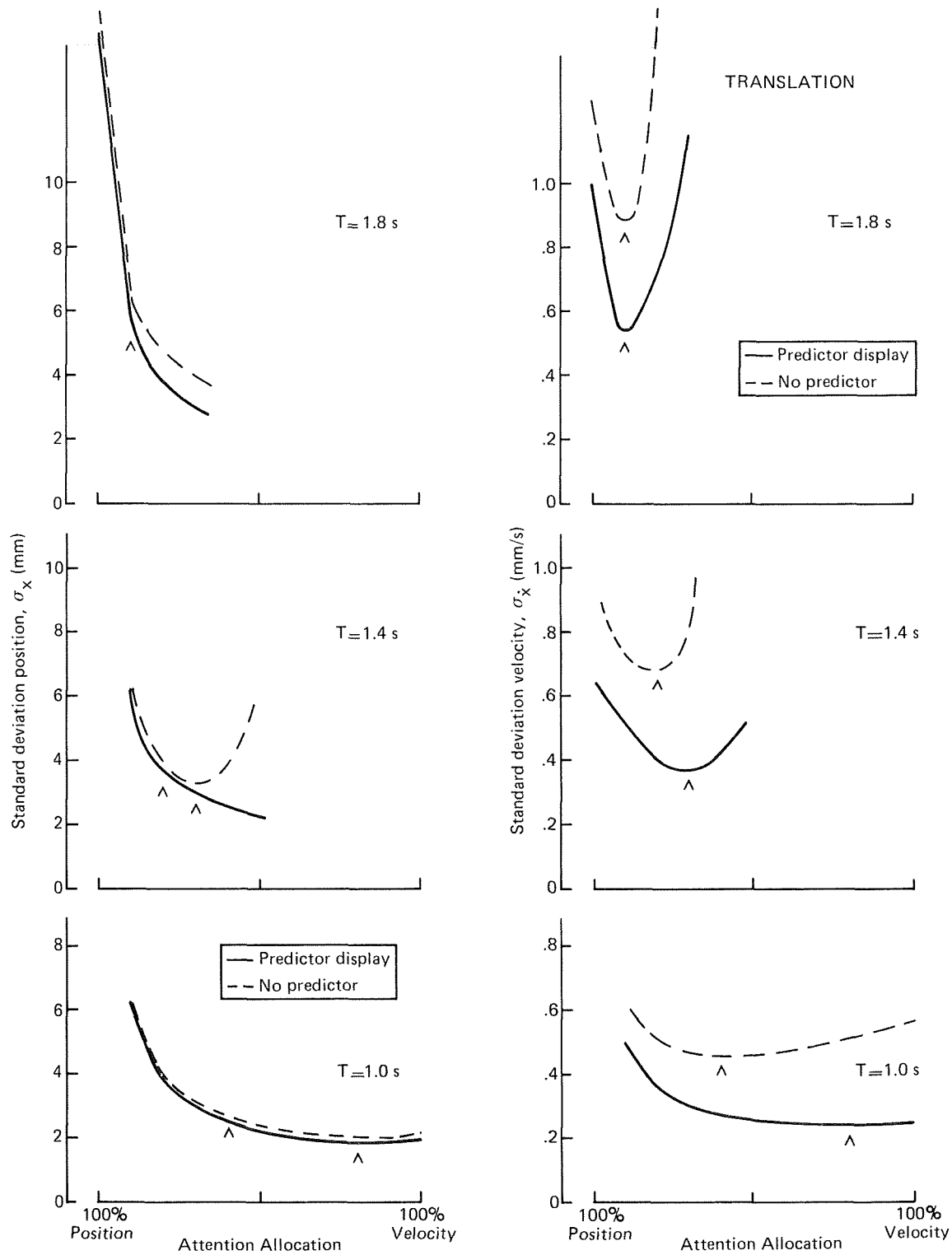
Fig. 8 Translational control: predictor and no-predictor display performance vs. attention allocation strategy for three time delays (3-axis tracking, $p_o = -17$ dB, $f_1 + f_2 = 1/3$)
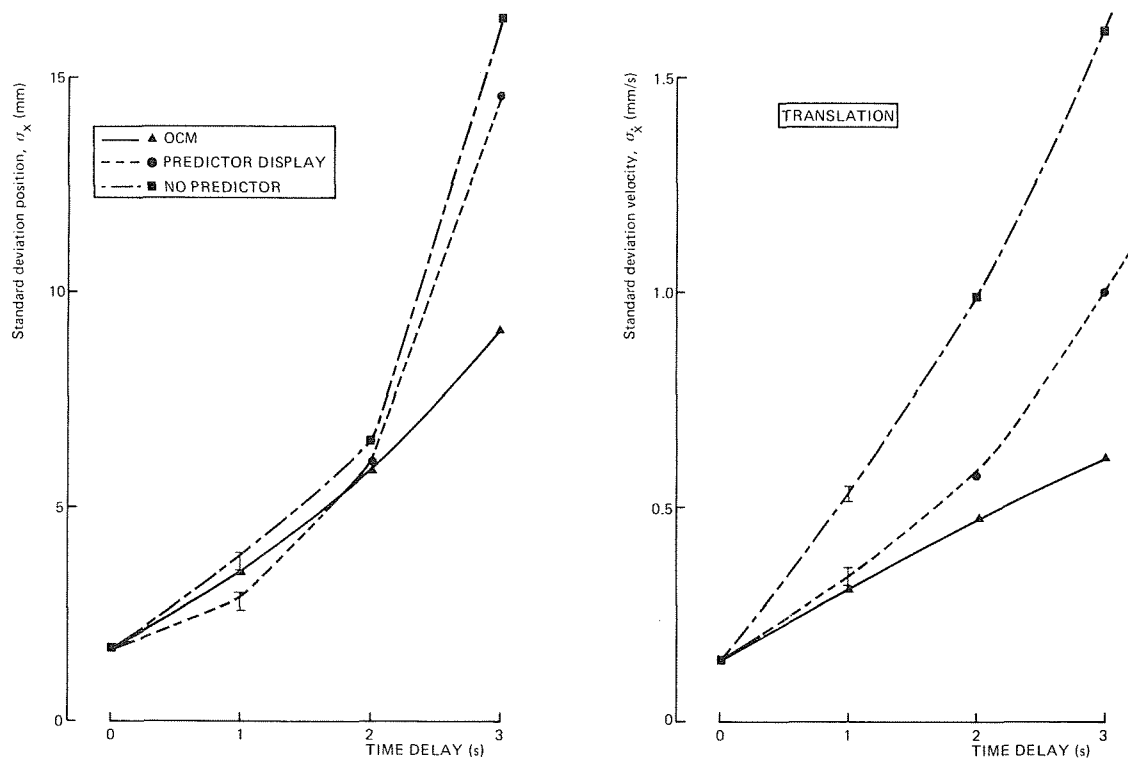
Fig. 9 Translational control: comparison of OCM, predictor display and no-predictor performance (3-axis tracking, optimal attention. $P_o = -17$ dB, $P_u = -14$ dB)
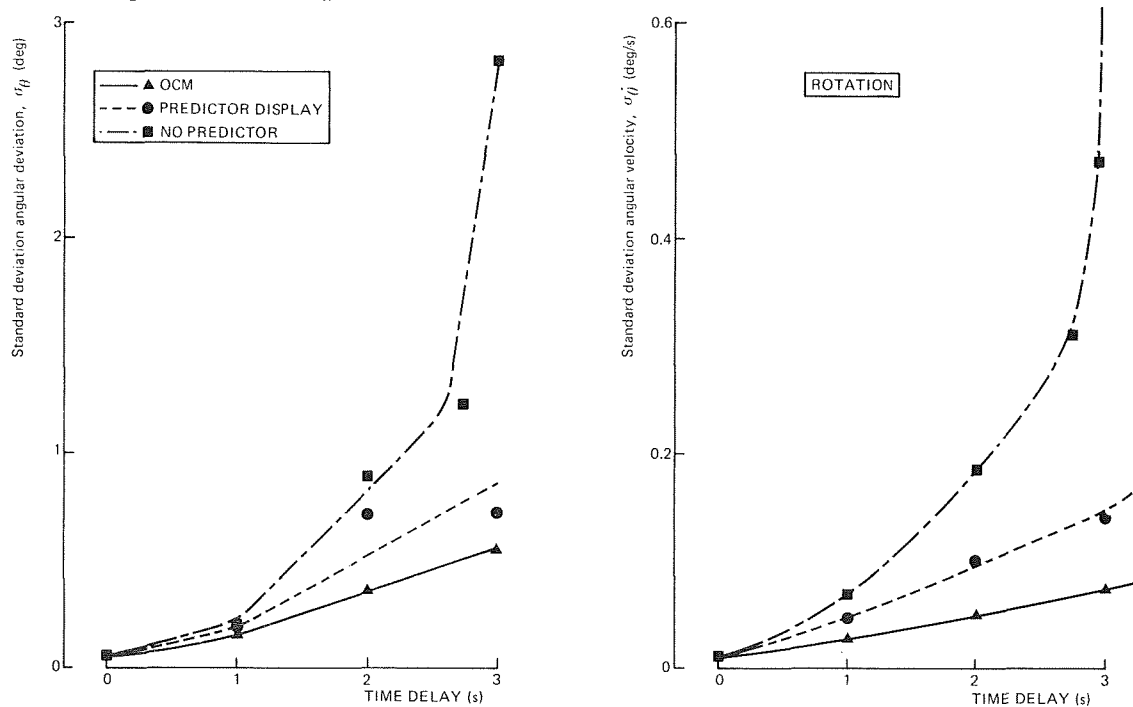
Fig. 10 Rotational control: comparison of OCM, predictor display and no-predictor performances (3-axis tracking, optimal attention, $P_o = -17$ dB, $P_u = -14$ dB)

A Model for the Human's Use of Visual Field
Cues in Nap-of-the-Earth Flight

Ronald A. Hess and Kam Chan
Division of Aeronautical Science and Engineering
Department of Mechanical Engineering
University of California, Davis, California

A simple model is developed which describes the manner in which
the human pilot may use visual field cues in controlling a vehicle in
nap-of-the earth flight. The model is based upon the feedforward of
information obtained from streamer patterns in the visual field to the
inner-most loop of a multi-loop pilot/vehicle model. In this framework, the
model is a logical extension of pursuit and preview models of the human
operator which have appeared in the literature. Simulation and flight test
data involving low-level helicopter flight tasks are applied to model
development and validation.

VIRTUAL SPACE AND TWO-DIMENSIONAL EFFECTS
IN PERSPECTIVE DISPLAYS

Michael Wallace McGreevy [1]

Cordell R. Ratzlaff [2]

Stephen R. Ellis [3]

ABSTRACT

When interpreting three-dimensional spatial relationships presented on a tw   ensional display surface, the viewer is required to mentally reconstruct the original informaticn. This reconstruction is influenced by both the perspective geometry of the displayed image and the viewer's eye position relative to the display. In a study which manipulated these variables, subjects judged the azimuth direction of a target object relative to a reference object fixed in the center of a perspective display. The results support a previously developed model which predicted that the azimuth judgement error would be a sinusoidal function of stimulus azimuth. The amplitude of this function was correctly predicted to be systematically modulated by both the perspective geometry of the image and the viewer's eye position relative to the screen. Interaction of the two components of our model, the virtual space effect and the 3D-to-2D projection effect, predicted the relative amplitudes of the sinusoidal azimuth error functions for the various conditions of the experiment. Mean azimuth judgements in some directions differed by as much as 25 degrees as a result of different combinations of eye position and image geometry. Our results illustrate the need to consider the effects of perspective geometry when designing spatial information instruments, and show our model to be a reliable predictor of average performance.

INTRODUCTION

An important result of the diffusion of computer technology into aerospace applications is a growing interest in new display methods (Getty, 1982; Jauer and Quinn, 1982; Roscoe, Corl and Jensen, 1981; Warner, 1979). Imaginative airbrushed artists' conceptions of proposed pictorial displays which are to replace the instument panels of futuristic aircraft and spacecraft are increasingly common in industry publications. Some researchers have even proposed that the traditional distinction between the outside scene and the panel instruments be replaced with a virtual scene that integrates information in a new, more interpretable format, one which can be spatially configured in any desired fashion.

Whether these proposals can be transformed into practical flight instruments remains to be demonstrated, of course. The task will require that the design of spatial information displays be based on human performance measures, so that the advertised improvement in interpretability is achieved.

Many information transfer questions are raised by spatial displays, and we have attempted to address a question raised in our work on airborne traffic displays. As part of a NASA/FAA study of airborne traffic display formats, McGreevy and Ellis developed a perspective format which was shown to be superior to planview formats for separation maintainence tasks (Ellis, McGreevy, and Hitchcock, 1984). What was not clear at the time, however, was whether the particular perspective parameters we had used in our research display were optimal for accurate spatial information transfer.

[1] NASA Ames Research Center, MS 239-3, Moffett Field, CA 94035.

[2] San Jose State University, San Jose, CA 95192

[3] NASA Ames Research Center, MS 239-3, Moffett Field, CA 94035.

In an exploratory study of direction judgements in similar perspective displays, we found that azimuth error is a sinusoidal function of azimuth direction, and that the amplitude of the sinusoid is modulated by the perspective of the image. From these results, we proposed a model which seemed to be able to account for the sinusoids. That model was tested in the experiment described in this paper.

In this experiment, we tested conditions that included some which are similar to those in the previous exper as well as some that are very different. In particular, we made predictions based on our model for conditions in which one component of our model was literally turned upside down. The results, even in these conditions, confirm the model.

Our model consists of two components, the virtual space effect and the 3D-to-2D projection effect. These are mathematical functions which represent suspected influences on direction judgements. They are derived from a combination of image geometry, viewing geometry, and some proposed interpretive behaviors. The 3D-to-2D projection effect arises from reasonable expectation that the judged magnitude of an angle depicted in a 3D scene will be influenced by the magnitude of the 2D projection of that angle in the perspective image. The virtual space effect is the result of a hypothesized interpretive behavior in which observers of perspective images assume that the geometry of the depicted space is like that seen through a window. If, however, the eye of the observer is not at the geometrically correct point, this assumption will lead to predictable errors. The two effects comprising the model are described in detail in McGreevy and Ellis (1985). Using our model, we have predicted how the visual angle subtended by a pictorial display screen and the geometric field of view of the displayed image influence direction judgements within the displayed scene.

METHOD
Subjects

Twelve male commercial pilots ranging in age from 29 to 62 served as subjects. Their flight experience varied from 8 to 45 years. Subjects were obtained through the NASA Ames Research Center subject pool and were paid for their participation.

Apparatus and Stimuli

The stimulus images were slides of computer generated perspective scenes which were rear-projected onto a large screen (104 cm square). These images were abstracted from a spatial display format (Figure 1) that has been developed and used in air traffic display research studies at NASA Ames (McGreevy, 1982; McGreevy, 1983; Ellis, et al., 1984; McGreevy and Ellis, 1985). Stimulus scenes consisted of a grid plane and two cubes. The "reference cube" always appeared in the center of the display while the "target cube" was displayed at various positions around the reference cube. The target cube was always at the same altitude as the reference cube, and lines connected each cube to the grid, as shown in Figure 2. Ninety-six different perspective images were used in the experiment: for each of four image geometries, the target cube was depicted in twenty-four different azimuth directions.

The perspective scenes were photographed directly from an Evans & Sutherland Picture System monitor. A Kodak carousel slide projector was used to project the images onto the screen which was positioned at various distances directly in front of the subject. An adjustable chair and chinrest kept the subject's central line of sight fixed at the center of the screen while allowing the subject to sit in a comfortable position. Subjects responded by using a stylus and digitizer pad to manipulate an angle indicator dial which appeared on a computer graphics display next to the projection screen. Programs to generate the dial image and record subjects' judgements ran on a PDP-11/40 computer under the RSX-11M operating system.

Design

The experiment utilized a fully crossed, within subjects design. Each subject was presented with a total of 384 stimulus images, viewing 96 images from each of four different distances (194, 90, 52, and 30 cm). The 96 images consisted of 24 scenes, each of which was calculated with four different geometric fields of view (30°, 60°, 90°, and 120°). Each of the 24 images depicted the target cube in one of 24 azimuth directions. This design allowed each subject to view depictions of 24 different directional stimuli under 16 combinations of image geometry and viewing distance, so that the viewer made direction judgements while his

eyepoint was at four different positions relative to four different geometric station points.

Figure 3 shows all sixteen eye point/geometric station point relationships. The station point is at the apex of each of the four triangles whose base is the screen, and is defined as the point through which all projectors pass. It is the mathematical analog of a pinhole lens, through which all imaged light rays pass. The angle at the apex is the geometric field of view, which we also refer to as the geometric FOV, and is encoded in the figures as g30, for example, to label the case of a geometric field of view of 30°. The visual subtense of the screen as seen from the eye position is the eye field of view, or eye FOV, and is labelled in the diagrams as e30, etc.

Figure 4 is a three-dimensional figure which shows the geometry of a 3D scene and corresponding 2D stimulus image, which is similar to the geometries used in this experiment. Each triangle of Figure 3 represents a top-view of the tip of a frustrum like that in Figure 4, which is a geometric analog of the cone of vision.

Target cube direction was reported in terms of the azimuth angle between the zero azimuth axis and the bearing of the target cube (see Figure 2). Judgement error was defined as the difference between the actual 3D angle depicted in the display and the judged angle. A positive azimuth error represents a clockwise (CW) error, where the response is clockwise in azimuth relative to the stimulus. For example, a positive error of 10° would result if a stimulus at 60° resulted in a response of 70°. A negative azimuth error represents a counterclockwise error.

Procedure

Each subject received instructions and was shown how to operate the equipment for recording judgements. Several practice trials were administered to ensure that the subject understood the task completely. The subject wore an eye patch over his non-dominant eye and made judgements while his chin was positioned in the chinrest, allowing control over the position of the subject's eye. To reduce extreme angles of eye movement and possible strain, the subject was allowed to swivel his head in the chinrest when looking from the screen to the angle indicator dial.

The task consisted of viewing a stimulus scene, manipulating the angle in the angle indicator dial until the subject felt it best represented the angle between the two cubes in the stimulus scene, and then activating a switch to record the judgement. Immediately after the judgement was recorded, the subject was presented with the next trial. The subject received no feedback concerning the accuracy of his judgements.

The experiment was comprised of 16 blocks of 24 trials. Stimulus scenes were randomly assigned to blocks and the order in which the blocks were viewed was randomized and counterbalanced for each subject. After a subject completed a block of trials, the screen was moved to a different distance. This allowed the subject a short rest period and helped prevent eye fatigue at the closer screen distances. At the halfway point of the experiment a longer rest break was provided. Total time for the experiment was approximately three and one-half hours.

RESULTS

The ANOVA results indicate that the three-way interaction of stimulus azimuth, geometric field of view, and eye field of view is statistically significant ($F=2.051$; $df=207,2277$; $p<0.0005$). Thus, the sixteen plots of the means which correspond to the sixteen field of view conditions of the experiment (Figure 5a) are significantly dissimilar. Based on results of a previous experiment (McGreevy and Ellis, 1984; McGreevy and Ellis, 1985), we had applied the 2D effect and virtual space effect to predict the nature of the individual plots of the azimuth error means which comprise the three-way interaction. The discussion section contains a detailed comparison of the predictions and results.

The two-way interactions, which are averages across either eye FOV or geometric FOV, are less useful for validating the model, but give insight into performance which is common to a particular class of conditions. The two-way interaction of geometric FOV and stimulus azimuth (Figure 5b) is significant ($F=18.257$; $df=69,759$; $p<0.0005$). The two-way interaction of eye FOV and stimulus azimuth (Figure 5c) is also significant ($F=6.790$; $df=69,759$; $p<0.0005$).

The so-called main effect of azimuth, which is an average across both eye FOV and geometric FOV, is even less useful in terms of the model,

29.3

and is included only for completeness. The main effect of azimuth (Figure 5d) is significant $(F=2.847; df=23,253; p=0.0005)$.

DISCUSSION

Error Function Equations

The plots of the three-way interaction of stimulus azimuth, geometric field of view, and eye field of view are distinctly sinusoidal. It is useful to fit analytic functions to the raw data so that the trends among the conditions of the experiment, as seen in the three-way interaction, may be described quantitatively.

In order to obtain an estimate of the shapes of the analytic functions, we fit polynomials of various degrees to the raw error data for each of the sixteen conditions of the experiment. The squared error of fit was reduced significantly for polynomials of degree greater than five, and polynomials of degree six to nine produced nearly identical plots. We used the shape of each sixth order polynomial to obtain estimates of the coefficients of a function consisting of a sine curve plus a line. These coefficients include the amplitude, frequency, and phase shift of the sine curve and the slope and intercept of the line. The estimated equations were input to a BMDPAR program (derivative-free non-linear regression) which adjusted the coefficients to obtain the "sine plus line" function for each condition which minimized the sum of the squared error of fit to the raw data.

The coefficients are shown in Tables 1-5 and are plotted next to each table, and the equations are shown in Table 6. Figure 6a shows the plots of the fitted analytic functions compared with the plots of the three-way interaction.

The **amplitudes** of the fitted sinusoidal azimuth error functions vary systematically among the conditions of the experiment (Table 1). For example, the average amplitude of the sinusoidal error is 12.34 ° when the image has a narrow geometric FOV of 30 ° (g30) and it is viewed such that it subtends a very wide visual angle of 120 ° (e120). At the other extreme, the average amplitude of the sinusoidal error is -6.72 ° when the image has a very wide geometric FOV of 120 ° (g120) and it is viewed such that it subtends a narrow visual angle of 30 ° (e30).

Notice that the minimum amplitudes of error are not obtained in those cases where the

eye is at the station point (ie. when the eye FOV equals the geometric FOV). For example, when the eye FOV is 30 ° (e30), the minimum amplitude among conditions tested is obtained with a geometric FOV of 60 ° (g60). This agrees with results of our previous experiment (McGreevy and Ellis, 1984; McGreevy and Ellis, 1985).

The **angular freqency** and **phase shift** of the sinusoidal azimuth error functions determine the azimuth directions which will be the peaks and valleys of the error functions. The frequency coefficients, Table 2, seem to be randomly scattered close to a value of 2.00 cycles of error function per 360 ° of target azimuth direction, for most conditions of the experiment.

In order to compare the phase shifts of functions with negative amplitudes with those whose functions have positive amplitudes, a -90 ° shift is added to the those phase shifts whose functions have negative amplitudes. This adjustment assumes a frequency of 2.00 cycles per 360 ° of target azimuth direction. Both the adjusted and unadjusted values are shown in Table 3. Phase shift shows a distinct pattern among the conditions of the experiment. In general, the error functions are shifted in the positive azimuth direction (clockwise) for the 30 ° geometric FOV (g30), and increasingly counterclockwise for the wider geometric fields of view. The effect is most pronounced for the eye FOV of 30 ° (e30). The effect decreases and shifts to the positive direction as eye FOV increases.

The **slope** of the linear component of the sinusoidal azimuth error function is near zero for all but the case of a geometric FOV of 30 ° (g30). In this case, the slope becomes more negative as the eye FOV increases. This can also be seen in the four curves of the g30 case in Figure 6a. The **intercept** is greatest, for all geometric fields of view, when the eye FOV is 30 ° (e30) and least for the eye FOV of 120 ° (e120). Note that in cases where the slope is zero, which is approximately true for all but the g30 case, the intercept is just a 'vertical' offset of the sinusoidal azimuth error function away from the zero error line.

3D-to-2D Projection Effect

The 3D-to-2D effect, or 2D effect for short, is a geometrical relationship which, we believe, influences viewers of 2D perspective images when they make angular judgements concerning the displayed 3D space. The magnitude of the effect,

for a given angle, is equal to the difference between the 2D angle on the image plane, and the 3D angle it represents. The effect is a function of image geometry, and when the plane of the angle is constant relative to the image plane, it varies in magnitude as a function of the size of the angle and the geometric field of view. These functions are shown in Figure 6b, for the conditions and image geometry of this experiment, as dashed lines. Note that the magnitude of the effect, for a given geometric FOV, is the same for all eye fields of view, and that the effect is strongest where the geometric FOV is smallest.

Virtual Space Effect

The virtual space effect is a geometrical relationship which is based on a suspected interpretive behavior. We have proposed (McGreevy and Ellis, 1984; McGreevy and Ellis, 1985) that viewers of perspective images make what we call the "window assumption," assuming that they are at the station point, and that they then distort the 3D space, creating a virtual 3D space, to conform to that assumption. The magnitude of the effect, for a given angle, is equal to the difference between the virtual 3D angle and the actual 3D angle. It is a function of the same image parameters as the 2D effect. and is also a function of the angular difference between the eye FOV and the geometric FOV of the image. Thus, there is no virtual space effect when the eye is at the station point, and the effect increases in magnitude as the distance between the eye and station point increases. The virtual space effect functions for the conditions and image geometry of this experiment are shown as solid lines in Figure 6b.

Combined Influence of the Two Effects

When the eye is closer to the screen than the station point, as when the eye FOV is greater than the geometric FOV, the virtual space effect and the 2D effect exert influences in the same direction. In this case, we say that the two effects are **in conjunction,** and that the virtual space effect is conjunctive with the 2D effect.

When the eye is farther from the screen than the station point, as when the eye FOV is less than the geometric FOV, the virtual space effect and 2D effect exert influences in opposite directions. In this case, we say that the two effects are **in opposition,** and that the virtual space effect is opposing the 2D effect.

Since there is no virtual space effect when the eye is at the station point, only the 2D effect is influential in these cases (according to our current model).

Predictions Confirmed

We predicted that the azimuth error functions would be sinusoidal, since the 2D effect and virtual space effect are sinusoidal, and this was borne out by the results of this experiment. The angular frequency of the error function was expected to be about 2 cycles of error per 360° of stimulus azimuth, since this is the frequency of the modelled effects, and this, too, was supported by the results. The amplitudes of the sinusoidal azimuth error functions were found to agree in great detail with those predicted by the expected interplay of the 2D effect and virtual space effect.

The following discussion relates information in three figures, Figure 3, in which the eye positions and geometric station points are graphically depicted and the predicted influences are explicitly noted; Figure 6a, which has the plots of the mean errors comprising the three-way interaction of stimulus azimuth, geometric field of view, and eye field of view, as well as the fitted sinusoidal error functions; and Figure 6b, with the virtual space effect and 2D effect functions which predict the azimuth error. Note that all three of these figures are in the same spatial format so that, for example, the upper right element in each of the figures represents the condition where the geometric field of view is 30° and the eye field of view is 120°.

Eye FOV = 30°

The four conditions in which the eye FOV was 30° (e30) involved geometric fields of view of 30° (g30), 60° (g60), 90° (g90), and 120° (g120). This set of conditions is quite similar to that used in our previous experiment, where the geometric fields of view were the same and the eye field of view was 18°; the results confirm those of the previous study. The amplitude of the error function is large and positive (6.82° error) when both the eye FOV and the geometric FOV are 30° (e30,g30), since the 2D effect is strong and the virtual space effect is zero. As the geometric FOV increases, the amplitude decreases, then reverses in sign, and then increases in the negative direction since the 2D effect becomes weaker and is gradually overcome by the increasing strength of the opposing virtual

space effect. Finally, when the geometric FOV reaches 120° (g120), with the eye FOV still 30° (e30), the amplitude of the error function reaches its largest negative value (-6.72° error), since the 2D effect is weakest and the opposing virtual space effect is strongest in this condition.

Eye FOV = 120°

The largest positive amplitude (12.34° error) of the sinusoidal error function occurs, as predicted, when the geometric FOV is 30° (g30) and the eye FOV is 120° (e120), since both the the 2D effect and the conjunctive virtual space effect are at their strongest. As the geometric FOV increases, with the eye FOV still 120°, the amplitude of the error function decreases since both the 2D effect and the conjunctive virtual space effect become weaker. Finally, when the geometric FOV reaches 120° (g120), with the eye FOV still 120° (e120), the amplitude is very small (1.97° error) since the virtual space effect is zero and the 2D effect is at its weakest.

Geometric FOV = 30°

In the set of conditions for which the geometric FOV is 30° (g30) and the eye FOV has values of 30° (e30), 60° (e60), 90° (e90), and 120° (e120), the 2D effect is at its strongest. Error amplitude changes from a large positive value (6.82° error) when the eye FOV is 30° (e30), to a very large positive value (12.34° error) when the eye FOV is 120°. This is as predicted, since the increasingly strong conjunctive virtual space effect adds its influence to the already strong 2D effect.

Geometric FOV = 120°

In the set of conditions for which the geometric FOV is 120° (g120) and the eye FOV has values of 30° (e30), 60° (e60), 90° (e90), and 120° (e120), the 2D effect is at its weakest. Error amplitude diminishes from a large negative value (-6.72° error) when the eye FOV is 30° (e30), to a small value (1.97° error) when the eye FOV is 120°. This is as predicted, since the gradually weakening magnitude of the opposing virtual space effect adds its influence to the weak 2D effect.

Eye FOV = Geometric FOV

In the four conditions where the eye FOV is the same as the geometric FOV the eye is positioned at the station point. For this reason, the virtual space effect is zero and only the 2D effect is influential. The largest error amplitude occurs when eye FOV and geometric FOV are both 30° since the 2D effect is strong when the geometric FOV is 30°. As both the geometric and the eye fields of view increase to 120°, the error amplitude decreases since the 2D effect becomes weaker as the geometric FOV approaches 120°.

Other Issues

In the two cases where the magnitudes of the opposing virtual space effect and the 2D effect are nearly identical, the greater strength of the 2D effect overcomes the opposition. These two conditions are those where the geometric FOV is 60° and the eye FOV is 30° (g60,e30) and where the geometric FOV is 90° and the eye FOV is 60° (g90,e60).

In our previous experiment, we found that the 2D and virtual space effect functions better matched the error data when they were both shifted counter-clockwise 22°. We suspected that this was caused by the fact that our zero azimuth axis, from which 3D azimuth judgements were measured, was rotated 22° counter-clockwise from straight ahead into the depicted scene. For that reason, we tried the opposite rotation in this experiment, and correctly predicted that the 2D and virtual space effect functions would best represent expected errors if they were correspondingly shifted 22° clockwise. This is how the two effects are plotted in Figure 6a.

Exceptions

While all of the predictions above apply to variations in the amplitude of the error function, no prediction was made regarding the optimum combination of the 2D effect and the virtual space effect. We have assumed, based on previous experimental results, that the two effects have positive weights, and that they are additive in some sense. It would appear that the relative weights of the two effects vary with stimulus azimuth. For example, in the condition where the geometric FOV is 30° and the eye FOV is 120° (see Figure 6), we correctly predicted that the two effects would combine to produce a sinusoidal error function with a large amplitude, but it is clear that the *varying* amplitude of the virtual space effect is not reflected in the data.

SUMMARY

Pictorial spatial instruments will continue to emerge in aerospace applications. Advanced computational and display technology will provide a *tabula rasa* for display designers with great potential for improvement in human-machine interaction. This great freedom, however, creates new and more difficult questions about information transfer. As more onboard systems are automated, mission operators will require a different class of instruments than those traditionally† used, in order to maintain overall situational awareness in complex and dynamic operational environments.

Our work in airborne traffic display research led us to study spatial information transfer issues related to the use of perspective display formats. In particular, we have studied how within-display-space direction judgements are affected by perspective geometry. We discovered that azimuth error is a sinusoidal function of stimulus azimuth and that the amplitude of the error function is modulated by the perspective of the image and the viewer's eye position relative to the display. To explain this result, we have developed a model which combines virtual space and 3D-to-2D projection effects. In this experiment, the model has been shown to be a reliable predictor of the amplitude of the error function under a wide variety of image geometries and viewing conditions.

REFERENCES

Ellis, S. R., McGreevy, M. W., and Hitchcock, Robert J. (1984, April). Influence of a perspective cockpit traffic display format on pilot avoidance maneuvers. In *Proceedings of the AGARD Aerospace Medical Panel Symposium on Human Factors Considerations in High Performance Aircraft*. Williamsburg, Virginia.

Getty, D.J. (Ed.). (1982, January). *3-D Displays: Perceptual research and applications to military systems*. Washington, D.C.: National Academy of Sciences.

Jauer, R.A., and Quinn, T.J. (1982, February). *Pictorial formats, vol 1: Format development* (Tech. Report AFWAL-TR-81-3156). Wright-Patterson AFB, OH: Flight Dynamics Laboratory.

Jones, L.F., Schrader, H.J., and Marshall, J.N. (1950, April). Pictorial display in aircraft navigation and landing. *Proceedings of the I.R.E.*, 391-400.

McGreevy, M.W. (1982, June). A perspective display of air traffic for the cockpit. In *Proceedings of the 18th Annual Conference on Manual Control* (pp. 514-521). (Tech. Report AFWAL-TR-83-3021), Wright-Patterson AFB, OH: Flight Dynamics Laboratory.

McGreevy, M.W. (1983). *A perspective display of air traffic for the cockpit*. Unpublished M.S. Report, University of California, Berkeley.

McGreevy, M.W., and Ellis, S.R. (1984, June). Direction judgement errors in perspective displays. In *Proceedings of the 20th Annual Conference on Manual Control* (pp. 531-549). NASA Conference Publication 2341.

McGreevy, M.W., and Ellis, S.R., (1985). *Format and basic geometry of a perspective display of air traffic for the cockpit*. (Tech. Memorandum 86680). Moffett Field, CA: NASA Ames Research Center, Aerospace Human Factors Research Division.

McGreevy, M.W., and Ellis, S.R., (1985). *The effect of perspective geometry on judged direction in spatial information instruments*. Submitted for publication.

Roscoe, S.N., Corl. L., and Jensen, R.S. (1981, June). Flight display dynamics revisited. *Human Factors, 23*, 341-353.

Warner, D.A. (1979, June). *Flight path displays* (Tech. Report AFFDL-TR-79-3075), Wright-Patterson Air Force Base, OH: Flight Dynamics Laboratory.

---

† For an early pictorial instrument idea which was ahead of the technology, see Jones, Schrader, and Marshall, 1950.
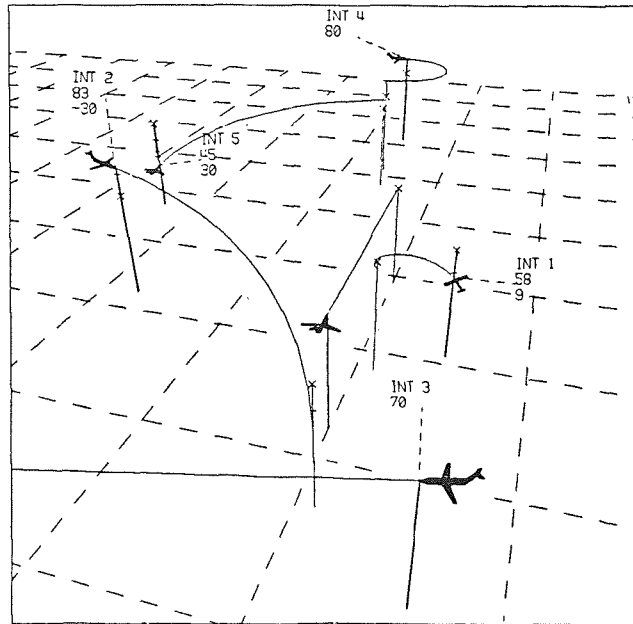
Figure 1. A perspective display of air traffic for the cockpit with ownship shown at the center of the image.
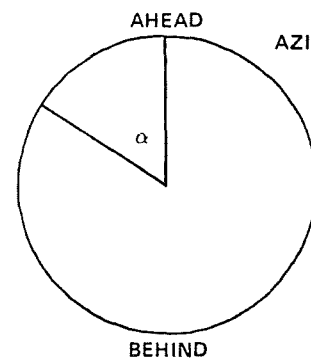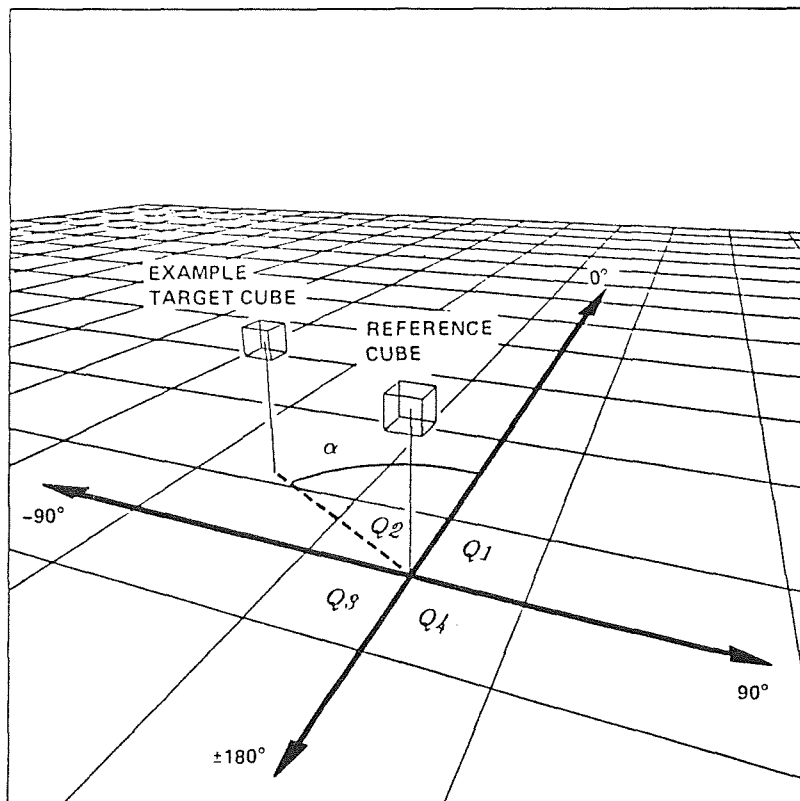

Figure 2. Diagram of a typical stimulus image. Bold axis lines, dashed line, angle arc, and text were not included in actual stimulus images. Response dial appeared on a separate screen.
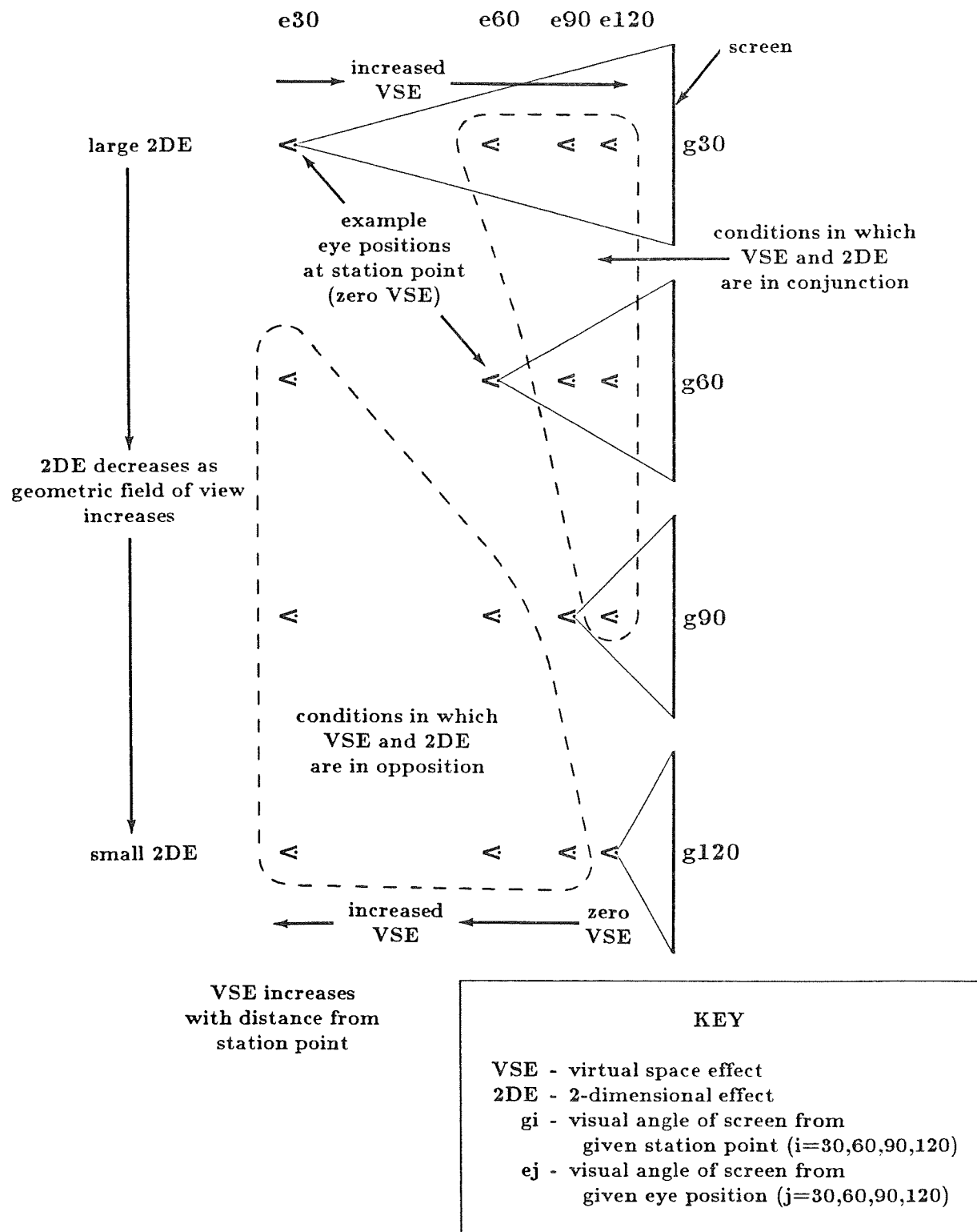
e30   e60   e90 e120

screen

increased
VSE

large 2DE

g30

example
eye positions
at station point
(zero VSE)

conditions in which
VSE and 2DE
are in conjunction

g60

2DE decreases as
geometric field of view
increases

g90

conditions in which
VSE and 2DE
are in opposition

small 2DE

g120

increased
VSE

zero
VSE

VSE increases
with distance from
station point

KEY

VSE - virtual space effect
2DE - 2-dimensional effect
  gi - visual angle of screen from
        given station point (i=30,60,90,120)
  ej - visual angle of screen from
        given eye position (j=30,60,90,120)

Figure 3. Conditions of the experiment: eye positions are crossed with geometric fields of view and shown relative to the screen (drawn to scale).
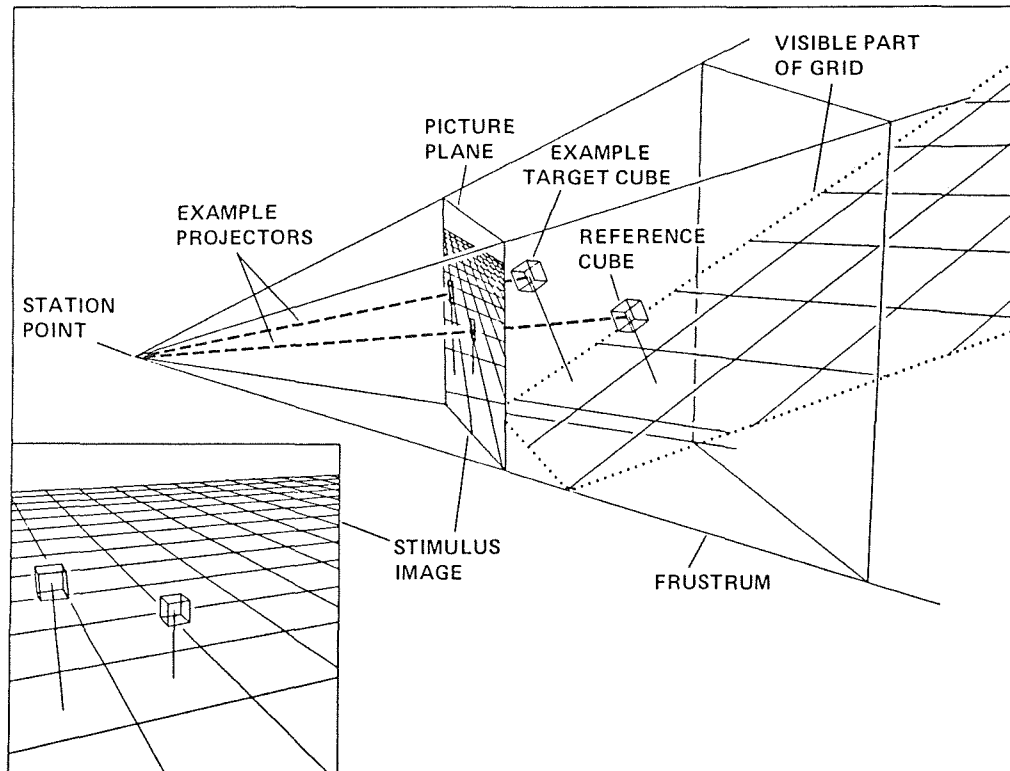
Figure 4. Example stimulus geometry showing relationship between 3D information and 2D projection.
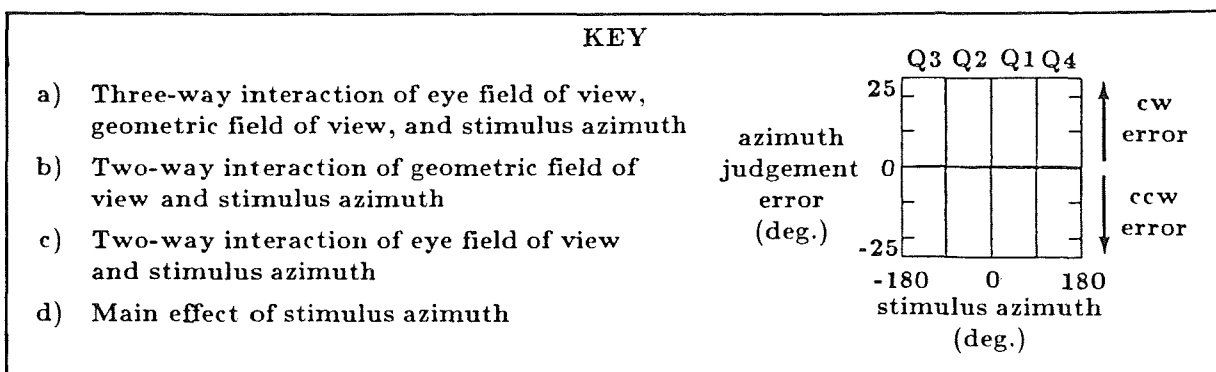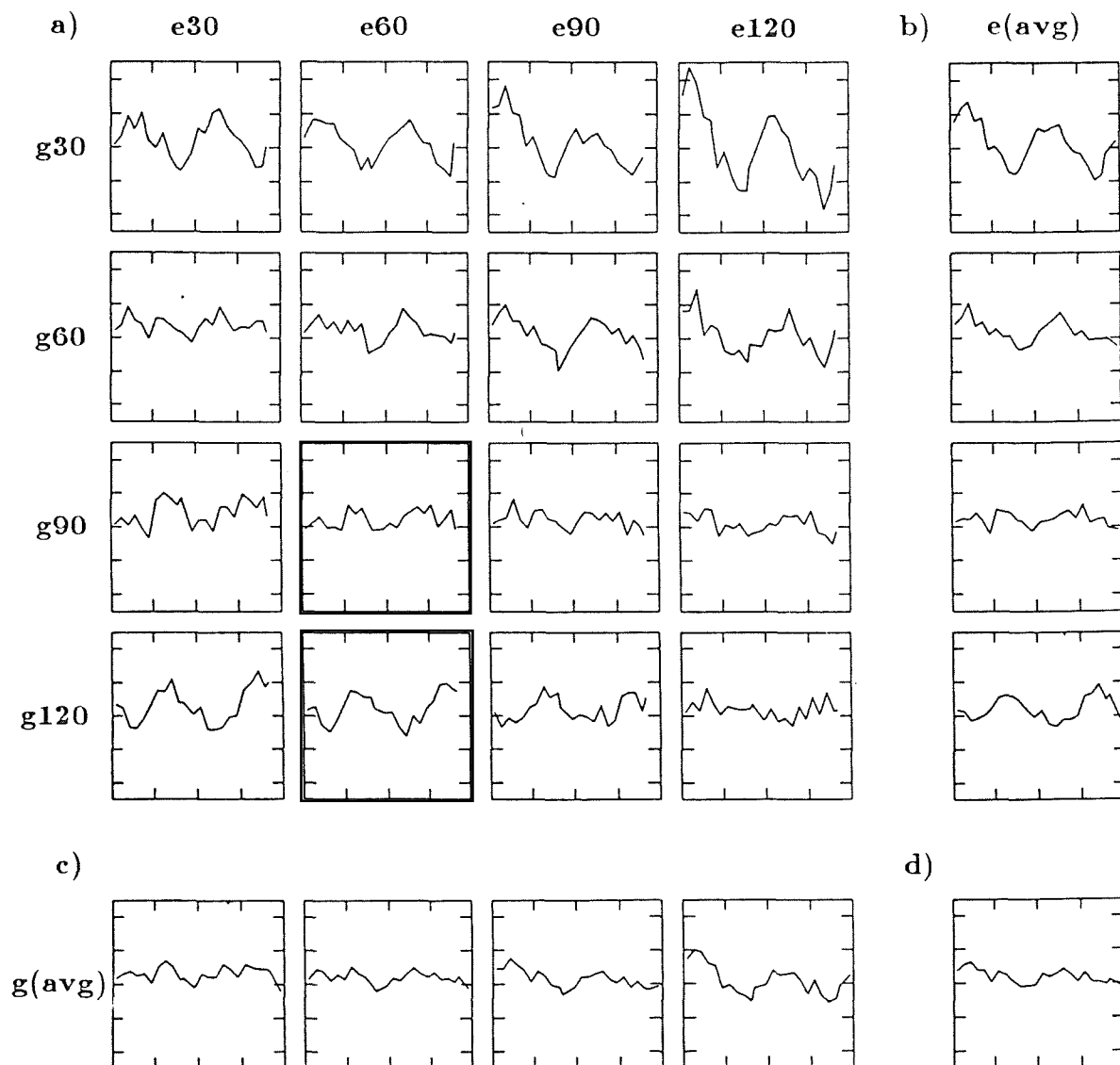
Figure 5. Average azimuth judgement error as a function of stimulus azimuth for the various perspective and viewing conditions of the experiment. Quadrants labelled in key correspond to those in Figure 2.
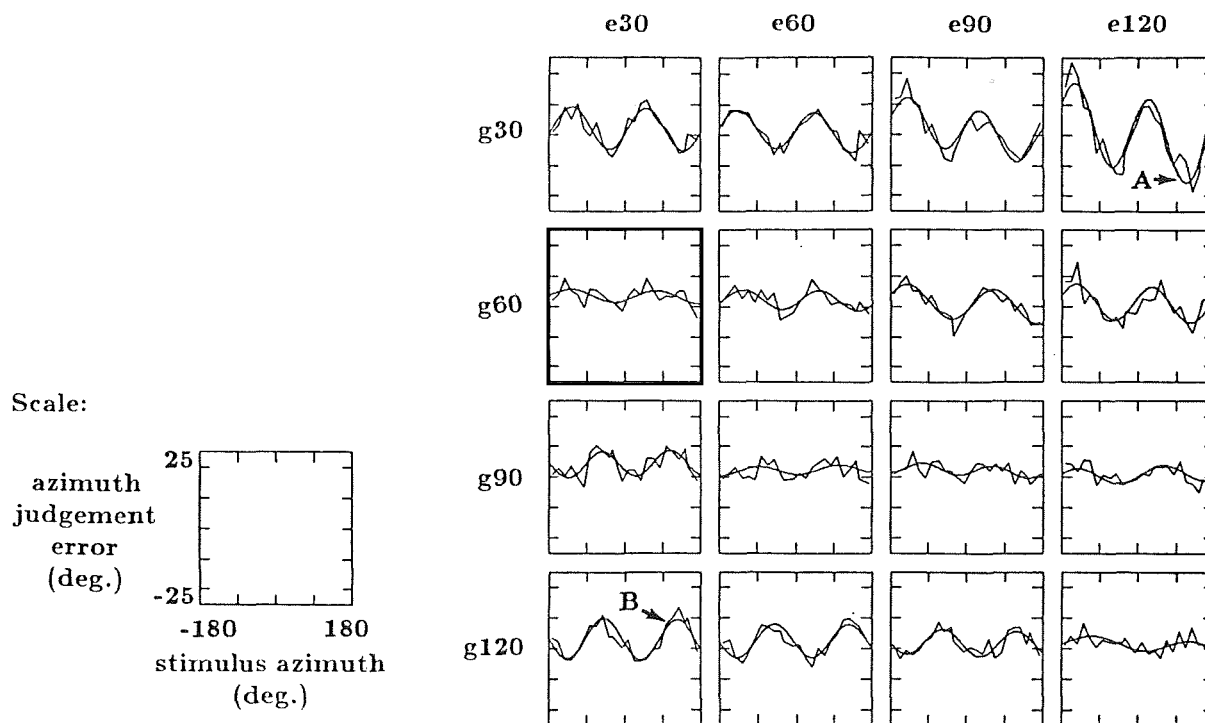
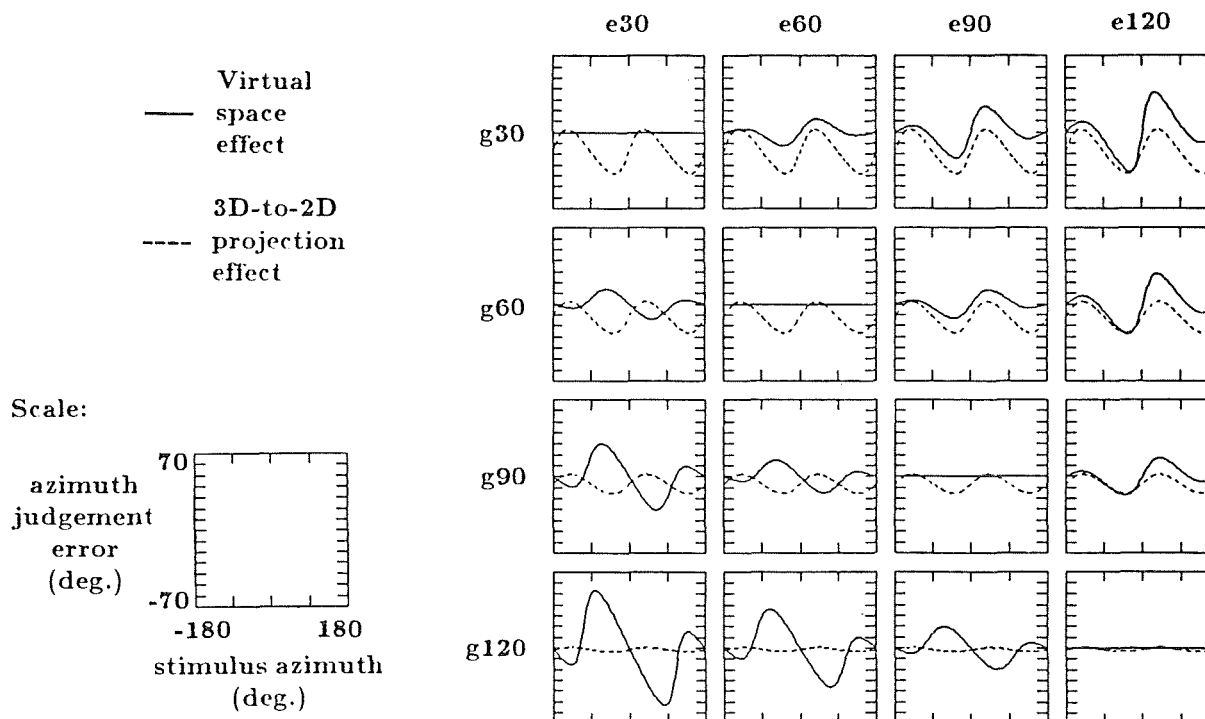Figure 6a. Mean azimuth error and fitted functions. Note that errors at A and B differ by about 25 degrees.

Figure 6b. Virtual space effect and 3D-to-2D projection effect difference functions for conditions of the experiment.
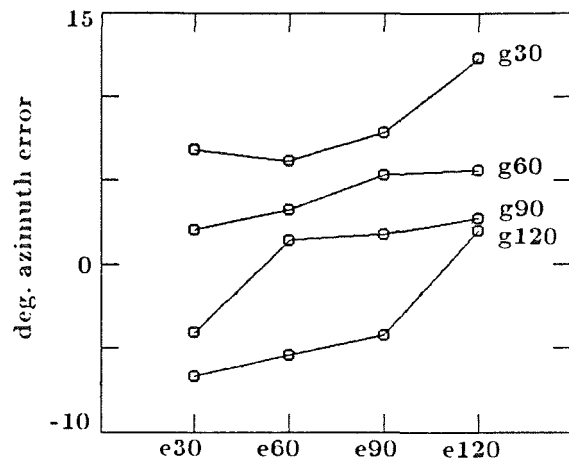
Table 1. Amplitude.

FOV	e30	e60	e90	e120
g30	6.82	6.15	7.28	12.34
g60	2.01	3.23	5.30	5.72
g90	-4.06	1.41	1.73	2.72
g120	-6.72	-5.45	-4.25	1.97
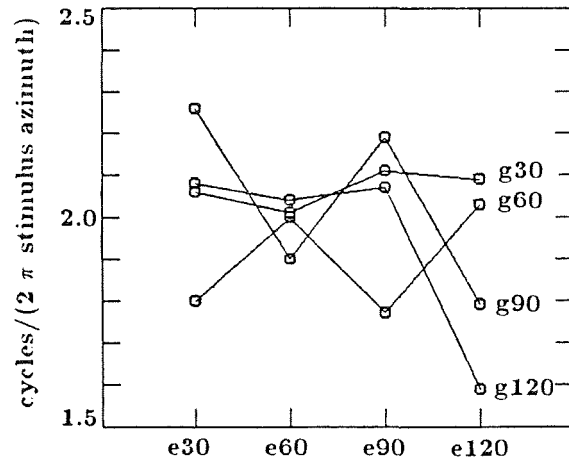
Table 2. Frequency.

FOV	e30	e60	e90	e120
g30	2.06	2.01	2.11	2.09
g60	1.80	2.00	1.77	2.03
g90	2.26	1.90	2.19	1.79
g120	2.08	2.04	2.07	1.59
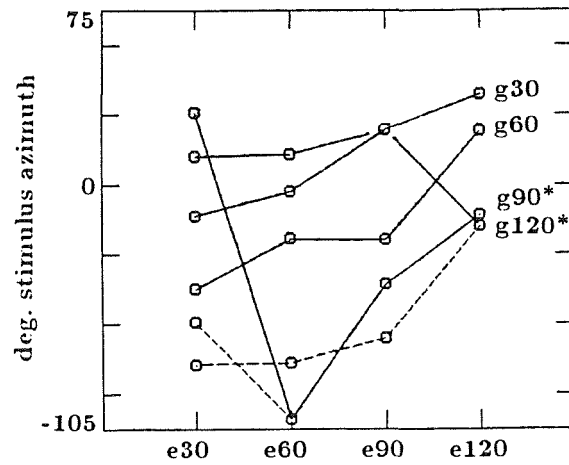
Table 3. Phase shift.

FOV	e30	e60	e90	e120
g30	-13.46	-3.04	23.61	38.85
g60	-44.92	-23.43	-24.06	23.09
g90*	30.83 (-59.17)	-101.07	-43.37	-13.06
g120*	12.38 (-77.62)	12.95 (-77.05)	23.66 (-66.34)	-17.98

* Numbers in parentheses and dashed graph lines represent alternative phase shift values for conditions in which negative amplitude values were obtained (See amplitude table). These alternative values are provided to facilitate comparison between phase shift values.
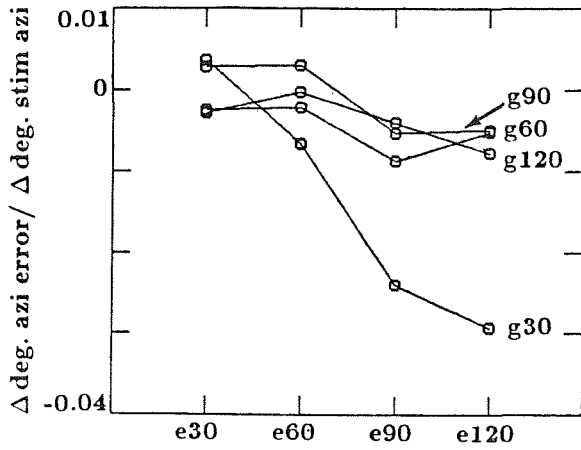
29.13

Top-left plot (y-axis: Δ deg. azi error / Δ deg. stim azi; values 0.01, 0, -0.04; x-axis: e30, e60, e90, e120). Curve labels: g90, g60, g120, g30.

Bottom-left plot (y-axis: deg. azimuth error; values 5, 0, -1; x-axis: e30, e60, e90, e120). Curve labels: g120, g90, g60, g30.

Table 4. Slope.

FOV	e30	e60	e90	e120
g30	0.0036	-0.0066	-0.0240	-0.0294
g60	-0.0025	-0.0022	-0.0087	-0.0052
g90	0.0027	0.0030	-0.0053	-0.0050
g120	-0.0028	-0.0003	-0.0041	-0.0077
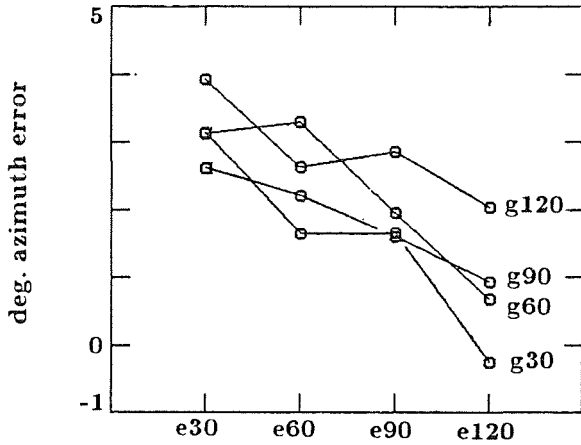
Table 5. Intercept.

FOV	e30	e60	e90	e120
g30	1.95	1.14	1.37	0.08
g60	3.31	2.04	0.71	0.87
g90	4.25	2.33	2.35	0.86
g120	3.06	2.39	1.79	1.38

Table 6. Sinusoidal azimuth error functions for all eye point/geometric station point conditions. θ = stimulus azimuth.

$$f\,(g\,30,e\,30,\theta) = 6.82\,\sin(2.06\theta - 13.46) + 0.0036\theta + 1.95$$
$$f\,(g\,30,e\,60,\theta) = 6.15\,\sin(2.01\theta - 3.04) - 0.0066\theta + 1.14$$
$$f\,(g\,30,e\,90,\theta) = 7.28\,\sin(2.11\theta + 23.61) - 0.0240\theta + 1.37$$
$$f\,(g\,30,e\,120,\theta) = 12.24\,\sin(2.09\theta + 38.85) - 0.0294\theta + 0.08$$

$$f\,(g\,60,e\,30,\theta) = 2.01\,\sin(1.80\theta - 44.92) - 0.0025\theta + 3.31$$
$$f\,(g\,60,e\,60,\theta) = 3.23\,\sin(2.00\theta - 23.43) - 0.0022\theta + 2.04$$
$$f\,(g\,60,e\,90,\theta) = 5.30\,\sin(1.77\theta - 24.06) - 0.0087\theta + 0.71$$
$$f\,(g\,60,e\,120,\theta) = 5.72\,\sin(2.03\theta + 23.09) - 0.0052\theta + 0.87$$

$$f\,(g\,90,e\,30,\theta) = -4.06\,\sin(2.26\theta + 30.83) + 0.0027\theta + 4.25$$
$$f\,(g\,90,e\,60,\theta) = 1.41\,\sin(1.90\theta - 101.07) + 0.0030\theta + 2.33$$
$$f\,(g\,90,e\,90,\theta) = 1.73\,\sin(2.19\theta - 43.37) - 0.0053\theta + 2.35$$
$$f\,(g\,90,e\,120,\theta) = 2.72\,\sin(1.79\theta - 13.06) - 0.0050\theta + 0.86$$

$$f\,(g\,120,e\,30,\theta) = -6.72\,\sin(2.01\theta + 12.38) - 0.0028\theta + 3.01$$
$$f\,(g\,120,e\,60,\theta) = -5.45\,\sin(2.04\theta + 12.95) - 0.0003\theta + 2.39$$
$$f\,(g\,120,e\,90,\theta) = -4.25\,\sin(2.07\theta + 23.66) - 0.0041\theta + 1.79$$
$$f\,(g\,120,e\,120,\theta) = 1.97\,\sin(1.59\theta - 17.98) - 0.0077\theta + 1.38$$

OPTIMAL COOPERATIVE CONTROL SYNTHESIS OF ACTIVE DISPLAYS

by

Sanjay Garg and David K. Schmidt
School of Aeronautics and Astronautics
Purdue University
West Lafayette, IN   47907

ABSTRACT

The utility of augmenting displays to aid the human operator in controlling high order complex systems is well known. Analytical evaluations of various display designs for a simple k/s^2 plant in a compensatory tracking task using an Optimal Control Model (OCM) of human behavior is carried out. This analysis reveals that significant improvement in performance should be obtained by skillful integration of key information into the display dynamics. The cooperative control synthesis technique previously developed to design pilot-optimal control augmentation is extended to incorporate the simultaneous design of performance enhancing augmented displays. The application of the cooperative control synthesis technique to the design of augmented displays is discussed for the simple k/s^2 plant. This technique is intended to provide a systematic approach to design optimally augmented displays tailored for specific tasks.

I. INTRODUCTION

With the advent of high performance aircraft, the amount of information to be processed by the pilot to successfully accomplish the assigned task has increased tremendously. It has, therefore, become critical to determine and limit information to the best informational set needed by the pilot so as to reduce his workload and improve his performance by reducing complex, unusual tasks to simpler, familiar ones. The need for providing augmented displays to the pilot to achieve this objective is very well understood. In the present paper, analytical evaluation of various display "quickening" control laws for a simple k/s^2 plant is carried out. The evaluation is done for a tracking task using an Optimal Control Model (OCM) [1] of human behavior.

A methodology to design pilot-optimal display/control augmentation systems which analytically takes into account the control and information processing limitations of the human controller is proposed. This methodology is an extension of the cooperative control synthesis technique previously developed to design pilot optimal control augmentation [2,3,4]. Though the proposed methodology has been developed so as to be applicable to simultaneous synthesis of pilot optimal control augmentation and display augmentation, the present discussion focuses on the application of the technique to display design only.

The cooperative display design technique is applied to synthesize performance enhancing augmented compensatory displays for the k/s^2 plant in the tracking task. The displays thus obtained show improved tracking performance for much reduced mean square pilot input when evaluated using the OCM. Moreover, the methodology offers considerable potential as a tool for providing a systematic approach to task tailoring of augmented displays.

II. DISPLAY DESIGN FOR k/s^2 PLANT

Consider the k/s^2 plant dynamics as discussed by Klienman et al. in [1]. The system state equations are

$$\left\{ \begin{array}{c} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{array} \right\} = \begin{bmatrix} -2 & 0 & 0 \\ 1 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \left\{ \begin{array}{c} x_1 \\ x_2 \\ x_3 \end{array} \right\} + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} u(t) + \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} w(t)$$

or in concise form

$$\dot{\bar{x}} = A_o \bar{x} + B_o u + D_o w \qquad (2.1)$$

Here $k=1$ in./in. and the state $x_1(t)$, a first order Markov process having a break frequency of 2 rads/sec, is the velocity of the command. $w(t)$ has intensity $W = 0.217$ to give $E\{x_1^2\} = 0.054$ in.^{2}

$$y_1 = x_2 = e: \quad \text{error}$$

$$y_2 = x_1 + x_3 = \dot{e}: \quad \text{error rate} \tag{2.2}$$

where the pilot is assumed to be able to reconstruct the error rate by observing the error itself. For the OCM model, the pilot's cost function is taken to be

$$J(u) = E\{e^2\} + rE\{\dot{u}^2\} \tag{2.3}$$

where "r" is chosen so as to give a neuromuscular lag time constant, $\tau_N = 0.1$ secs.

For all the analysis carried out in this section, the following parameters were set for the OCM pilot model

a. Pilot's observation time delay set to 0.2 seconds

b. Observation noise ratio was set at -20 dB

c. Motor noise ratio was set at -25 dB

d. The weighting on the control rate in the pilot's cost function was always adjusted to yield $\tau_N = 0.1$ secs.

e. Very low values of thresholds were used for the observations made available to the pilot.

With the above parameter settings, the OCM analysis of system (2.1)-(2.3) gave results that are compatible with those given in [1]. These results are as shown in the last row of Table 1.

Next consider the display dynamics having the form

$$\dot{x}_d = a_d x_d + u_d \tag{2.4}$$

with the display quickening control u_d given by

$$u_d = G_d \bar{y}_d \tag{2.5}$$

where $\bar{y}_d$ is the vector of plant outputs which are available for driving the display and G_d is the set of display control gains being determined, or

$$\bar{y}_d = C_d \bar{x} \tag{2.6}$$

The dynamics of the display augmented system can then be written as

$$\left\{ \begin{array}{c} \dot{\bar{x}} \\ \hline \dot{x}_d \end{array} \right\} = \left[\begin{array}{c|c} A_o & 0 \\ \hline G_d C_d & a_d \end{array} \right] \left\{ \begin{array}{c} \bar{x} \\ \hline x_d \end{array} \right\} + \left[\begin{array}{c} B_o \\ \hline 0 \end{array} \right] u + \left[\begin{array}{c} D_o \\ \hline 0 \end{array} \right] w \tag{2.7}$$

The pilot's observations for the display augmented system are

$$y_1 = x_d$$

$$y_2 = \dot{x}_d \tag{2.8}$$

where it is again assumed that the pilot is able to reconstruct the rate of display by observing the displayed variable itself. The pilot's performance objective for the display augmented system is to minimize the cost

$$J_d(u) = E\{x_d^2\} + rE\{\dot{u}^2\} \tag{2.9}$$

With the above formulation in mind, the performance of the display augmented system is evaluated using the OCM model for various values of a_d and various combinations of the display control gains G_d. Two cases of $\bar{y}_d$ are considered. The first is when the display state is driven only by the error, i.e. state x_2, and the second is when $\bar{y}_d$ consists of both the error as well as the plant velocity state x_3.

Case(a) $y_d = x_2$:

This is the simplest possible case for display of the form (2.4). For this case the displayed variable is just lagged error. The display dynamics are given by

$$\dot{x}_d = a_d x_d + g_{d2} x_2 \tag{2.10}$$

where g_{d2} is the display control gain on state x_2. For $g_{d2} = -a_d$ (2.10) can be written in transfer function form as

$$x_d(s) = \frac{-a_d}{s - a_d} x_2(s) \tag{2.11}$$

Since $x_2 = e$, it is clear from (2.11) that in the steady state the displayed variable will closely approximate the error.

The OCM results for various values of a_d are presented in Table 1. The results of Table 1 are also plotted in Fig. 1 and Fig. 2, and correspond to the curve marked ① . From these plots it is clear that with only error driving the display, the pilot's performance is worse than the idealized no-display case. As $a_d \rightarrow -\infty$, the pilot's performance approaches that of the case with no display augmentation. The no-display case which then corresponds to an infinetly fast display is not desirable because of the inherent limitations on the pilot's ability to perceive fast changing signals, and the need to provide filtering of the noisy outputs. It might be reasonable to select a display which has a slightly higher bandwidth than the pilot, so a_d in the range -10 to -20 sec^{-1} is desirable since the pilot's minimum neuro-muscular lag time constant is approximately 0.1 secs.

Case(b) $\bar{y}_d = [x_2, x_3]^T$:

For this case, the display dynamics have the form

$$\dot{x}_d = a_d x_d + g_{d2} x_2 + g_{d3} x_3 \qquad (2.12)$$

where g_{di}, $i = 2, 3$ is the display gain on the state x_i.

Since $x_3(t)$ is the plant velocity state, the above form of display will provide lead information to the pilot. The pilot's performance can then be expected to improve as the gain g_{d3} is increased.

OCM analysis is carried out for two values of a_d: -10 and -20 sec^{-1}. For each of these values of a_d, $g_{d2} = -a_d$ and g_{d3} is varied from 1 to 6 in steps of 1. The results of this analysis are presented in Table 2 and are also plotted in Fig. 1 and Fig. 2 so as to compare them with the case of $g_{d3} = 0$. In the two figures, the curve marked ② corresponds to $a_d = -10$ sec^{-1} and that marked ③ to $a_d = -20$ sec^{-1}.

Fig. 1 is a plot of mean square error vs. mean square control rate ($\dot{u}$) for the various display cases discussed above and Fig. 2 is a plot of mean square error vs. the mean square control input (u). The point marked A corresponds to the no display case in the two figures. From these two figures it is clear that the mean square input and the mean square control rate both decrease as the display control gain g_{d3} is increased. What is most interesting is that the mean square error initially decreases as g_{d3} is increased and then starts increasing beyond a certain value of g_{d3} that depends on the choice of the display bandwidth and the display gain g_{d2}. Noting that earlier work [5,6] has shown that the pilot's workload is directly related to the mean square control rate, this means that it is possible to improve performance (of which mean square error is a measure) while at the same time decreasing pilot's workload and the control energy required by a skillful integration of key information into the display dynamics. Moreover, the results indicate that for a given display bandwidth there is an optimal choice of display control gains which leads to the best possible performance. For instance, in Figures 1 and 2, point C is such an optimal display design for $a_d = -20$ sec^{-1}, and for this case the performance is slightly better than the no-display case. Meanwhile the pilot's workload and the control effort required are both significantly reduced.

It then appears desirable to develop a systematic approach to display augmentation which will make it possible to directly synthesize the optimal display design without having to resort to trial and error. In the following sections an extension of the optimal cooperative control synthesis technique is proposed as a methodology to synthesize pilot-optimal display/control augmentation systems.

TABLE 1:   OCM RESULTS FOR VARYING DISPLAY BANDWIDTH ($g_{d2} = -a_d$)

a_d (sec^{-1})	r ($\tau_N = 0.1$ secs)	M.S. Error (in.^{2})	M.S. Input (in.^{2})	M.S. Control rate (in.^{2}/sec^2)
-5	4.6×10^{-5}	0.0215	2.176	106.91
-10	5.8×10^{-5}	0.0177	1.543	76.47
-20	6.2×10^{-5}	0.0157	1.353	67.44
-50	6.25×10^{-5}	0.0142	1.261	62.9
-100	6.3×10^{-5}	0.0135	1.223	60.98
NO DISPLAY	7.0×10^{-5}	0.0131	1.141	54.73

TABLE 2:   OCM RESULTS FOR VARYING DISPLAY CONTROL GAINS

g_{d3}	$a_d = -10$ (sec^{-1}), $g_{d2} = 10$			$a_d = -20$ (sec^{-1}), $g_{d2} = 20$		
	M.S. Error (in.^{2})	M.S. Input (in.^{2})	M.S. Control rate (in.^{2}/sec^2)	M.S. Error (in.^{2})	M.S. Input (in.^{2})	M.S. Control rate (in.^{2}/sec^2)
1	0.0144	1.113	54.75	0.014	1.175	58.06
2	0.0138	0.733	35.92	0.013	0.968	47.47
3	0.0143	0.486	23.71	0.0127	0.789	38.49
4	0.0157	0.339	16.52	0.0128	0.639	30.97
5	0.0175	0.248	12.05	0.0131	0.521	25.15
6	0.0195	0.187	9.01	0.0136	0.427	20.46

II.   OPTIMAL COOPERATIVE CONTROL/DISPLAY DESIGN MEHODOLOGY

PROBLEM FORMULATION:

In this section the mathematical formulation of the cooperative control synthesis technique is presented, and necessary conditions for the simultaneous optimality of the display and control augmentation systems are developed. The procedure followed here is very similar to that of [3, 4].

Consider the dual controller system described by the linear time invariant set of first order differential equations

$$\dot{\bar{x}} = A_o \bar{x} + B_{1o} \bar{u}_1 + B_{2o} \bar{u}_2 + D_o \bar{w} \tag{3.1}$$

with $\bar{x} \epsilon R^n$, $\bar{u}_1 \epsilon R^{m_1}$, $\bar{u}_2 \epsilon R^{m_2}$ and $\bar{w}$ a zero-mean Gaussian white noise process with intensity W. The two controls represent two physically independent controllers.

The display dynamics are assumed to be of the form

$$\dot{\bar{x}}_d = A_d \bar{x}_d + B_{do} \bar{u}_d \tag{3.2}$$

with $\bar{x}_d \in R^d$, $u_d \in R^{m_d}$, and $\bar{u}_d$ is the display quickening controller. The objective is to find the optimal cooperative controllers 1 and 2 ($\bar{u}_1$ and $\bar{u}_2$) along with the optimal display control law $\bar{u}_d$.

Controller 1 ($\bar{u}_1$) has noisy observations available for feedback given by

$$\bar{y}_1 = C_{1o}\bar{x} + C_{d1}\bar{x}_d + C_u\bar{u}_d + \bar{v}_y \qquad (3.3)$$

where $\bar{v}_y$ is also a zero-mean Gaussian white noise process with intensity V_y.

The augmentation controller $\bar{u}_2$ and the display control law $\bar{u}_d$ are assumed to have noise-free system outputs $\bar{y}_2$ and $\bar{y}_d$, respectively, available for feedback, where

$$\bar{y}_2 = C_{2o}\bar{x}; \qquad \bar{y}_d = C_d \left\{ \begin{array}{c} \bar{x} \\ -- \\ \bar{x}_d \end{array} \right\} \qquad (3.4)$$

Note that the above formulation does not allow feedback of the display states into the augmentation controller $\bar{u}_2$.

Finally, these two controllers are constrained to have the direct output feedback form

$$\bar{u}_2 = C_2\bar{y}_2 = G_2 C_{2o}\bar{x}$$

$$\bar{u}_d = G_d\bar{y}_d = G_d C_d \left\{ \begin{array}{c} \bar{x} \\ -- \\ \bar{x}_d \end{array} \right\} \qquad (3.5)$$

which is consistent with the desire for simple, easy to implement control laws.

The interaction between the different controllers is shown in the block diagram of Figure 3.

DESIGN OBJECTIVES:

Controller 1 is to be optimal with respect to the cost

$$J_1 = E\{\lim_{T\to\infty} \frac{1}{T} \int_o^T (\bar{x}^T Q_{1o}\bar{x} + \bar{x}_d^T Q_{1d}\bar{x}_d + \bar{u}_1^T R_1\bar{u}_1 + \bar{u}_2^T F_1\bar{u}_2)dt\} \qquad (3.6)$$

in the presence of the action of control inputs $\bar{u}_2$ and $\bar{u}_d$. Here $E\{\cdot\}$ indicates the expected value operator and the weighting matrices are $Q_{1o} \geq 0$, $Q_{1d} \geq 0$, $R_1 > 0$, $F_1 \geq 0$.

Conversely, Controller 2 ($\bar{u}_2$) and the display control law $\bar{u}_d$ are to be optimal with respect to the cost

$$J_2 = E\{\lim_{t\to\infty} \frac{1}{T} \int_o^T (\bar{x}^T Q_{2o}\bar{x} + \bar{x}_d^T Q_{2d}\bar{x}_d + \bar{u}_1^T R_2\bar{u}_1 + \bar{u}_2^T F_2\bar{u}_2 + \bar{u}_d^T F_{2d}\bar{u}_d)dt\} \qquad (3.7)$$

in the presence of the control action $\bar{u}_1$. The weighting matrices are $0_{2o} > 0$, $0_{2d} > 0$, $R_2 > 0$, $F_2 > 0$, $F_{2d} > 0$. Augmenting the system dynamics (3.1) with the display dynamics (3.2), the state-space description of this augmented system is obtained to be

$$
\left\{ \begin{array}{c} \dot{\bar{x}} \\ \hline \dot{\bar{x}}_d \end{array} \right\} = \left[\begin{array}{c|c} A_o & 0 \\ \hline 0 & A_d \end{array} \right] \left\{ \begin{array}{c} \bar{x} \\ \hline \bar{x}_d \end{array} \right\} + \left[\begin{array}{c} B_{1o} \\ \hline 0 \end{array} \right] \bar{u}_1 + \left[\begin{array}{c} B_{2o} \\ \hline 0 \end{array} \right] \bar{u}_2 + \left[\begin{array}{c} 0 \\ \hline B_{do} \end{array} \right] \bar{u}_d + \left[\begin{array}{c} D_o \\ \hline 0 \end{array} \right] \bar{w} \quad (3.8)
$$

Defining $\bar{\chi} = COL\ (\bar{x},\ \bar{x}_d)$, (3.8) can be written in a compact form with appropriate definitions for the matrices as

$$
\dot{\bar{\chi}} = A\bar{\chi} + B_1\bar{u}_1 + B_2\bar{u}_2 + B_d\bar{u}_d + D\bar{w} \quad (3.9)
$$

The outputs can similarly be written as

$$
\bar{y}_1 = C_1\bar{\chi} + C_u\bar{u}_d + \bar{v}_y
$$

$$
\bar{y}_2 = [C_{2o} \quad 0]\bar{\chi} = C_2\bar{\chi} \quad (3.10)
$$

The two cost functions can then be expressed in terms of the augmented state vector $\bar{\chi}$ as

$$
J_1 = E\{\lim_{T\to\infty} \frac{1}{T} \int_0^T (\bar{\chi}^T 0_1 \bar{\chi} + \bar{u}_1^T R_1 \bar{u}_1 + \bar{u}_2^T F_1 \bar{u}_2)dt\}
$$

$$
J_2 = E\{\lim_{T\to\infty} \frac{1}{T} \int_0^T (\bar{\chi}^T 0_2 \bar{\chi} + \bar{u}_1^T R_2 \bar{u}_1 + \bar{u}_2^T F_2 \bar{u}_2 + \bar{u}_d^T F_{2d} \bar{u}_d)dt\} \quad (3.11)
$$

where the weighting matrices 0_1 and 0_2 are appropriately defined.

SOLUTION FOR $\bar{u}_1$:

In the presence of the action of control inputs $\bar{u}_2$ and $\bar{u}_d$, as given by (3.5), the dynamics of the augmented system are obtained to be

$$
\dot{\bar{\chi}} = A_{aug}\bar{\chi} + B_1\bar{u}_1 + D\bar{w}
$$

$$
\bar{y}_1 = C_{aug}\bar{\chi} + \bar{v}_y \quad (3.12)
$$

where

$$
A_{aug} \overset{\Delta}{=} (A + B_2 G_2 C_2 + B_d G_d C_d)
$$

$$
C_{aug} \overset{\Delta}{=} (C_1 + C_u G_d C_d) \quad (3.13)
$$

and the performance index J_1 becomes

$$
J_1 = E\{\lim_{T\to\infty} \frac{1}{T} \int_0^T (\bar{\chi}^T (0_1 + C_2^T G_2^T F_1 G_2 C_2)\bar{\chi} + \bar{u}_1^T R_1 \bar{u}_1)dt\} \quad (3.14)
$$

30.8

Equations (3.12) and (3.14), in the case of uncorrelated process and measurement noises and for $V_y > 0$, describe the standard non-singular linear quadratic Gaussian regulator problem. When stabilizability and detectability conditions for the system are satisfied, the optimal controller is known [7] to have the form

$$u_1 = k_1 \hat{\overline{\chi}} \tag{3.15}$$

where $\hat{\overline{\chi}}$ is the minimum mean-square estimate of the system state vector $\overline{\chi}$.

The gain matrix k_1 is given by

$$k_1 = -R_1^{-1} B_1^T P \tag{3.16}$$

with $P > 0$ the symmetric solution of the algebraic Ricatti equation

$$A_{aug}^T P + P A_{aug} + (Q_1 + C_2^T G_2^T F_1 G_2 C_2) - P B_1^T R_1^{-1} B_1 P = 0 \tag{3.17}$$

The dynamics of the state estimator are

$$\dot{\hat{\overline{\chi}}} = A_{aug} \hat{\overline{\chi}} + B_1 \overline{u}_1 + M_1 (\overline{y}_1 - C_{aug} \hat{\overline{\chi}}) \tag{3.18}$$

where the Kalman filter gain matrix M_1 is given by

$$M_1 = \Sigma \, C_{aug}^T V_y^{-1} \tag{3.19}$$

with $\Sigma > 0$ the symmetric solution of the algebraic equation

$$A_{aug} \Sigma + \Sigma A_{aug}^T + DVD^T - \Sigma C_{aug}^T V_y^{-1} C_{aug} \Sigma = 0 \tag{3.20}$$

SOLUTION FOR $\overline{u}_2$ AND $\overline{u}_d$:

The optimal controller $\overline{u}_1$ as derived above has the form

$$\overline{u}_1 = k_1 \hat{\overline{\chi}}; \qquad \dot{\hat{\overline{\chi}}} = A_1 \hat{\overline{\chi}} + M_1 \overline{y}_1 \tag{3.21}$$

where $A_1 \overset{\Delta}{=} (A_{aug} + B_1 k_1 - M_1 C_{aug})$.

Then in the presence of the control action $\overline{u}_1$, the system dynamics can be written in terms of the augmented state vector $\overline{q} \overset{\Delta}{=} COL(\overline{\chi}, \hat{\overline{\chi}})$ as

$$\dot{\overline{q}} = \begin{bmatrix} A & B_1 k_1 \\ \hline M_1 C_1 & A_1 \end{bmatrix} \overline{q} + \begin{bmatrix} B_2 \\ \hline 0 \end{bmatrix} \overline{u}_2 + \begin{bmatrix} B_d \\ \hline M_1 C_u \end{bmatrix} \overline{u}_d + \begin{bmatrix} D & 0 \\ \hline 0 & M_1 \end{bmatrix} \left\{ \begin{array}{c} \overline{w} \\ \hline \overline{v}_y \end{array} \right\} \tag{3.22}$$

which can further be written in a compact form with appropriate definitions of matrices as

$$\dot{\overline{q}} = A_1' \overline{q} + B_2' \overline{u}_2 + B_d' \overline{u}_d + D' \overline{w}' \tag{3.23}$$

The intensity of the process $\bar{w}'$ is $W' = \begin{bmatrix} W & | & 0 \\ \hline 0 & | & V \\ & | & y \end{bmatrix}$.

The index of performance then becomes

$$J_2 = E\{\lim_{T \to \infty} \frac{1}{T} \int_0^T (\bar{q}^T O' \bar{q} + \bar{u}_2^T F_2 \bar{u}_2 + \bar{u}_d^T F_{2d} \bar{u}_d) dt\} \qquad (3.24)$$

with

$$O' \overset{\Delta}{=} \begin{bmatrix} O_2 & | & 0 \\ \hline 0 & | & k_1^T R_2 k_1 \end{bmatrix}.$$

The design objective can then be stated as to find the optimal controller $\bar{u}_2$ and optimal display control $\bar{u}_d$ which minimize the cost J_2 as given by (3.24).

Proceeding in a way as detailed in [4], it can be shown that the gains G_2 and G_d which correspond to the simultaneous optimality of the two controllers $\bar{u}_2$ and $\bar{u}_d$ are given by

$$G_2 = -F_2^{-1} [B_2^T \; | \; 0] HL \begin{bmatrix} C_2^T \\ \hline 0 \end{bmatrix} ([C_2 \; | \; 0] \; L \begin{bmatrix} C_2^T \\ \hline 0 \end{bmatrix})^{-1} \qquad (3.25)$$

and

$$G_d = -F_{2d}^{-1} \begin{bmatrix} B_d \\ \hline M_1 C_u \end{bmatrix}^T HL \begin{bmatrix} C_d^T \\ \hline 0 \end{bmatrix} ([C_d \; | \; 0] \; L \begin{bmatrix} C_d^T \\ \hline 0 \end{bmatrix})^{-1} \qquad (3.26)$$

with $L = E\{\bar{q} \; \bar{q}^T\}$ satisfying the relation

$$A_c L + L A_c^T + D' W' D'^T = 0 \qquad (3.27)$$

and H satisfying

$$A_c^T H + H A_c + \bar{O} = 0 \qquad (3.28)$$

where the following definitions have been used

$$A_c \overset{\Delta}{=} \begin{bmatrix} A_{aug} & | & B_1 k_1 \\ \hline M_1 C_{aug} & | & A_1 \end{bmatrix}; \quad \bar{O} \overset{\Delta}{=} O' + \begin{bmatrix} C_2^T G_2^T F_2 G_2 C_2 + C_d^T G_d^T F_{2d} G_d C_d & | & 0 \\ \hline 0 & | & 0 \end{bmatrix}$$

Though the methodology developed above is applicable for simultaneous synthesis of optimal control and display augmentation, only the application to display design will be discussed in this paper. For the case of display design only, the controller $\bar{u}_2$ is inactive and the system dynamics and corresponding conditions for optimality are accordingly simplified.

IV. APPLICATION OF DISPLAY DESIGN METHODOLOGY TO k/s^2 PLANT

A computer code was developed to determine the optimal display control gains using the above methodology. The details of a similar computer code are documented in [4]. The algorithm is iterative using a gradient search technique. Given a starting display gain matrix (including the null matrix) the display gains that satisfy the conditions of optimality as stated in Section III are determined.

The dynamics of the k/s^2, plant augmented with the display, are as in Section II. The application of the methodology to optimal display design for the case of $a_d = -20\sec^{-1}$ with both x_2 and x_3 driving the display will be discussed.

The controller $\bar{u}_1$ is analogous to the control rate $\bar{u}_p$ of the OCM, so in order to be consistent with the pilot's stated objective of regulating the display, the cost J_1 is defined as

$$J_1 = E\{x_d^2\} + R_1 E\{u_1^2\} \qquad (4.1)$$

where R_1 is chosen so as to satisfy the requirement of $\tau_N = 0.1$ seconds for the pilot's neuro-muscular lag. Also the process noise and the measurement noise in the problem formulation are chosen such that the controller $\bar{u}_1$ for the beginning display dynamics is compatible with the OCM model corresponding to the dynamics. (The reader is referred to [4] for details of how to achieve this).

The cost J_2 is defined as

$$J_2 = Q_e E\{e^2\} + R_2 E\{u_1^2\} + F_{2d} E\{u_d^2\} \qquad (4.2)$$

which is reflective of the overall objective of reducing the tracking error through the means of an "intelligent" display. Note that in (4.2), F_{2d} needs to be positive definite in order to get a finite optimal solution to the problem. However, since the display control does not reflect any measure of energy, the weighting F_{2d} may be chosen small such that its contribution to the cost J_2 is not significant. For the results presented in this section $F_{2d} = 0.001$ was used.

The results obtained using the optimal cooperative design methodology for various values of Q_e and R_2 are presented in Table 3. For all these cases the starting display gains were taken to be $G_d = [20, 0]$. In Table 3 the optimal display gains are listed as well as the results of evaluation of the corresponding augmented dynamics using the OCM. The parameters that define the OCM were set to the values stated in Section II. The OCM analysis results for the no-display case and the cases

of G_d = [20, 0] and G_d = [20, 3] for a_d = -20sec^{-1}, are also listed in
Table 3 to provide a comparison. The results of Table 3 are also plotted in Figures 4 and 5.

Note that as the relative weighting on the error is increased in
the cost J_2, the optimal cooperative display design methodology does
lead to display gains which give improved performance at the expense of
increased control activity. Thus this methodology, through a proper
choice of weightings in the cost function J_2, provides a systematic
approach to design of task-tailored display augmentation.

Also note that for all the 5 cases of display design using this
methodology, the final optimal display gains were such that the perfor-
mance is significantly improved as compared to the beginning display and
at the same time the workload ($\dot{u}$) and control effort (u) are consider-
ably reduced. If the weighting on the error is made high enough (cases
4 and 5), performance comparable to the no display case and the best
case corresponding to G_d = [20, 3] of Section II is obtained for signi-
ficantly reduced workload and control effort. Moreover it is clear that
for the display bandwidth such that a_d = -20sec^{-1}, performance better
than that of case 5 cannot be obtained. Increasing the weight on error
in the cost function J_2 any further would only have the effect of lead-
ing to a display design requiring higher control effort without any
noticeable improvement in performance.

TABLE 3: OCM RESULTS FOR OPTIMAL DISPLAYS FOR k/s^2 PLANT
a_d = -20sec^{-1}, G_d = [g_{d2}, g_{d3}]

S.N.	Q_e	R_2	Optimal G_d	M.S. ($in.^2$) Error	M.S. ($in.^2$) Input	M.S. ($in.^2/sec^2$) Control Rate
1	1	20x10^{-5}	[36.6, 11.9]	0.014	0.389	18.64
2	2	20x10^{-5}	[49.2, 13]	0.0132	0.492	23.72
3	2	10x10^{-5}	[64.4, 16.3]	0.0131	0.514	24.78
4	4	10x10^{-5}	[88.1, 17.3]	0.01272	0.650	31.58
5	4	5x10^{-5}	[118.2, 22.6]	0.01269	0.665	32.40
A	NO DISPLAY		—	0.0131	1.141	54.73
B	BEG. DISPLAY FOR DESIGN		[20, 0]	0.0157	1.353	67.44
C	BEST PER. DISPLAY (II)		[20, 3]	0.0127	0.789	38.49

CONCLUSIONS

Through OCM analysis of a simple k/s^2 plant it was shown that the
performance of a human controller can be improved and his workload sig-
nificantly reduced by providing him an active display which integrates
information for the system dynamics being controlled. A methodology
based on the optimal cooperative control synthesis technique was sug-
gested as a means to synthesize optimal display gains, tailored for
specific tasks. The application of this methodology to the k/s^2 plant
was discussed and the results presented show that the methodology has
potential for providing a systematic approach to display design.

The results obtained for the k/s^2 plant need to be experimentally verified with man in the loop simulation in order to validate the display design methodology. Research in the area of applying the proposed design methodology to high order dynamical systems in a complex multi-control task scenario is presently ongoing and the preliminary results are quite encouraging.

ACKNOWLEDGEMENT

REFERENCES

[1] Kleinman, D.R., Baron, S., and Levison, W.H., "An Optimal Control Model of Human Response," Parts I and II, Automatica, Vol. 6, pp. 357-383, 1970.

[2] Schmidt, D.K., "Optimal Flight Control Synthesis via Pilot Modeling," AIAA Journal of Guidance and Control, July-August 1979.

[3] Schmidt, D.K., and Innocenti, M., "Pilot Optimal Multivariable Control Synthesis by Output Feedback," NASA CR-16312, July 1981.

[4] Innocenti, M., "Cooperative Pilot-Optimal Augmentation System Synthesis for Complex Flight Vehicles," Ph.D. Thesis, Purdue University, West Lafayette, Indiana, May 1983.

[5] Schmidt, D.K., "On the Use of the OCM's Quadratic Objective Function as a Pilot Rating Metric," 17th Annual Conference on Manual Control, Los Angeles, CA, June 1981.

[6] Wierwille, Walter W., and Connor, Sydney A., "Evaluation of 20 Workload Measures Using a Psychomotor Task in a Moving-Base Aircraft Simulator", Human Factors, Vol. 25, pp. 1-16, 1983.

[7] Kwakernaak, H., and Sivan, R., "Linear Optimal Control Systems," Wiley-Interscience, 1972.
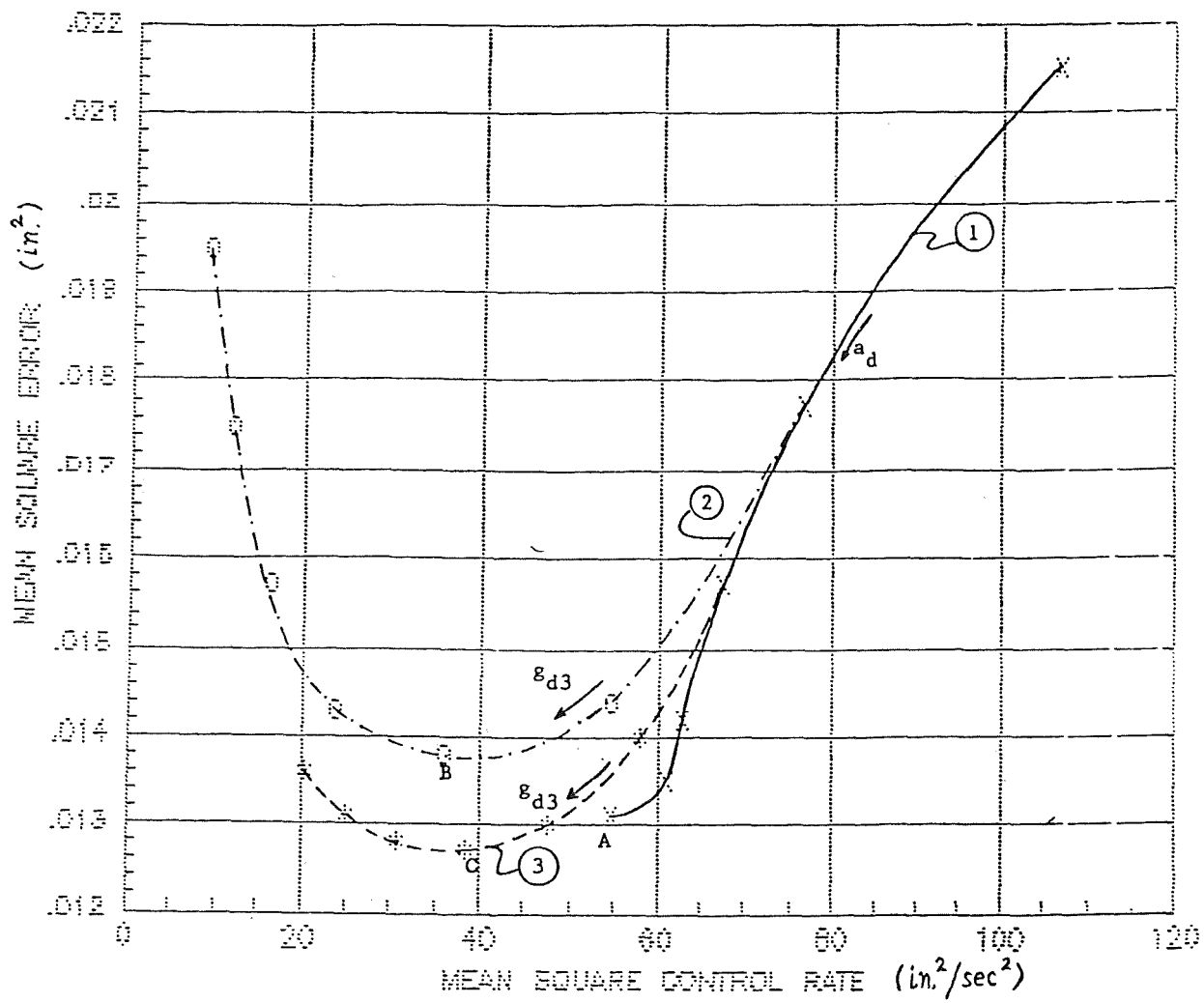
NOMENCLATURE

① Only x_2 driving display, a_d = -5, -10, -20, -50, -100

   A - no display

② a_d = -10, g_{d2} = 10, g_{d3} = 1, 2, 3, 4, 5, 6

   B - g_{d3} = 2

③ a_d = -20, g_d = 20, g_{d3} = 1, 2, 3, 4, 5, 6

   C - g_{d3} = 3

FIG. 1   PERFORMANCE VS WORKLOAD
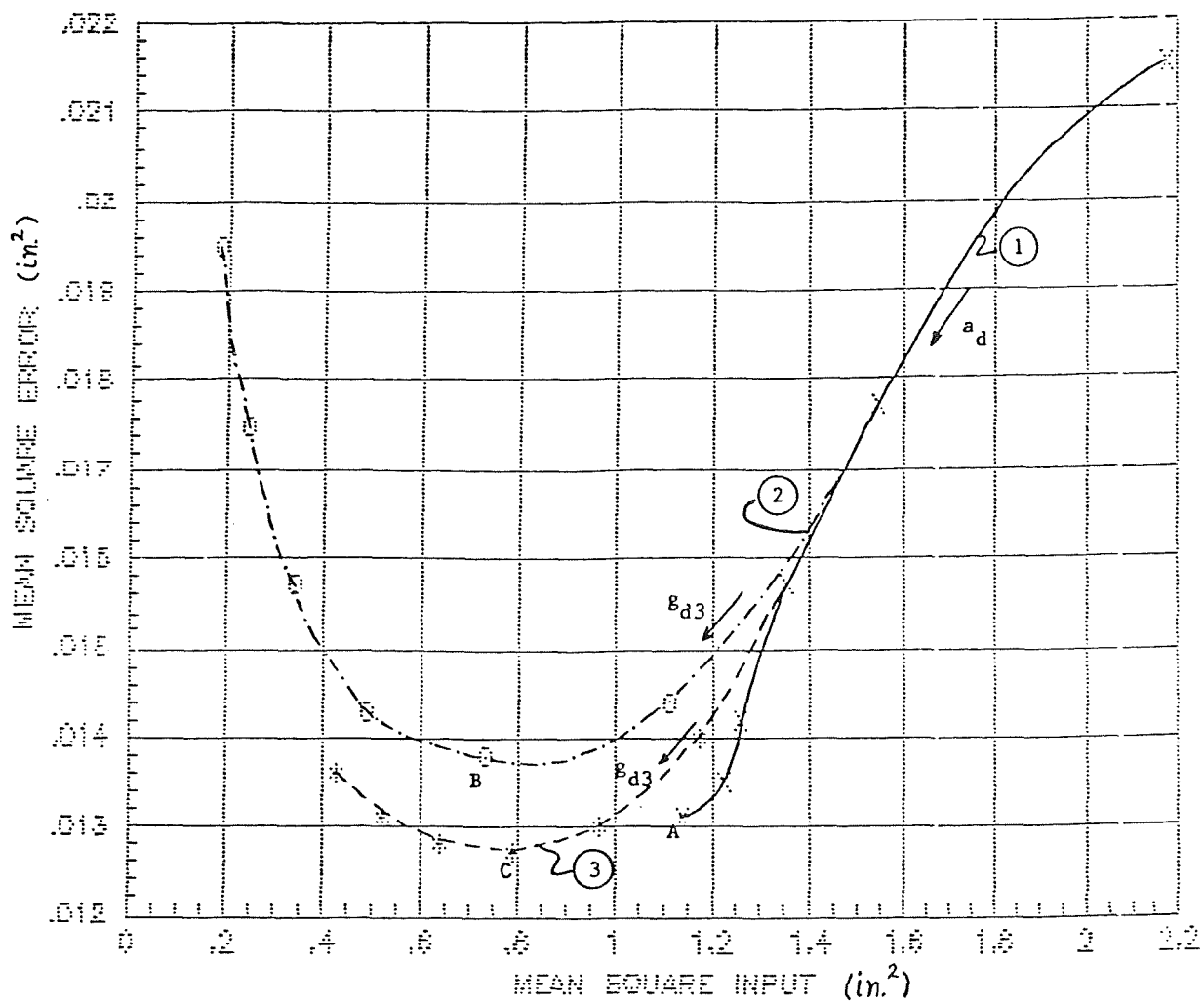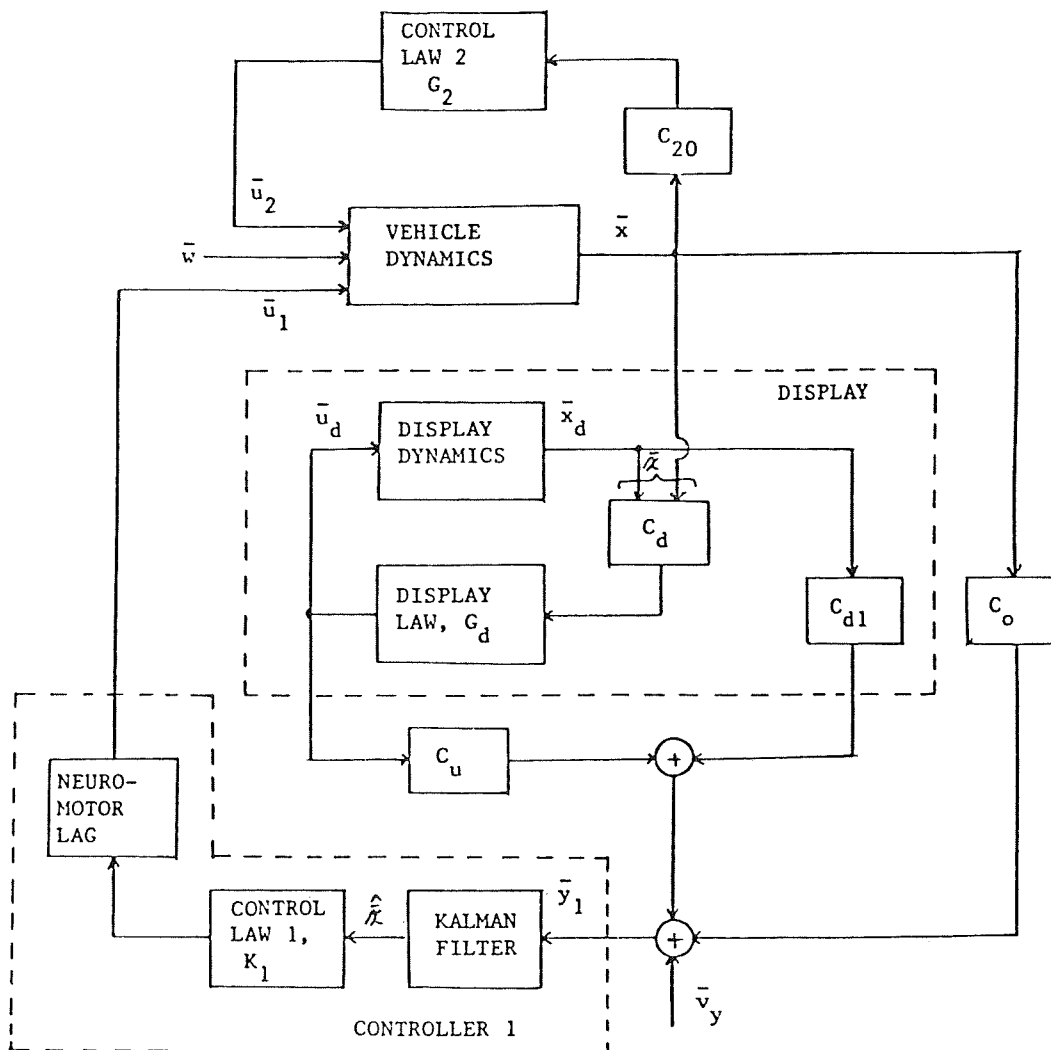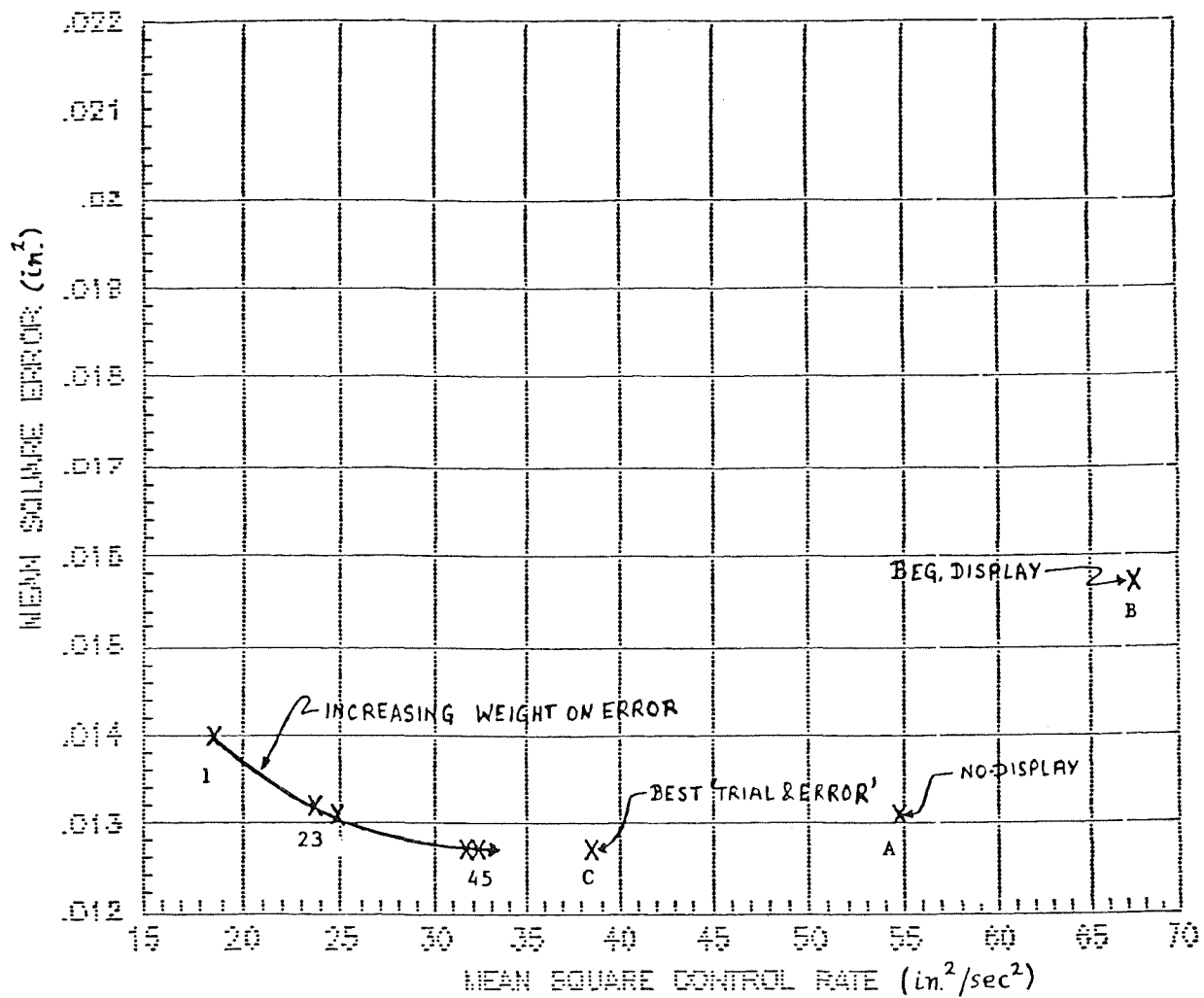
30.14

FIG. 2 PERFORMANCE VS CONTROL EFFORT

NOMENCLATURE

(1) Only x_2 driving display, a_d = -5, -10, -20, -50, -100

   A - no display

(2) a_d = -10, g_{d2} = 10, g_{d3} = 1, 2, 3, 4, 5, 6

   B - g_{d3} = 2

(3) a_d = -20, g_d = 20, g_{d3} = 1, 2, 3, 4, 5, 6

   C - g_{d3} = 3

30.15

FIG. 3 BLOCK DIAGRAM FOR CONTROL AND DISPLAY AUGMENTATION

30.16

FIG. 4. PERFORMANCE VS WORKLOAD FOR OPTIMAL DISPLAYS

Plot axes: vertical — MEAN SQUARE ERROR ($in.^2$); horizontal — MEAN SQUARE CONTROL RATE ($in.^2/sec^2$).

Vertical axis values: .022, .021, .02, .019, .018, .017, .016, .015, .014, .013, .012

Horizontal axis values: 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70

Annotations on plot: INCREASING WEIGHT ON ERROR; BEST 'TRIAL & ERROR'; NO-DISPLAY; BEG. DISPLAY; points labeled 1, 23, 45, A, B, C.
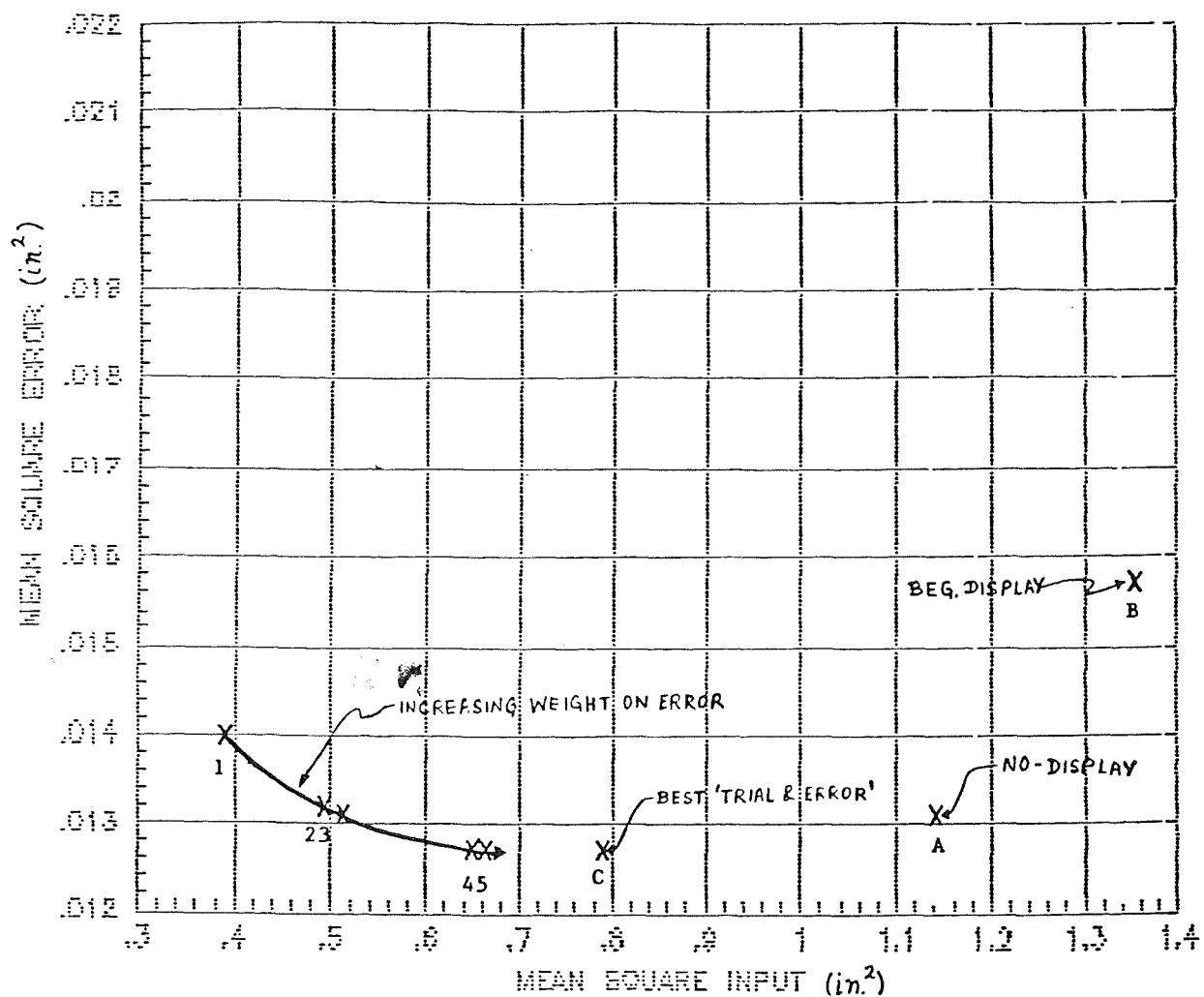
NOMENCLATURE

A: NO-DISPLAY

B: $G_d = [20, 0]$

C: $G_d = [20, 3]$

1: $Q_e = 1$, $R_2 = 20 \times 10^{-5}$

2: $Q_e = 2$, $R_2 = 20 \times 10^{-5}$

3: $Q_e = 2$, $R_2 = 10 \times 10^{-5}$ OPTIMAL DISPLAYS

4: $Q_e = 4$, $R_2 = 10 \times 10^{-5}$

5: $Q_e = 4$, $R_2 = 5 \times 10^{-5}$

30.17

NOMENCLATURE

A: NO-DISPLAY

B: G_d = [20, 0]

C: G_d = [20, 3]

1: Q_e = 1, R_2 = 20 x 10^{-5} ⎫
2: Q_e = 2, R_2 = 20 x 10^{-5} ⎪
3: Q_e = 2, R_2 = 10 x 10^{-5} ⎬   OPTIMAL DISPLAYS
4: Q_e = 4, R_2 = 10 x 10^{-5} ⎪
5: Q_e = 4, R_2 = 5 x 10^{-5} ⎭

FIG. 5.   PERFORMANCE VS CONTROL EFFORT FOR OPTIMAL DISPLAYS

SOME COMPUTATIONAL TECHNIQUES FOR ESTIMATING
HUMAN OPERATOR DESCRIBING FUNCTIONS

(Informal Paper)

by

William H. Levison
BBN Laboratories Incorporated
10 Moulton St.
Cambridge, MA 02238

ABSTRACT

Computational procedures for improving the reliability of human operator describing functions are described. Special attention is given to the estimation of standard errors associated with mean operator gain and phase shift as computed from an ensemble of experimental trials. This analysis pertains to experiments using sum— of—sines forcing functions. Both open—loop and closed—loop measurement environments are considered.

INTRODUCTION

Linear analysis of human operator response behavior is complicated by the presence of operator "remnant"; i.e., by response components that cannot be related to the input signal by a time—invariant linear process. Remnant may arise from a multiplicity of sources, such as nonlinearities in the response strategy, time variations in the linear aspect of the response strategies, and purely stochastic response behavior.

Experiments designed to measure and model operator behavior in closed—loop control tasks have made considerable use of external forcing functions constructed as sums of sinusoids. This technology has recently been applied to the measurement of physiologic response as well.

Among the potential advantages of the sum—of—sines (SOS) technique are:

1. Describing functions can be obtained without averaging cross—spectral quantities.

2. Concentration of input power at a few select frequencies enhances the reliability of the describing function measurements at those frequencies.

3. Estimation of remnant power is enhanced.

4. Comparison of spectral estimates at input and non—input frequencies provides an indication of the reliability of the describing function estimate.

SOS techniques can yield reliable performance estimates over a relatively wide

31.1

frequency bandwidth for idealized laboratory tasks in which a high bandwidth system is controlled and in which the operator is paying close attention to the tracking task [1]. Measurement bandwidth may be seriously reduced, however, when the tracking dynamics contain significant lags or delays [2]; when the operator is involved in an operational task, or in a realistic simulation thereof, requiring attention to tasks other than continuous control [3]; or when measurements are made of inherently "noisy" physiologic response mechanisms such as evoked electrocortical response [4,5]. In these situations ensemble averaging procedures are required to maximize the bandwidth over which reliable performance estimates can be obtained.

The purpose of this article is to suggest a particular method for computing the average operator describing function from an ensemble of experimental trials, and for estimating the reliability of the ensemble mean, in both open-loop and closed-loop measurement environments. Compared to analysis methods used in the recent past by this author and others, the methods suggested here are expected to increase the bandwidth over which reliable performance measures can be obtained.

The method suggested here makes use of trial-to-trial variations in the describing function to determine the reliability of the describing function estimates. This method, of course, requires that a number of experimental replicates be obtained. If there are only a few replicates -- or only a single trial -- reliability must be determined from remnant measurements as outlined above.

The following discussion is confined to experiments using SOS inputs. The reader is directed to two review articles [6,7] for a more detailed discussion of SOS analysis techniques, and for a comparison of SOS with alternative techniques for identifying operator response parameters.

CURRENT PRACTICE

Given sufficient time for transients to damp out, a noise-free linear system driven by a sum-of-sines (SOS) input will respond only at frequencies contained in the forcing function. Describing function estimates, therefore, are obtained only at input (i.e., SOS) frequencies. Conversely, system response power at non-input frequencies is defined as "remnant".

Experimental data are usually digitized for either online or offline analysis by digital computer. The resultant time histories, then, are sampled, and analysis techniques appropriate to sampled data are employed. Discrete Fourier transform (DFT) techniques are employed to compute Fourier coefficients of relevant time histories, and to compute estimates of power spectra (actually, squared magnitudes of Fourier coefficients).

The following procedure has often been used to estimate human operator describing functions and remnant:

1. By means of the DFT, compute Fourier coefficients for the time histories representing the operator's input (e.g., tracking error) and output (e.g., control response).

2. At each SOS frequency, compute the estimate of the operator's describing function as the (complex) ratio H of the Fourier coefficient of the output signal to the Fourier coefficient of the input signal. Express this estimate in terms of "gain" and "phase shift", where

31.2

$$\text{Gain} \quad = \quad 10 \ \log \ (|\text{H}|^2) \quad dB$$

$$\text{Phase} \ = \quad 57.3 \ \tan^{-1} \ (\text{Im}\{\text{H}\}/\text{Re}\{\text{H}\}) \quad \text{degrees}$$

3. Compute the "spectra" for the input and output signals as the magnitude-squared of the Fourier coefficients.

4. For both the input and output signals, compute the average remnant power in a small frequency band about each SOS frequency. Assume the remnant power varies smoothly with frequency, and consider this average power to be an estimate of the remnant power at the corresponding SOS frequency.

5. Compute signal–to–noise (S/N) ratios for both signals by dividing the power actually measured at a given SOS frequency by the estimated remnant power at that frequency. If the S/N ratios for both input and response signals are above some criterion level (typically, 6 or 7 dB) at a given SOS frequency, consider the corresponding describing function estimate computed in Step 2 to be valid. If the S/N for either the input or the output signal falls below the criterion, we conclude that a valid describing function cannot be obtained at that particular frequency.

The above procedure is a reasonable one to follow when considering a single experimental trial, as it prevents the acceptance of a describing function estimate that is likely to be seriously corrupted by operator remnant. When performing experiments with human test subjects, however, we generally attempt to improve measurement reliability by ensemble–averaging the results from a number of replications of a given test condition.

To compute ensemble statistics of the operator describing function, we first compute the describing function (in terms of gain and phase) for each experimental trial, retaining only those measurements considered valid by the signal–to–noise test. Using only these valid measurements, we then compute the mean and standard deviation of the gain, and the mean and standard deviation of the phase shift at each SOS frequency.

While this method is straightforward, it is deficient in a number of respects. First, it tends to be pessimistic in that it tests the reliability of each individual measurement rather than of the ensemble mean. As a result, certain measures are unnecessarily discarded. Second, it may yield a frequency response curve that has an inconsistent data base. That is, measurements will be retained from all experimental trials at frequencies where remnant is relatively small, whereas measures from only a subset of trials will generally be retained at frequencies where remnant is significant. Finally, this method tends to overestimate the mean gain, because it retains measurements where remnant power has tended to reinforce the input–correlation portion of the response, and it discards measurements where remnant has tended to counteract the input correlated component.

The analysis methodology described in the remainder of this document circumvents these particular difficulties by using all the available data to compute the ensemble mean, and then directly estimating the reliability of the mean. Thus, one retains or rejects all the describing function data at a given SOS frequency.

KEY ASSUMPTIONS AND RELATIONSHIPS

Before proceeding with the development of the describing function analysis techniques, we first make certain assumptions concerning the nature of operator remnant, and we then present certain mathematical results that are used in the subsequent development.

Remnant

The following discussion concerns Fourier coefficients of the remnant processes as might be determined by a DFT. In general, a number of experimental trials are analyzed and, for each trial and each signal analyzed, remnant coefficients are computed at each DFT frequency.

In general, a remnant-related DFT coefficient will be a complex number. Let

$$R_{i,k} = X_{i,k} + jY_{i,k}$$

Where R is a complex quantity having real part X and imaginary part Y, and "i" and "k" are the frequency and ensemble (i.e., experimental replication) indices, respectively.

The following key assumptions are made concerning the remnant process:

Assumption 1: Remnant is linearly uncorrelated with external signals and system functions.

Assumption 2: The Fourier coefficients are zero-mean Gaussian variables. Thus

$$\varepsilon\{X_{i,k}\} = \varepsilon\{Y_{i,k}\} = 0$$

where is the expectation operator.

Assumption 3): The real and imaginary components of R are linearly uncorrelated across frequency and across replications: Thus,

31.4

$$\varepsilon\{X_{i,k} \cdot Y_{j,\ell}\} = 0 \qquad \text{all } i,j,k,\ell$$

Assumption 4:   The autocovariance of the real part is equal to the autocovariance of the imaginary part.   The real and imaginary parts of the Fourier coefficient are otherwise uncorrelated across frequencies and across replications.   Thus

$$\varepsilon\{X_{i,k} \cdot X_{j,\ell}\} = \varepsilon\{Y_{i,k} \cdot Y_{j,\ell}\}$$

$$= \begin{cases} 0 & i \neq j \text{ or } k \neq \ell \\ \sigma_{r/2}^2 & i = j \text{ and } k = \ell \end{cases}$$

where

$$\sigma_r^2 = \varepsilon\{|R_{i,k}|^2\}$$

$$= \varepsilon\{X^2_{i,k}\} + \varepsilon\{Y^2_{i,k}\}$$

Assumption 5:   The remnant process is a smoothly−varying function of frequency. The spectrum contains no "spikes" or "holes", and the power density spectrum may be considered locally stationary over any sufficiently narrow frequency band.   Expressed mathematically,

$$\varepsilon\{X^2_{i,k}\} = \varepsilon\{X^2_{j,\ell}\}$$

for all values of k and ℓ and for i "close" to j.

In summary, the remnant is assumed to be a zero−mean Gaussian process whose real and imaginary coefficients have zero cross−correlation, zero covariance across frequency and replication, and equal autocovariance.   We shall refer to this process as a "stationary incoherent" process, as it implies that remnant power is statistically constant, whereas phasing is randomly distributed between 0 and 2π across frequencies and across replications.

Other than for local stationarity, we make no assumptions concerning the frequency shaping of the remnant process.   In general, the remnant process will be non−white, and the frequency dependencey will depend on the internal state of the operator and on the external task environment.

31.5

Key Relationships

The following key relationships, which follow from the assumptions stated above and from the properties of linear systems, form the basis of the error analysis to follow.

1. Linear Transformation of the Remnant

A Fourier coefficient obtained by transforming the remnant coefficient by a linear system is a stationary incoherent variable. Thus if

$$F = AR = X_f + jY_f$$

where A is the system function (at a given frequency) of some linear process, then

$$\varepsilon\{X_f^2\} = \varepsilon\{Y_f^2\} = |A|^2 \frac{\sigma_r^2}{2}$$

and

$$\varepsilon\{|F|^2\} = |A|^2 \sigma_r^2$$

2. Effects of Averaging

The Fourier coefficient obtained by averaging multiple samples of linearly transformed remnant is a stationary incoherent variable, with variance reduced by the number of samples. Let

$$\bar{F} = \frac{1}{N} \sum_{n=1}^{N} F = \frac{1}{N} \sum_{n=1}^{N} AR$$

$$= X_{\bar{f}} + jY_{\bar{f}}$$

then

$$\varepsilon\{X_{\bar{f}}^2\} = \varepsilon\{Y_{\bar{f}}^2\} = \frac{|A|^2}{N} \cdot \frac{\sigma_r^2}{2}$$

and

$$\varepsilon\{|\bar{F}|^2\} = \frac{|A^2|}{N} \sigma_r^2$$

31.6

3. Error Analysis of Gain and Phase Estimates

Let

$$H = H_o \cdot (1+R')$$

where H is the describing function (or average describing function) measured in some experiment, H_o is the "true" describing function that one wishes to estimate, and R' is a stationary incoherent noise process (typically, the operator's remnant R linearly transformed and averaged), such that

$$\varepsilon\{|R'|^2\} = \sigma_{r'}^2,$$

We define the operations of computing gain and phase as follows:

$$G(H) \equiv 10 \cdot Log(|H|^2) \qquad dB$$

$$\phi(H) = 57.3 \ tan^{-1}(Im\{H\}/Re\{H\}) \quad \cdot degrees$$

If $\sigma_r^2 << 1$, the following approximations (see the Appendix) may be used for estimating the gain and phase and their standard errors:

$$\hat{G}_o \simeq G(\hat{H}_o)$$

$$\hat{\phi}_o \simeq \phi(\hat{H}_o)$$

$$\sigma_{\tilde{G}} = 6.14 \ \sigma_{r'}$$

$$\sigma_{\tilde{\phi}_o} = 40.5 \ \sigma_{r'} = 6.60 \ \sigma_{\tilde{G}}$$

where the symbols " ^ " and " ~ " signify estimation and estimation error, respectively. Note that a fixed relationship obtains between the estimation errors (i.e., standard error) for gain and phase.[1]

---

[1]The number 6.14 is a three-digit approximation to $\sqrt{2} \cdot 10 \cdot log(e)$; the number 40.5 is an approximation to $(180/\pi)/\sqrt{2}$.

ANALYSIS OF OPEN–LOOP SYSTEMS

We define an "open–loop" measurement environment as one in which the system to be investigated is driven by an external forcing function whose characteristics can be controlled and measured exactly. With respect to human operator response, the measurement of a physiologic response such as a visually–evoked electrocortical response [4] falls into this category. A block diagram of such a system is given in Figure 1, where E represents the Fourier coefficient of the external input (or "error"), C the Fourier coefficient of the system response, H the linear system function to be estimated, and R the Fourier coefficient of the remnant added to the linear portion of the system response. The goal of analysis procedure is to estimate the gain and phase shift of the describing function H, and to estimate the associated standard errors.
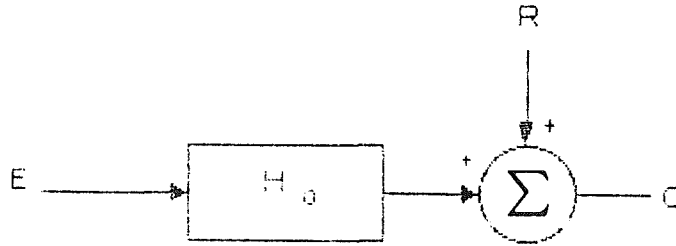
Figure 1: Block Diagram of the Open–Loop Measurement Situation

At non–SOS frequencies, the response C will consist simply of remnant. At SOS frequencies, the response will consist of the sum of remnant plus an input–correlated component. Thus, at input frequencies;

$$C = H_O E + R$$

We assume that E is statistically stationary for a given SOS frequency. That is

$$\left| E_n \right|^2 = E$$

for all n. In general, however, the phasing of E will vary from trial–to–trial.

Let H_n be the describing function (a complex number) measured on the n_{th} experimental trial for some SOS frequency. Thus

$$H_n \equiv \frac{C_n}{E_n} = H_O + \frac{R_n}{E_n}$$

31.8

The describing function may be expressed equivalently as

$$H_n = H_0 + \tilde{H} = H_0(1+\tilde{H}/H_0)$$

where $\tilde{H} = R/E$. Since R is (by assumption) an incoherent random process, H_0 is (by assumption) a constant, and E is (by design) a complex number having a fixed magnitude, the processes $\tilde{H}$ and $\tilde{H}/H_0$ are also stationary incoherent processes.

Because we are considering stationary incoherent process, statistics of the complex variable $\tilde{H}/H_0$ may be computed the same way one would compute statistics for a Gaussian random variable. Specifically, an unbiased estimate of the population mean is the experimental sample mean, and an unbiased estimate of the population variance may be computed as the ensemble sum of the magnitude-squared, minus the sum of squared magnitude of the experimental mean, divided by the number of samples minus one. Thus

$$\hat{h}_0 = \bar{H}$$

$$\sigma^2_{\tilde{H}/H} = \frac{1}{|H_0|^2} \cdot \frac{1}{N-1} \cdot \left[\sum |H|^2 - \sum |\bar{H}|^2 \right]$$

$$= \frac{1}{|H_0|^2} \frac{N}{N-1} \left[\overline{|H|^2} - |\bar{H}|^2 \right]$$

where $\bar{H}$ is the empirical average of H over the N experimental trials and $\overline{|H|^2}$ is the average squared magnitude.

Since the averaging process reduces the variances by 1/N, the estimated variance of the error in the mean describing function (i.e., the standard error) is

$$\sigma^2_{\tilde{H}/H_0} = \frac{1}{N-1} \left[\frac{\overline{|H|^2}}{|\bar{H}|^2} - 1 \right]$$

where $|\bar{H}|^2$ is taken as an estimate of $|H_0|^2$.

If we define $\sigma^2_{\tilde{H}/H_0}$ as σ^2_r , then, from the relationships of (1) developed earlier, we obtain the following expressions for the estimated mean gain and phase, and associated standard errors:

$$\hat{G}_O = G(\bar{H})$$

$$\hat{\phi}_O = \phi(\bar{H})$$

$$\sigma_{\tilde{G}} = 6.14 \left[\frac{1}{N-1} \left(\frac{\overline{|H|^2}}{|\bar{H}|^2} - 1 \right) \right]^{1/2}$$

$$\sigma_{\tilde{\phi}} = 6.60 \ \sigma_{\tilde{G}}$$

Since the mean—squared value of a quantity is never less than the square of the sample mean, the expression for $\sigma_{\tilde{G}}$ is guaranteed to be mathematically well—behaved in that the quantity to be square—rooted is always non—negative.

In the case where one has a sufficient number of experimental replications (say, more than four), the following procedure is recommended for estimation of describing functions at each SOS frequency:

1. Compute the describing function for each replicate.

2. Average the describing function measurements (as complex coefficients) across trials.

3. Compute the estimated gain and phase shift from the average (complex) describing function.

4. Estimate the standard errors of the gain and phase estimates from the relationships given above.

5. Accept the gain and phase estimates as "valid" if the standard error for gain is below some criterion level (say, 2 or 3 dB); otherwise, reject the estimates as "invalid".

There are two reasons for testing the estimated standard error against some criterion for validity. First, a large standard error would tend to render the estimated gain and phase of minimal usefulness for further analysis (such as averaging with the results of other test subjects, or for performing model analysis). Second, the procedures given here for estimating standard errors are valid only if these errors are relatively small. Thus, if we compute a relatively large standard error, we are in doubt not only about the mean gain and phase, but we are also unsure of the reliability of these estimates.

ANALYSIS OF CLOSED—LOOP SYSTEMS

A closed—loop system is defined diagrammatically in Figure 2. In this situation, the input E to the system of interest is not an independent variable, but a linear function of an external SOS input I and the system response C. Again, the operator's response

31.10

is assumed to contain a component linearly related to his input, plus a remnant component.   A minus sign is associated with H_o in Figure 2 to conform to the conventions used when analyzing systems with negative feedback.
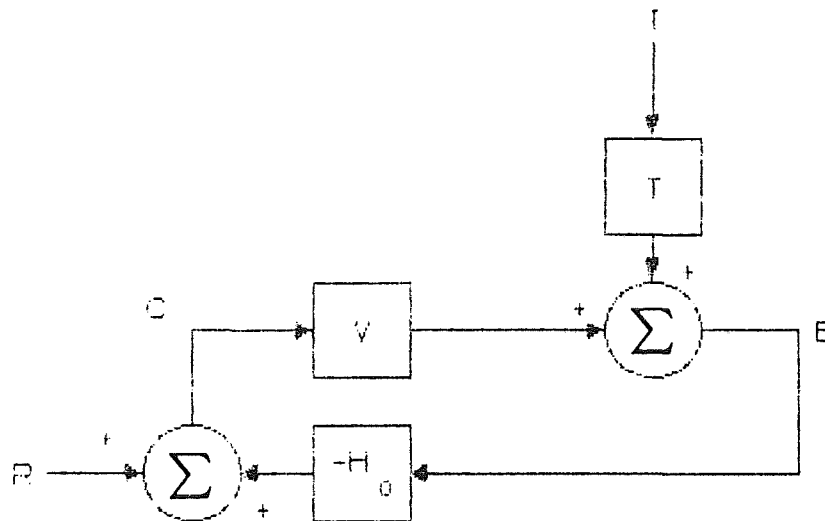
Figure 2:   Block Diagram of the Closed—Loop Measurement Situation

The transfer function T is included in the system diagram to allow general treatment of the external SOS forcing function I. For example, T is unity when I is a simple command or target input, whereas T is equivalent to the vehicle transfer function V when the input is added directly to the operator's control response.   On the other hand, if the system to be analyzed consists of a simulated flight task with gusts interacting in an aerodynamically realistic fashion, the transfer function T will be neither V nor unity.

Again, the measurement goal is to obtain an estimate of the operator's describing function H_o, expressed in terms of gain and phase, and to estimate the associated standard errors.   The situation is complicated, however, by the fact that the input to the operator is not an independent variable under the complete control of the experimenter, but is determined in part by the closed—loop system response and is therefore corrupted by operator remnant.

From Figure 2 we derive the following relationships between the "error" and "control" signals (i.e., input and output to the human operator) and the independent forcing function:

$$E = \frac{T}{\Delta} I + \frac{V}{\Delta} R$$

$$C = -\frac{H_o T}{\Delta} I + \frac{1}{\Delta} R$$

where

$$\Delta = 1 + H_o V$$

31.11

If we were to compute the operator's describing function from a single experimental trial, we would obtain

$$H = -\frac{C}{E} = H_o \frac{1 - \frac{R}{H_oTI}}{1 + \frac{VR}{TI}}$$

Because remnant appears in both the denominator and numerator terms, ensemble-averaging the describing function estimates computed in this manner will not necessarily reduce the measurement error due to remnant. In the case of sufficiently large remnant, the describing function computed as shown above will be approximately the negative inverse of the vehicle dynamics –– not the desired quantity H_θ –– no matter how many replications are averaged.

An alternative approach is to perform the averaging process <u>before</u> performing the division. This can be accomplished by using the following average cross–power coefficients:

$$\overline{EI^*} = \frac{T}{\Delta} |I|^2 + \frac{V}{\Delta} \overline{RI^*}$$

$$\overline{CI^*} = - \frac{H_oT}{\Delta} |I|^2 + \frac{1}{\Delta} \overline{RI^*}$$

where the overstrike represents an average computed over N experimental replications. As the input power is assumed to be stationary across replicates, average input power is equal to the input power at any replicate (for a given SOS frequency).

We now compute the average (complex) transfer function from the average cross-spectral components as:

$$\hat{H}_o = \frac{\overline{CI^*}}{\overline{EI^*}} = H_o \frac{1 - \frac{\overline{RI^*}}{H_oT|I|^2}}{1 + \frac{V \overline{RI^*}}{T|I|^2}}$$

The expression for the estimated transfer no longer contains the "raw" remnant signal, but rather the average of the cross–correlation of the remnant with the external forcing function. Because the remnant is, by definition, theoretically uncorrelated with the forcing function, the error due to remnant will tend toward zero

31.12

as the number of replicates increases, even when the remnant power is relatively large.

The following effective "remnant variance", derived in the appendix, is needed for estimating standard errors:

$$\sigma_{r'}^2 = \frac{1}{(N-1)} \left[\frac{\overline{|CI*|^2}}{|\overline{CI*}|^2} + \frac{\overline{|EI*|^2}}{|\overline{EI*}|^2} - 2 \cdot Re \left\{ \frac{\overline{(CI*)(EI*)^*}}{\overline{(CI*)} \cdot (EI*)^*} \right\} \right]$$

where averaging operations are performed across an ensemble of N experimental trials. The third term in the above expression accounts for the linear correlation (through the vehicle dynamics) between the remnant—related components of the error and control signals. The variance term σ_r^2, is guaranteed to be non—negative. The square root of the above quantity is used to compute the following standard errors:

$$\sigma_{\tilde{G}} = 6.14 \ \sigma_{r'} \qquad dB$$

$$\sigma_{\tilde{\phi}} = 40.5 \ \sigma_{r'} = 6.60 \ \sigma_{\tilde{G}} \qquad \text{degrees}$$

The following procedure is recommended for estimating operator describing functions obtained from a closed—loop control environment:

1. For a given SOS frequency, compute the cross—power coefficients CI* and EI*, and the magnitude—squared of these complex coefficients, for each experimental trial in the ensemble.

2. Average the cross—power coefficients across the ensemble, and compute the average (complex) transfer function as $-\overline{CI*}/\overline{EI*}$.

3. Compute the gain and phase from the average describing function computed in Step 2.

4. Estimate the standard errors of the gain and phase estimates as shown above.

5. Discard measurements for which the standard error of the gain exceeds some allowable maximum level.

DISCUSSION

Modifications to current methods for estimating human operator describing functions have been suggested. Although the techniques proposed here are similar to those employed when inputs are continuous in frequency, they are not generally employed when sum—of—sinusoids inputs are used.

31.13

The key assumption underlying the method is that operator remnant is a "stationary incoherent" process -- that is, a zero-mean Gaussian process whose real and imaginary DFT coefficients have zero cross-correlation, zero covariance across frequency and experimental replication, and equal autocovariance at a given frequency. With this assumption, along with the known properties of linear systems, we can compute approximations to the estimation errors for both gain and phase shift.

Some of the key features of the method are:

1. Statistics are performed on Fourier coefficients or ratios of Fourier coefficients; these statistics are then transformed to the amplitude ratio ("gain") and phase-shift domain.

2. For closed-loop systems, the estimated average describing function is computed as the ratio of the (a) ensemble-averaged cross-power spectral densities between input and control response, and (b) the cross-power density between input and error. Describing function estimates are not directly computed for individual trials. For open--loop systems, however, the average describing function is computed by averaging the describing function computations for individual trials.

3. The standard errors of the average gain and phase estimates are computed as transforms of the standard errors of Fourier coefficients; they are not computed by first determining the standard deviations of gain and phase and then normalizing by the square root of the number of trials.

4. Data from all experimental trials are used in computing the describing function statistics. Reliability criteria are applied to the resulting averages, not to individual describing function estimates.

5. The standard error associated with the phase-shift estimate at a given measurement frequency is related by a known constant to the standard error of the corresponding gain estimate.

The assumption behind this proposed technique is that one is interested primarily in analyzing a given subject's average behavior. This is usually the case when subjects have been trained to asymptotic behavior, or where model analysis is to be performed. In this case, the reliability of the experimental mean (i.e., the "standard error") is of direct concern, not the reliability of measurements that might be obtained in individual experimental trials.

If trial-to-trial variations are of interest, however, as might be the case in studies of training effectiveness, it may be necessary to estimate operator performance on single experimental trials, using remnant-based methods for determining measurement reliability.

The derivation of the methods presented here were motivated by difficulties in obtaining reliable estimates of describing functions for physiologic systems and for pilot response behavior in simulations of operational situations. The method has recently been used to analyze visual evoked electrocortical responses, [4,5] , and is contemplated for application to data obtained from a simulated air-combat tracking task [7].

REFERENCES

1. Levison, W.H., "Measurement of Human Operator Response Behavior", Report No. 4307, Bolt Beranek and Newman Inc., Cambridge, MA, February 1980.

2. Levison, W.H., "The Effects of Display Gain and Signal Bandwidth on Human Controller Remnant," AMRL-TR-70-93, Aerospace Medical Research Laboratory, Wright-Patterson Air Force Base, Ohio, March 1971.

3. Levison, W.H., Muralidharan, R., "Analysis of Manual Control Tasks in Support of the Chemical Defense Program", Report No. 5168, Bolt Beranek and Newman Inc., Cambridge, MA, December 1982.

4. Junker, A.M., Kenner, K.M., Kleinman, D.L., McClurg, T.D., "Comparison of Steady-State and Transient Visual Evoked Response Potential", Twenty-First Annual Conference on Manual Control, June 17-19, Ohio State University, Columbus, Ohio.

5. Levison, W.H., Junker, A.M., Kenner, K.M., "Descriptive Linear Modelling of Steady-State Visual Evoked Response", Twenty-First Annual Conference on Manual Control, June 17-19, Ohio State University, Columbus, Ohio.

6. Levison, W.H., "Methods for Identifying Pilot Dynamics", in Frazier, M.L., and Crombie, R.B., (eds): Proceedings of the Workshop on Flight Testing to Identify Pilot Workload and Pilot Dynamics, AFFTC-TR-82-5, Air Force Flight Test Center, Edwards Air Force Base, CA, May 1982.

7. "F-14 Modelling Study", NASA Langley Research Center, Contract No. NAS1-17648.

ACKNOWLEDGEMENT

APPENDIX

Error Analysis for Closed-Loop Describing Function

In the main text we developed the following expression for estimating describing functions in closed-loop control tasks:

$$\hat{H}_O = \frac{\overline{CI^*}}{\overline{EI^*}}$$

where H is the estimated describing function at a given input frequency; I,E, and C are the complex Fourier coefficients of the input, error, and control signals, respectively;

and the overstrike indicates ensemble averaging across experimental replications. Note that the average cross-power products are obtained before the ratio is taken.

It is convenient to represent the computed average describing function as

$$\overline{H} = H_o (1 + r')$$

where H is the "theoretical" or "true" describing function (i.e., the describing function one would measure if the operator were totally linear, noise-free, and consistent), and r' is the deviation of the empirical average from this value. In the following development we derive the variance (expected squared magnitude) of the complex quantity r'.

To simplify the notation, we define the following quantities:

$$C' \equiv \overline{CI^*}$$

$$E' \equiv \overline{EI^*}$$

We now write the average cross-power spectral quantities as

$$C' = C'_o + \tilde{C}'$$

$$E' = E'_o + \tilde{E}'$$

The average describing function may then be represented as

$$\overline{H} = \frac{C'_o + \tilde{C}'}{E'_o + \tilde{E}'} = H_o \frac{1 + \tilde{C}/C'_o}{1 + \tilde{E}/E'_o}$$

If we assume that $\tilde{E}'/E'_o \ll 1$, the above expression may be approximated as:

$$\overline{H} \simeq E_0 \left(1 + \tilde{C}'/ C_0' - \tilde{E}' /E_0' \right)$$

The "measurement error" term r' is thus identified as

$$r' \simeq \tilde{C}'/ C_0' - \tilde{E}'/ E_0'$$

Because the quantities $\tilde{E}'$ and $\tilde{C}'$ are, by assumption, "stationary incoherent" processes as defined in the main text, the error term r' is also a stationary incoherent process.

We cannot measure the error quantities $\tilde{C}$ and $\tilde{E}$. In order to work with quantities that can be measured (or estimated from measured quantities), we use the relationships

$$\tilde{E}' = E' - E_0'$$

$$\tilde{C}' = C' - C_0'$$

to derive the following equivalent expression for the error term:

$$r' = \frac{C' - C_0'}{C_0} - \frac{E' - E_0'}{E_0}$$

$$= \left(\frac{C'}{C_0} - 1\right) - \left(\frac{E'}{E_0} - 1\right) = \frac{C'}{C_0} - \frac{E'}{E_0}$$

The expected magnitude—squared of the error term is thus computed as

31.17

$$\hat{\sigma}^2_{r'} = \frac{N}{N-1}\left[\frac{\overline{|C'|^2}}{|C_o'|^2} + \frac{\overline{|E'|^2}}{|E_o'|^2} - 2\cdot\text{Re}\left\{\frac{\overline{C'\ E'^*}}{C_o'\ E_o'^*}\right\}\right]$$

In terms of the cross-power quantities computed from the experimental data, the above expression may be written as

$$\hat{\sigma}^2_{r'} = \frac{1}{N-1}\left[\frac{\overline{|C1*|^2}}{|\overline{C1*}|^2} + \frac{\overline{|EI*|^2}}{|\overline{EI*}|^2} - 2\cdot\text{Re}\left\{\frac{\overline{(CI*)\ (EI*)^*}}{(\overline{CI*})\cdot(\overline{EI*})^*}\right\}\right]$$

Note that the factor "N" has been dropped from the numerator, because the variance of interest is the standard error of the mean, not the trial-to-trial standard deviation. Also, the empirical calculation of $|\overline{CI*}|^2$ is used for $|C_o'|^2$, etc.

Transformation from Complex to Gain/Phase Domain

The methodology presented in this paper requires that statistics of the describing function measures (mean and standard error of the mean) be obtained in the complex-number domain, then transformed into the gain/phase domain. This transformation is derived below.

Let the estimated average describing function (complex quantity) be expressed as

$$\overline{H} = H_o(1 + r')$$

where H is the "true" describing function and r' is a stationary incoherent error term having a variance as derived above. Let X and Y represent the imaginary parts of r; the above expression may be written as:

$$\overline{H} = H_O (1 + X + jY)$$

Gain Computation

The gain G is defined as

$$G = 10 \; \text{Log} \; (|H|^2) = 4.34 \; \text{Ln}(|H|^2)$$
$$= G_O + G_e$$

where

$$G_O = 10 \; \text{Log} \; (H_O)$$
$$G_e = 4.34 \; \text{Ln} \; (1 + 2X + X^2 + Y^2)$$

We note that the natural logarithm of $(1+z)$ may be expressed by the series $z - z^2/2 + z^3/3 - (\text{etc.})$. If X and Y are Gaussian variables or otherwise have symmetric statistics, expected value of odd powers are zero. If, in addition, the magnitude of the error term is small compared to unity, we may ignore powers greated than 2 when computing expected values. Thus,

$$G_e \simeq 2X + Y^2 - X^2$$

31.19

Since X and Y are assumed equi-variant, the mean of the error term is negligibly different from zero. Therefore, performing the "gain operation" on the magnitude of the average describing function yields an unbiased estimate of the gain of the "true" describing function.

If we drop terms higher than second-order, the variance (expected mean-squared magnitude) of the error term is approximately

$$\sigma^2_{G_e} \simeq (4.34)^2 \; \varepsilon \left\{ 4 \cdot X^2 \right\}$$

Noting that the expected value of X is half the expected magnitude-squared of the variable r', we obtain

$$\sigma^2_G = 2 \; (4.34)^2 \; \sigma^2_{r'}$$

and

$$\sigma_G = 6.14 \; \sigma_{r'} \quad dB$$

Phase Computation

The phase shift of H may be expressed as

31.20

$$\phi = \phi_0 + \phi_e$$

where

$$\phi_e = \tan^{-1} \left[Y/(1+X) \right]$$

If we assume X,Y << 1, the phase shift of the error term is approximately

$$\phi_e \simeq Y (1 - X + X^2)$$

Because X and Y are assumed to be linearly uncorrelated, the expected mean error is zero. Thus, performing the "phase operation" on the average describing function yields an unbiased estimate of the average phase shift. If we square the error term, ignore terms higher than second order, and take the expected value (recalling that X and Y are equi—variate), we obtain

$$\sigma_\phi^2 = \sigma_r^2 / 2 \qquad \text{rad}^2$$

$$= (57.3)^2 \ \sigma_r^2 / 2 \qquad \text{deg}^2$$

and

$$\sigma_\phi = 40.5 \ \sigma_r \qquad \text{deg}$$

Note that the standard error of the phase bears a fixed relationship to the standard error of the gain.

31.21

MANUAL CONTROL OF UNSTABLE SYSTEMS

R. Wade Allen
Jeffrey R. Hogue
Zareh Parseghian
Systems Technology, Inc.
13766 S. Hawthorne Blvd.
Hawthorne, CA 90250
(213)679-2281

ABSTRACT

Under certain operational regimes and failure modes, air and ground vehicles can present the human operator with a dynamically unstable or divergent control task. Research conducted over the last two decades has explored the ability of the human operator to control unstable systems under a variety of circumstances. This paper will review past research and summarize human operator control capabilities. A current example of automobile directional control under rear brake lockup conditions is also reviewed. A control system model analysis of the driver's steering control task is summarized, based on a generic driver/vehicle model presented at last year's Annual Manual. Results from closed course braking tests are presented that confirm the difficulty the average driver has in controlling the unstable directional dynamics arising from rear wheel lockup.

INTRODUCTION

Unstable vehicle dynamics present a rather specific task demand on the human operator. Vehicle system states tend to diverge exponentially, and the human controller must be alert and attentive enough to counteract this divergent system behavior. In many situations, due to a transition in vehicle behavior (e.g., component failures or a change in operating conditions), unstable dynamics may occur unexpectedly. In this case the human operator must detect the change and adapt to the vehicle's new response characteristics. Some attention has been devoted to control of unstable dynamic systems at past manual control conferences (e.g., Refs. 1-3).

In this paper we will start off with a simple analysis of the response of unstable vehicles. Next we will consider the ability of the human operator to control unstable dynamics. Then we will analyze an unstable vehicle control problem, i.e., a car with the rear wheels locked up during braking. Following this, the closed-loop stability properties of cars with and without rear wheel lockup are analyzed. Finally, field test data is presented which illustrates the ability of the average driver in controlling unstable automobile dynamics.

BACKGROUND

It is important to focus on the nature of unstable vehicle dynamics in order to appreciate the task difficulty imposed on the human operator.

32.1

Basically, simple unstable vehicle dynamics result in the exponential divergence of state variables and their derivatives:

$$X = Ke^{t/T\lambda}$$

$$\dot{X} = \frac{K}{T\lambda} e^{t/T\lambda} \qquad (1)$$

$$\ddot{X} = \frac{K}{T\lambda^2} e^{t/T\lambda}$$

where

$$t = \text{time}$$

$$K = \text{multiplying constant}$$

$$T_\lambda = \text{divergence time constant}$$

This effect occurs without any forcing function, and it should be noted that all variables have the same characteristic exponential time response, differing only by a multiplying constant as indicated above.

This exponential divergence characteristic is apparent in both field test and simulation data associated with simple unstable dynamics. To observe this, first note that the time required for an exponential curve to double in amplitude is related to the divergence time constant of the exponential as derived in Table 1:

$$T_\lambda = 1.44 \; \Delta t_{2/1}$$

Given vehicle response test data, this relationship can be used to identify divergence time constants as will be discussed subsequently.

HUMAN OPERATOR CAPABILITY

Given that unstable vehicle dynamics result in an exponential state variable divergence, can the human operator be expected to control such an occurrence? This question has been addressed extensively in the literature, involving a variety of situations including aircraft piloting, tracking task research, and a vehicle mounted task for screening drunk drivers. A summary of this research is given in Table 2 including the limiting divergence time

TABLE 1.   RELATIONSHIP BETWEEN TIME TO DOUBLE
AND DIVERGENCE TIME CONSTANT

A system with an unstable root $s = \lambda$ will have an exponentially divergent response given by

$$X = Ke^{t/T_\lambda}$$

where $T_\lambda = 1/\lambda$

Now evaluate X at two time points

$$X_1 = Ke^{t_1/T_\lambda} \quad ; \quad X_2 = 2X_1 = Ke^{t_2/T_\lambda}$$

then

$$\frac{X_2}{X_1} = 2 = \frac{e^{t_2/T_\lambda}}{e^{t_1/T_\lambda}} = e^{(t_2 - t_1)/T_\lambda}$$

$$\therefore \quad \frac{t_2 - t_1}{T_\lambda} = \ln 2$$

and $T_\lambda = \dfrac{t_2 - t_1}{\ln 2}$

Finally

$$T_\lambda = 1.44 \, (t_2 - t_1) = 1.44 \, \Delta t_{2/1}$$

32.3

TABLE 2. SUMMARY OF RESEARCH ON HUMAN CONTROL OF TASKS
WITH UNSTABLE DYNAMICS

REF.	STUDY	UNSTABLE CONTROL LIMIT		
		$\Delta t_{2/1}$ (sec)	$T\lambda$ (sec)	λ (rad/sec)
4	Cheatham (1954): study of the characteristics of human pilot control response to simulated aircraft lateral motions using rudder pedals	0.3	0.43	2.3
5	Jex, et al. (1960): correlation of theoretical limits with past experimental results	0.23	0.33	3.0
6	Sadoff, et al. (1961): experimental study of aircraft longitudinal control problems limit unacceptable	0.58 1.4	0.835 2.0	1.2 0.5
7	Taylor & Day (1961): controllability limits determined from simulator and flight tests Long. practiced inexperienced	0.3 0.5	0.43 0.72	2.3 1.4
	Lat. practiced inexperienced	0.28 0.46	0.40 0.66	2.5 1.5
8	Jex & Cromwell (1962): theoretical and experimental study of aircraft longitudinal handling qualities parameters	0.23	0.33	3.0
9	Young & Meiry (1965): manual control of unstable systems with visual and motion cues	0.3	0.43	2.3
10	Washizu & Miyajima (1965): theoretical and experimental study of human pilot lateral controllability limits practiced inexperienced	0.17 0.20	0.24 0.29	4.1 3.5
11	Jex, et al. (1966): studied well practiced limits of human controllability using a laboratory tracking task (Critical Tracking Task or CTT) and isometric control stick	0.11	0.15	6.6
12	Allen, et al. (1983): CTT mounted in a car, used as a drunk driver detection system practiced inexperienced	0.14 0.28	0.2 0.4	5 2.5

constant that subjects were able to control. This summary suggests the following regarding human operator capability:

1) inexperienced operators can nominally handle divergence time constants greater than 0.5 sec.

2) well-practiced vehicle operators can handle divergence time constants on the order of 0.3 sec.

3) the well-practiced human operator's ultimate limit is on the order of 0.2 sec when a car steering wheel is used as a control device. When stiff "fly-by-wire" aircraft sticks are used as a control device, the controllable divergence limit can be reduced to 0.15 sec.

The above results are overly optimistic (i.e., time constants are too low) for cases where operators are surprised by a sudden change in vehicle response properties. There is a body of literature that relates to this situation. This literature is summarized in Ref. 13, along with the following summary statement:

> "The process of adaptive control is thought to consist of four phases: retention of prefailure dynamics, detection of the failure, identification of the failure and adaptation of appropriate dynamic form for the postfailure situations, and, finally, optimization of postfailure control. ... Typical detection times for laboratory experiments with sudden changes in gain or velocity range from 0.5 to 3 sec. Times to detect failures involving higher order plants are increased to several seconds and may be considerably longer if emergency training is insufficient.

In the case where the human operator is controlling a vehicle that transitions to unstable operation, any delay in counteracting divergent state variables can be critical. As noted from Table 1, the state variable for a first-order unstable plant will double in less than one divergent time constant (i.e., $\Delta t_{2/1} = 0.69\ T_\lambda$). Thus, state variables could easily diverge over several doubling times for a system with a divergence time constant of less than one second before the human operator detects and recognizes the problem and takes appropriate action. Whether or not the operator can then regain control depends on whether the system has diverged to an uncontrollable state before corrective action is taken.

A CAR DRIVING EXAMPLE

As a common example of a potentially unstable vehicle consider hard braking in an automobile. If the rear brakes should lock first (as can happen in cars with misbalanced brakes or pickup trucks with no cargo), then the vehicle will exhibit a directional instability. A simple approximation for this vehicle behavior can be derived as follows:

1) Assume a simple free body diagram as shown in Fig. 1. This is similar to several approaches that have been discussed in the literature (e.g., Refs. 14, 15).

2) Develop two degree of freedom force and moment equations from the free body diagram as shown in Table 3.

3) Derive the yaw rate transfer function from the Laplace transform of the Table 3 force and moment equations as given in Table 4. Now, for rear wheel lockup, since a locked and sliding wheel cannot develop any side force, set the rear side force coefficient ($Y_{\alpha 2}$) to zero. Then the transfer function reduces to an unstable form as shown in Table 4.

32.5
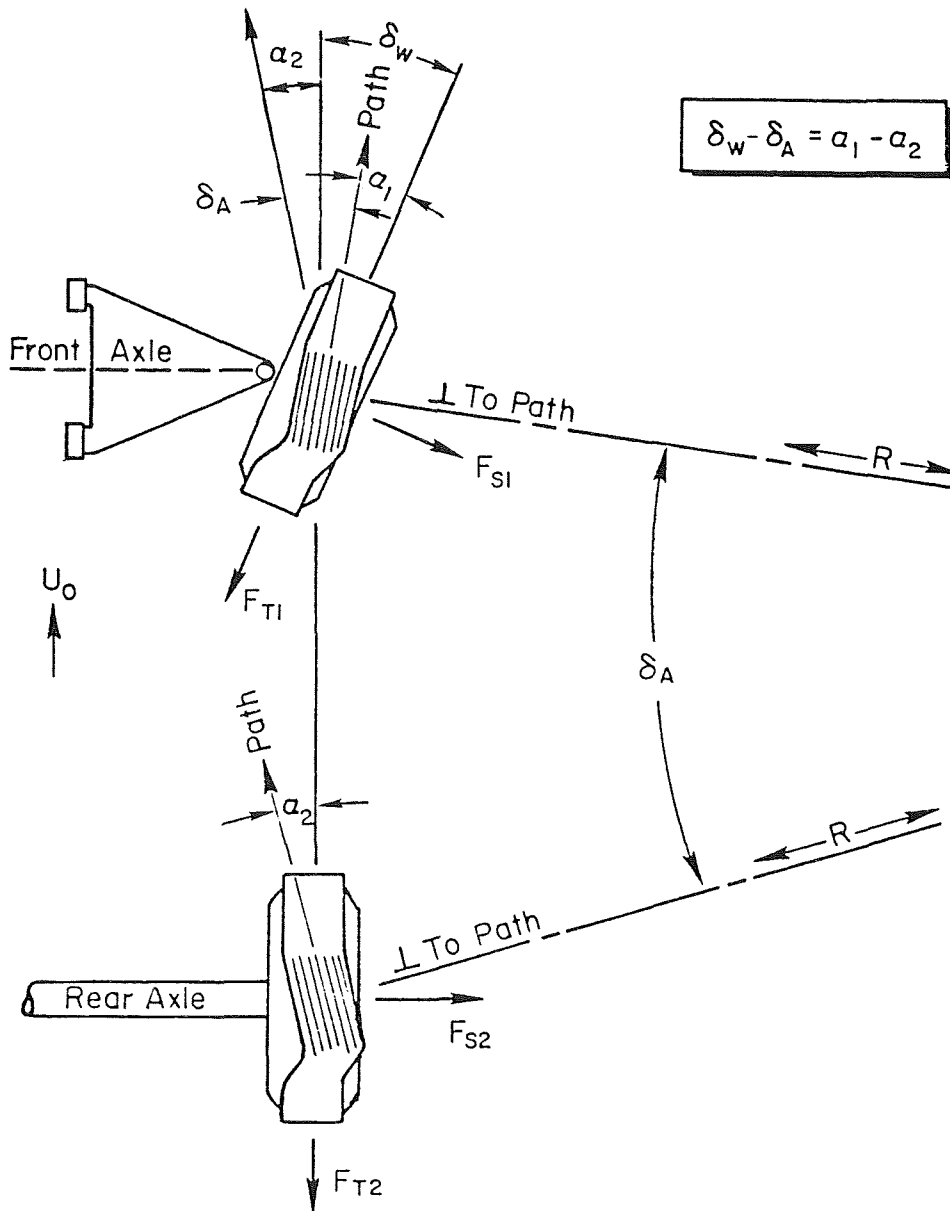
Figure 1. Free Body Diagram and Constant Radius Turn Definitions

Force Equation:

$$m(\dot{v} + U_0 r) = -\left(\frac{Y_{\alpha 1} + Y_{\alpha 2}}{U_0}\right) v - \left(\frac{a\, Y_{\alpha 1} - b\, Y_{\alpha 2}}{U_0}\right)$$

$$+ (Y_{\alpha 1} - F_{T1})\, \delta_w$$

Moment Equation:

$$I\dot{r} = -\left(\frac{a\, Y_{\alpha 1} - b\, Y_{\alpha 2}}{U_0}\right) v - \left(\frac{a^2\, Y_{\alpha 1} + b^2\, Y_{\alpha 2}}{U_0}\right) r$$

$$+ a\, (Y_{\alpha 1} - F_{T1})\, \delta_w$$

v = side slip velocity	r = yaw rate
m = mass	I = moment of inertia
U_0 = longitudinal speed	
$Y_{\alpha 1}$ = front axle side force coeff. (left + right)	
$Y_{\alpha 2}$ = rear axle side force coeff. (left + right)	
F_{T1} = front axle traction force	
a = distance from front axle to c.g.	
b = distance from real axle to c.g.	

TABLE 4. LAPLACE TRANSFORM TRANSFER FUNCTIONS FOR YAW RATE
RESPONSE TO STEERING COMMANDS DEVELOPED
FROM TABLE 3 EQUATIONS

Complete Transfer Function:

$$\frac{r}{\delta_w} = \frac{\left(\frac{Y_{\alpha 1} - F_{T1}}{Y_{\alpha 1}}\right) \bullet \left(\frac{m \, a \, U_0}{\ell \, Y_{\alpha 2}} s + 1\right)}{\frac{m \, U_0 \, I}{\ell \, Y_{\alpha 1} \, Y_{\alpha 2}} s^2 + \left[\frac{m}{\ell}\left(\frac{a^2}{Y_{\alpha 2}} + \frac{b^2}{Y_{\alpha 1}}\right) + \frac{I}{\ell}\left(\frac{1}{Y_{\alpha 1}} + \frac{1}{Y_{\alpha 2}}\right)\right] s + \frac{\ell}{U_0} + \frac{m U_0}{\ell}\left(\frac{b}{Y_{\alpha 1}} - \frac{a}{Y_{\alpha 2}}\right)}$$

where ℓ = wheelbase = a + b

setting $Y_{\alpha 2} = 0$:

$$\frac{r}{\delta_w} = \frac{(Y_{\alpha 1} - F_{T1}) \bullet \frac{a}{I} \bullet s}{s^2 + \frac{Y_{\alpha 1}}{U_0}\left[\frac{a^2}{I} + \frac{1}{m}\right] s - \frac{a \, Y_{\alpha 1}}{I}}$$

Negative constant term in denominator characteristic equation
indicates basic dynamic instability

4) Find roots of the unstable transfer function using the quadratic formula as shown in Table 5. Using typical front wheel drive/passenger car parameters it is apparent from Table 5 that speed only has a minor effect on the divergent time constant, and that typical values for T_λ are in the region of 0.3 seconds.

CLOSED-LOOP VEHICLE CONTROL

Now consider a closed-loop vehicle control model including visual and motion cue feedbacks shown in Fig. 2 that was presented at this conference last year (Ref. 16). Operating in this mode, the car driver ordinarily has a rather easy control task. Past analysis (Ref. 17) has shown that the driver's control parameters can be derived in a fairly straightforward manner. What we wish to consider here is what happens to closed-loop stability when the rear wheels lock up and how must the driver change his/her behavior to maintain stable closed-loop operation.

As has been derived in the past (Ref. 16) the closed-loop stability properties of the Fig. 2 model can be assessed by considering an opened-loop transfer function for the loop broken at the equivalent of the visual feedback point:

$$
G_{OL}(s) = \underbrace{\frac{s + K'}{s}}_{\substack{\text{Trimming} \\ \text{Function}}} \cdot \underbrace{\frac{K_\psi e^{-\tau_v s}}{s}}_{\substack{\text{Yaw Rate} \\ \text{Integration +} \\ \text{Visual Time} \\ \text{Delay}}} \cdot \underbrace{\frac{s + U_o/R}{s}}_{\substack{\text{Kinematic Trans-} \\ \text{fer Function for} \\ \text{Look Ahead} \\ \text{Angular Error}}} \cdot \underbrace{G_{MOT}}_{\substack{\text{Closed-Loop} \\ \text{Transfer Function} \\ \text{for Motion Feed-} \\ \text{back Loop}}}
$$

$$(2)$$

where G_{MOT} is the closed-loop transfer function for the motion feedback loop:

$$
G_{MOT} = \frac{G_{NM} \cdot G_V}{1 + G_{NM} \cdot G_V \cdot e^{-\tau_m s}} \tag{3}
$$

and

G_{NM} = neuromuscular dynamics
G_V = vehicle directional control dynamics
τ_m = motion feedback delay

TABLE 5. ROOTS OF CHARACTERISTIC EQUATION

Equation: $s^2 + Bs + C$

where

$$B = \frac{Y_{\alpha_1}}{U_0}\left[\frac{a^2}{I} + \frac{1}{m}\right] \quad ; \quad C = -\frac{a\,Y_{\alpha_1}}{I}$$

Quadratic Roots:

$$s = \frac{1}{2}\left(-B \pm \sqrt{B^2 - 4C}\right)$$

Typical Front Wheel Driver Passenger Car Parameters:

$$m = 89\ \text{lb-sec}^2/\text{ft} \quad ; \quad I = 1475\ \text{lb-ft-sec}^2$$

$$a = 3\ \text{ft} \quad ; \quad b = 5.75\ \text{ft} \quad ; \quad \ell = a + b = 8.75\ \text{ft}$$

$$\therefore \quad B = 0.0173 \cdot \frac{Y_{\alpha_1}}{U_0} \quad ; \quad C = 0.00203 \cdot Y_{\alpha_1}$$

for $Y_{\alpha_1} = 15,000$

$$s = \frac{1}{2}\left(\frac{260}{U_0} \pm \sqrt{\frac{67340}{U_0^2} + 122}\right)$$

SPEED, U_0		ROOTS (rad/sec)		DIVERGENCE TIME CONST. T_λ (sec)
mph	ft/sec	STABLE	UNSTABLE	
30	44	−9.21	+3.3	0.30
40	58.7	−8.16	+3.74	0.27
50	73.4	−7.57	+4.06	0.25
60	88	−7.19	+4.24	0.24
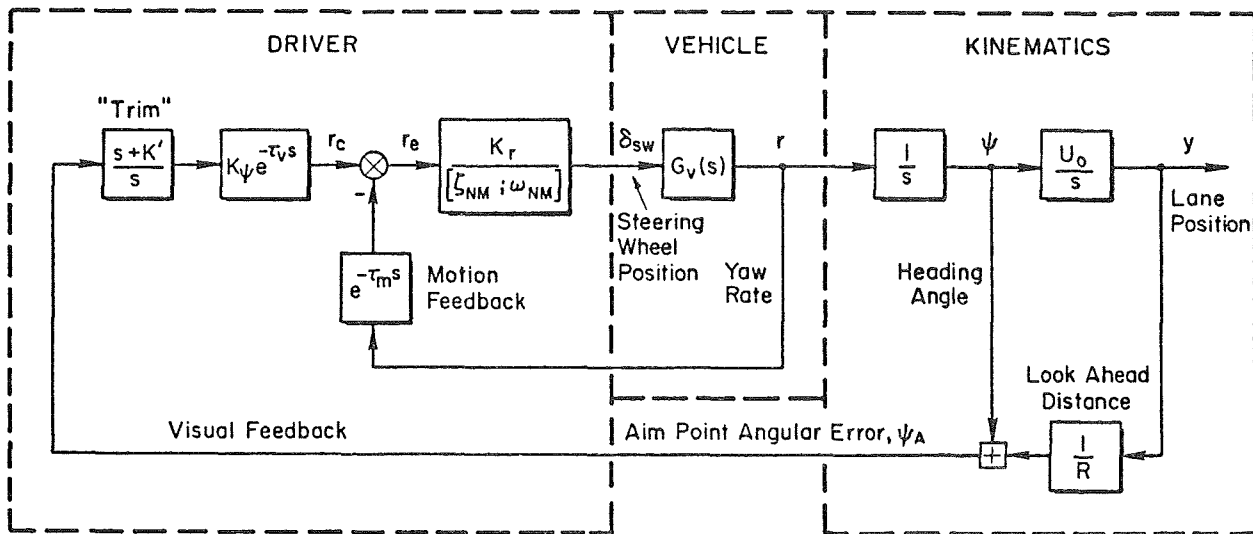
K' = trimming gain, K_ψ = visual feedback gain, K_r = motion feedback loop gain

ζ_{NM} ; ω_{NM} = neuromuscular damping and natural frequency, $G_v(s)$ = vehicle transfer function

τ_v ; τ_m = visual and motion feedback time delay, U_0 = vehicle speed, R = aim point (look aheah) distance

Figure 2.   Driver/Vehicle Stability Analysis Model

The properties of the Eq. 2 transfer function have been discussed in the past (Refs. 16-18), and for nominal driver behavior with stable vehicle dynamics, a Bode plot of Eq. 2 appears as shown in Fig. 3. In order to maintain stable operation the driver must adjust his visual feedback gain K_ψ to lie within the stable phase region as shown. Now consider unstable car dynamics due to rear wheel lock. The driver/vehicle transfer function in Fig. 4 assumes that the driver has maintained his pretransition behavior, and it is obvious that under these circumstances the closed-loop operation will be unstable for any level of visual feedback gain K_ψ because the open-loop phase curve never has less than 180° phase lag!

It is clear from the above results that the driver must change behavior and adapt to rear wheel lockup conditions in order to maintain stable closed-loop vehicle control. Basically the driver must reduce system open-loop phase lag, and this can be accomplished in several phases as follows:

1) Change gain in the motion feedback loop (K_r) to reduce high frequency phase lag shown in Fig. 4.

2) Eliminate trimming behavior (K' = 0) to reduce low frequency phase lag as shown in Fig. 4.

3) Increase lookahead distance R (equivalent of reducing outer loop gain) in order to further reduce low frequency phase lag as shown in Fig. 4.

32.11

$$G_v(S) = \frac{2.07(S+5.61)}{[S^2+9.7S+39.2]}$$

Region of Stable
Crossover Gains

$K' = 0.4$ $K_r = 550$ $K_\psi = 12.8$

$\tau_V = .05$ sec $\tau_m = .05$ sec $U_o = 80.7$ fps

$\zeta_{NM} = .774$ $\omega_{NM} = 6.26$ $U_o/R = 1.5$ r/s
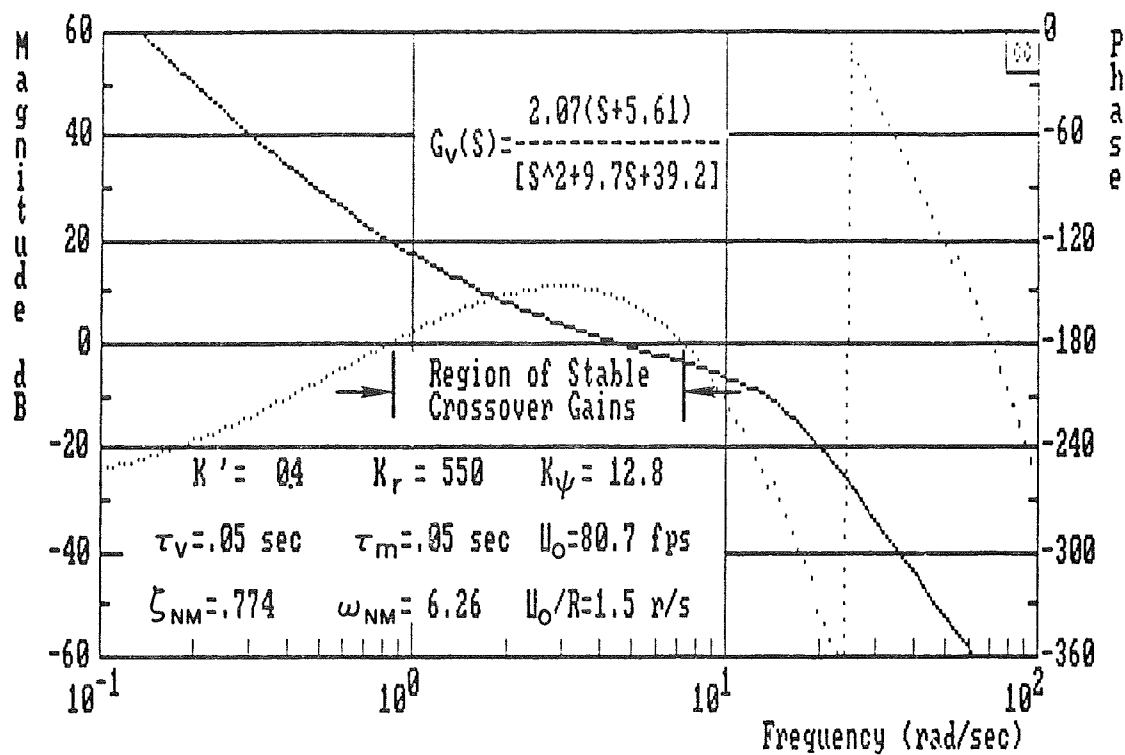
Figure 3. Bode Plot for Normal Stable Driver/Vehicle Control

Stable Gains:

     $K' = 0$ $K_r = 1650$

     $K_\psi = 8.53$ $U_o/R = 0.75$

Region of Stable
Crossover Gains

Unstable Behavior
(Fig. 3 driver gains)

$$G_v(S) = \frac{1.66(S)}{(S-4.11)(S+7.27)}$$
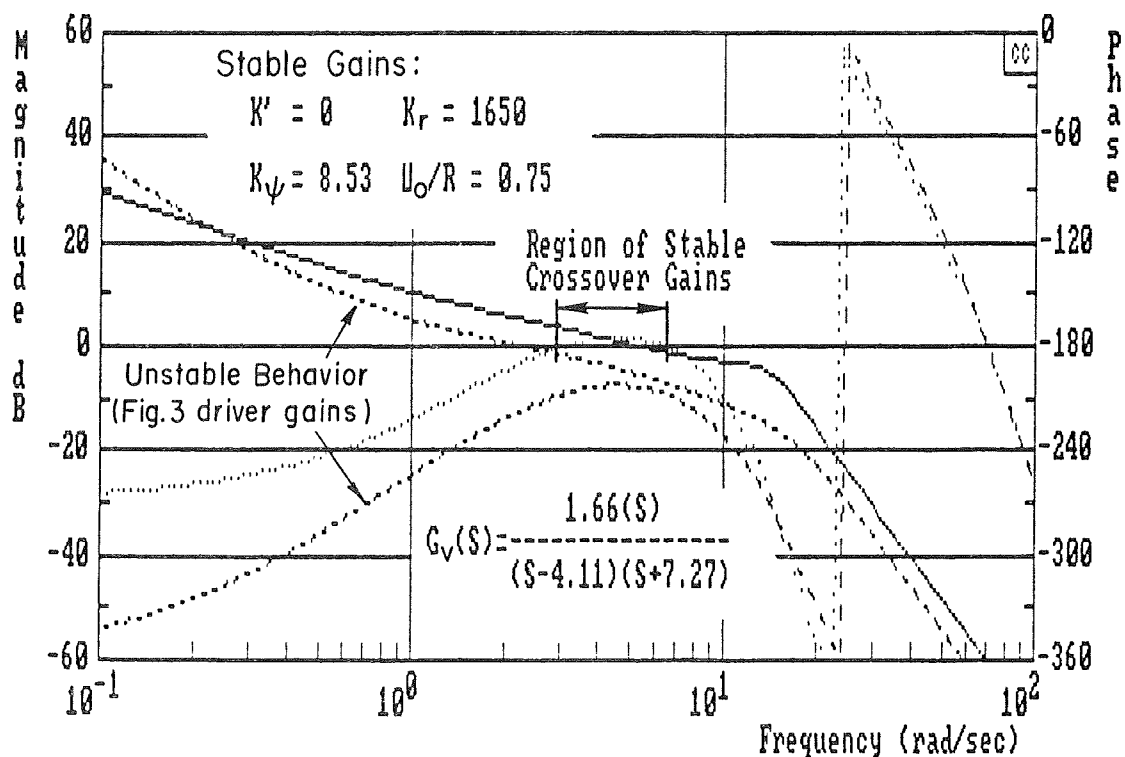
Figure 4. Bode Plot for Driver/Vehicle Control with Unstable
Vehicle Dynamics Due to Rear Wheel Lockup. Figure 3
Driver Gains Give Unstable Closed-Loop Response

32.12

With this adapted driver behavior it can be seen in Fig. 5 that a small phase angle region of K_ψ stability is allowed. At this stage, any further improvement in stability is limited by the driver's time delay and neuromuscular lag. The closed-loop control will not be very good under these circumstances because the closed-loop phase margin will be very low, but since the driver is slowing rapidly (for rear wheels locked, deceleration can be on the order of 0.3-0.4g's) he/she only has to maintain control until the vehicle comes to rest. Also, based on the Table 5 analysis, the vehicle becomes less unstable as speed decreases.

FIELD TEST EXPERIMENT

Methods and Procedures

A field test was conducted to determine driver behavior under actual wheel lockup condition's. The test course layout which defined the task to be performed by the drivers is illustrated in Fig. 5. The basic task was for the driver to stop safely and quickly within the 180 ft stopping zone as defined by the sets of orange cones indicated in Fig. 5. The approach speed to the test course was nominally 40 miles an hour, which would permit the driver to stop in 180 ft at a nominal deceleration of 0.3g. Drivers were told to imagine that the stopping barrier indicated by two orange cones was a car that had pulled out in front of them or possibly pedestrians that had moved into their path and that they were to do their best to stop within the lane before reaching this barrier. Subjects were not told anything about the objectives of the tests other than that we were testing stopping behavior and would be making some variations in the car characteristics.

Lanes edges delineated by white cones spaced 20 ft apart

|◄——————— 180 ft Stopping Zone ———————►|

12 ft Lane

Beginning of stopping zone indicated with two orange cones

Stopping barrier indicated by two orange cones
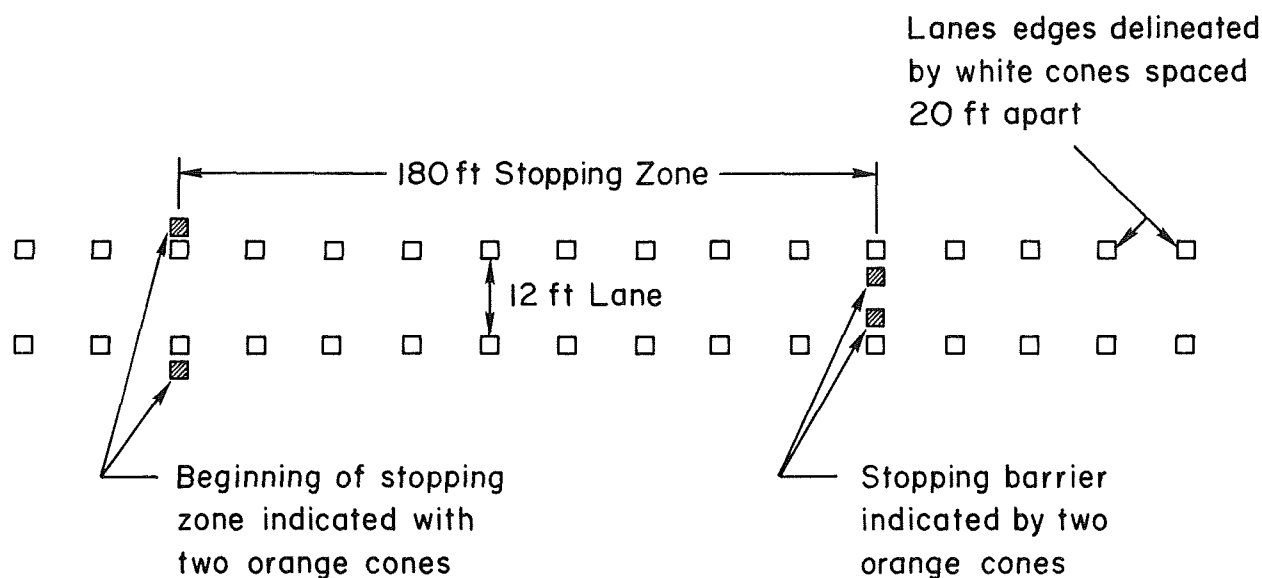
Figure 5. Test Course Layout

The test car was outfitted with a special valve that permitted changing the proportioning of brake pressure going to the rear brakes. Valve settings were setup to achieve three experimental conditions:

A - Significant tendency for front brakes to lockup
B - Moderate tendency for rear brakes to lockup
C - Significant tendency for rear brakes to lockup

In braking, driver's do have the option to modulate their brakes and avoid or at least minimize wheel lockup, and the above experimental condition's allowed for observing this behavior over a range of possible brake balance conditions.

The above three brake bias conditions were tested for each subject in the design indicated in Table 6. The conditions were tested on consecutive runs for each subject. In order to avoid biasing the results, the ordering of the test conditions was changed between subjects as indicated in Table 6.

TABLE 6. EXPERIMENTAL DESIGN

Test Conditions:

    A - 1:1 valve setting (front bias)

    B - 1:2 valve setting (smaller rear bias)

    C - 1:3 valve setting (larger rear bias)

Condition Orders Assigned to Subjects in Sequential Order:

    1) A, B, C   4) B, A, C

    2) C, B, A   5) C, A, B

    3) B, C, A   6) A, C, B

Results

In Fig. 6 distributions of directional control performance metrics are given. For final heading deviations it is noted that the worst performance was encountered under condition C. The best or smallest heading angle deviations were achieved under the front bias condition (A) as might be expected since front wheel lockup does not tend to excite the directional mode of the vehicle or result in unstable dynamics. Final heading angle deviation is an overall directional control metric and it should be noted that only the

a) Dist. of Abs. Final Heading

Brake Balance Conditions:

A - Significant front wheel lock tendency
B - Moderate rear wheel lock tendency
C - Significant rear wheel lock tendency
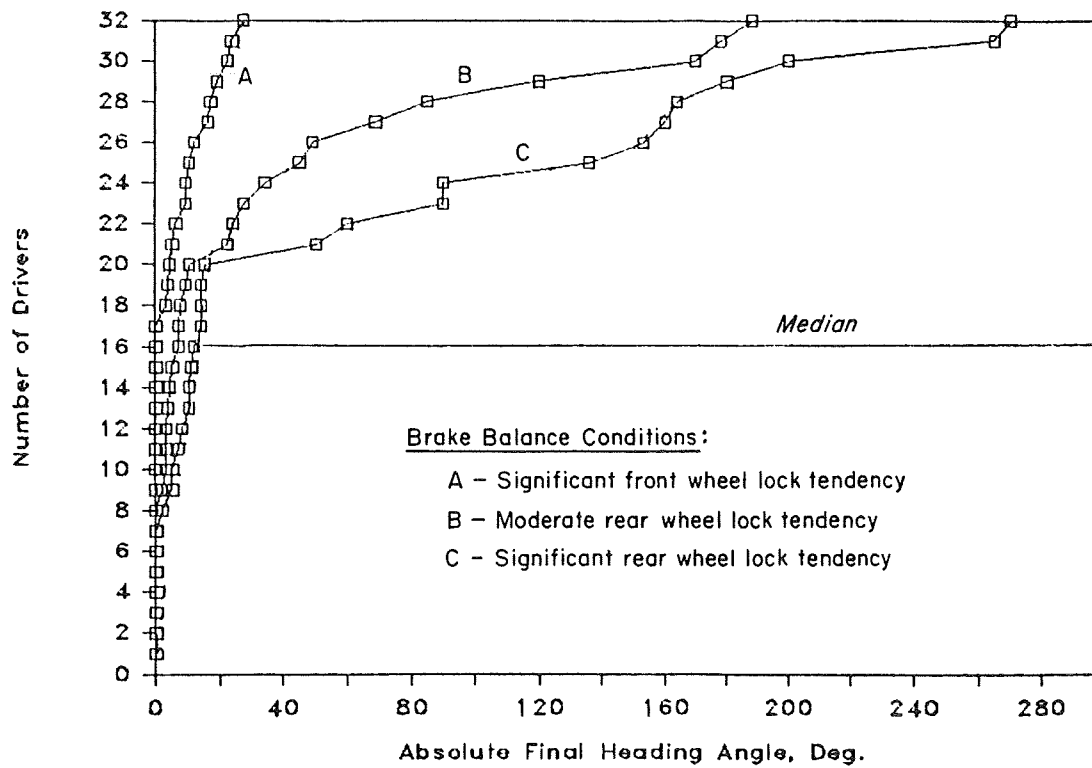
b) Dist. of Abs. Peak Yaw Rate (sc)
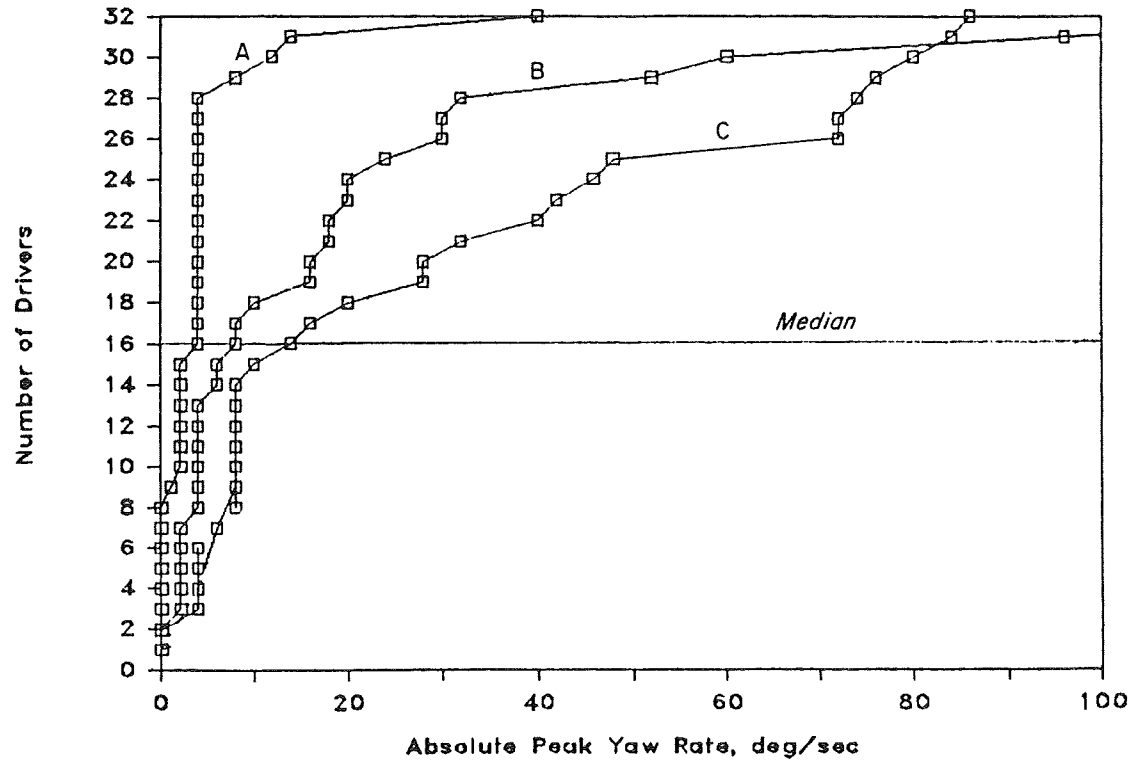
Figure 6. Distributions of Directional Control Performance Measures

32.15

poorest third of the subjects are having a significant control problem. Referring to peak yaw rate distributions in Part b of Fig. 6, note that this intermediate directional control metric gives the same ranking of the brake bias conditions as did heading angle, but tends to be a more sensitive measure in that now fully half of the subject population is having trouble with the rear bias brake conditions.

In general the field test results tend to confirm that rear brake lockup leads to directional control problems, which will cause problems for some portion of the driving public. Although the vehicle dynamics alone represent a dynamic instability which is characterized by an exponentially divergent heading mode, the driver can exert some influence over vehicle heading through steering actions. In many cases even though the rear brakes were locked up and the vehicle itself was unstable drivers were able to exert positive steering control on the vehicle and maintain adequate directional control. There were a few runs, however, where drivers exerted little or no steering action and the vehicle spinout was basically a classical exponential divergence. There were 12 such runs and from yaw rate gryo strip chart records of these few runs we were able to measure a divergent time constant. The distribution of these divergent time constants is illustrated in Fig. 7. Note that one half of these runs or 6 runs in total were near the theoretical vehicle only divergence time constant given in Table 5.
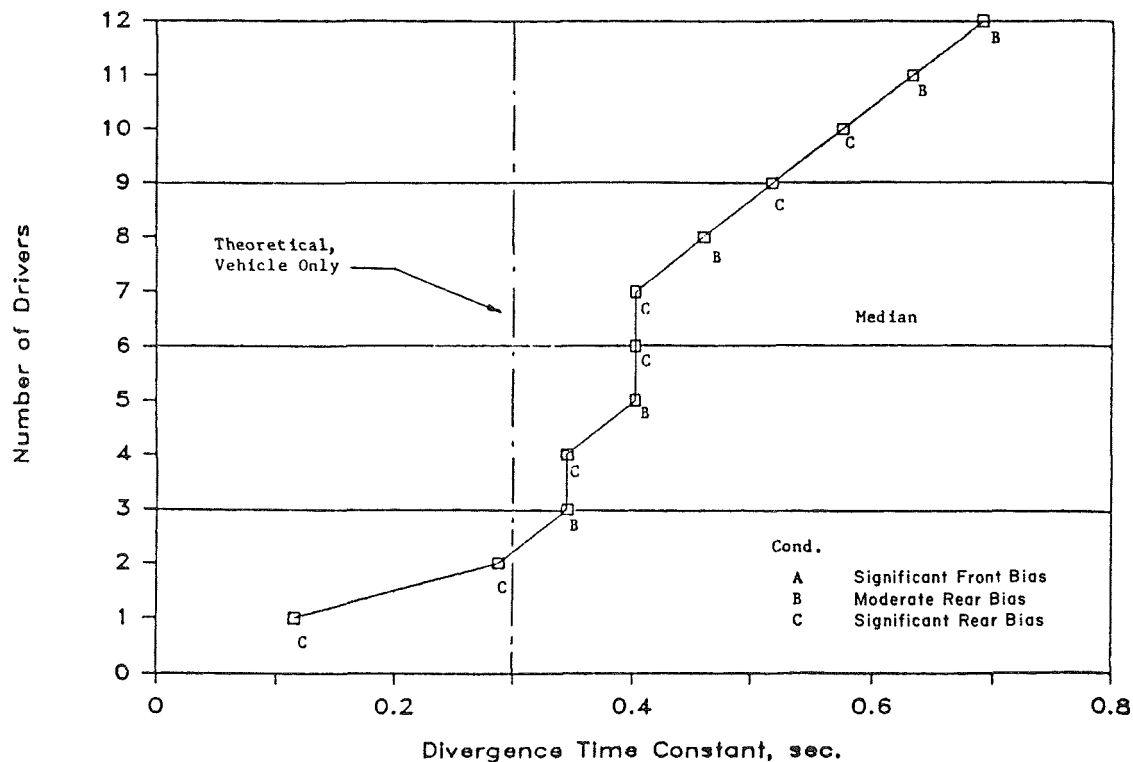
Figure 7. Driver/Vehicle System Divergence Time
Constants Measured from Yaw Rate Recordings for
Runs Exhibiting Little or No Driver Control

32.16

CONCLUDING REMARKS

Human operator control of unstable vehicle dynamics is a fairly well understood problem based on over two decades of research. Limiting human operator capability is constrained to a large extent by internal perceptual and processing time delays. Training and other system characteristics have some influence on limit performance. Analysis of driver/vehicle behavior under rear wheel lockup conditions shows a classical unstable vehicle control problem which leads to loss of control for some portion of the driver population. Experimental results are consistent with a driver/vehicle system stability analysis and past research on limit control capabilities and unexpected transition of vehicle dynamics.

ACKNOWLEDGEMENTS

REFERENCES

1. Weir, David H., and Walter A. Johnson, "Pilot's Response to Stability Augmentation System Failures and Implications for Design," 4th Annual NASA-University Conference on Manual Control, NASA SP-192, 1969, pp. 341-360.

2. "Session XII. Theory and Application of the 'Critical Tracking Task'," Proceedings of the 8th Annual Conference on Manual Control, AFFDL-TR-72-92, 1972, pp. 529-612.

3. Pew, R. W., "Performance Assessment via Critical Task and Dowel Balancing," Proceedings of the 9th Annual Conference on Manual Control, NASA CR-142295, 1973, p. 215.

4. Cheatham, Donald C., A Study of the Characteristics of Human-Pilot Control Response to Simulated Aircraft Lateral Motions, NACA Rept. 1197, 1954.

5. Jex, H. R., C. H. Cromwell, and R. K. Siskind, Correlation of Experimental and Theoretical Limits for Pilot Control of Unstable Second Order Systems, Systems Technology, Inc., TM-56, July 1960.

6. Sadoff, Melvin, Norman M. McFadden, and Donovan R. Heinle, A Study of Longitudinal Control Problems at Low and Negative Damping and Stability with Emphasis on Effects of Motion Cues, NASA TN D-348, Jan. 1961.

7. Taylor, Lawrence W., Jr., and Richard E. Day, Flight Controllability Limits and Related Human Transfer Functions as Determined from Simulator and Flight Tests, NASA TN D-746, May 1961.

8. Jex, H. R., and C. H. Cromwell, III, Theoretical and Experimental Investigation of Some New Longitudinal Handling Qualities Parameters, ASD-TR-61-26, June 1962.

9. Young, Laurence R., and Jacob L. Meiry, "Manual Control of an Unstable System with Visual and Motion Cues," 1965 IEEE International Convention Record, Part 6, Vol. 13, pp. 123-127.

10. Washizu, Kyuichiro, and Katsuyuki Miyajima, "Controllability Limit of a Human Pilot," AIAA Journal, Vol. 3, No. 5, May 1965, pp. 941-947.

11. Jex, H. R., J. D. McDonnell, and A. V. Phatak, "A 'Critical' Tracking Task for Man/Machine Research Related to the Operator's Effective Delay Time," Systems Technology, Inc., P-44, presented at the 7th IEEE Symp. on HFE, May 5-6, 1966.

12. Allen, R. Wade, Anthony C. Stein, Leland G. Summers, and Marcia L. Cook, Drunk Driving Warning System (DDWS). Volume I: System Concept and Description, DOT HS-806 459, Nov. 1983.

    Allen, R. Wade, Anthony C. Stein, Marcia L. Cook, and Leland G. Summers, Drunk Driving Warning System (DDWS). Volume II: Field Test Evaluation, DOT HS-806 460, Dec. 1983.

13. Bioastronautics Data Book, James F. Parker, Jr., and Vita R. West (eds.), NASA SP-3006, 1973.

14. McRuer, D. T., "Simplified Automobile Steering Dynamics for Driver Control," Systems Technology, Inc., P-165, presented at the SAE Aerospace Control & Guidance Systems Committee Mtg. No. 35, Palo Alto, CA, 19-21 Mar. 1975.

15. Weir, D. H., C. P. Shortwell, and W. A. Johnson, Dynamics of the Automobile Related to Driver Control, Systems Technology, Inc., TR-157-1, July 1966.

16. Allen, R. Wade, and Richard J. DiMarco, "Effects of Transport Delays on Manual Control System Performance," 20th Annual Conference on Manual Control, NASA Ames Research Center, 12-14 June 1984, NASA CP-2341-Vol. 1, Sept. 1984, pp. 185-201.

17. Allen, R. Wade, "Stability and Performance Analysis of Automobile Driver Steering Control," SAE Paper No. 820303, presented at the 1982 SAE International Congress and Exposition, Cobo Hall, Detroit, MI, 22-26 Feb. 1982.

18. McRuer, Duane T., David H. Weir, Henry R. Jex, et al., "Measurement of Driver/Vehicle Multiloop Response Properties with a Single Disturbance Input," IEEE Trans., Vol. SMC-5, No. 5, Sept. 1975, pp. 490-497.

Modified Superposition: A Simple
Time Series Approach to Closed-Loop Manual
Controller Identification

Daniel J. Biezad, Lt. Col, USAF
Professor David K. Schmidt
Mr. Frank Leban
Susan Mashiko, Capt. USAF

Purdue University
School of Aeronautics and Astronautics
West Lafayette, Indiana 47907

Air Force Institute of Technology
Department of Electrical and Computer Engineering
Wright-Patterson AFB, Ohio 45433

ABSTRACT

Single-channel "pilot" manual control output in closed-tracking tasks is modeled in terms of linear discrete transfer functions which are parsimonious and guaranteed stable. The transfer functions are found by applying a modified super-position time series generation technique. A Levinson-Durbin algorithm is used to determine the filter which prewhitens the input and a projective (least squares) fit of pulse response estimates is used to guarantee indentified model stability. Results from two case studies are compared to previous findings, where the source of data are relatively short data records, approximately 25 seconds long. Time delay effects and pilot seasonalities are discussed and analyzed. It is concluded that single-channel time series controller modeling is feasible on short records, and that it is important for the analyst to determine a criterion for "best time domain fit" which allows association of model parameter values, such as pure time delay, with actual physical and physiological constraints. The "purpose" of the modeling is thus paramount.

SHORT TITLE: AUTOREGRESSIVE PILOT MODELS

KEY WORDS:

       pilot modeling
       autoregressive process
       closed loop system identification
       prewhitening
       superposition
       single input, single output
       manual control

NOMENCLATURE

$a(z)$	numerator discrete polynomial in z
a_k	coefficient of z^{-k} in $a(z)$
$b(z)$	denominator discrete polynomial in z
b_k	coefficient of z^{-k} in $b(z)$
$g_p(z)$	discrete pilot model pulse response
$e(t)$	error displayed to pilot at instant t
g_k	coefficient of z^{-k} in $g(z)$
$G_p(z)$	pilot transfer function as a ratio of polynomials
i.i.d.	independent, identically distributed
k	lag implying "$k\Delta$" seconds
K_p	pilot gain expressed in degrees per degree
N	total points available
$R(t)$	pilot input uncorrelated with $y(t)$ in degrees at instant t
$w(t), v(t)$	white noise sequence (i.i.d.) at instant t
$y(t)$	controlled element output signal in degrees pitch angle at instant t
Δ	sample interval (seconds)
$\delta(t)$	pilot output in degrees of elevator deflection at instant t
τ	number of sample times in pure time delay
ν	transformed frequency
ω	frequency
$\Omega(z)$	prewhitening filter in z

1. INTRODUCTION

The key question of how the human being will be inserted in the control loop of complex processes remains an issue throughout our society (Rosenbrock, 1983), but nowhere is it more urgent than in flight control systems design and analysis (Harper, 1983). The fact that a pilot of a modern aircraft is becoming a sophisticated systems monitor (Rouse, 1983) in no way implies his demise as a controller (Rouse, 1980; Sheridan, 1974), and a fundamental assumption in this work is that the interaction between man and machine should be understood much better than it is today (Palmer, 1983).

Although describing function (McRuer, 1965) and optimal control (Kleinman, 1969-1974) pilot models have been ingeniously used to provide insight into piloting strategy (Schmidt, 1979; Bacon, 1983; Hess, 1977), they are now supplemented with pilot models derived from the emerging field of time series analysis. Time series modeling of pilot behavior offers tremendous potential for discerning key system characteristics and relationships, such as the actual effect of instabilities (Goto, 1974), pilot stress (Shinners, 1974), or task effects (Agarwal, 1980).

The key questions in time series models involve not only the parsimony of parameters, well established by Breddermann et al (1978), but of identified model stability and the model's practical application in analysis (Baron, 1980). Shinner (1974) seriously discussed the closed-loop identification problem, but the manipulation of transfer functions in his fitting procedure contains no guarantee of final model stability. The primary purpose of this work is to present a theoretically sound and relatively simple closed-loop fitting procedure, still based firmly in the common sense methods of Box and Jenkins (1976), which guarantees model stability without sacrificing model accuracy.

2. MODEL

The linear discrete closed-loop model structure is shown in Figure 1. Each block represents a discrete pulse response sequence which, when convolved with the discrete input sequence, yields the discrete output sequence. Stable pulse sequences, even though infinite in durations, eventually must decay for a stable system. When the pulse sequence is expressed as a ratio of polynomials, stability is guaranteed if the denominator roots are less in magnitude than one. The goal

is to identify the pulse response sequence $g_p(z)$ and approximate its discrete (z domain) transfer function from actual data sets $\{\delta(t)\}$, $\{y(t)\}$, and $\{W(t)\}$ which are equispaced in time with their means removed.

The assumptions are model linearity, time invariance, causality, uncorrelated inputs W(t) and R(t), and prewhitenable input W(t); that is, W(t) is a linear function of previous values plus a white noise "shock." Previous values are mathematically linked by the backward shift operator $z^{(-1)}$.

3. MODIFIED SUPERPOSITION TECHNIQUE

First, every signal in Figure 1 is decomposed conceptually into a part linearly correlated with command disturbance W(t), the remainder uncorrelated with W(t). For example, output y(t) is the sum of $Y_L(t)$, which is correlated with W(t), and of $Y_R(t)$, considered the effect of an additional unknown input R(t), termed "Remnant," uncorrelated with W(t). The pulse response to be found relates, for constant sampling interval "Δ" seconds, the linearly correlated pilot output $\delta_L(t)$ to the correlated error signal $e_L(t)$; that is,

$$e_L(z)g_p(z) = \delta_L(z) \tag{1}$$

This pulse response may be expressed as an infinite sequence or as a ratio of polynomials:

$$G_p(z) \stackrel{\Delta}{=} Kz^{-\tau}a(z)/b(z) \stackrel{\Delta}{=} z^{-\tau}\left(\sum_{k=0}^{\infty} g_k z^{-k}\right) \tag{2}$$

where

$$a(z) = \left(1 + \sum_{k=1}^{k=\ell} a_k z^{-k}\right) \tag{3}$$

$$b(z) = \left(1 + \sum_{k=1}^{k=s} b_k z^{-k}\right) \tag{4}$$

and $s \geq \ell$ imposed constraint

If the integer "k" in Equation (2) is allowed all values ($-\infty < k < +\infty$), then Equation (2) defines the discrete transfer function relating the z-transform of input sequence $e_L(t)$ to the z-transform of output sequence $\delta_L(t)$ (Franklin and Powell, 1980, p.15).

Although the signals $\delta_L(t)$ and $e_L(t)$ are not directly available, they must be "generated" if loop closure effects are properly taken into account. To do this apply superposition to signals $Y(t)$ and $W(t)$ of Figure 1:

$$Y(t) = G_1(z)W(t) + G_2(z)R(t) \tag{5}$$

$$Y_L(t) \overset{\Delta}{=} G_1(z)W(t) \tag{6}$$

where

$$G_2(z) = -G_a(z)/[1-G_a(z)G_p(z)] \tag{7}$$

$$G_1(z) = G_p(z)G_2(z) \tag{8}$$

Since $W(t)$ is prewhitenable (defined above) and uncorrelated with $R(t)$ the cross correlation identification technique of Box and Jenkins (details in Appendix) may be applied to find an estimate of the initial portion of the pulse response sequence $g_1(z)$, between $y(t)$ and $W(t)$. Then $a_1(z)$ and $b_1(z)$ may be determined as shown in the section on model stability, such that

$$G_1(z) = a_1(z)/b_1(z) \tag{9}$$

The essence of modified superposition is now to <u>generate</u> the time series $Y_L(t)$ using the autoregressive relation

$$b_1(z)Y_L(t) = a_1(z)W(t) \tag{10}$$

Where $a_1(z)$ and $b_1(z)$ are numerator and denominator polynomials,
respectively, with the structure of Equations (3) and (4). The linearly correlated signal $e_L(t)$ is then generated from

$$e_L(t) = W(t) - Y_L(t) \tag{11}$$

The above process is then repeated by reapplying superposition to obtain the following relation between $\delta(t)$ and $W(t)$:

$$\delta(t) = G_3(z)W(t) + G_4(z)R(t) \tag{12}$$

$$\delta_L(t) \overset{\Delta}{=} G_3(z)W(t) \tag{13}$$

The cross correlation identification (Appendix) applied to the sequence $\delta(t)$ and $W(t)$ yields the initial segment of pulse response sequence $g_3(z)$, and the polynomials $a_3(z)$ and $b_3(z)$ may be determined (see next section) such that

$$G_3(z) = a_3(z)/b_3(z) \qquad (14)$$

Pilot output linearly correlated with $W(t)$ is generated from the autoregressive relation

$$b_3(z)\ \delta_L(t) = a_3(z)W(t) \qquad (15)$$

Finally, the cross correlation technique (Appendix) is applied to $\delta_L(t)$ and $e_L(t)$ to find the initial segment of $g_p(k)$, defined by the coefficient set $\{g_{PK},\ 0 \le k < N\}$, of the pilot model pulse response. Numerator and denominator polynomials are then found (see next section) which yields

$$G_p(z) = \delta_L(z)/e_L(z) \qquad (16)$$

No multiplication or divisions of transfer functions occurs throughtout the above procedure.

4. MODEL STABILITY

As mentioned above, the pulse response sequence identified $[g_1(z), g_3(z)$ and $g_p(z)]$ will be truncated at some finite lag "k" final task is to find a parsimonious numerator polynomial stable denominator polynomial which together are equivalent mathematically to the identified pulse response. These polynomials are chosen to have the structure shown in Equation (2), which is re-arranged into the following form:

$$\left(1 + \sum_{i=1}^{s} b_i z^{-i}\right)\left(\sum_{k=0}^{kmax} g_k z^{-k}\right) = K\left(1 + \sum_{k=1}^{\ell} a_k z^{-k}\right) \qquad (17)$$

$$kmax > s \ge \ell$$

Since the pulse response g_k is known for $0 < k < N$, by equating coefficients for the operator "z" at each exponential power "ℓ", o relationships may be found between numerator and denomiator coefficients a_k and b_k. Moreover, by equating coefficients for the operator z above power "s", for which the right side of equation (15) vanishes, one obtains for every $j > 0$

$$g_{s+j} + g_{s+j-1}b_1 + \cdots + g_{s+j-s}b_s = 0 \qquad (18)$$

The above relation exists for a finite but large number of "$J > 0$", so projection theory (least squares) may be used to coefficients b_k ($0 < k \leq s$). Bringing term "$g_{(s+j)}$" to the other side of equation (18) and divided by "$g_{(s+j)}$" one may write

$$A[b_1, \ b_2, \ \ldots, \ b_s]^T = [-1, \ \ldots, \ -1]^T \qquad (19)$$

and the "j"th row of A is given by

$$\left| \frac{g_{s+j-1}}{g_{s+j}}, \ \frac{g_{s+j-2}}{g_{s+j}}, \ \ldots, \ \frac{g_{s+j-s}}{g_{s+j}} \right| \qquad j > 0 \qquad (20)$$
$$1 \text{ by } s$$

The solution from linear algebra is

$$[b_1, \ b_2, \ \ldots, \ b_s]^T = -(A^T A)^{-1} A^T [1, \ \ldots, \ 1]^T \qquad (21)$$

To provide a parsimonious denominator, the solution of Equation (21) is accepted for the lowest order "s" which has both a stable characteristic equation (i.e. roots less than 1.0 in magnitude) and which yields a model pulse response similar in shape to the truncated pulse response identified from the data. Once a stable denominator is found the numerator a (z) and the gain K may be determined by once again matching coefficients in Equation (17);

$$K = g_0 \qquad (22)$$

$$a_k = -\frac{1}{K}[g_k + \sum_{i=1}^{i=k} b_i g_{k-i}] \qquad 0 < k \leq \ell \qquad (23)$$

33.7

By defining error residual to be the actual output time series minus the pilot model output series at each sample instant, the gain K may be adjusted by a suitable minimization technique to minimize the error residual variance. Alternatively, it may be adjusted to provide a steady state response of unity when the input to the transfer function is a unity pulse train, a constraint recommended by Agarwal (1980).

If a time delay "τ" is to be included, the final form of Gp(z) will be as shown in Equation (2), and the indices for the pulse responses in Equations (17)-(23) should be incremented by the integer "τ" during identification (for example the gain K from Equation (2) equals g_τ identified from the data).

Validation tests may also be applied to the model. There are two types of tests: acceptability and statistical significance. Acceptability tests are common sense checks which compare model output series verses actual autocorrelation estimates from the data, autocorrelation of residuals for whiteness properties, and checks for negligible cross-correlation between the noise inputs.

Statistical significance tests may be performed after acceptability tests indicate the model is reasonable. Chi-squared statistics are available from the w(t) and v(t) prewhitened series (discussed in the Appendix and shown in Figure 13). Assuming one can safely neglect correlations beyond a lag of 20, for example, the statistics to be computed are, for "whiteness" of v(t)

$$(N-p) \sum_{k=1}^{20} \left\{ \frac{1}{(N-k)} \sum_{t=k}^{N} v(t-k)v(t) \right. \tag{24}$$

and, for uncorrelated w(t) and v(t)

$$(N-p) \sum_{k=1}^{20} \left\{ \frac{1}{(N-k)} \sum_{t=k}^{N} w(t-k)v(t) \right\} \tag{25}$$

p = order of $\Omega_v(z)$ filter

N = total points in data set

which should pass the chi-squared significance test for degrees of freedom (20-p) and (20-1-s-1) respectively (Box and Jenkins, 1976, p.394). Failure of either significance test is evidence of a faulty assumption or a modeling inadequacy.

To summarize the modified superposition technique

a) Find a finite pulse sequence relating y(t) and W(t) using cross correlation identification (Appendix).

b) Determine a parsimonious, stable transfer function $G_1(I)$ which is mathematically equivalent, in the least squares sense, to the sequence identified from the data $g_1(z)$ [Equation (9)].

c) Generate time series realizations $\{y_L(t)\}$, $\{e_L(t)\}$ using Equations (10) and (11).

d) Find a finite pulse sequence, $g_3(z)$, relating $\delta(t)$ and W(t) using cross correlation identification, and determine a stable transfer function $G_3(\tau)$ for this pulse response (Equation (14)).

e) Generate time realization $\delta_L(t)$ using Equation (15).

f) Find a finite sequence of the pulse response $g_p(z)$, from $\delta_L(t)$ and $\delta_L(t)$ using cross correlation identification, and fit a stable pilot model transfer function $G_p(\tau)$ to this pulse response (Equation (16)).

g) Adjust K if desired and validate the model.

5. PILOTED LABORATORY SIMULATION

Single-channel "piloted" simulations in the Flight Simulation Laboratory at Purdue University were accomplished with a pilot performing pursuit tracking tasks using a single and double integrator (K/s and K/s^2 respectively) controlled element dynamics. The task involved a command disturbance input of a random appearing forcing function, and a standard pursuit (McRuer, 1974) display using a CRT Monitor. Data sets were obtained at a 20 hertz sample rate and 500 points were used for modeling, providing a record length of only 25 seconds (although the data run itself exceeded 60 seconds).

For the single-integrator controlled element many low-order transfer functions provided excellent "fits," and the lowest order model is shown in Table 1. A "direct iden-tification" neglecting the closed-loop structure was also per-formed by merely fitting signals $\{\delta(t)\}$ and $\{e(t)\}$, and a comparison of those results in Table 1 shows little variation in parameter values between direct and indirect identification in this case. This implies a small value for pilot injected noise relative to stick output (See Figure 1), a reasonable

deduction for a "simple" controlled element such as K/s. An a priori selected time delay of 0.2 seconds yielded the lowest error residual variance and is consistent with previous results (Bredderman, 1976).

A frequency response of the identified transfer function is shown in Figures 2 and 3 where it is clear that a delay in series with a pure gain effectively describes pilot behavior. This is consistent with classical pilot modeling results (McRuer, 1974). Since a conventional Bode interpretation and analysis using these frequency responses is not valid over all frequencies in discrete systems z-domain analysis, a transformation of variables from z to w' was accomplished using (Franklin and Powell, 1980, p.114)

$$w' \overset{\Delta}{=} \frac{2}{\Delta} \frac{(z-1)}{(z+1)}$$

$$\upsilon \overset{\Delta}{=} \frac{2}{\Delta} \tan \frac{\omega \Delta}{2} \tag{27}$$

Figures 4 and 5 show the transformed frequency (υ) response in the w' domain, where a conventional Bode interpretation is allowed. By comparing Figures 4 and 5 with Figures 2 and 3, one can find no discernable difference between the responses over the frequency range of interest (0   ω   25 rps).

The time histories are shown in Figure 6. Only the first 500 points (25 seconds) were used to develop the model, and the model output remains reasonably accurate beyond this time. This verifies stationarity and avoids an overfit (Kashyap, 1976), which would be evidenced by increased error residual when the model is applied to data independent of model derivation(in this case beyond 25 seconds).

For the double-integrator controlled element a more complex transfer function was identified and is shown in Table 2 for two values of a priori selected time delay (0.05 seconds and 0.2 seconds).

From the frequency response plot in Figure 7 there is some resonance near 2.0 Hz. The phase plots are shown in Figures 8 and 9 for two different values of time delay (0.2 and 0.05 seconds respectively). The transformation to w' domain yields no discernable difference from these responses and they are not shown.

In contrast to control of a "simple" K/s, the "best fit (minimizing residual error variance) was obtained when time delay was set to 0.05 control for of K/s^2. The phase contribution (from the poles and zeros) of the discrete transfer function is apparent as time delay changes between 0.2 seconds and 0.05 seconds, as may be seen by the phase plots of Figures 10 and 11 in which the pure time delay has been removed from the discrete transfer function. Selecting the larger pure time delay for the model exposes the considerable lead generation from the transfer function poles and zeros. This lead generation is not as apparent when pure time delay is reduced for the "best time domain" fit, but the resulting 0.05 seconds might be judged too fast to associate with a lumped physiological delay for a human operator. A possible explanation is unmodeled pilot anticipation; that is, a possible anticipatory loop closure not accounted for in Figure 1.

Further evidence of this is provided in the time history for the best fitting model in Figure 12. Note that a seasonal pilot residual (where pilot output "leads" model output) occurs during some of the longer intervals of large slope. This could be caused by momentary anticipatory behavior arising from the "pursuit" display including commanded input, a factor not accounted for in a time invariant model. Thus in determining the "best" model using time series analysis, the purpose of the model must be given as much consideration as tests for "best fit."

In summary for the K/s^2 controlled element, an a priori time delay in series with a rate sensitive gain describes "pilot" behavior over his usable bandwidth, in agreement with classical results (McRuer, 1974). When pure time delay is not set a priori but allowed to vary in obtaining the "best time domain fit," the minimization of an error variance criterion results in a math model where the time delay is perhaps too small to be associated with physiological operator delays. This case is associated with a pursuit task in which the command as well as the plant output is displayed.

33.11

6. CONCLUSIONS

A modified superposition technique was described for obtaining a parsimonious and stable discrete transfer function, along with statistical tests for model validation. Results provide evidence that the time series technique appears feasible to implement on "short" data records. The analyst needs, however, to determine the criterion for a "best time domain fit" which allows association of parameter values, such as pure time delay, with actual physical and physiological constraints. Seasonalities in pilot residual, possibly caused by anticipatory behavior, were observed as first noted by Shinners (1974), and are not well modeled with a time invariate model.

Future work should concentrate on the full potential of these time series models for analyses, especially their ability to provide stable and accurate power spectral densities, and on their application to multi-channed closed-loop pilot modeling.

7. ACKNOWLEDGMENT

8.   APPENDIX: Cross Correlation Identification
(Box and Jenkins, 1976)

Given the situation in Figure 13, the goal is to find the pulse response relating Y(t) and W(t), which is prewhitenable by $\mathcal{U}_w(z)$. The prewhitening is accomplished by applying the Levinson-Durbin algorithm as given by Kay and Marple (1981, pp. 1388-1389). By reversing the order of the blocks in the forward path of Figure 13, and multiplying each signal at the summer by

$\Omega_w^{-1}(z)$, the following equation results:

$$G(z)w(t) + \Omega_w^{-1}(z)V(t) = \beta(t) \tag{28}$$

$$\beta(t) = \Omega_w^{-1}(z)y(t) \tag{29}$$

Now multiply Equation (28) by w(t-k) and take the expectation, recalling that w(t) is uncorrelated by assumption with v(t):

$$G(z) \ E[w(t)w(t-k)] = E[\beta(t)w(t-k)] \tag{30}$$

By expanding G(z) using shift properties of z one obtains

$$(g_0 + g_1 z^{-1} + g_2 z^{-2} + \ldots) \ E[w(t)w(t-k)] = E[\beta(t)w(t-k)] \tag{31}$$

Since w(t) is an independent, identically distributed sequence of

random numbers with variance σ_w^2, one obtains for every lag k

$$g_k \sigma_w^2 = E[\beta(t)w(t-k)] \tag{32}$$

$$k > 0$$

Conventional estimation relations may now be used to estimate the terms in Equation (32) and solve for gk; for example, from Box and Jenkins (1976, pp. 32-33) one obtains

$$\hat{g}_k \ \left\{ \frac{1}{N} \sum_{t=1}^{N} w(t)w(t) \right\} = \left\{ \frac{1}{N-k} \sum_{t=k}^{N} \beta(t)w(t-k) \right\} \tag{33}$$

which detemines the pulse response sequence estimate g_k.

REFERENCES

1. Bacon, B. and Schmidt, D., "An Optimal Control Approach to Pilot Vehicle Analysis and the Neal Smith Criteria, "AIAA Journal of Guidance, Control, and Dynamics, Vol. 6, No. 5, Sept. -Oct. 1983, pp. 321-330.

2. Baron, Muralidharan, and Kleinman, "Closed Loop Models for Analyzing Engineering Requirements for Simulators," NASA Rept. 2965, Feb. 1980.

3. Box, G. and Jenkins, G., Time Series Analysis: Forecasting and Control, San Francisco, Holden Day, 1976.

4. Breddermann, Glockner, and Henninset, "On the Identifiability of the Human Controller in a Closed Loop System," Identifica-and System Parameter Identification, Rajbman ed., North Holland, 1978.

5. Franklin, G. and Powell, J., 'Digital Control of Dynamic Systems," Addison-Wesley, 1980.

6. Goto and Washizu, "On the Dynamics of Human Pilots in Marginally Controllable Systems," AIAA Journal of Aircraft, Vol. 12, NO. 3, March 1974, pp. 310-315.

7. Harper, R., "Handling Qualities of Flight Vehicles," internal memo, Flight Research Department, CALSPAN Advanced Technology Center, Buffalo, New York.

8. Hess, R., "Prediction of Pilot Opinion Rating Using an Optimal Pilot Model," Human Factors, Vol. 10, No. 5, Oct. 1977, pp.459.

9. Hoh and Mitchell, "Low-Order Approaches to High Order Systems Problems and Promises," AIAA Journal of Guidance and Control, Vol. 5, No. 3, Sept.-Oct. 1982, pp. 482.

10. Kashyap and Roa, Dynamic Stochastic Models from Empirical Data, Academic Press, 1976.

11. Kleinman, D., "A Predictive Pilot Model for STOL Aircraft Landing, "NASA CR-2374, March 1974.

12. Kleinman, Baron, and Levinson, "Optimal Control Model of Human Response," Automatica, Vol. 6, No. 3, 1970.

13. McRuer, Graham, Krendel, and Reisener, "Human Dynamics in Compensatory Systems," AFFDL-TR-65-15, Wright-Patterson AFB, Ohio, July 1965.

14. McRuer and Krendel, "Mathematical Model of Human Pilot Behavior," AGARD AG-188, January 1974.

15. Palmer, "Large Scale Systems: Ststems, Man, and Cybernetics Overview," IEEE Transactions on Automatic Control, Vol. AC-28, No. 6, June 1983.

16. Rosenbrock, H., "The Proper Use of Human Ability: A Challenge to Engineers," IFAC Newsletter, No. 5, Oct. 1983.

17. Rouse, W., Systems Engineering Models of Human-Machine Inter-action, North Holland, 1980.

18. Schmidt, D., "Optimal Flight Control Syntheses Via Pilot Modeling," AIAA Journal of Guidance and Control, Vol. 2, No. 4, July 1979.

19. Sheridan and Ferrell, Man Machine Systems, MIT Press, 1974.

20. Shinners, S., "Modeling of Human Operator Performance Utilizing Time Series Analysis," IEEE Transactions on SMC, Vol. SMC-4, No. 5, pp. 446-458, September 1974.

C - 6

Table 1 Discrete Transfer Function Identification
Results for Controlled Element K/s (K = 1)

$$\text{Model Structure } G_p(z) = \frac{K_p^{-\tau}(1+a_1 z^{-1})}{(1+b_1 z_p^{-1})}$$

Signal to noise ratio ≈ 50

N = 500 points   Δ = 0.05 seconds   τ = 4 (0.2 seconds)

Parameter	Modified Superposition Value	Direct Identification
K_p^*	0.64	0.69
K_p^{**}	0.79	0.72
a_1	0.71	0.69
b_1	0.32	0.37

* Gain which minimizes error residual variance

** Gain yields steady state step response of unity

33.16

Table 2 Modified Superposition Identification Results
for Controlled Element K/s^2 ($K = 1$)

$$\text{Model Structure } G_p(z) = \frac{Kz^{-\tau}(1+a_1z^{-1})}{(1+b_1z^{-1}+b_2z^{-2}+b_3z^{-3})}$$

Signal to noise ratio ≈ 30

$N = 500$ $\Delta = 0.05$ seconds $\tau = 4$ (0.2 seconds)

Parameter	$\tau = 0.05$ sec	$\tau = 0.2$ seconds
K_p*	0.03	0.89
K_p**	0.033	1.22
a_1	10.9	-0.67
b_1	- 1.42	-1.41
b_2	0.91	0.88
b_3	- 0.1	-0.06
Roots	0.14	0.08
	0.64 $\pm$j 0.57	0.67 $\pm$j 0.57

* Gain which minimizes error residual variance

** Gain yields steady state step response of unity

Figure 1 Linear Discrete Closed-loop Pilot Model

33.18
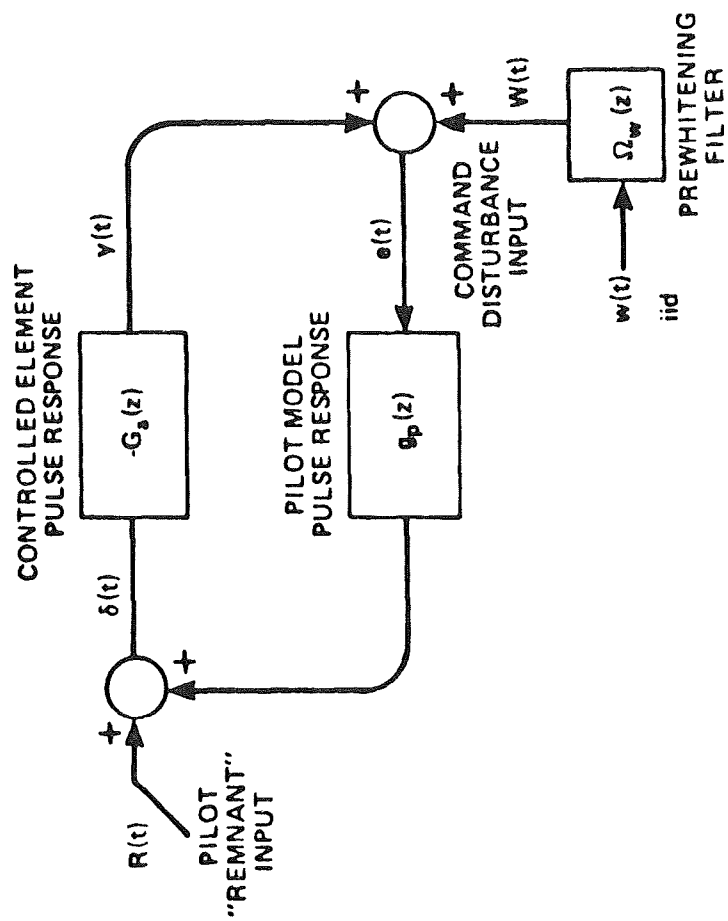
Figure 2   Manual Controller Frequency Response Magnitude:
Controlled Element K/s
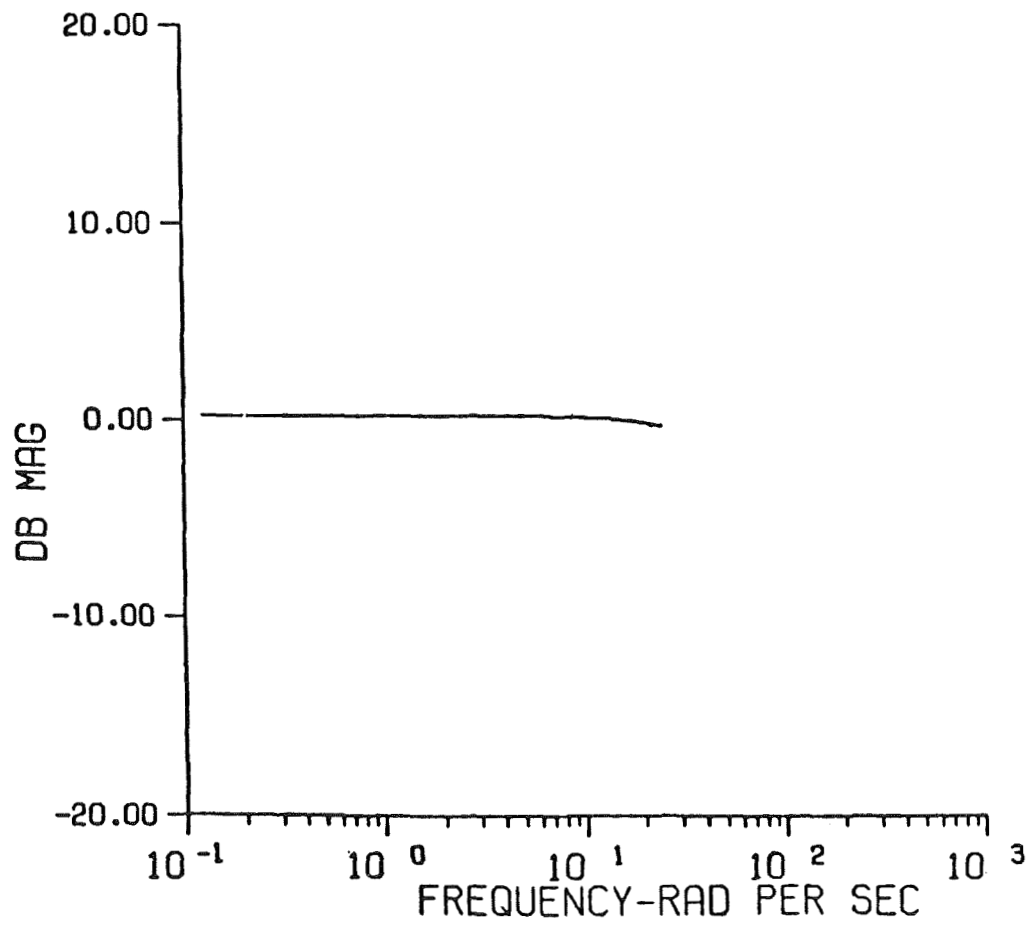
Figure 3   Manual Controller Frequency Response Phase:
Controlled Element K/s

33.20
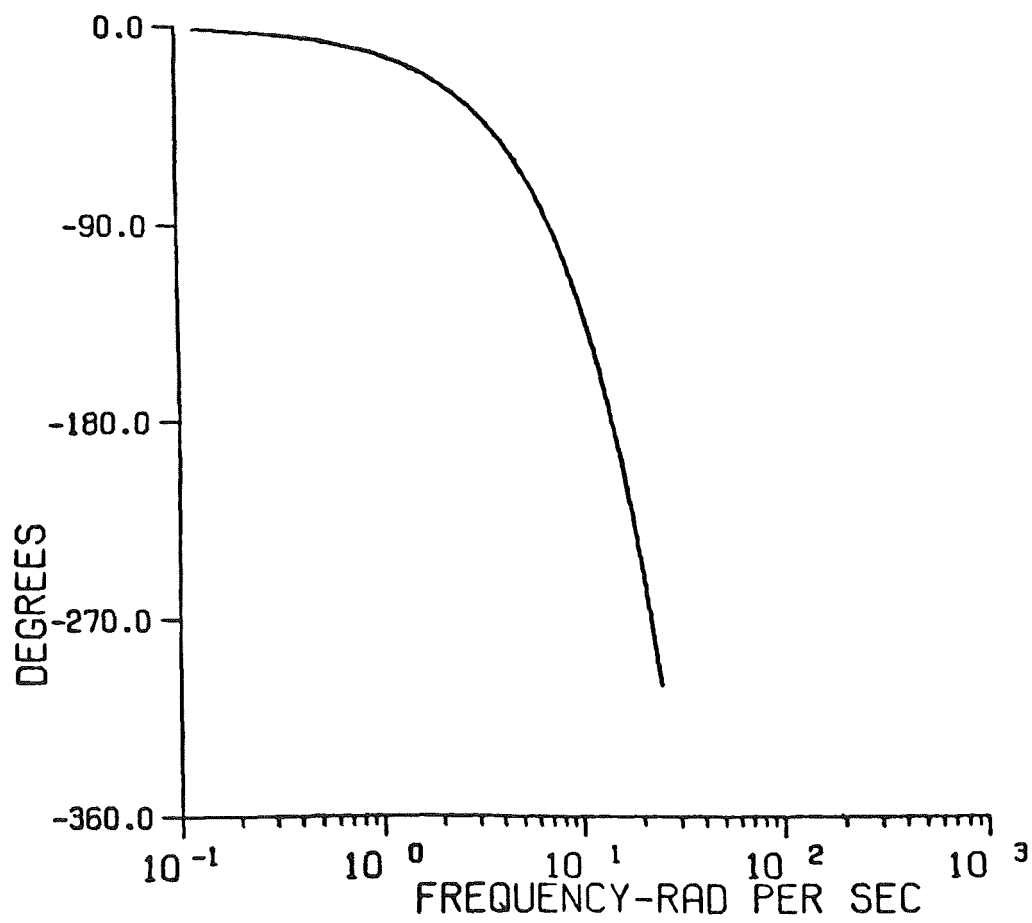
Figure 4   W' Response Magnitude: Controlled Element K/s

33.21
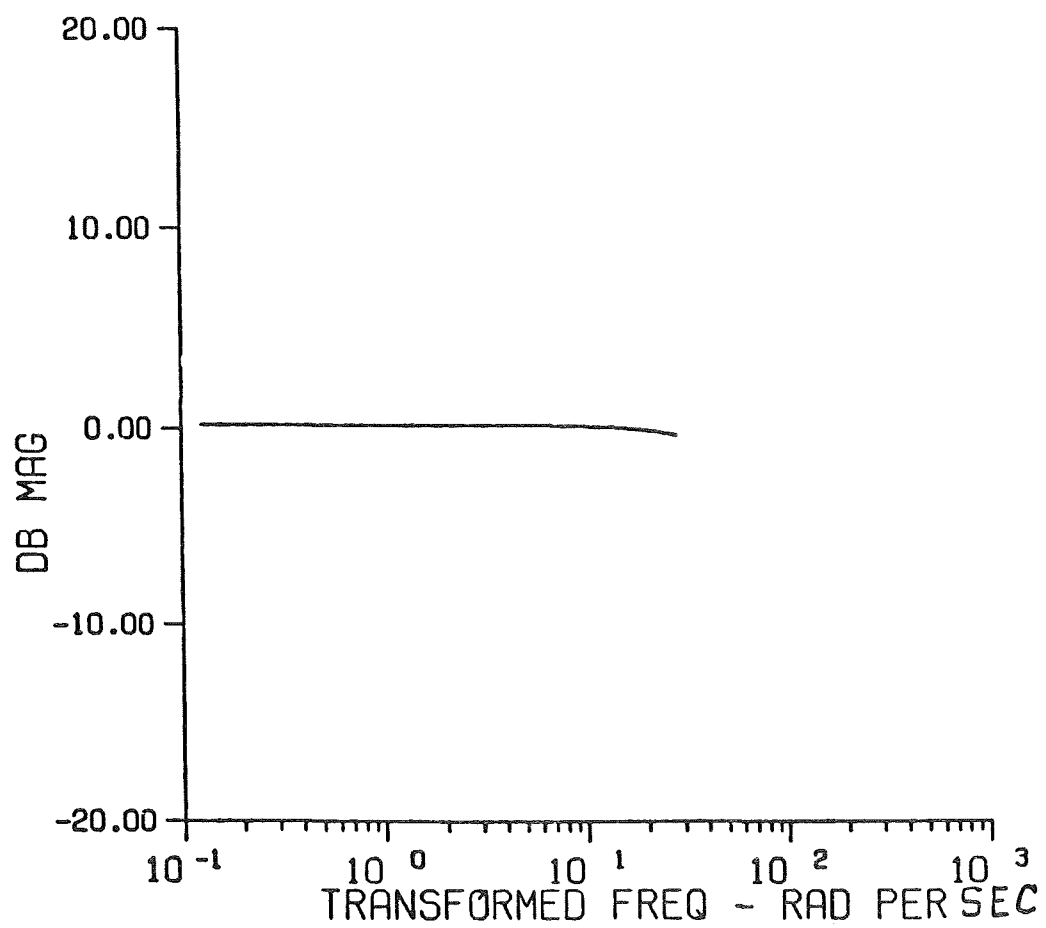
Figure 5   W' Response Phase: Controlled Element K/s
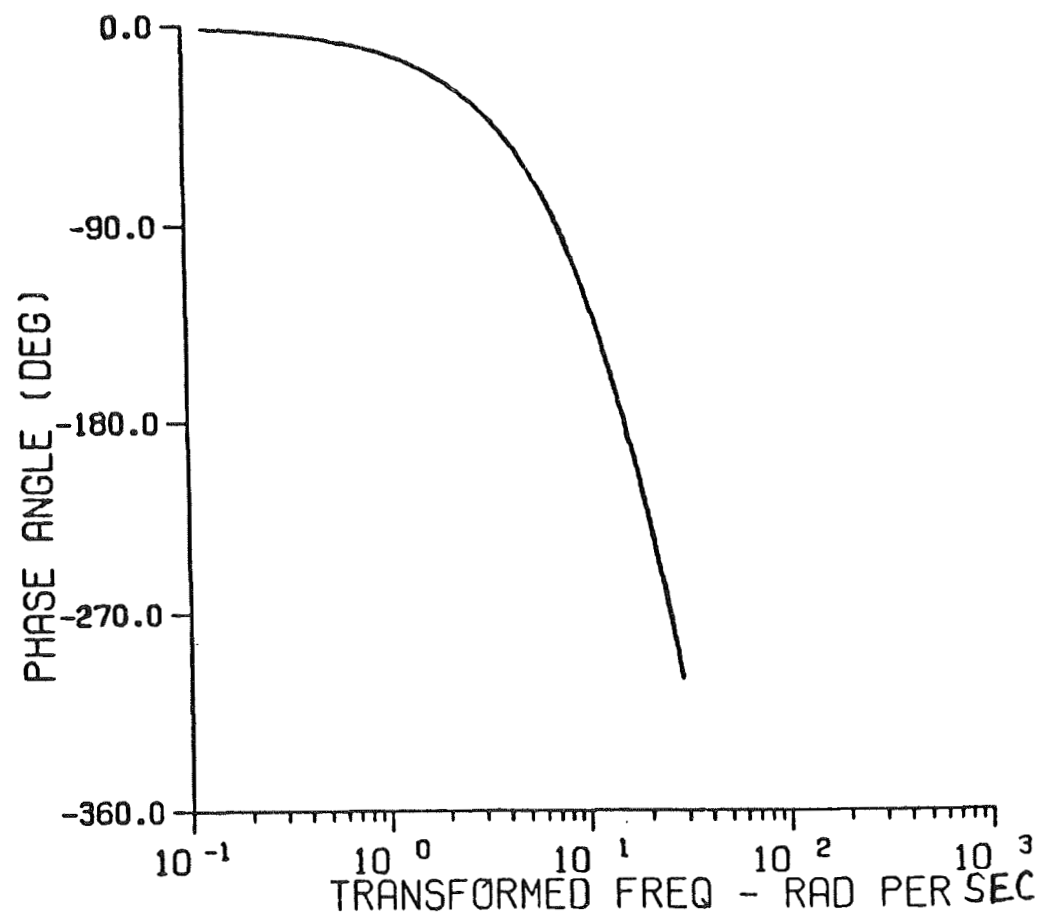
Figure 6 Model Output vs. Pilot Output: Controlled Element K/S

----- predicted
——— actual

Data Not Used
for modeling

Data Source
for Modeling

TIME - SECONDS X 20.0

STICK DEF (DEGREES)

2.500
1.500
.500
-.500
-1.500
-2.500

0.0
150.0
300.0
450.0
600.0
750.0
900.0

33.23
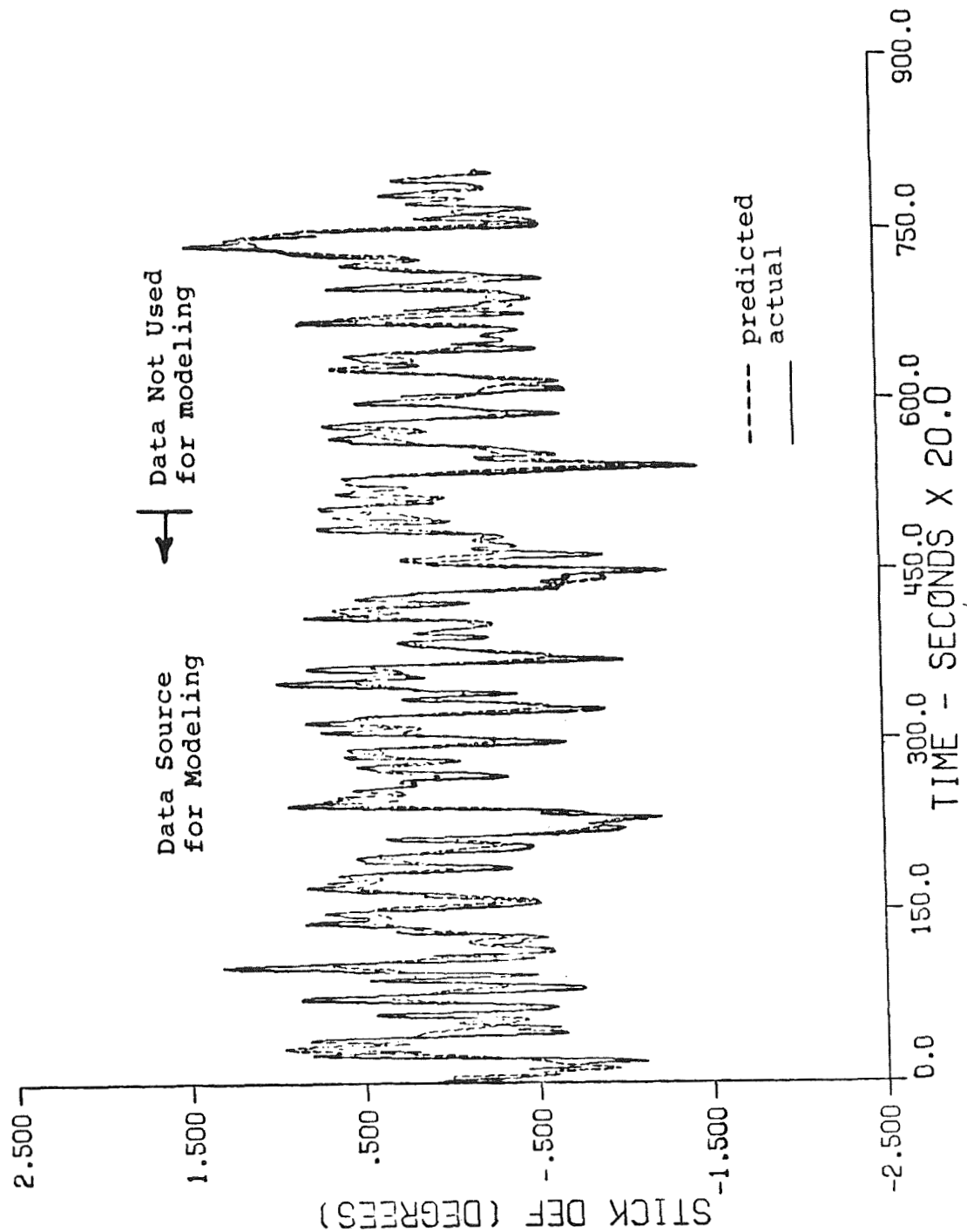
Figure 7   Manual Controller Frequency Response Magnitude:
K/s^2, $\tau = 0.2$ seconds

33.24
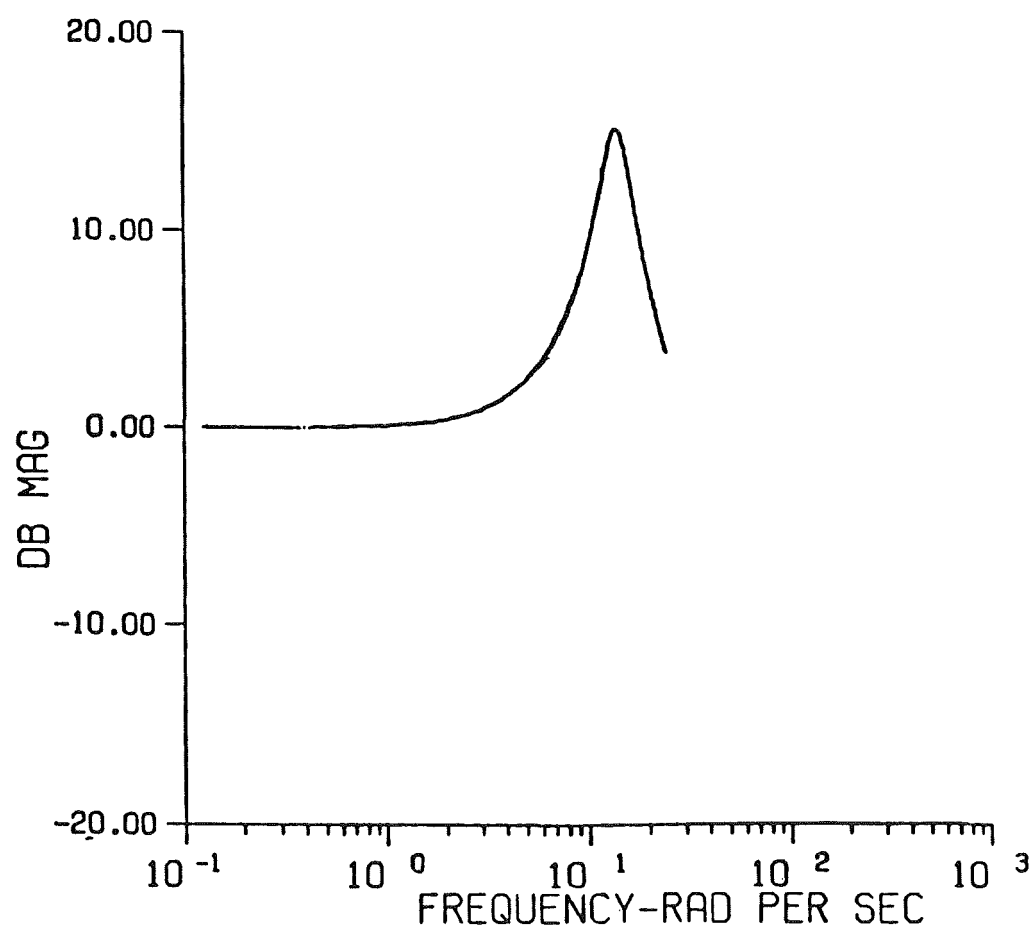
Figure 8   Manual Controller Frequency Response Phase:
K/s^2, $\tau = 0.2$ seconds
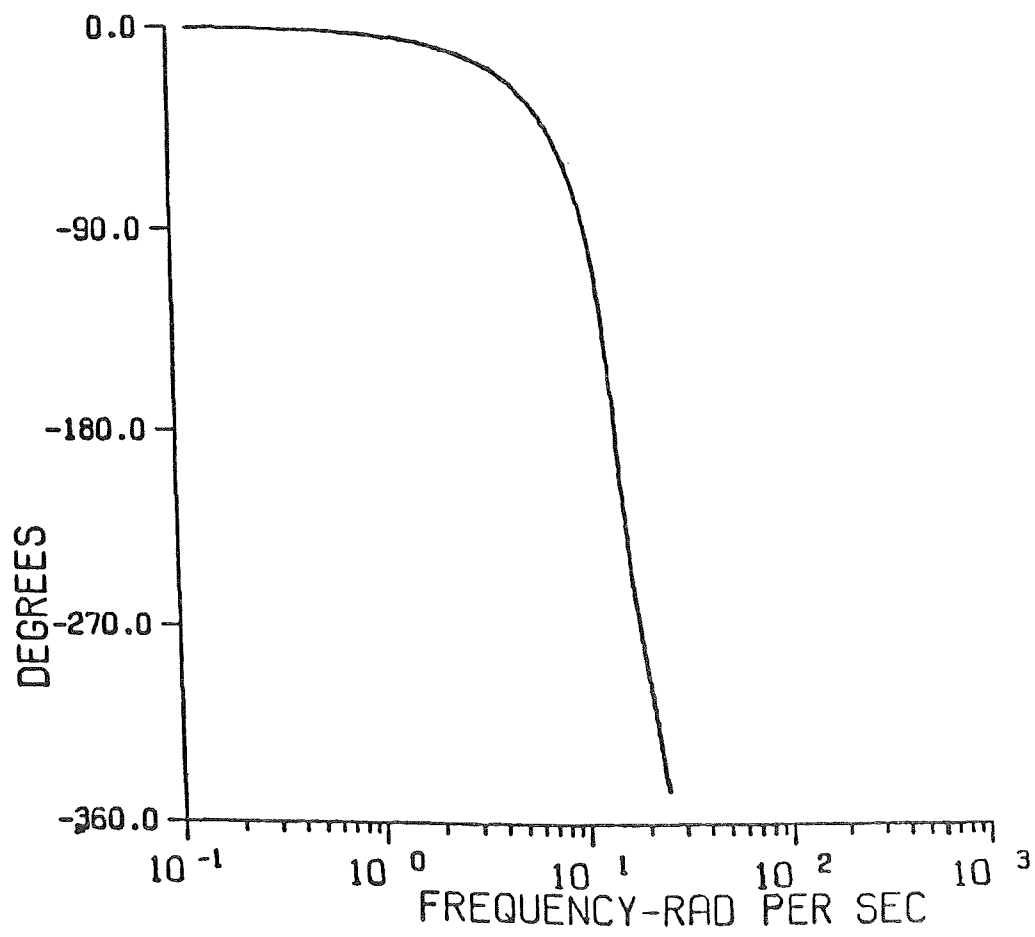
Figure 9 Manual Controller Frequency Response Phase:
K/s^2, $\tau = .05$ seconds

33.26
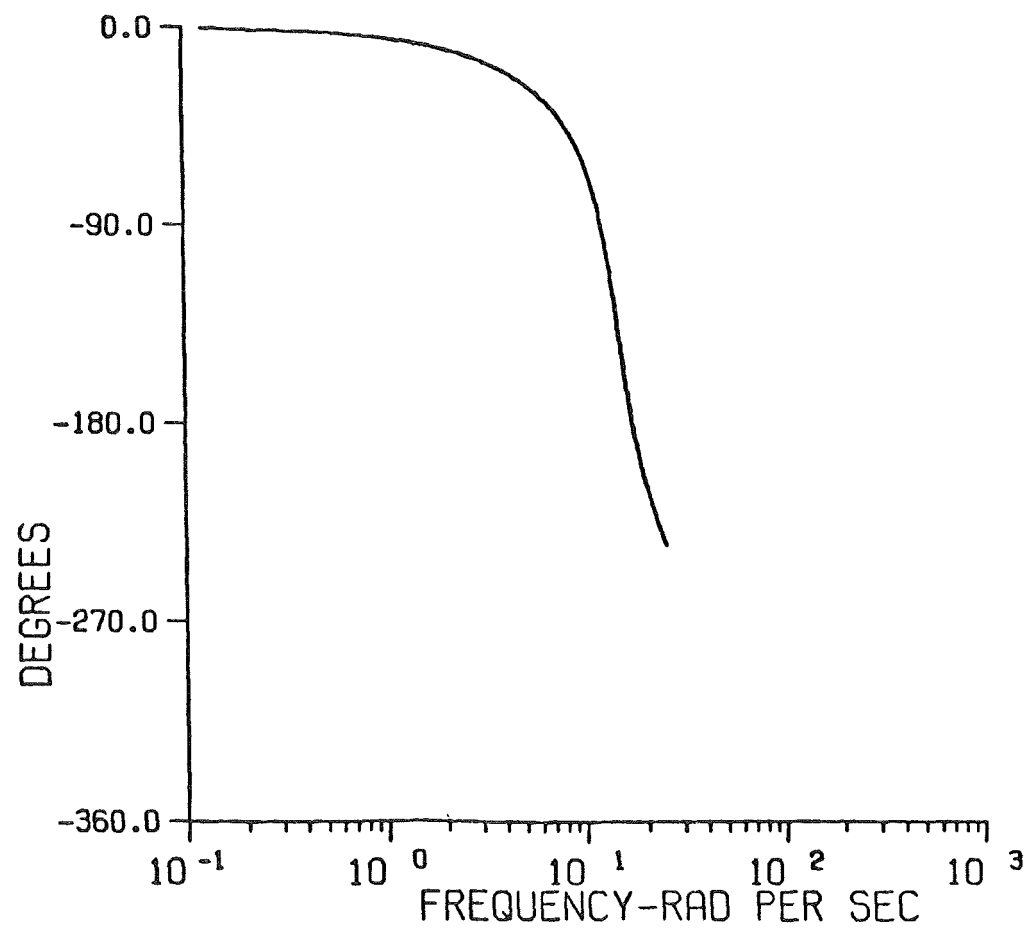
Figure 10 Manual Controller Frequency Pole-zero Response Phase:
K/s^2, $\tau = 0.2$ seconds
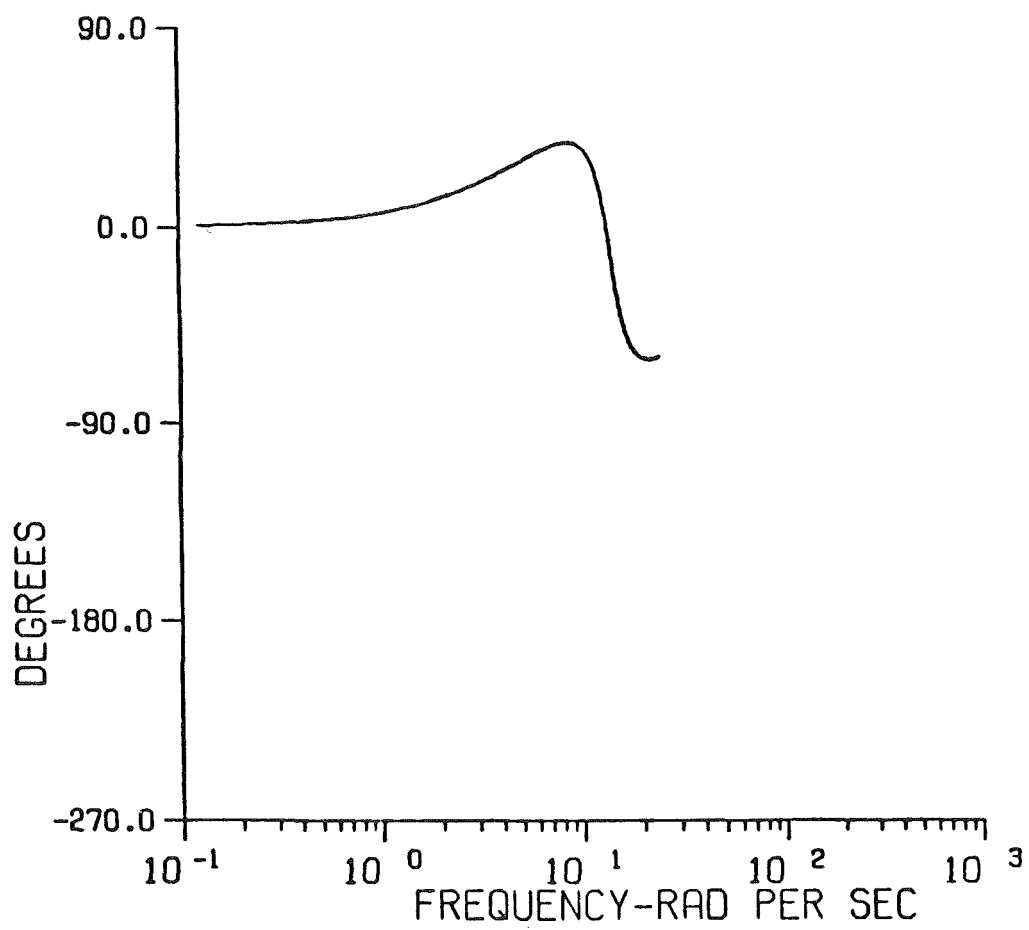
Figure 11 Manual Controller Frequency Pole-zero Response Phase:
K/s^2, $\tau = 0.5$ seconds

33.28
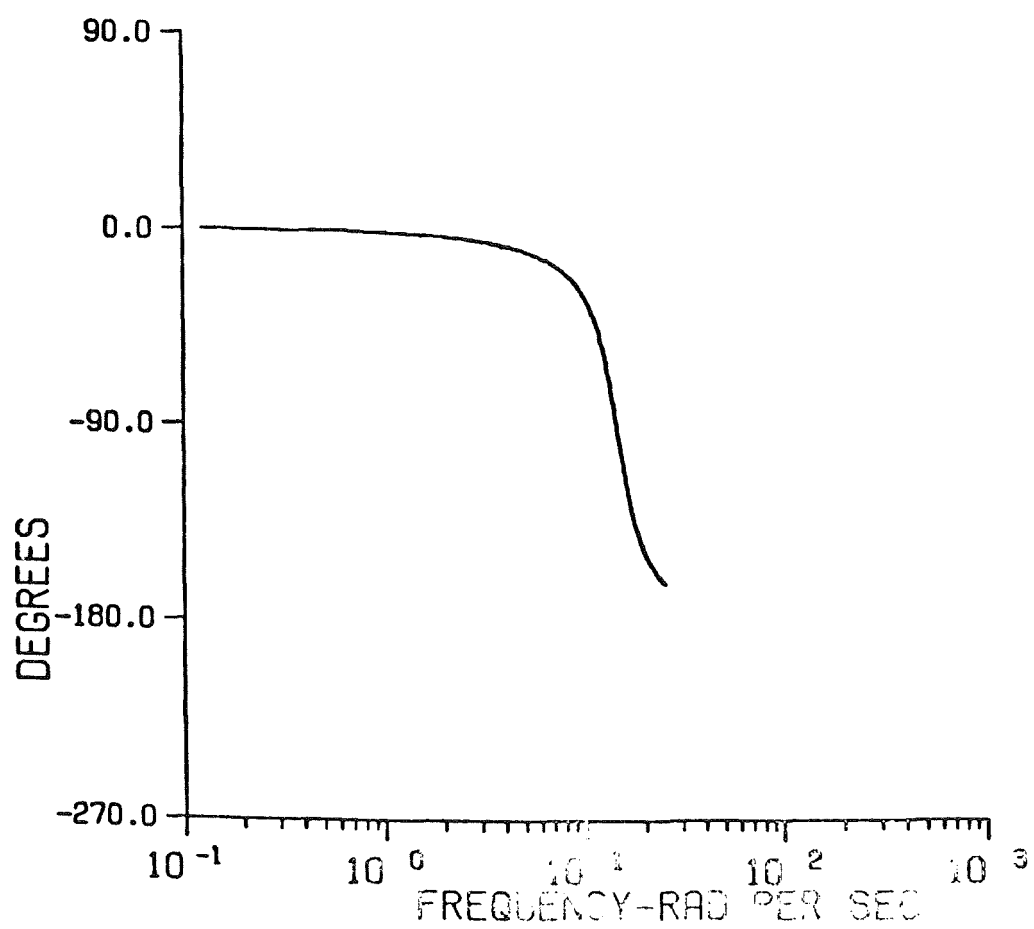
Figure 12 Model Output vs. Pilot Output: K/s^2, $\tau = 0.5$ seconds

predicted
actual

Data not used
for Modeling

Data Source
for Modeling

STICK DEF (DEGREES X 0.1)

3.000
1.500
0.000
-1.500
-3.000

TIME - SECONDS X 20.0

0.0   150.0   300.0   450.0   600.0   750.0   900.0
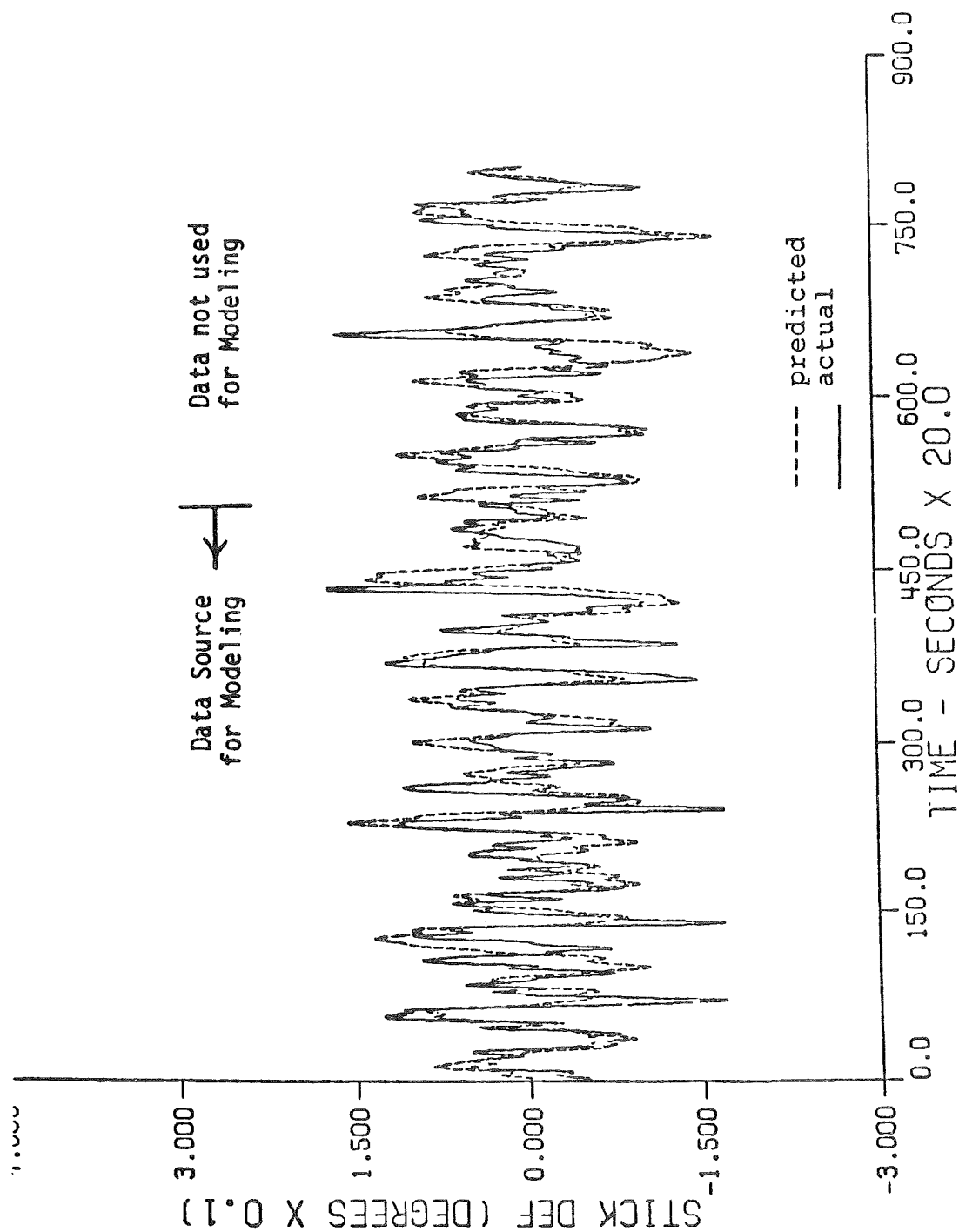
FIGURE 13: Linear Discrete Model for Cross
Correlation Identification

33.30
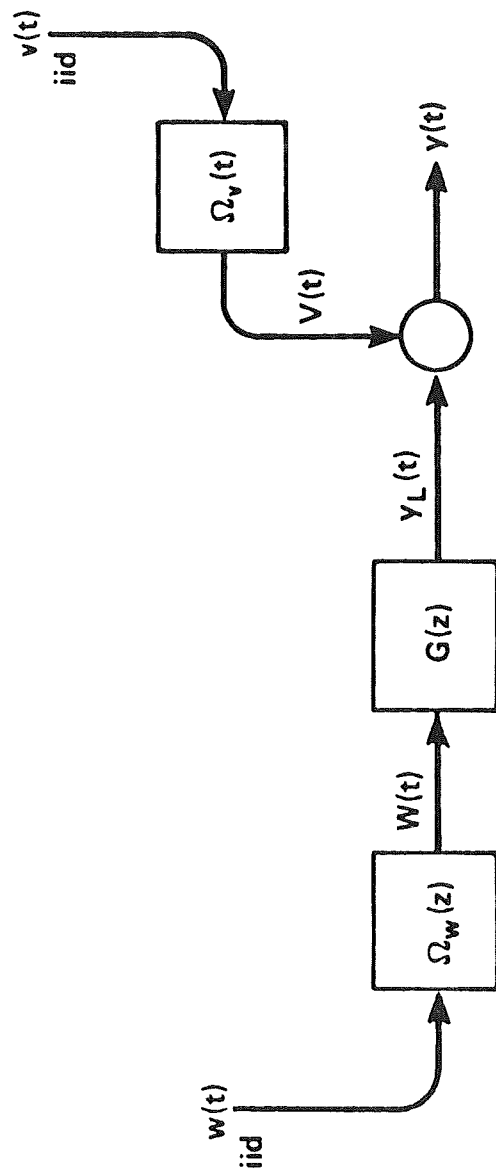
1985 ANNUAL MANUAL

Informal Paper — 10 Minutes

COMPARISON OF THE STI "NIPIP" TRACKING
DYNAMICS IDENTIFICATION WITH THE
ON-LINE FOURIER ANALYZER "DFA" RESULTS
INCLUDING A TIME VARYING CASE

Henry R. Jex
Greg Hanson (formerly at STI)

Systems Technology, Inc.
13766 S. Hawthorne Blvd.
Hawthorne, CA 90250
(213) 679-2281

The Non-Intrusive Pilot Identification Procedure (NIPIP) recently developed at STI and described at the 1981 Annual Manual has been used to identify operators who were compensatory tracking a "sub-critical-instability" task; i.e., the controlled element: $Y_c = K/(s-2)$. NIPIP uses a time domain least squares procedure converting to frequency domain coefficients. The forcing function was a sum of sinusoids supplied by the STI Mark II Describing Function Analyzer, which computes on-line Fourier coefficients of the operator's error/input describing function. The resulting open-loop and operator dynamics computed by each procedure are compared, and they are shown to be reasonably close when there is reasonable power in the error signal at the measurement frequencies.

A special run was made in which the operator abruptly reduced gain within 1 sec, and the ability of the NIPIP to identify this step time variation in the operator is illustrated.

THE DESCRIBING FUNCTION ANALYZER - D.F.A.
SIMPLIFIED FOURIER ANALYSIS

34.2
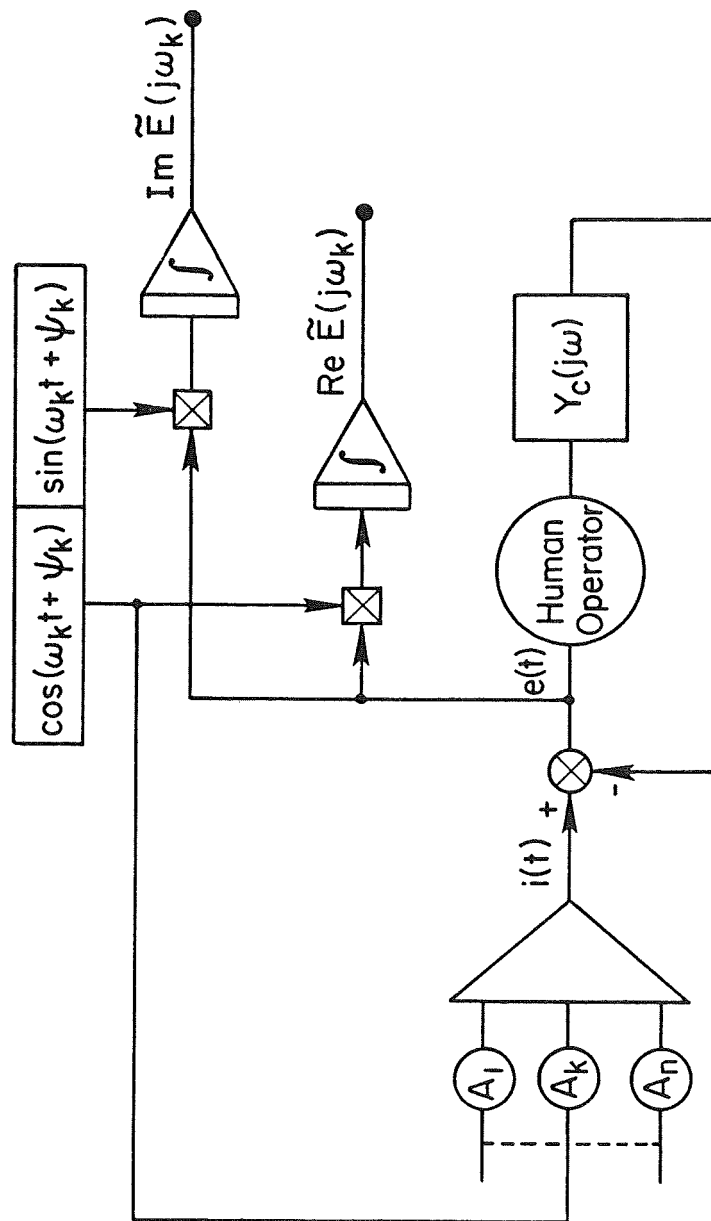
DESCRIBING FUNCTION RELATIONSHIPS

MEASURE

ERROR/INPUT : $\quad Y_{ie} = \dfrac{E(j\omega_k)}{I(j\omega_k)}$ \hfill (1)

DERIVE

OPEN LOOP $\quad$: $\quad Y_{OL} = \dfrac{1}{Y_{ie}} - 1$ \hfill (2)

CLOSED LOOP : $\quad Y_{CL} = 1 - Y_{ie}$ \hfill (3)

HUMAN OPERATOR: $\quad Y_p = \dfrac{Y_{OL}}{Y_c}$ \hfill (4)

OPERATOR'S DESCRIBING FUNCTIONS FROM THE D.F.A.
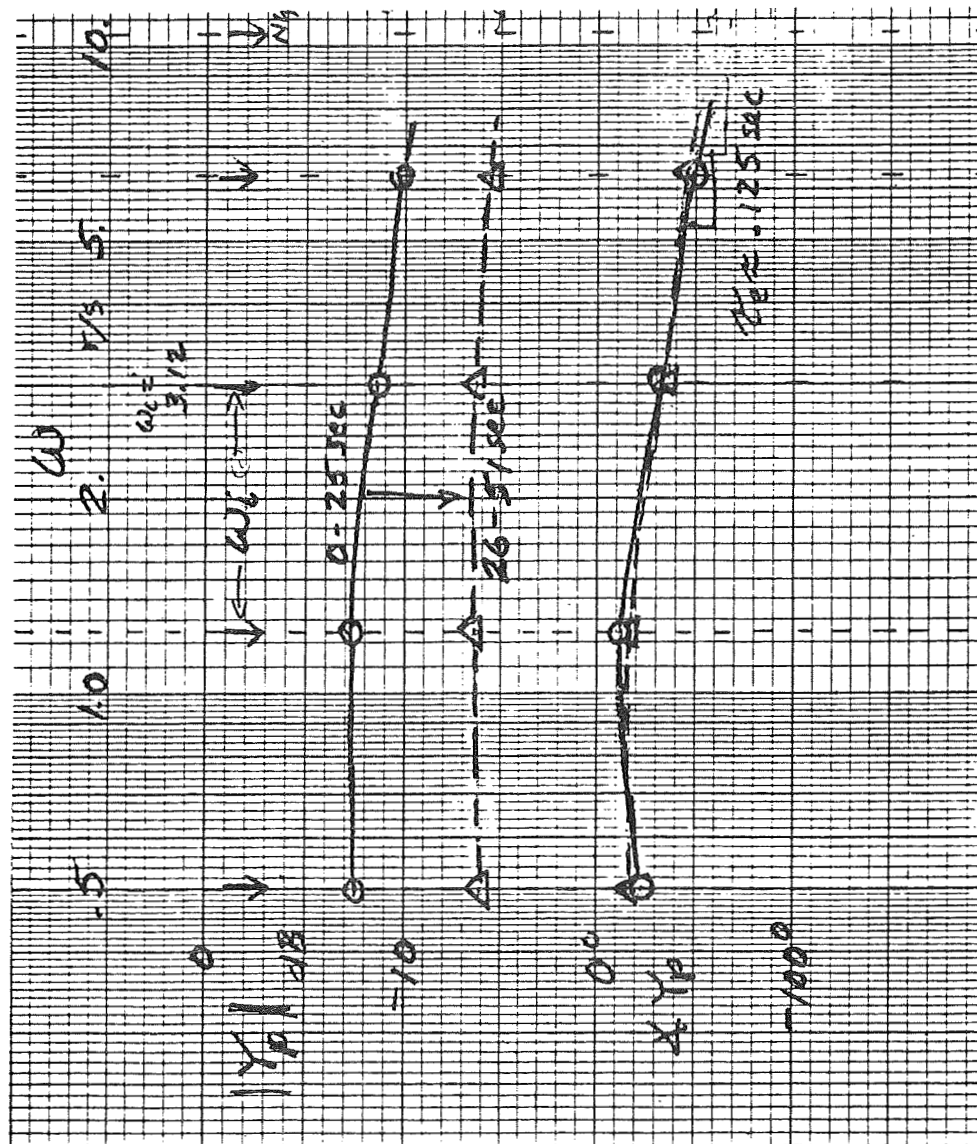
34.4

OPERATOR'S DESCRIBING FUNCTIONS FROM N.I.P.I.P.

$|Y_p|$ dB

$\angle Y_p$

ω_1

ω_2

0-25 sec

25 to 50 sec

$\tau_e = .10$ sec

V_{sine}

3.12

6 db on factor of 2 change in gain

.2   .5   1.0   2.   4/s   5.   10.

0   -10   0°   -100°
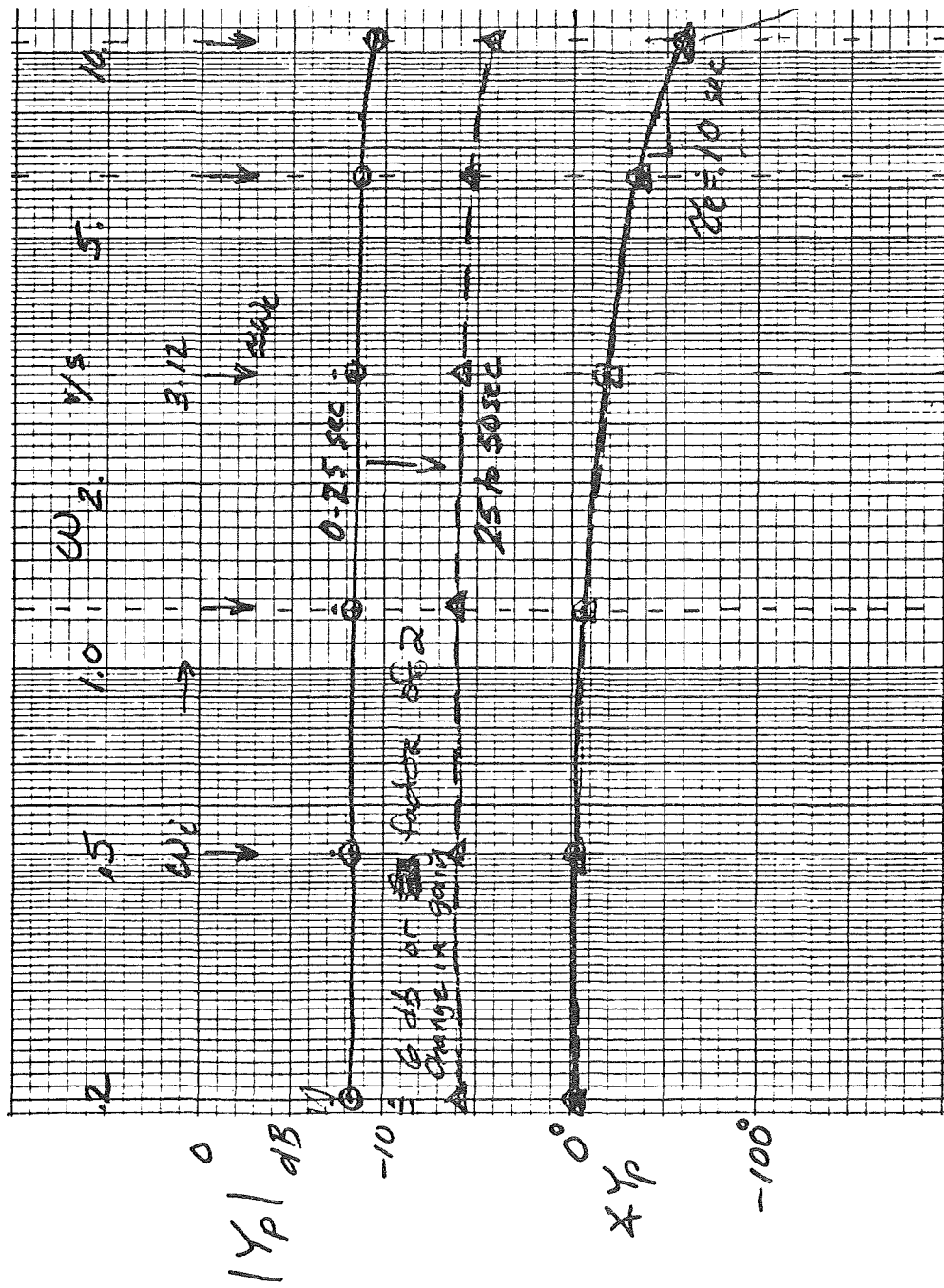
TIME COURSE OF NIPIP's $Y_p @ \omega = 3.12 Y_s$ FOR VARIOUS "WINDOWS"
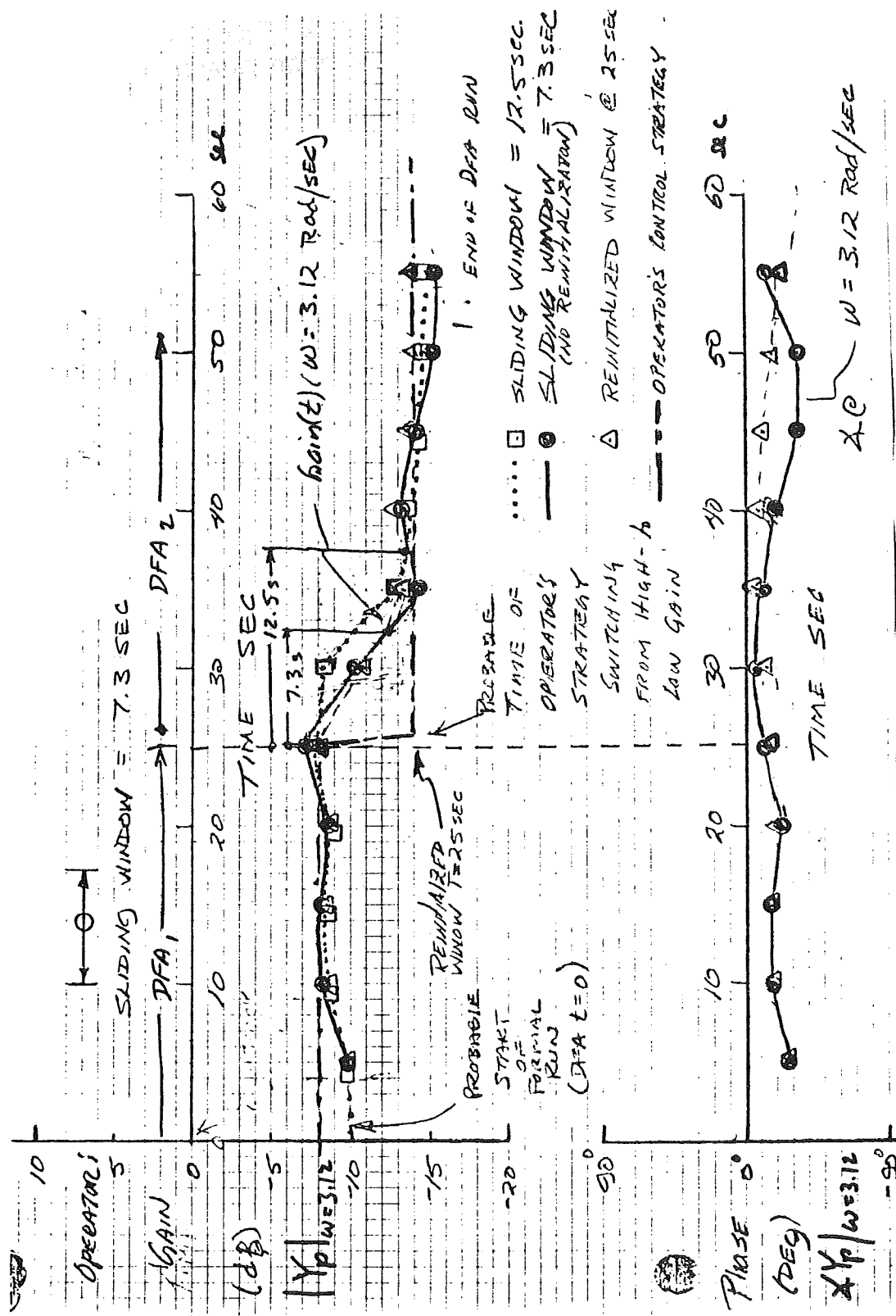
CONCLUSIONS

- NIPIP RESULTS CLOSELY MATCH DFA FOR FREQUENCIES BELOW GAIN CROSSOVER, WHERE $|E/I (j\omega)| < 1.0$

- ABOVE ω_c, IN NOISY CASES (LOW A_i, FATIGUED OPERATOR), DFA DATA ARE UNRELIABLE

- CONVERSION OF TIME DOMAIN COEFFICIENTS TO FREQUENCY DOMAIN DESCRIBING FUNCTIONS <u>AT</u> ω_i's, GIVES VERY RAPID AND ACCURATE DATA

- NIPIP CAN "FOLLOW" TIME A VARYING OPERATOR WITH AN EFFECTIVE LAG $\approx$ SLIDING-WINDOW TIME

A FLIGHT TEST METHOD FOR PILOT/AIRCRAFT ANALYSIS

Dipl.-Ing. Ruthard Koehler
Flight Mechanics Branch, Institut für Flugmechanik der DFVLR
Flugplatz, D-3300 Braunschweig, FRG

and

Dr.-Ing. Ernst Buchacker
Head Handling Qualities Branch, Handling and Performance Section;
Federal Office of Military Technology and Procurement,
German Forces Flight Test Centre, D-8072 Manching, FRG

1. INTRODUCTION

In high precision flight manoeuvres a pilot is a part of a closed loop pilot/aircraft system. The assessment of the flying qualities is highly dependent on the closed loop characteristics related to precision manoeuvres like approach, landing, air-to-air tracking, air-to-ground tracking, close formation flying and air-to-air refueling of the receiver.

The object of a research program at DFVLR is the final flight phase of an air to ground mission. In this flight phase the pilot has to align the aircraft with the target, correct small deviations from the target direction and keep the target in his sights for a specific time period (Fig. 1).

To investigate the dynamic behaviour of the pilot/aircraft system a special ground attack flight test technique with a prolongued tracking manoeuvre has been developed (Fig. 2).

By changing the targets during the attack the pilot is forced to react continously on aiming errors in his sights. Thus the closed loop pilot/aircraft system is excited over a wide frequency range of interest, the pilot gets more information about mission oriented aircraft dynamics and suitable flight test data for a pilot/aircraft analysis can be generated.

This report includes

o general description of the test equipment
o input signal design
o flight test program
o first results of an evaluation.

2. TECHNICAL ARRANGEMENT

The test set up of the Ground Attack Target Equipment (GRATE) shows that it consists of onboard and ground systems (Fig. 3). Main part of the ground system are nine light targets.

The overall arrangement of the ground system (all lamps switched on) is shown in Fig. 4 from the point of view of a pilot during a simulated attack.

Each target is a lamp cross with eight halogen lamps switched on and off by a microprocessor according to the signal received via connecting cables from the telemetry ground station (Fig. 5).

3. INPUT SIGNAL DESIGN

When the pilot has to align the aircraft to light targets switched on and off at different positions on the ground, a multi-step input signal to the pilot/aircraft system is generated. This input signal should be designed to obtain suitable flight test data for system identification.

3.1 General Characteristics

A period of a step signal which changes its value in accordance with constant time intervals Δt is shown in Fig. 6.

The power spectrum of the signal indicates the frequency ranges of the system which can be analysed.

The spectrum $|Z(\omega)|^2/T$ is a function of the interval Δt and the amplitudes v_i and consists of two factors.

The first factor $2 \Delta t (1 - \cos \Omega)/\Omega^2$, where $\Omega = \omega \Delta t$, is a function of the duration of the interval Δt and the frequency ω and is not affected by the switching amplitudes v_i. Changing from the power spectrum to the amplitude spectrum, the root must be taken, resulting in a clearer diagram for the first factor (Fig. 6).

In dependence on Δt amplitude values vanish at equidistant frequencies $\Omega_N = 2 k\pi$ ($k = 1, 2, 3, \ldots$). At these frequencies the power spectrum disappears, independently of the second factor.

The peaks of the functions shown steeply decrease with increasing frequency Ω. Since the second factor is periodic with $\Omega = 2\pi$ the drop of the amplitude spectrum at higher frequencies cannot be prevented by a special selection of the amplitudes v_i. These characteristics of the spectrum counteract the effort of generating signals with an approximately constant spectrum. But the possibilities existing within the limits discussed should be utilized.

For the tests the effort should first be made to keep the duration of the interval Δt as small as possible. This expands the region to the first null towards higher frequencies.

In the flight test, limits based on the characteristics of the pilot/ aircraft system are set to the selection of short interval periods Δt.

If the change-over times are too short, the pilot is not able to perform the attack. On the other hand interval periods which are too long are meaningless since the alignment of the aircraft is followed by an approximately steady-state process which contains only few information and the pilot is not motivated to pay full attention.

The range of meaningful interval periods was determined in the test and is approximately

$$2.25 \leq \Delta t \leq 3.15.$$

Step signals with different duration of intervals Δt can be applied in the tests to alleviate the effects of the nulls to a certain extent.

3.2 Input Signals for Ground Attack

In order to be able to investigate various parts of a pilot model, they are subjected to separate tests.

To investigate the pilot characteristics with respect to compensation in longitudinal motion an excitation by setting up the lamps in longitudinal direction is provided (Fig. 7).

The lamps are switched whilst the aircraft flies along a predetermined flight path.

A computer program was generated for selecting suitable signals from the multitude of all possible input signals. It supplies a predetermined number of signals $z(t)$ which

- ○ change over to a new value after each interval Δt
- ○ exceed a minimum limit for the standard deviation to cause sufficiently large jumps in pitch angle
- ○ do not exceed a predetermined size of visual angle steps ($1°$).
- ○ have a power spectrum of the visual angle ε_x which is constant within the frame of possibilities.

The step signal in Fig. 7 was determined by this program. Its spectrum is also shown in the figure.

To investigate the pilot characteristics with respect to compensation in lateral direction an excitation by setting up the lamps in lateral direction is provided (Fig. 8).

Although this arrangement also generates an excitation in the longitudinal motion, the excitation in the lateral motion remains dominant.

The input signals used in the test were also determined by the program mentioned above.

For the investigation of the overall pilot model an apparently arbitrary excitation in any direction is required. A favourable arrangement is given if nine lamps are set up in a nine pins game-like pattern (Fig. 9).

Each lamp is provided with a position number as specified in the Figure. It is switched on when the signal $z(t)$ has the value of the position number.

From the signal $z(t)$ the x- and y-position number can be derived. These signals $z_x(t)$ and $z_y(t)$ can be treated in the same manner as the input signals previously discussed.

Also for these tests a computer program was written to generate input signals which have power spectra of the visual angles ε_x and ε_y which are constant within the frame of possibilities.

4.   FLIGHT TEST PROGRAM

A flight test program was performed utilizing a modified Alpha-Jet with a multi mode control system.

A total of 10 flights with 183 attacks were executed. The test pilots rated the task to be well suited for evaluating air-to-ground handling qualities. Pilot comments were documented using the well known Cooper Harper rating scale along with related scales for turbulence, pilot induced oscillation susceptibility and buffet. Good correlation was obtained between pilot comments and ratings and the apparent behaviour of the system.

5.   EVALUATION OF FLIGHT TEST DATA

The evaluation of the flight test data has been concentrated on the investigation of tracking performance parameters. In particular the initial line-up time was evaluated. For these investigations flight test data were measured from head-up display camera film including position of pipper and the illuminated lamp.

In the time histories of an example shown in Fig. 10 the steps in pitch and azimuth angle when the lamps are switched, and the changes initiated by the pilot in order to track the targets are clearly visible.

The star like pattern in the cross plot of target minus pipper position shows four loops which correspond to the four steps of the light signal.

35.4

The time histories of these four sequences can be treated as four isolated characteristic motions with different initial conditions of the pilot aircraft system. When the light jumped in the negative direction of pitch or yaw, the subsequent time histories of the deviations in pitch or yaw were turned over (multiplied by -1) to deliver a characteristic motion with a positive initial condition.

The mean values calculated from the four characteristic motions and curves of limits of confidence are shown in Fig. 11.

Thus the influence of noise on the time histories can be reduced.

The characteristic motions of pitch and yaw were used to compute a mean radial deviation which decreases over time.

The time up to the moment, when the mean radial deviation passes the value of 3 mrads was determined and increased by 10 %. This result was defined to be the initial line-up time.

Fig. 12 shows the dependence of this time to the serial run number of a flight and the excitation mode.

Each symbol is a result of one attack run. Two approaches signed in the diagram were affected by windshear effects and resulted in comparatively large initial line-up times. Therefore they should be unconsidered in further contemplations.

In the diagrams a slight increase of the initial line-up time with respect to the serial run number is visible.

The time to align the aircraft after a vertical step of the target is shorter than after lateral or combined displacements.

In general the time to line-up the aircraft is very short and did not exceed the value of 1.5 sec.

Further investigations will include a flight path reconstruction utilizing camera and tape recorded data. Tracking performance parameters will be determined and an identification of the pilot/aircraft system will be initiated.

For the enlargement of the data base additional flight tests were performed.

6. CONCLUSIONS

A flight test method was presented that has some advantages compared to previous methods.

o The pilot is engaged in an operational task of a flight phase, which is very important for military missions.

o Input signals act directly on the pilot, no modification in the control system of the aircraft is necessary and the pilot can immediately interrupt the test.

o The applied input signals are predesigned and reproducible in the flight tests. They can be adapted to the manoeuvrability of the pilot/ aircraft system. Thus the pilot and the aircraft can be excited up to high frequencies.

o The test method is well accepted by test pilots and proved very effective for flying quality assessments by pilot comments and ratings.

o The input and output signals of all subsystems can be measured.

o The data are suitable for pilot/aircraft analysis.

o A preliminary evaluation of test data was concentrated on the determination of initial line-up times. The dependence of the results on windshear effects, serial run number of a flight, and excitation mode was discussed.

Further investigations will be concentrated on determining more tracking performance parameters and on system identification of pilot/aircraft systems.

7. REFERENCES

Onstott, E.D.	Prediction, Evaluation and Specification of Closed
Faulkner, W.H.	Loop and Multiaxis Flying Qualities.
	AFFDL-TR-78-3 (1978)

Buchacker, E.	Flight Testing an Alpha Jet with a Direct Side Force
Koehler, R.	Control System in the Air-to-Ground Role.
Meyer, H.	Flight Test Symposium 1984, CFB Cold Lake, Medley,
Skudridakis, J.	Alberta (Canada), 11./12.04.1984.

Koehler, R.	Design and Implementation of Input Signals for Identification of Pilot/Aircraft Models.
	ESA-Translation of DFVLR-FB 84-08, EA-TT-880.
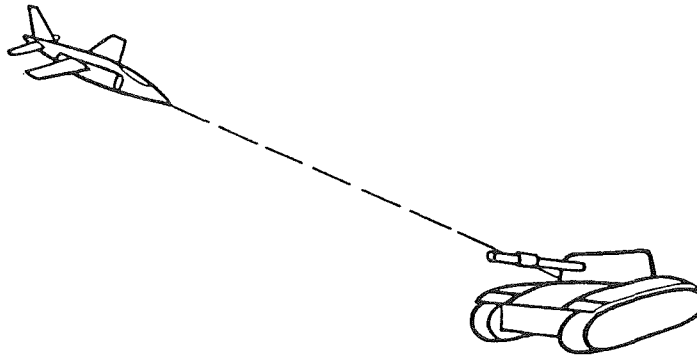
Fig. 1 Operational Air to Ground Tracking Task

LIGHT TARGETS

SEQUENCE OF
SHINING TARGETS

T
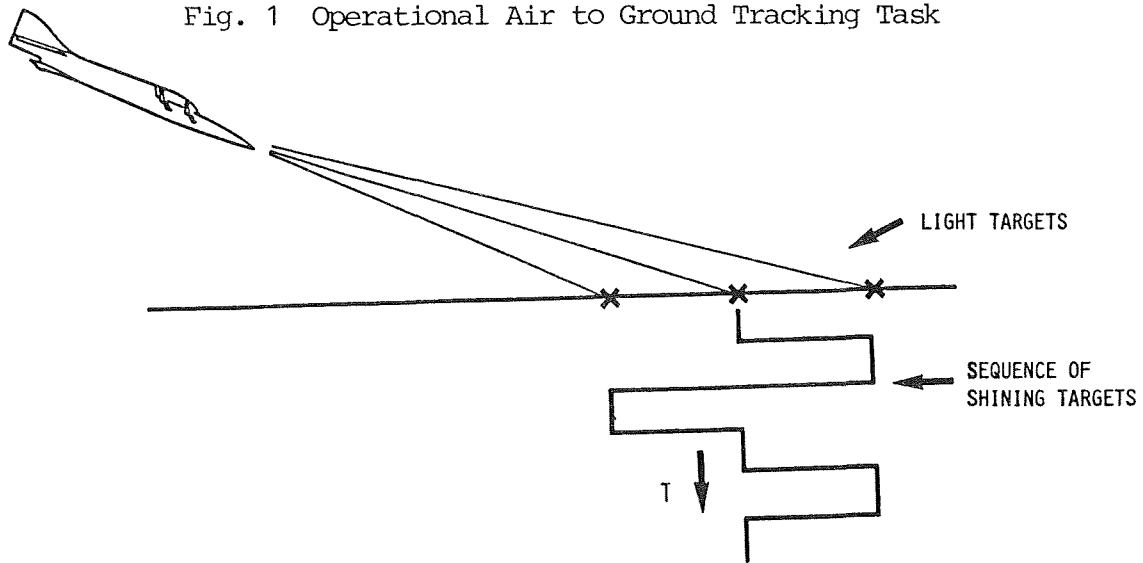
Fig. 2 Modified Air to Ground Tracking Task

| AIRBORNE FTI TAPE | AIRBORNE SIGNAL GENERATOR | HUD CAMERA |

AIRBORNE
TELEMETRY
TRANSMITTER

LIGHT
BEAM

Fig. 3 Test Setup

TELEMETRY
GROUND STATION

TARGET
LIGHTS

35.7
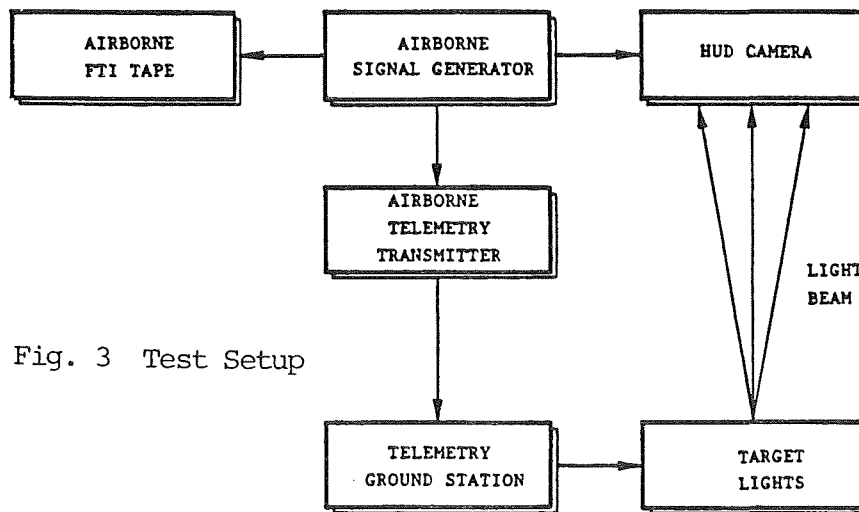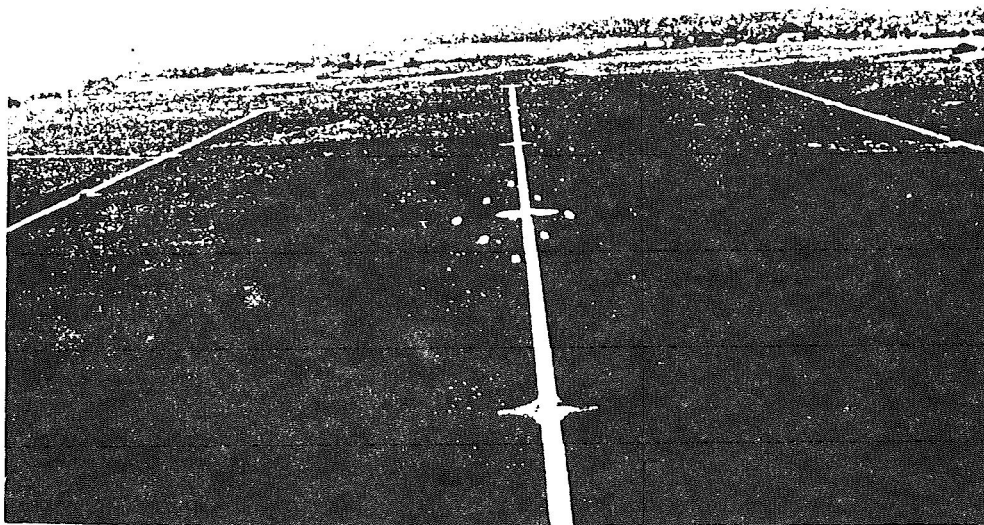
Fig. 4 Arrangement of Target Lights on Ground

Fig. 5 A Lamp Cross

35.8
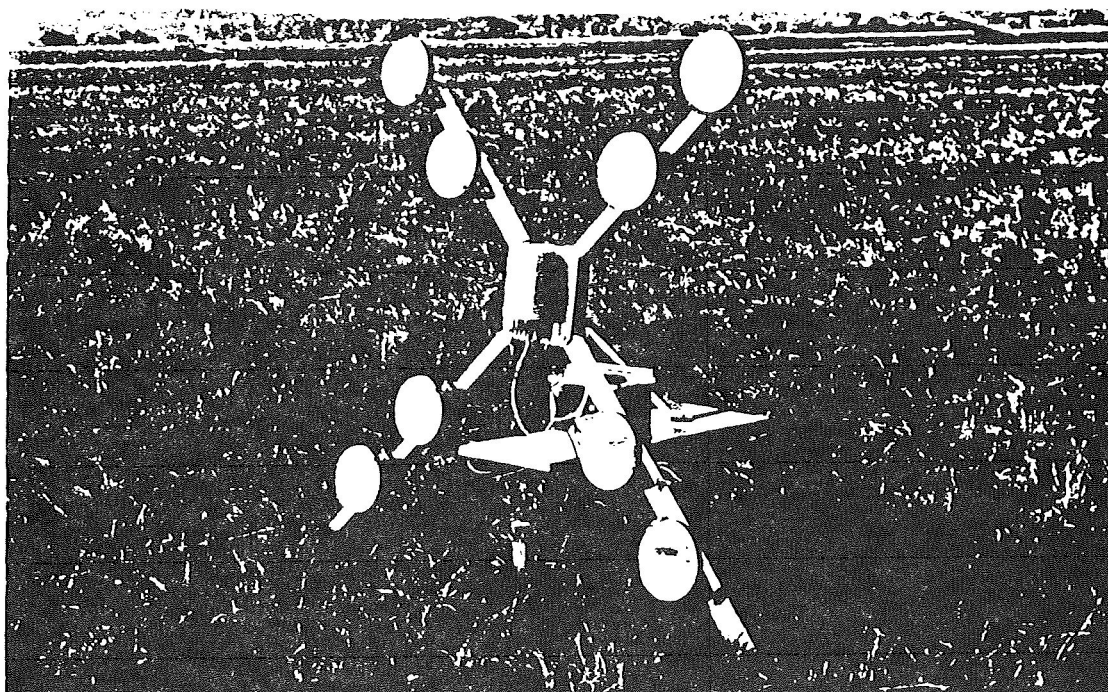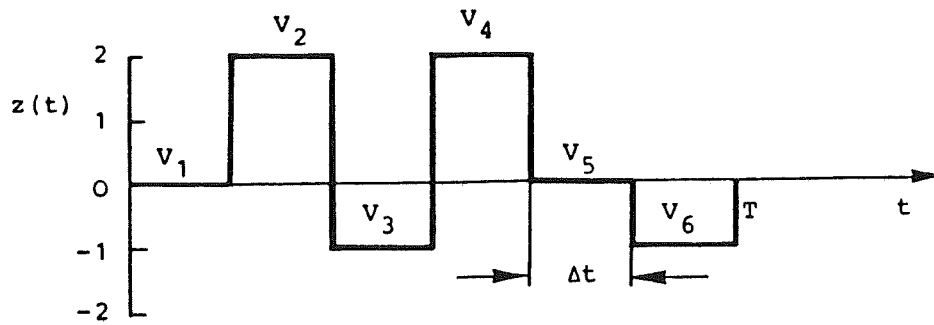
$$\frac{|Z(\omega)|^2}{T} = \underbrace{2\Delta t \ \frac{1-\cos\Omega}{\Omega^2}}_{} \left[\frac{1}{N} \sum_{i=1}^{N} V_i^2 + \frac{2}{N} \sum_{i=1}^{N-1} \cos i\Omega \sum_{j=1}^{N-i} V_{i+j} \ V_j \right]$$

with $\Omega = \omega\Delta t$
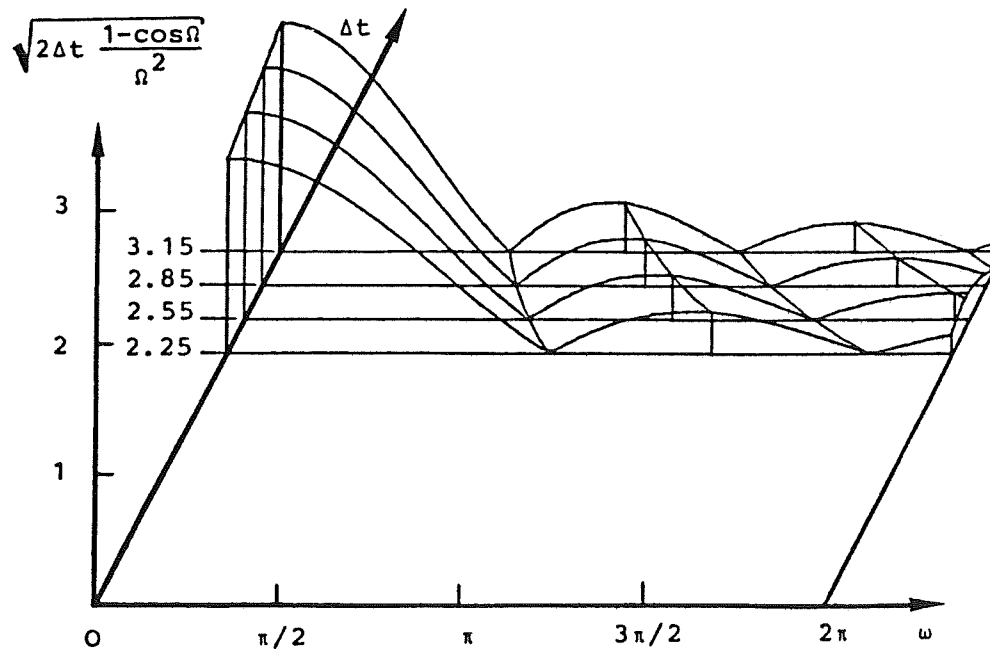
Fig. 6   Power Spektrum of a Multi Step Input Signal and
a Presentation of the Dominant Factor

35.9
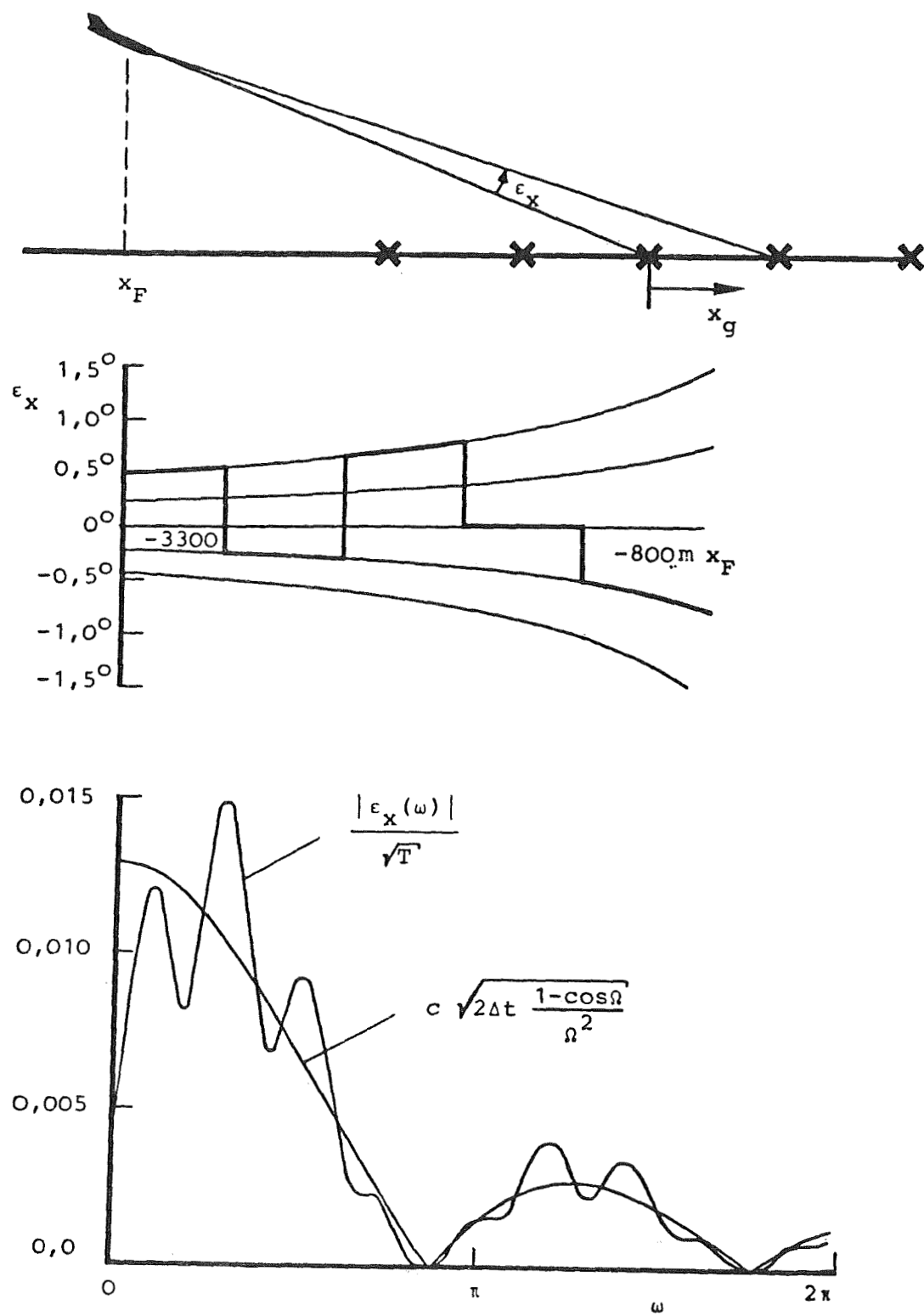
Fig. 7 Time History and Spectrum of Line of Sight Angle $\epsilon_x(t)$

(Δt = 2.25 sec)   35.10

$$\Delta y \ = \ 14 \ m$$
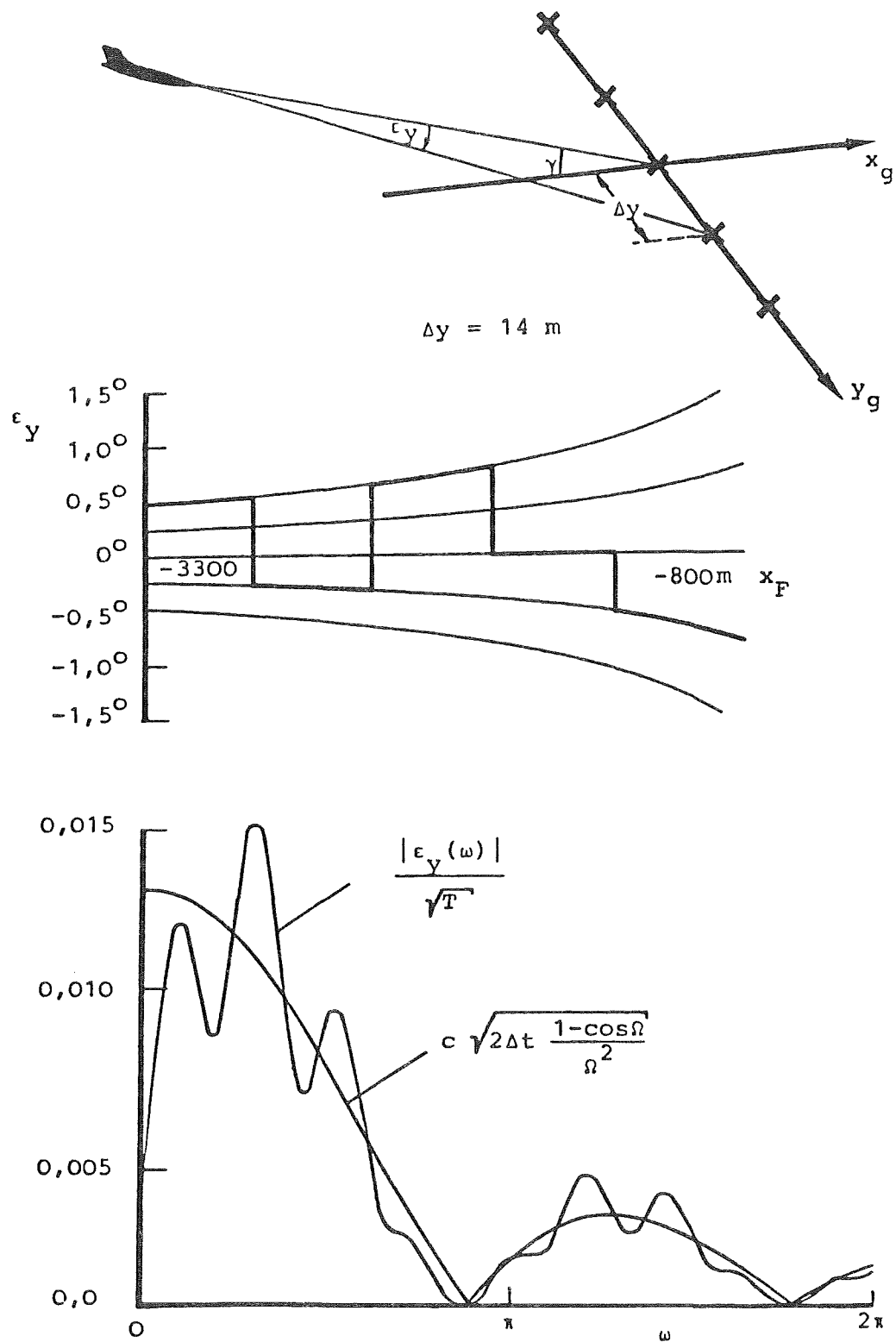
Fig. 8 Time History and Spectrum of Line of Sight Angle $\varepsilon_y(t)$

($\Delta t = 2.25$ sec)

35.11
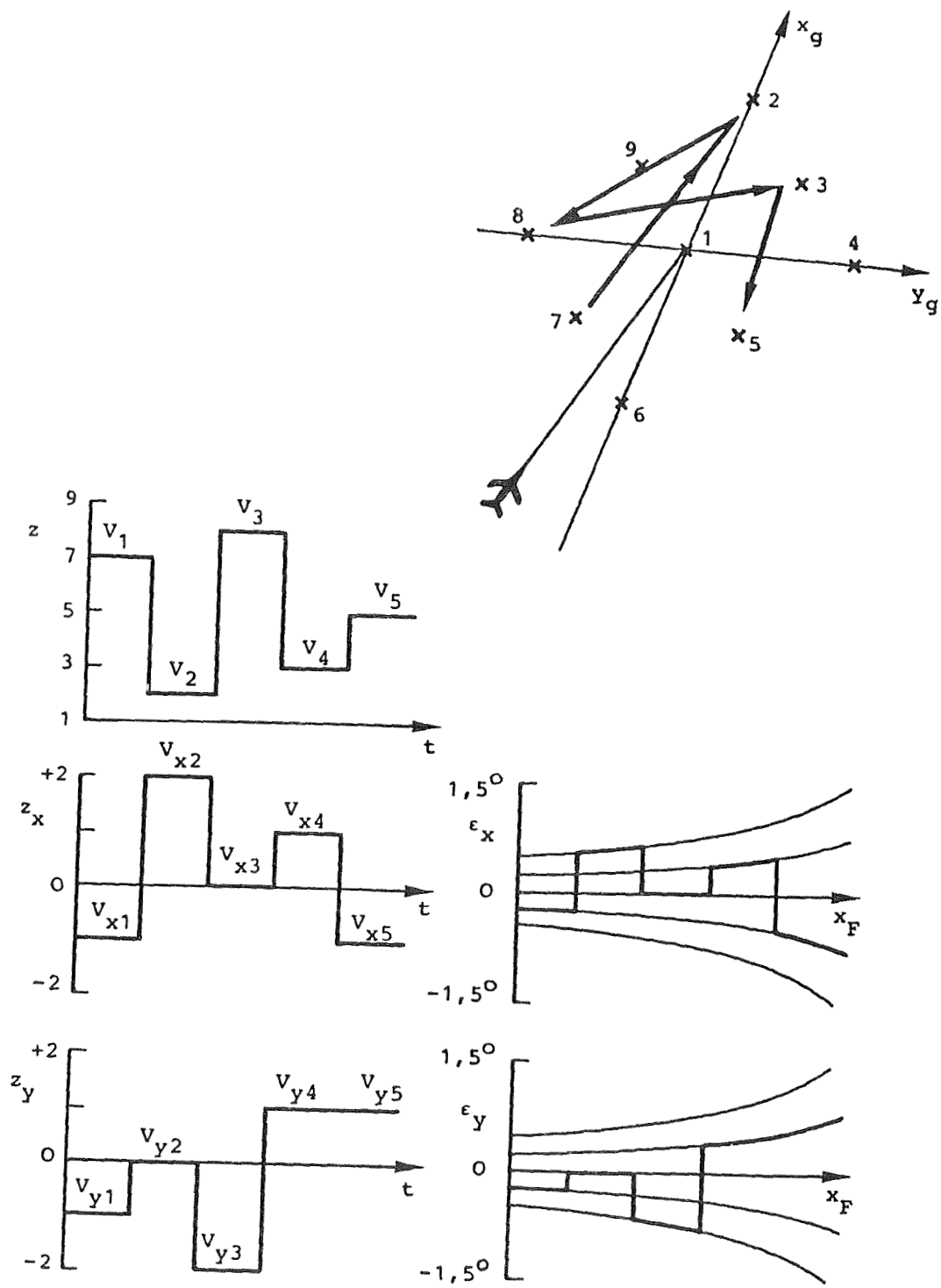
Fig. 9   Input Signals of the Combined Pattern

35.12
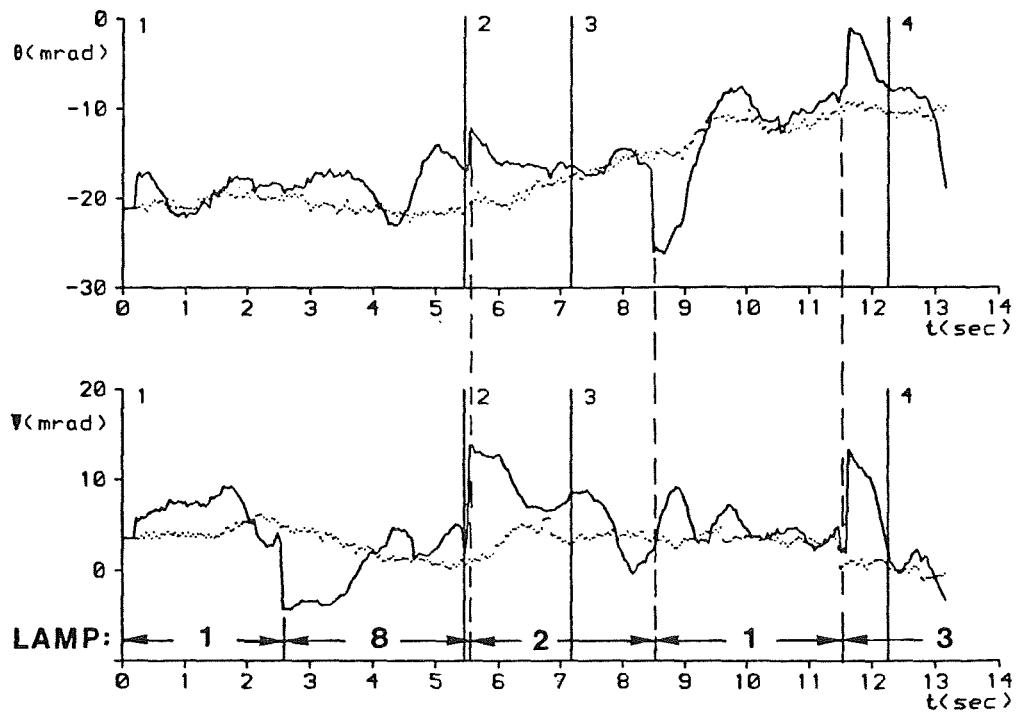
TIME HISTORIES OF FILM DATA

——————— LIGHT   — — — — PIPPER

|3 RANGE OF PIPPER SYMBOL

$\theta_L - \theta_P$ (mrad)
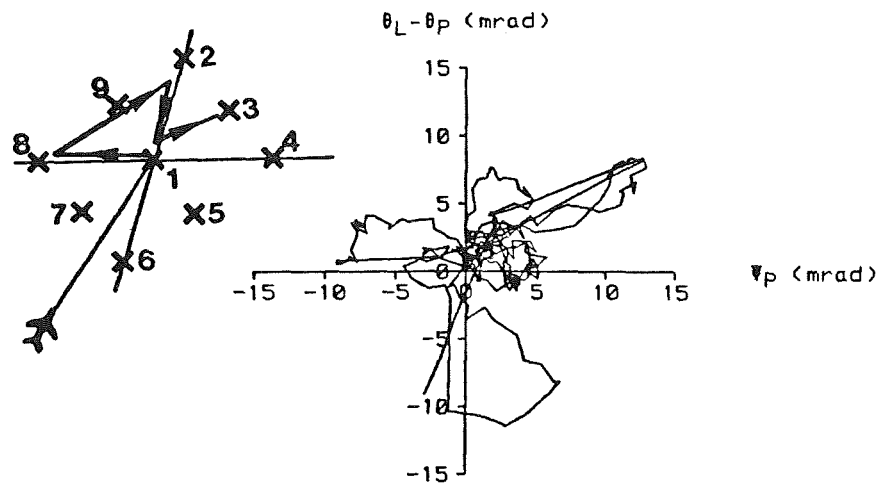
AIMING ERROR

Fig. 10 Plots of Film Data

FLIGHT   FILM   RUN
786   1   4

35.13

MEAN DEVIATION IN VERTICAL DIRECTION

MEAN DEVIATION IN LATERAL DIRECTION
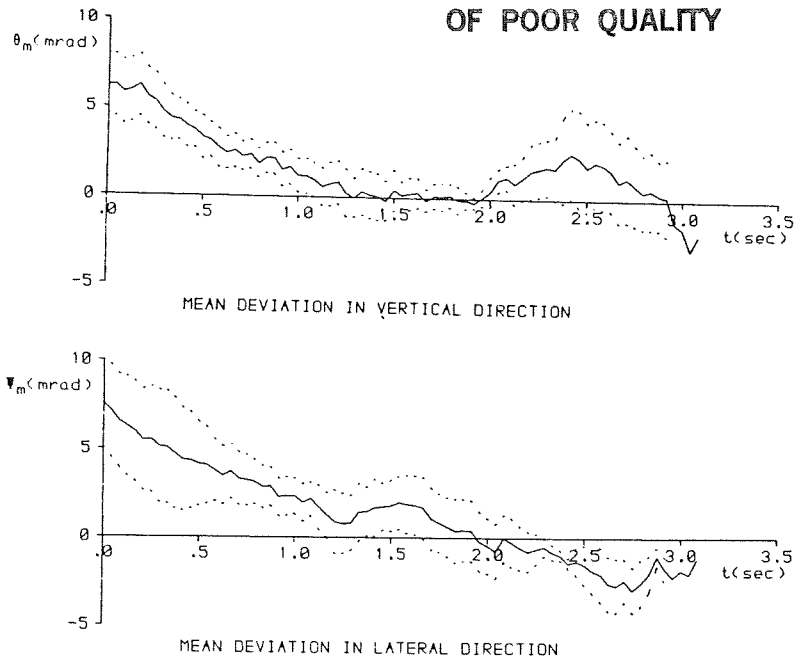
Fig. 11   Time Histories of the Characteristic Motions
from Fig. 10

———— Mean Values
– – –   Limits of Confidence
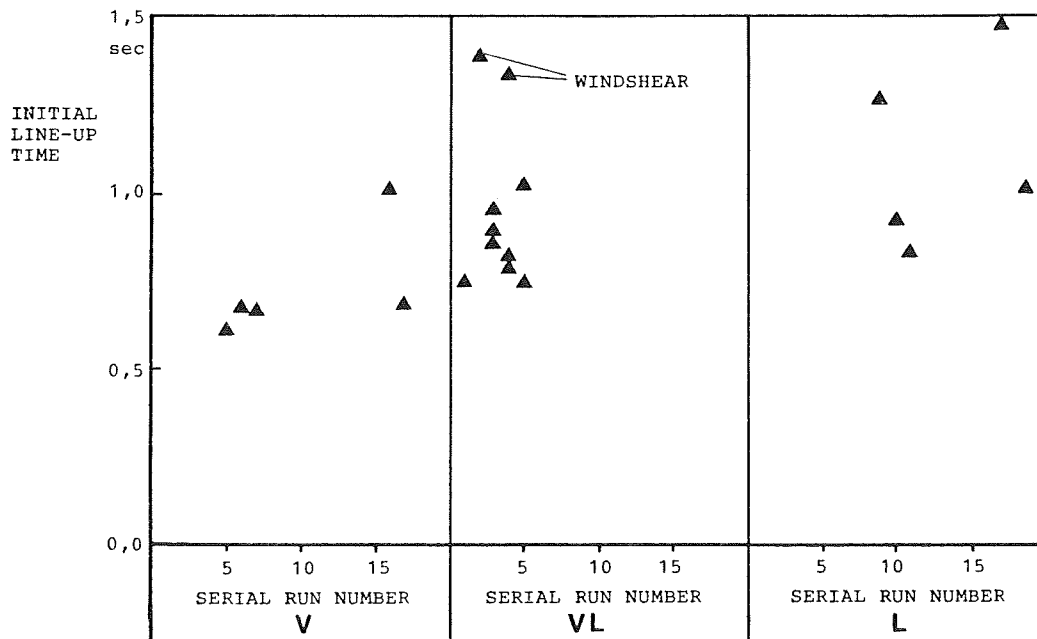
Fig. 12   Initial Line-Up Time Evaluated from Camera Data

V Vertical, L Lateral, VL Vertical and Lateral Excitation

MAXIMUM NORMALIZED ACCELERATION
AS A FLYING QUALITIES PARAMETER

Joel S. Warner
Edward D. Onstott

Northrop Corporation, Aircraft Division
Hawthorne, California 90250

INTRODUCTION

In 1984, Maximum Normalized Rate (MNR) was presented as a Flying Qualities parameter [1]. Subsequent analysis of data from ground based simulation and flight test revealed the utility of a companion parameter, Maximum Normalized Acceleration (MNA). MNR and MNA profiles reveal the presence of both continuous and pulsed compensation strategies during discrete attitude tracking. In addition, MNR appears to be a suitable metric for pilot opinion in the LATHOS data base, while the MNR/MNA relationship is sensitive to pilot-induced-oscillation (PIO) and roll ratcheting problems.

Although the lateral roll mode of a conventional aircraft is perhaps the easiest dynamic mode to comprehend, there remain several poorly understood aspects of piloted control in this axis. For example, analytical prediction and fixed-base flight simulation tend to indicate that the shortest possible roll mode time constant is best. However, moving-base and in-flight simulations show clear disadvantages in such highly damped aircraft: pilot-induced-oscillations and roll-ratcheting often result during these cases [2]. Thus, real-world considerations, such as ride qualities effects on pilot compensation strategies, need to be accounted for.

Step Target Method

As part of an investigation of this problem, Northrop has developed an analysis technique known as the Step Target Method [3]. The Step Target method is essentially a one degree-of-freedom simulation, where an attitude command in the form of a step function is presented to a closed-loop pilot/aircraft model, as shown in Figure 1.

The aircraft dynamics model can be as simple or complex as the investigation warrants. Although discrete pilot modeling technology is largely still in development, an effective tracking model consisting of proportional blends of error and error rate are used as the basis of the Step Target method. An essential feature of this model is that it consists of two stages; the first stage contains values of gain and lead which are appropriate for gross target acquisition, while the second stage

is tuned for fine tracking. The model automatically switches from the first stage to the second when the attitude tracking error is brought within 25% of the commanded attitude change.

PILOT MODEL FOR ACQUISITION:

$\text{TIME} < D,\ \delta S_{P_I} - (\text{DELAY } \tau)\ \left\{ K_{P_I} (\theta_e(t)\ \cdot\ T_{L_I}\ \dot{\theta}_e(t)) \right\}$

PILOT MODEL FOR TRACKING:

$\text{TIME} \geqslant D,\ \delta S_{PF} = \text{DELAY } \tau)\ \left\{ K_{P_F} (\theta_e(t) + T_{L_F}\ \dot{\theta}e(t) + K_{IC} \int_0^t \theta_e(s)\ ds) \right\}$
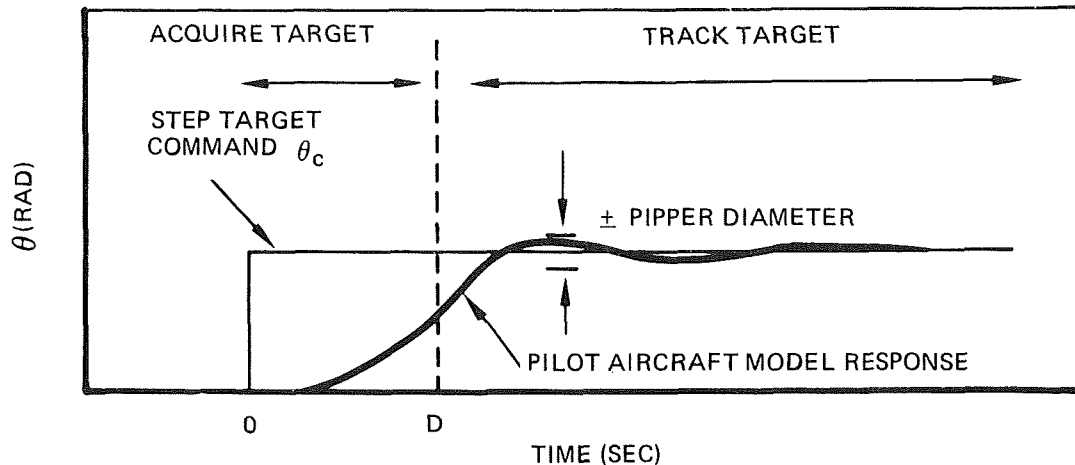
Figure 1. Definition of Step Target Tracking Task.

Previous Analysis of Neal-Smith Data

Onstott and Faulkner used the Step Target method to analyze an in-flight simulation performed by Neal and Smith [3]. The Neal-Smith simulation involved discrete pitch step attitude tracking, using the NT-33 variable stability aircraft [4].

The dynamic configurations modeled with the NT-33 were analyzed using the Step Target method. The two primary output parameters examined were Time-On-Target (TOT), and the Root-Mean-Square of the tracking error (RMS). RMS reflects the ability to maneuver the vehicle, while TOT, specified with respect to a tolerance of 2.5% of the commanded step magnitude, reflects freedom from overshoot and oscillation. Pilot model coefficients were adjusted to obtain maximum TOT for each individual vehicle configuration, which forces the quickest acquisition of the target, with low overshoot and oscillation. The resulting TOT and RMS values were compared to Pilot Opinion Ratings from the Neal-Smith experiment, and are shown in Figure 2 [5].
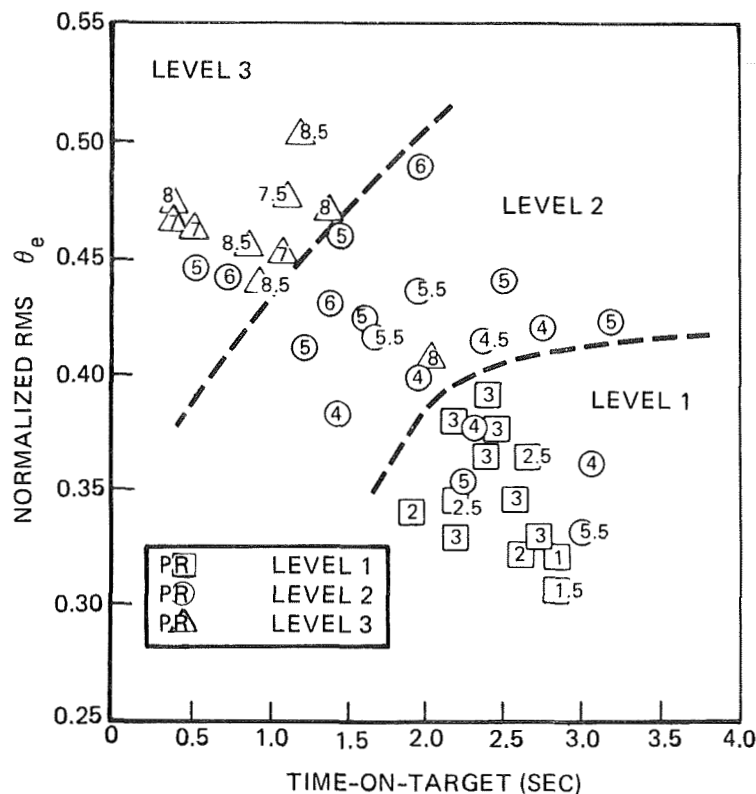
LEVEL 3

LEVEL 2

LEVEL 1

NORMALIZED RMS θ_e

0.55
0.50
0.45
0.40
0.35
0.30
0.25

0   0.5   1.0   1.5   2.0   2.5   3.0   3.5   4.0

TIME-ON-TARGET (SEC)

PR ☐   LEVEL 1
PR ◯   LEVEL 2
PR △   LEVEL 3

Figure 2. Pilot Opinion Ratings from the Neal-Smith Study
as Functions of RMS and TOT.

Analysis of LATHOS Data

Further analysis with the Step Target method was conducted using additional in-flight simulation data from the NT-33. The Lateral Flying Qualities of Highly Augmented Fighters study (LATHOS) was used as a source of time histories and pilot comments [6].

Using a first order lag/delay aircraft model to simulate various LATHOS configurations, attempts were made to optimize the two-stage pilot model in the same manner used to generate the Neal-Smith correlations. Unlike the routine used in the Neal-Smith problem, the automatic optimizing algorithm proved to be badly behaved, resulting in very large values of gain and lead during a short first stage, followed by second stage coefficients which were stable but very small. In short, the model seemed to approximate, as well as it could, a time optimal pulsed solution. However, for the first order lag/delay aircraft models simulated in the LATHOS study, such an optimization problem for maximizing TOT is ill-posed in the absence of constraints imposed by higher order dynamics and nonlinearities; the model could be made to do arbitrarily well at the expense of sufficiently large control inputs.

36.3

As the two-stage model in its current state was shown to be ill behaved, analysis was performed using the single-stage model. Correlations between LATHOS pilot comments and the continuous pilot models were therefore sought. This effort yielded two conclusions:

1) Strong linear correlations were obtained between Pilot Opinion Ratings (POR), RMS, and TOT, as shown in Figure 3 [1]. This correlation is stronger than in Figure 2, which displays regional but not linear correlations.

2) For the LATHOS discrete task simulation, Pilot Opinion Rating is correlated well by a parameter called Maximum Normalized Rate (MNR), as shown in Figure 4 [1]. MNR is defined as the maximum roll rate achieved during the maneuver, normalized by the magnitude of the step command:

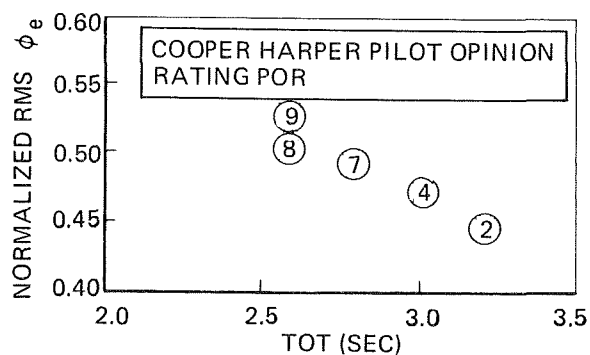$$\text{MNR} = \max\ (\ d\phi /\ dt\ )\ /\ \phi\ \text{command}$$

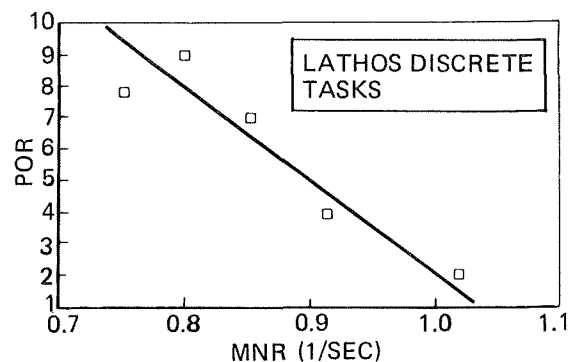Figure 3. Pilot Opinion Rating, Correlated with TOT and RMS Bank Angle Response

Figure 4. Relationship of POR to MNR

These results indicate that the LATHOS pilots were evaluating the configurations in terms that correlate with MNR derived from continuous constant control, even though time histories from [6] exhibit pulsed pilot control strategies. This left two questions to be resolved: 1) what control strategies were the NT-33 pilots using when they flew this discrete step problem in LATHOS, and 2) could simulator pilots achieve the extremely large TOT's that the model was indicating. In the case of the LATHOS simulations, pilots were given performance standards which did not require extremely large TOT. Nevertheless, the pilots often maneuvered very aggressively, as shown in published LATHOS time histories.

Northrop Ground Based Simulation

To investigate the mechanics of discrete maneuver attitude tracking, an investigative simulation study was performed on the Northrop Flight Controls Research Simulator (FCRS).

The FCRS consists of a generic single-place cockpit, representative of a single seat fighter aircraft. Out-the-window and instrument symbology display capabilities are provided by a Megatek 7000 video monitor. Computation was performed by a pair of Gould/SEL 32/55 minicomputers, configured to operate in parallel.
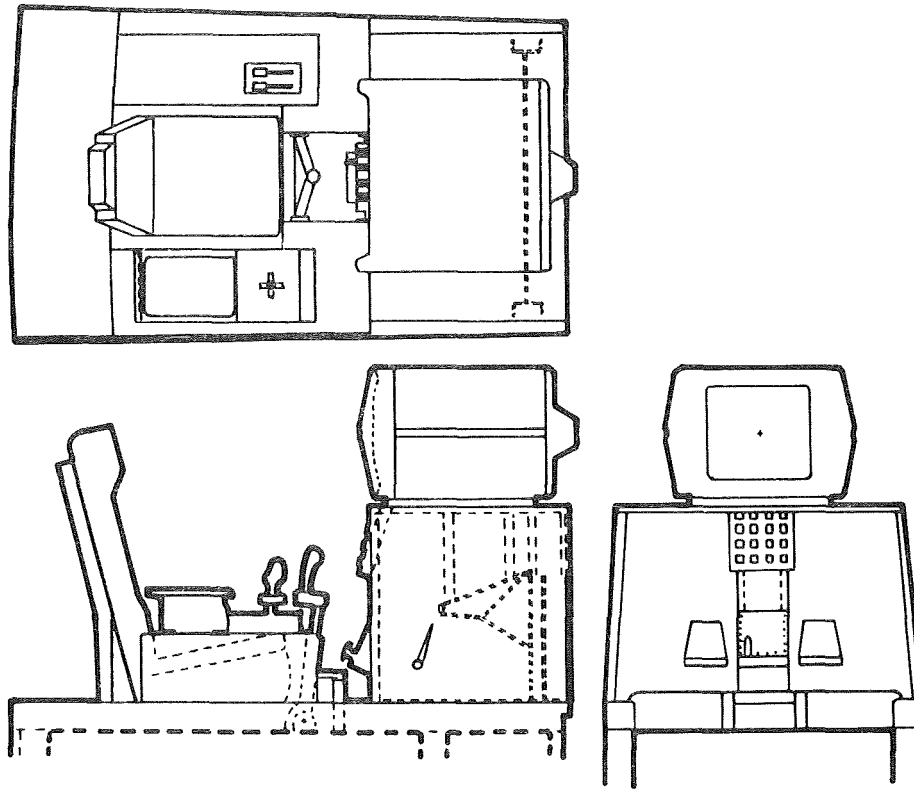

Figure 5. Northrop Flight Controls Research Simulator

A series of six discrete bank-angle attitude step-tracking commands was presented to the pilot during a 30 second trial. Commanded bank angles were randomly varied between 0.3 and 0.6 radians. After a brief pause, another set of six steps was presented. After ten sets, statistics were computed and printed. Data collected included RMS, TOT, MNR, and a new parameter, Maximum Normalized Acceleration (MNA). MNA is defined to be the maximum roll acceleration achieved during the maneuver, normalized by the magnitude of the step command:

$$MNA = \max (d^2\phi / dt^2) / \phi \text{ command}$$

Simulator results exhibited abrupt pulsed pilot control. In fact, the pilots were utilizing full stick deflections during the tracking, producing large values of MNR, MNA, and TOT. Nevertheless, this type of aggressive control activity was identified in LATHOS time history data.

Comparison of LATHOS Data with Northrop Simulation Data

In order to allow a meaningful comparison of LATHOS time histories with the Northrop ground-based simulation data, a set of LATHOS cases was chosen using the following criteria:

1) The maneuver had to be flown with rapid acquisition of the commanded target value, with minimal overshoot and oscillation.

2) There could be no reported problems with control harmony, adverse force gradients, or other contaminating influences.

In comparing these selected cases against the ground based simulation data, similarities were observed: ground-based and in-flight simulation pilots both were able to push their TOT performance in a manner reminiscent of the automatic optimization algorithm. Data from both sources have been plotted together as functions of MNR versus MNA, as shown in Figures 6 and 7.
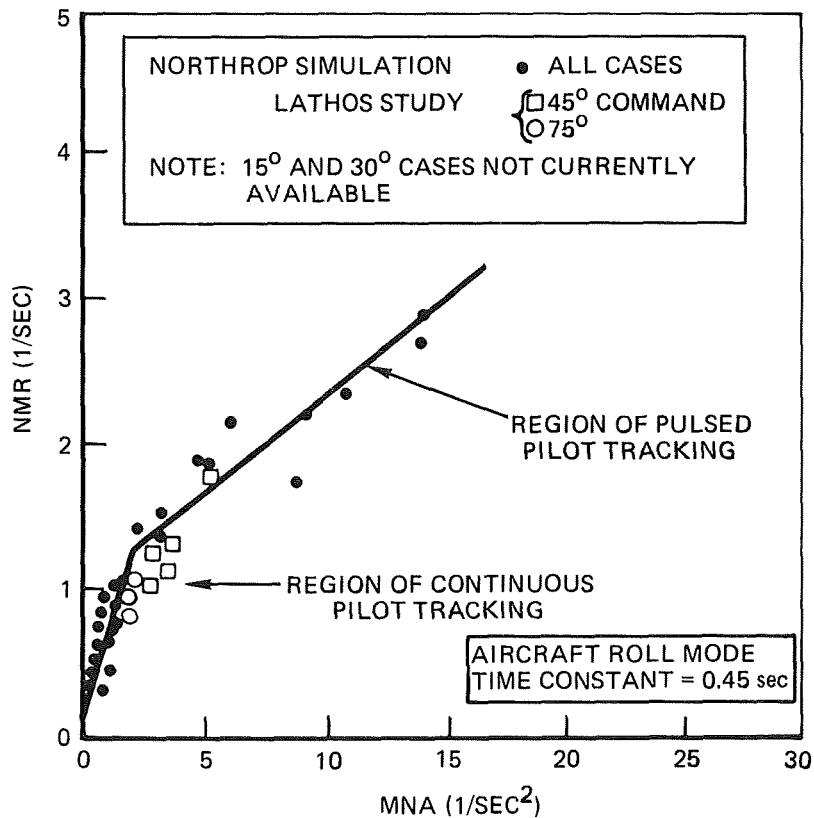

Figure 6. Correspondence Between LATHOS and Northrop Simulation Data for Roll Mode Time Constant of .45 seconds.
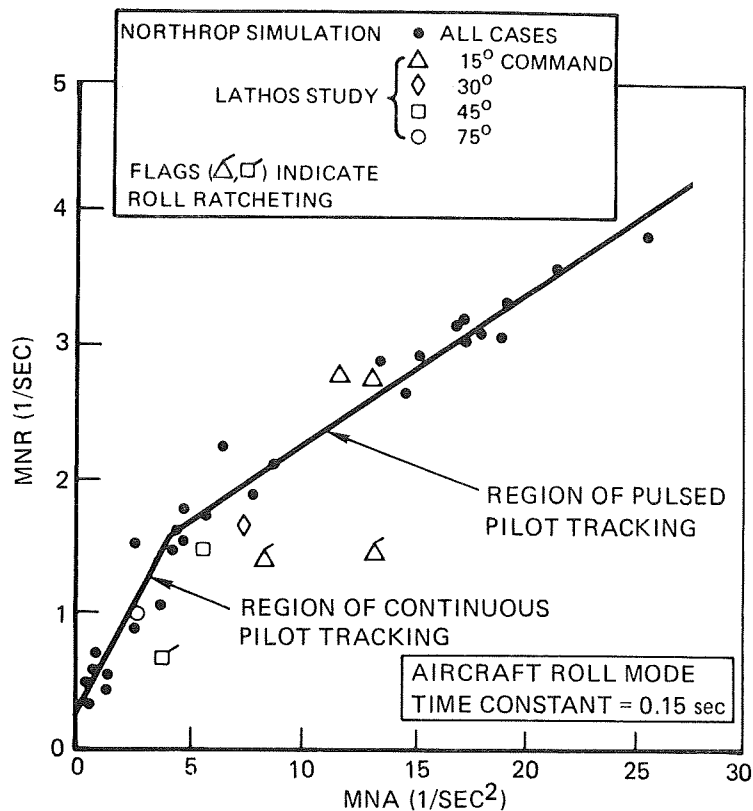
Figure 7.   Correspondence Between LATHOS and Northrop Simulation
Data for Roll Mode Time Constant of .15 seconds.

On these figures, the solid bent line approximates the Northrop
flight simulation data. For each configuration, the step target
method predicts where this abrupt change of slope occurs, in
terms of MNR generated by the single stage model. For this
reason, it appears that for lower MNR values, the pilot has
adopted a continuous compensatory tracking strategy, while the
higher MNR cases represent pulsed piloting techniques.

Figure 7 contains three points where LATHOS pilots experienced
undesirable oscillations, called ratcheting. These points fall
well to the right of the remainder of the data, indicating that,
for this experiment, ratcheting is characterized by considerably
higher values of MNA than the resulting MNR warrants.

Unfortunately, there are too few time histories currently
available from the LATHOS study to allow validation of the above
results. Even so, the following observations seem to be
justified:

1) MNR versus MNA profiles indicate the presence of both
continuous and pulsed control strategies.

2) MNR is a suitable metric for pilot opinion in the LATHOS data
base, while the MNR/MNA relationship appears to be sensitive
to PIO and roll racheting problems.

3) The relative distributions of MNR/MNA data for LATHOS step commands indicates that during in-flight roll maneuvering, pilots tend to limit lateral acceleration at the pilot station.

Comparison of Discrete and Continuous Tracking Tasks

Another simulation study was performed at Northrop, in order to compare continuous closed-loop tracking with step target tracking. The experiment, which was intended to refine the test matrix for an in-flight simulation involving the NASA Digital-Fly-By-Wire F-8, involved testing a number of lateral dynamics configurations, using both discrete and continuous tracking tasks [8,9].

Again, the FCRS simulator was utilized, and the step bank angle command task was used to provide a discrete compensatory task. The continuously varying bank angle command signal was formed from a sum-of-sines equation. The equation contained ten frequency terms, arranged to have an overall period of 50 seconds. In addition, the signs of the relative amplitudes were randomized for each run, in order to minimize pilot familiarity and task learning effects. The absolute magnitude of the sum-of-sines equation was scaled to be plus/minus one radian. The frequency and amplitude characteristics used are shown in Figure 8.

INDEX	FREQUENCY $\left(\frac{RAD}{SEC}\right)$	RELATIVE AMPLITUDE
1	0.251	1.00
2	0.377	1.00
3	0.628	1.00
4	0.880	1.00
5	1.382	1.00
6	2.136	0.50
7	3.267	0.10
8	5.152	0.10
9	7.791	0.10
10	12.189	0.10

Figure 8. Frequency and Amplitude Data for the Continuous Tracking Task.

The experiment test matrix was composed of three pure time delays (0.100 , 0.175 , and 0.250 seconds) versus six roll mode time constants (TR) (0.2 , 0.3 , 0.4 , 0.6 , 0.8 , and 1.0 seconds).

Figure 9 shows the results of the CONTINUOUS tracking experiment. Points corresponding to each of the three values of delay are connected by straight lines. The apparent effect of delay is to displace the tracking data, for each roll mode time constant, toward greater values of tracking error normalized by the RMS of

the tracking command. This figure also illustrates that an increased roll mode time constant will result in a decreased tracking capability. An exception to this trend occurs in the very highly damped case of TR = 0.2, where rate perception effects are encountered in the simulation.
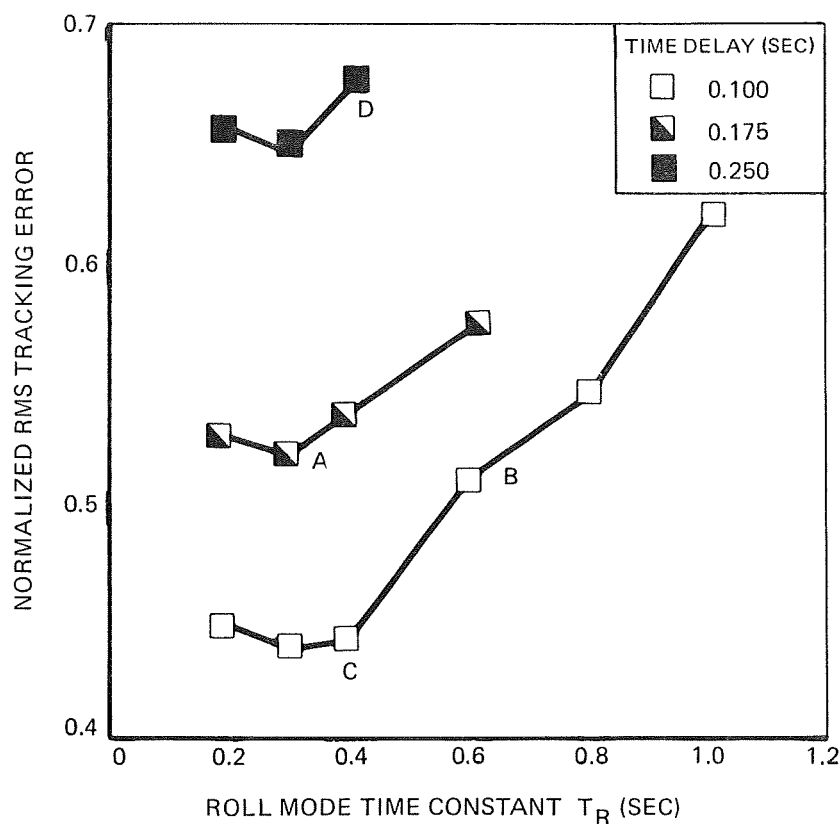

Figure 9.   Summary of Continuous Tracking Averages, Showing Effects of Time Delay and Roll Mode Time Constant

The same test matrix was used in the DISCRETE tracking experiment. Figure 10 presents the profiles of MNR versus MNA produced by tracking the discrete Step commands. Clearly, there is a trend for higher values of MNR to be associated with higher values of MNA. The greater values of time delay lead tend to result in lower values of both MNR and MNA for a given value of TR.
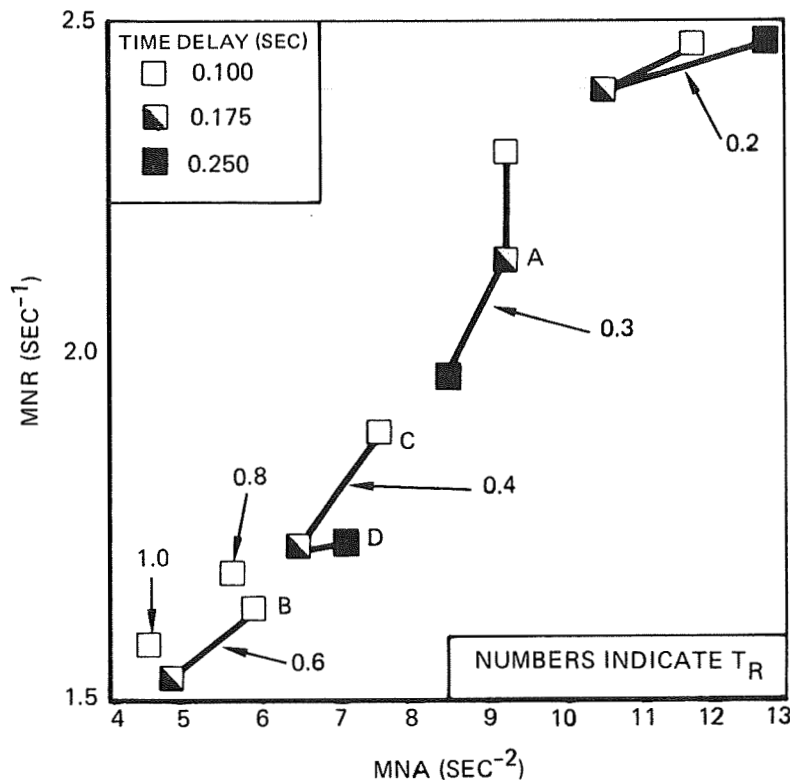
Figure 10. MNR vs. MNA for Discrete Tasks, Showing
Effects of Time Delay and Roll Mode
Time Constant.

Figures 9 and 10 both show smooth variations in plotted
parameters, with respect to the corresponding aircraft dynamics.
However, it should be noted that the associated sensitivities do
not necessarily correspond. This can be observed through
comparison of the two points labeled 'A' and 'B' on both figures.
In Figure 9, these points are associated with roughly the same
RMS tracking errors, while on Figure 10, 'A' and 'B' are greatly
separated in both MNR and MNA parameters. Thus, as the previous
experiment revealed a correlation between MNR and Pilot Opinion
Ratings, one would have anticipated that 'A' and 'B' would
receive quite different POR's, even though they exhibit nearly
identical RMS tracking error scores in the continuous tracking
task. Conversely, the points labeled 'C' and 'D' appear quite
dissimilar in terms of RMS tracking error, as shown in Figure
9, while the same two points are close together in terms of MNR
and MNA, as shown in Figure 10.

SUMMARY OF ANALYSIS

The two parameters MNR and MNA have been shown to be useful in
Flying Qualities analysis. MNR was shown to correlate with Pilot
Opinion Rating in the LATHOS data base, while MNA reflects PIO
and roll ratcheting. Profiles of MNR versus MNA reveal the
presence of pulsed compensation strategies in both ground based
and in-flight simulation. Furthermore, comparison of continuous
and discrete attitude tracking simulation data reveals that these

two tracking tasks exhibit independent sensitivities to aircraft characteristics.

REFERENCES

1) Onstott, E.D., Warner, J.S., and Hodgkinson, J., Maximum Normalized Rate as a Flying Qualities Parameter, Twentieth Annual Conference on Manual Control, NASA-CP-2341, NASA Ames Research Center, Moffett Field, California, June 1984.

2) Chalk, C.R., Excessive Roll Damping Can Cause Roll Rachet, Paper 82-1606, AIAA Guidance and Control Conference, San Diego, California, August 9-11, 1982.

3) Onstott, E.D., and Faulkner, W.H., Discrete Maneuver Pilot Models for Flying Qualities Evaluation, Journal of Guidance and Control Volume 1, Number 2, March-April 1978.

4) Neal, T.P., and Smith, R.E., An In-Flight Investigation to Develop Control System Design Criteria for Fighter Airplanes, AFFDL-TR-70-74, Air Force Flight Dynamics Laboratory, Wright-Patterson Air Force Base, Ohio, February 1971.

5) Moorhouse, D.J., and Woodcock, R.J., Background Information and User Guide for MIL-F-8785C, Military Specification - Flying Qualities of Piloted Airplanes, AFWAL-TR-81-3109, Air Force Flight Dynamics Laboratory, Wright-Patterson Air Force Base, Ohio, June 1982.

6) Monagan, S.J., Smith, R.E., and Bailey, R.E., Lateral Flying Qualities of Highly Augmented Fighter Aircraft, AFWAL-TR-81-3171, Air Force Flight Dynamics Laboratory, Wright-Patterson Air Force Base, Ohio, June 1982.

7) Warner, J.S., and Onstott, E.D., A Description of the Northrop Flight Controls Research Simulator and Step Target Software, NOR-84-135, Northrop Corporation Aircraft Division, August 1984.

8) Onstott, E.D., and Warner, J.S., Ground-Based Flight Simulation of the Proposed F-8 Continuous Task Test Matrix, NOR-84-148, Northrop Corporation Aircraft Division, September 1984.

9) Onstott, E.D., and Warner, J.S., Ground-Based Flight Simulation of the Proposed F-8 Discrete Task Test Matrix, NOR-84-149, Northrop Corporation Aircraft Division, September 1984.

CLOSED-LOOP, PILOT/VEHICLE ANALYSIS
OF THE APPROACH AND LANDING TASK

by

Mark R. Anderson* and David K. Schmidt†

School of Aeronautics and Astronautics
Purdue University
West Lafayette, IN 47906

Extended Abstract

Recently, Bacon and Schmidt[1] presented an integrated optimal-control, frequency-domain approach for pilot/vehicle analysis of the precision attitude control task. When applied to the flight test results of Neal and Smith[2], the optimal control approach was shown, not only to agree extremely well with the original technique developed by Neal and Smith, but also to yield additional information on the achieveable closed-loop bandwidth in the task. This task was essentially modeled as a single-input, single-output, closed-loop task.

In the case of approach and landing, however, it is universally accepted that the pilot uses more than one vehicle response, or output, to close his control loops. Therefore, to model this task, a multi-loop analysis technique is required. The analysis problem has been in obtaining reasonable analytic estimates of the describing functions representing the pilot's loop compensation. Once these pilot describing functions are obtained, appropriate performance and workload metrics must then be developed for the landing task.

The optimal control approach[1,3] provides a powerful technique for obtaining the necessary describing functions, once the appropriate task objective is defined in terms of a quadratic objective function. In this discussion, we will present such an approach through the use of a simple,

---

* Graduate Student.
† Professor, Associate Fellow, AIAA.

reasonable objective function and model-based metrics to evaluate loop performance and pilot workload. We will also present the results of an analysis of the LAHOS (Landing and Approach of Higher Order Systems) study performed by R.E. Smith[4].

In flare or near touchdown, precision flight-path control is required. Assuming a "frontside" landing technique is used, the pilot can control flight path or sink rate through elevator commands. Including inner pitch-attitude and flight-path-angle feedback loops, this situation leads to a block diagram of the approach and landing task shown in Figure (1). A reasonable task objective function would then reflect the pilot's desire to minimize flight-path error, γ_{error}, by using pitch-attitude, flight-path, and flight-path-error information in the following form,

$$J_p(u_p) = E\{\lim_{T\to\infty} \frac{1}{T}\int_0^T (q\gamma_{error}^2 + g\dot{u}_p^2)dt\} \qquad (1)$$

where u_p is the pilot's stick force input.

The pilot describing functions, $P_{(\cdot)}$, shown in the closed-loop structure of Figure (1) can then be obtained using the optimal-control approach. These describing functions represent those required to achieve the best loop performance, subject to the task definition and inherent pilot limitations modeled. Once determined, they can also be manipulated using block diagram algebra to obtain, for example, an equivalent unity feedback single-loop structure shown in Figure (2).

Neal and Smith, as well as Bacon and Schmidt, described the pilot/vehicle handling-quality criteria problem as a trade-off between the pilot workload required to achieve acceptable task performance and a subsequent measure of the pilot/vehicle closed-loop performance. The most important aspect of closed-loop performance, furthermore, is stability and robustness (or insensitivity to small changes in pilot compensation). These loop characteristics are clearly reflected in the open-loop, γ/γ_{error}, frequency response. In fact, for good closed-loop stabilty properties, the desirable "shape" of this frequency response in the crossover region is well known (i.e. constant -20 dB/decade slope). Any deviation from the desirable frequency response is defined herein as a reduction in *loop quality*.

A model-based measure of the "loop quality" has been developed and is entitled the "open loop peak", obtainable from the open-loop frequency response plots after the pilot/vehicle system has been modeled. Also a model-
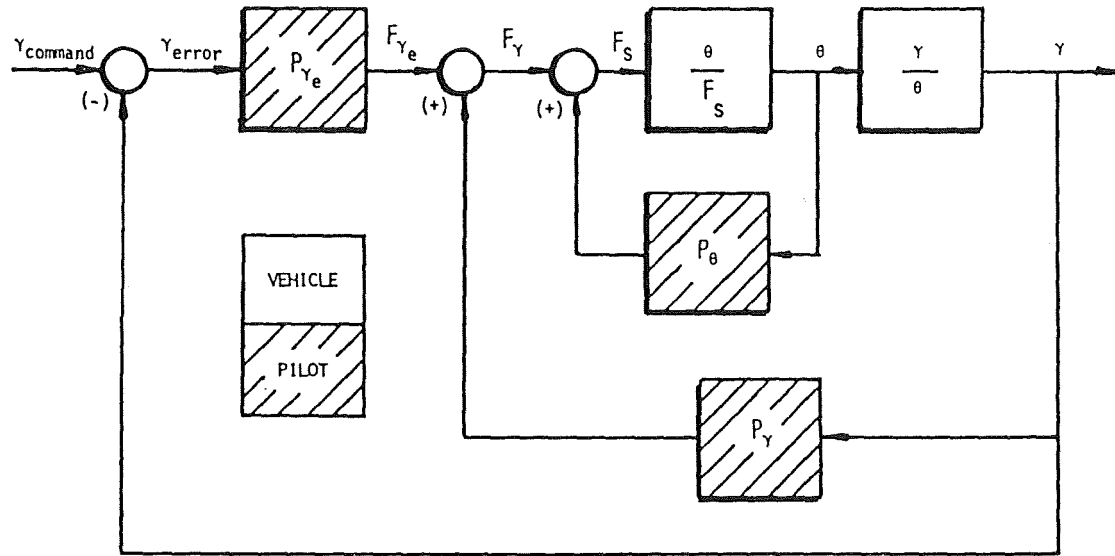
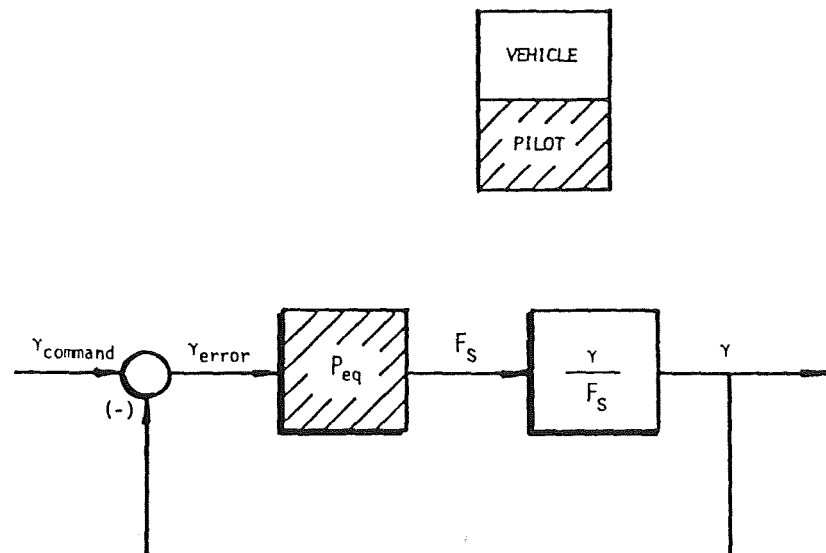Figure 1 The Multi-Loop Flight Path Tracking Task

Figure 2 Flight Path Tracking with
Equivalent Pilot Function

based metric has been identified that reflects the pilot workload necessary to achieve closed-loop stability. This workload metric is expressed in terms of a pilot phase compensation angle.

When thirty-two of the aircraft configurations flight tested in the LAHOS study were modeled and analyzed, the results are as shown in Figure (3). Recalling that the " open-loop peak" is a measure of stability robustness, and the "pilot compensation" is a measure of workload, we see a characteristic grouping of the results not unlike that presented in References [1] and [2]. However, in these references, the task modeled was precision attitude control, and two different (though similar) model-based metrics were used in the related plots.

It is also noted from Figure (3), that those configurations rated best (Cooper-Harper Level 1) in the *approach and landling task* were appropriately grouped together, in terms of "performance" and "workload". Those rated worse were the result of excessive pilot phase lead or lag conpensation required or a reduction in "loop quality". Other results concerning loop characteristics such as achieveable loop bandwidths, pilot comments, and pilot behavior can be found in Reference [5].

References

[1]   Bacon, B.J. and Schmidt, D.K., "An Optimal Control Approach to Pilot/Vehicle Analysis and the Neal-Smith Criteria", *Journal of Guidance, Control, and Dynamics*, Vol. 6, No. 5, Sept-Oct., 1983, pp. 339-347.

[2]   Neal, T.P. and Smith, R.E., *An In-Flight Investigation to Develop Control System Design Criteria for Fighter Airplanes*. AFFDL-TR-70-74, Vol. I, December, 1970.

[3]   Kleinman, D.L., Baron, S., and Levison, W.H., "An Optimal Control Model of Human Response Part I: Theory and Validation", *Automatica*, Vol. 6, 1970, pp. 357-369.

[4]   Smith, R.E., *Effects of Control System Dynamics on Fighter Approach and Landing Longitudinal Flying Qualities*, AFFDL-TR-78-122, Vol. 1, March, 1978.

[5]   Anderson, M.R. and Schmidt, D.K., "Closed-Loop, Pilot/Vehicle Analysis of the Approach and Landing Task",AIAA Paper 85-1851-CP, AIAA Guidance and Control Conference, Snowmass, CO, 1985.
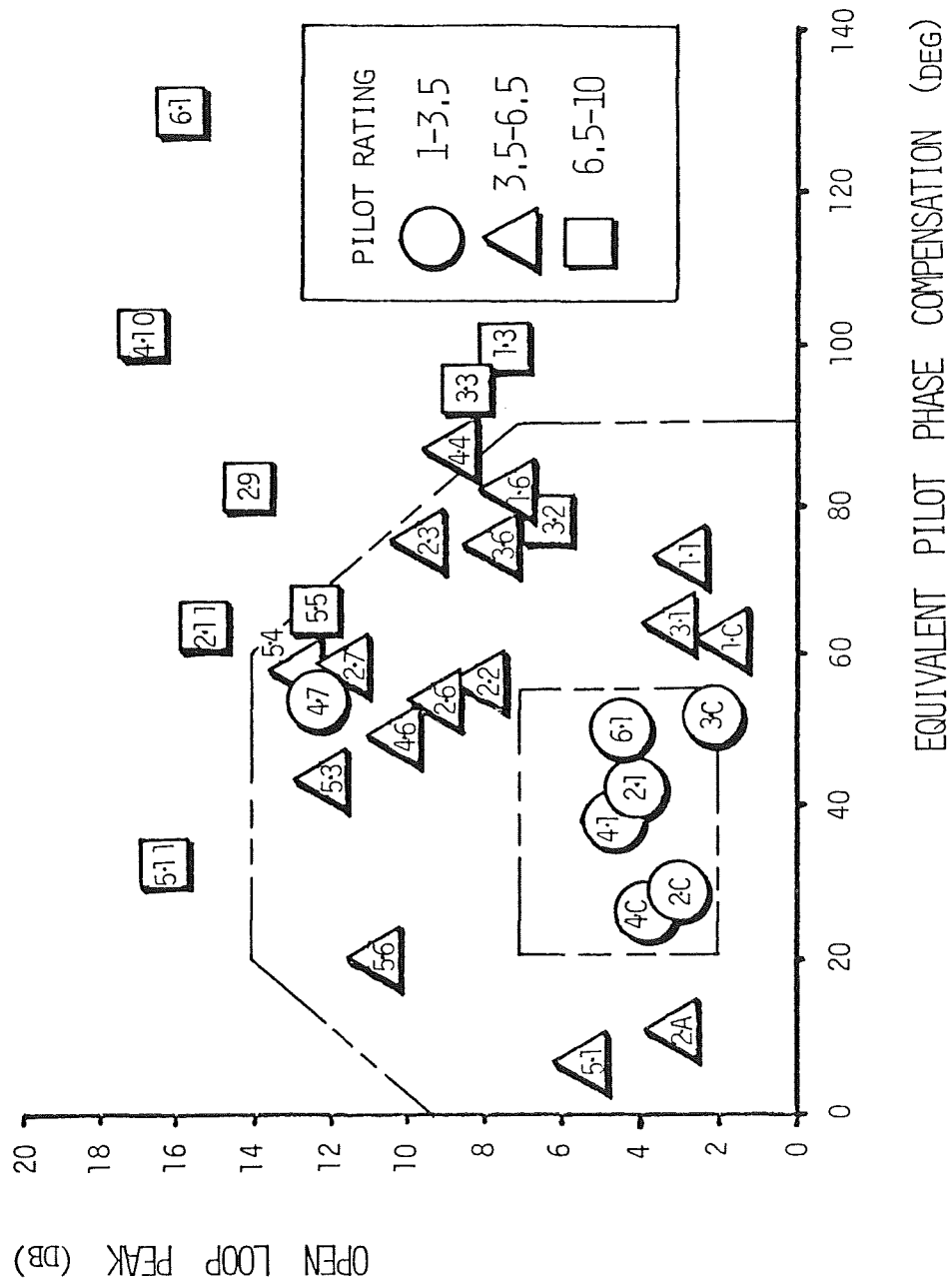
Figure 3 OCM Results for the Flight Path Tracking Task

1. Report No. NASA CP 2428	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle TWENTY-FIRST ANNUAL CONFERENCE ON MANUAL CONTROL JUNE 17-19, 1985		5. Report Date May 1986
		6. Performing Organization Code
7. Author(s) Compiled by Richard A. Miller and Richard J. Jagacinski		8. Performing Organization Report No. 86218
		10. Work Unit No.
9. Performing Organization Name and Address Ames Research Center,. Moffett Field, CA 94035 Ohio State University, Columbus, OH 43210		11. Contract or Grant No.
		13. Type of Report and Period Covered Conference Publication
12. Sponsoring Agency Name and Address National Aeronautics and Space Administration Washington, DC, 20546		14. Sponsoring Agency Code 505-35-11

15. Supplementary Notes

Point of contact: R. A. Miller or R. J. Jagacinski, Ohio State University, Columbus, OH 43210; (614) 422-7067 or (614) 422-4131

16. Abstract

This volume contains the proceedings of the Twenty-First Annual Conference on Manual Control, held in Columbus, Ohio, June 17-19, 1985. It contains 29 manuscripts and 8 abstracts pertaining to Workload, Attention and Errors, Controller Evaluation, Movement Skills, Coordination and Decision Making, Display Evaluation and Human Operator Modelling and Manual Control.

17. Key Words (Suggested by Author(s)) Human-machine interaction Human modelling Manual control Decision making	Displays Workload Simulation Attention	18. Distribution Statement Unlimited Subject Category: 54		
19. Security Classif. (of this report) Uncl.	20. Security Classif. (of this page) Uncl.	21. No. of Pages 527	22. Price* A23

*For sale by the National Technical Information Service, Springfield, Virginia 22161