Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

AD A059477

DDC FILE COPY

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: AFOSR-TR- 78-1239	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle): THE DEVELOPMENT AND EVALUATION OF OBJECTIVE FREQUENCY DOMAIN BASED PILOT PERFORMANCE MEASURES IN ASUPT		5. TYPE OF REPORT & PERIOD COVERED: Final Rept. March 1977 to Feb 1978
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Edward D. McDowell		8. CONTRACT OR GRANT NUMBER(s): AFOSR-77-3294
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Engineering Experiment Station, Oregon State University, Corvallis, Oregon 97331		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 61102F 2313D9 D9
11. CONTROLLING OFFICE NAME AND ADDRESS: Air Force Office of Scientific Research (NL), Bollings, AFB, D. C. 20332		12. REPORT DATE: 30 April 1978
		13. NUMBER OF PAGES: 40
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report): Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

LEVEL

New 410874

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

41 p.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

pilot performance measurement
proficiency measurement
objective measurement
frequency domain measurement

20. ABSTRACT (Continue on reverse side if necessary and identify by block number) To realize its full research potential a need exists for the development of an automated objective pilot performance evaluation system for use in the Advanced Simulator for Undergraduate Pilot Training (ASUPT) facility. The present report documents the approach taken for the development of frequency domain pilot control movement based measures and presents data collected from an evaluation study. The results indicate that frequency domain based measures of a pilot's control movements do discriminate between pilot experience levels. These finds are encouraging and demonstrate the potential of this approach for developing an automated objective pilot performance measurement system.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

410 874

AFOSR-TR- 78 - 1239

THE DEVELOPMENT AND EVALUATION OF OBJECTIVE FREQUENCY DOMAIN BASED PILOT PERFORMANCE MEASURES IN ASUPT

Grant No.

AFOSR 77-3294

Edward D. McDowell
Department of Industrial
& General Engineering
Oregon State University
Corvallis, Oregon 97331

April 1978

Prepared for:

Air Force Office of Scientific Research
Air Force Systems Command
USAF, Bollings AFB, D.C. 20332

78 09 05 053

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

The author expresses his appreciation to the personnel of the Human Resources Laboratory Flying Training Division and the 82nd Flying Training Wing, Williams Air Force Base, Chandler, AZ. The success of this study was due, in large measure, to their outstanding efforts.

The author is especially grateful for the assistance of Dr. Elizabeth L. Martin and Mr. Gary Reed for the extensive assistance throughout the project.

He is also indebted to Messrs. T. R. Templeton, and R. D. Greatorex, Human Resources Laboratory and L. Thompson, Systems Engineering Laboratory for developing the computer software necessary to conduct the study and analyze the results.

His appreciation is also extended to Captain L. Lesher 82nd Flying Training Wing for his extensive assistance throughout the course of the study.

The author is especially grateful to Mr. M. L. Cyrus for developing the Butterworth Filters and whose invaluable skill and effort made this research possible.

I. BACKGROUND

Introduction

Since the Arab oil boycott, the Air Force has significantly reduced its fuel consumption from 155 million barrels in 1973 to an estimated 107 million barrels in 1976. This reduction has been achieved primarily by a concommitant reduction in flying time.

One potential area of opportunity for obtaining additional fuel saving is through the extended use of flight simulators in flying training. If simulator time can be substituted for flying time, both during the initial training and later during maintenance training phases, then significant fuel saving might be achieved. The greater use of flying simulators in training is, however, contingent upon the development of advanced simulation techniques.

Similar saving might be realized by devising more fuel efficient flying training programs. This could be accomplished, for example, by substituting flying time in more fuel efficient aircraft for time in aircraft with high fuel consumption rates. Of course, simple reductions in flying time during initial training could also produce significant fuel savings. These savings might be realized by developing flying training programs that incorporate fixed criteria for completion instead of the fixed instruction time concept presently used, or by using adaptive or self-paced instruction programs. These types of instructional programs are dependent upon the availability of adequate student performance measures.

Performance Measurement in Flying Training

Performance measures serve two major functions in training. First, they provided the instructor, be it man or machine, with the information necessary for the guidance of the student training program. Typical examples of this use of performance measures are provided by Ping and Gill (1976) and Brown, Waag, and Eddowes (1975), where performance measures are used in an automated adaptive training system for training pilots the proper execution of ground controlled approaches (GCA's). The performance measures in these cases are used to control the difficulty level of the next training block as a function of the student's past performance.

Performance measures also serve as a source of feedback for the student, thus indicating the areas in which his performance is inadequate, and often suggesting the type of adjustments the student must make in order to improve.

Another important use of performance measurements is in research on and the evaluation of training innovations. Generally for a training innovation to be deemed effective it must either result in higher levels of trainee performance for the same training cost or a reduction in the training cost necessary for an average student to reach a criterian performance level or both. The absence of adequate performance measures has often lead to uncertainty and conflict between authorities regarding the desirability of certain instructional features.

Desirable Qualities for Performance Measures

In the development of performance measurement systems, it is necessary to recognize that performance measures should have certain qualities. The relative importance of these qualities are, however, a function of the intended use of the measurement system. For example, temporal invariance may be especially important in training innovations evaluation, but of lesser importance when the performance measures are used as a source of feedback in actual training programs. This list of desirable performance measurement qualities is neither meant to be exclusive or exhaustive, but is offered only as a sample of several qualities performance measures should possess.

i) Temporal Invariance or Repeatability: This simply suggests that a specified score today represents the same level of performance as it did at a previous time. Objective performance measures on absolute scales generally possess this quality.

One example of a performance measure with high temporal invariance is typing speed. A rate of say 55 words per minute today represents about the same level of performance today as it did five years ago. A counter example would be student gpa's. Because of "grade inflation" it is quite likely that a 3.00 gps today does not represent the same level of achievement as it did several years ago. Because of this temporal variance it is difficult to compare the performance of today's student with that of his predecessor.

ii) Criteria Based: Although they are not always explicitly stated, most tasks involving training have specific criteria that must be met for their satisfactory performance. Performance measures should reflect the degree to which these criteria are met by the trainee's performance. Moreover, improvements in performance due to practice should be reflected by the performance measures.

iii) Interpretability: Performance measures, especially those used in actual training as opposed to those used exclusively for evaluation, should be easily interpreted by both the instructor and student. If they are not, they they cannot be used to show the student how to improve his performance as well as reflect his level of improvement.

iv) Immediately Available: Additionally, if performance measures are to serve as an effective source of feedback, they they must be available to the student as soon after completing a performance trial as possible.

v) Peer Group Performance: Performance measures should reflect the actual level of performance achieved and not be influenced by the class performance. That is, the use of norm reference measures should be avoided. Obviously class rank is one popular performance measure that does not possess this quality. Also, when instructors are asked to subjectively rate students, their ratings may be influenced by the general level of performance of the group, thus are not necessarily insensitive to peer group performance.

vi) Instrument Invariance: The derived measure of performance should not be a function of the particular instrument used. That is, if written examinations are used, the score a student receives should not be a function of the particular examination given. In flying training, the students score should not be a function of the particular instructor pilot rating him. This is a common threat to internal validity (see Goldstein (1974)).

vii) Transfer Effectiveness: When performance measures are used in programs where the student trains on a task other than the actual task, it is important that the performance measure reflect his initial performance on the actual task as well as his performance level on the training task. With respect to flying training this simply implies we are interested in training pilots to fly airplaces not simulators, and performance in the airplane is the important criteria.

viii) Task Related: Because of several of the qualities cited above it is desirable, but not necessary, that performance measures be applicable to the task as well as the training situation. Therefore, it would be desirable to have performance measures that could be used in the aircraft as well as the simulator.

Approaches to Performance Measurement

Performance measures can in general be classified as either subjective or objective in nature. Subjective performance measures usually call for an "expert" to judge the performance level demonstrated by a trainee through an introspective process. Conversely, objective performance measures measure directly how well the trainee is performing on an objective physical scale, such as the words per minute measure of typing speed cited above.

Subjective performance rating tends to have several decided advantages over objective systems. First, it is relatively easy to develop subjective performance measurement systems for many complex tasks such as flying or driving, since you need only hire an "expert" and allow him to observe trainee performance. Also, since the rater is usually an acknowledged expert at the task, his rating has a high face validity. More important, however, is the specific feedback usually provided by this type of a performance measurement system. The rater, at least in training situations, provides the trainee with very specific feedback with respect to how his performance may be improved, feedback of a type usually not found in objective performance measurement systems.

Unfortunately, subjective performance measures have several distinct disadvantages, especially when used within research paradigms. Although Waag et al. (1975) found high correlations between raters, this is not generally the case. Knoop (1973) reports two instructor pilots (IP's) subjective ratings were each correlated with certain objective performance measures, but that nowhere were the same objective measures involved within the significant correlations. Even more significantly, Knoop and Welde* (1975) found lack of agreement between pilots on the specific criteria for successful performance of certain aerial maneuvers.

Additionally, the individual rater's concept of what level of performance a specific rating is indicative of usually varies significantly from rater to rater. Even an individual rater's concept of performance levels will change over time, and most probably is influenced by the group performance level, societal factors and others as mentioned previously.

For these reasons, subjective performance rating may be satisfactory for training purposes, but not for research on training innovations. Both the inter and intra rater variations add considerable noise to the experimental results, thus obscuring many significant effects, or requiring the use of large sample sizes.

Conversely, objective performance measures are not subject to these sources of error, making them more desirable for research purposes. However, for complex tasks, the performance is not always easily quantifiable.

However, it would seem highly desirable to have objective performance measures available for use in flying training research involving flight simulators. Because of its high operating cost it would appear particularly desirable to have such measures available for the USAF Human Resources Laboratories Achieved Simulator for Pilot Training (ASPT). The availability of such performance measures would enhance the capabilities of this device for research into training innovations. The increased sensitivity of objective performance measures would allow more efficient experiments to be performed, thus allowing smaller sample sizes. One basis for developing such measures would be the "successive organization of perception" hypothesis.

The "Successive Organization of Perception" hypothesis as proposed by Magdaleno, Jex and Johnson (1969) suggests in performing control tasks the human operator first acts as a simple error controller. As he learns the task he adds operations on the system as a simple error controller. As he learns the task he adds operations on the system input and finally adds a preprogrammed pattern to his control responses. Thus the operator moves from essentially a compensatory to primarily a pursuit and then to a pre-cognitive control mode. Such changes in the operators behavior should be reflected by a shift in his control input power spectra toward lower frequencies.

If this hypothesis may be shown to hold for flying, then objective frequency domain based performance measures sensitive to this shift could be developed. These performance measures might provide a parsimonious non-maneuver specific pilot performance measurement system.

Research Objectives

The first objective of the research was to determine if the pilots control movement relative power spectra would shift toward higher frequencies with increasing pilot experience as the successive organization of perception hypothesis suggests. A second and closely allied objective was to identify the pilot's control movement measures, including measures sensitive to shifts in the relative power spectra, which vary significantly as a function of pilot experience level. If it is assumed that pilot skill and experience closely co-vary, then these measures would become candidate measures for an automated performance measurement system.

Estimating the possible discriminating power of these variables, that is the degree to which they could be used to identify pilot experience (skill) level then became a research objective.

An ancillary objective of the research program was to determine if the standard embeeded figures test (EFT) scores were related to pilot skill. If it was as hypothesized, then EFT scores could be used to match subjects for future research involving flight simulation.

An embedded figures test measures an individuals ability to disembed a figure from a complex background. The ability to perform such a task is hypothesed to be closely related to a novice pilot's ability to disembed, that is separate, relevant information from his complex visual field and to a lesser extent to his ability to perceptually organize a task. Since low EFT scores imply field independence we would expect novice pilots EFT scores to be negatively correlated with their initial performance. As the pilot learns to fly, he becomes more organized perceptually in the sense that he learns to efficiently extract visual information. Thus we would expect his field dependence-independence to play a diminishing role as experience increases.

II. EXPERIMENTAL METHODOLOGY

General Approach

The general research objective was to determine if a shift in the pilot's control movement power spectra can be used as the basis for discriminating between pilot skill levels and to identify those functions of a pilot's control movements which show the greatest potential for determining his proficiency level. For the purpose of this research it is assumed that a pilot's proficiency level is synonymous with his experience. Thus performance measurements were taken on selected undergraduate pilot training (UPT) maneuvers in ASPT for each of three subject populations ranging from naive to highly experienced. Only the pilot experience level was systematically manipulated as an independent factor. The three levels of this factor included in the design are pre-flight T-37 student pilots as the novice group, post T-37 pre-T-38 students as the intermediate group and T-37 Instructor Pilots (IP1) as the experienced group.

Maneuvers: Considering the nature of the subject populations, flight maneuvers specific to the UPT-T-37 syllabus were required, since presumably the novice and intermediate groups would differ only on the maneuvers the intermediate groups had an opportunity to practice. Moreover, the maneuvers the novice group could not be expected to complete were excluded from the maneuver set. Finally, since the type of maneuver selected was expected to significantly interact with the discriminating power of the performance measures, the maneuvers selected should be representive of the UPT program. Using these criteria the maneuvers selected were straight and level flight (approximately 2 min.), turn to heading (90° turn 30° bank), vertical S Delta, and formation flight. It is believed these maneuvers adequately span the range of difficulty and complexity encountered in UPT training. In addition they involve varying degrees of aileron, elevator, and throttle involvement.

Subject Selection and Screening

Three subject populations were used within the research paradigm, each containing ten subjects. The first or novice group consists of ten pre-flight T-37 students. Foreign students or students with over fifty hours of previous flying time were excluded from this group.

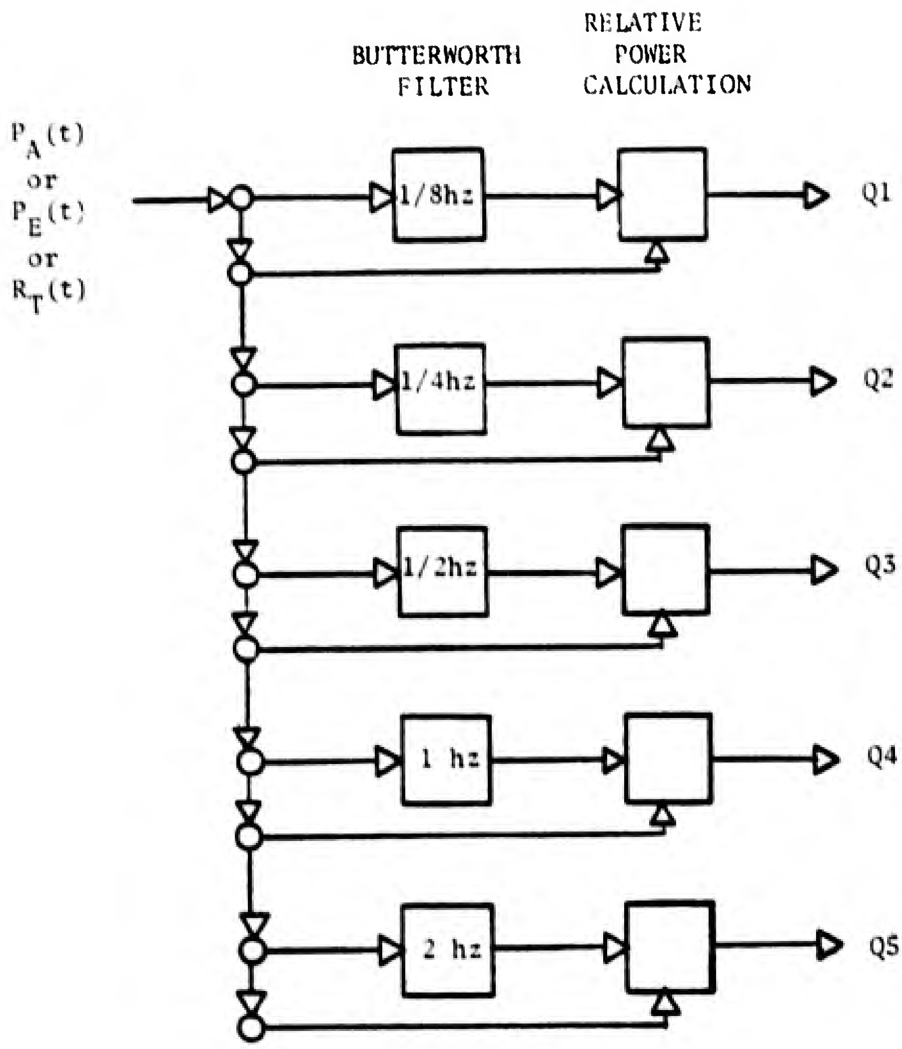
The intermediate group consists of ten student pilots who had completed T-37 training but had not yet begun T-38 training, while the experienced group consists of ten T-37 instructor pilots. Foreign students were not used in either of these groups.

Performance Measurement Systems

All performance measures used in the research with the exception ASPT automated performance measures (APM) total score, were based upon the pilots control inputs, specifically the aileron, elevator and throttle positions. Each of these three variables were sampled fifteen times per second. The minimum, the maximum, and the first four moments, V_1 - V_4, were then calculated for each of the discrete sample records, $P_A(t)$, $P_E(t)$, and $P_T(t)$. In addition, each series was passed through five secussive low pass butterworth filters of the form

$$x_i(t+1) = \alpha_i X_i(t) + \beta_i*(P(t+1) + P(t))$$

were the coefficients α_i and β_i were selected to give approximate cutoff frequencies of 1/8, 1/4, 1/2, 1 and 2 Hz respectively as outlined in Figure 1. The power in each filter output, $x_i(t)$, was then calculated and divided by the power of the original series, P(t). The resulting quantities, Q1-Q5, then provided estimates of the fraction of the original series power below each of the cut-off frequencies. This provided a parsimonious real time method for approximating the input series relative power spectra. The five values Q1-Q5 should



FILTER SYSTEM

Figure 1

be sensitive to shifts in the power spectra. In addition a modified S-omega (MSO) was calculated. This is a single variable which is indicative of the distribution of the power spectra. The exact equation for calculating MSO is given in Appendix A along with the other equations used and the filter constants. The minimum, the maximum, the four moments, the five filter values and modified S-omega resulted in twelve control measures for each control input or a total of thirty-six measures in all. Five more variables were used to indicate group membership, subject, maneuver, and trial number for a total of forty-one variables.

The ASPT APM total score was essentially the percent of the total time the pilot maintains his aircraft within prescribed tolerance limits for each of several state variables simultaneously. The specific variables and associated tolerance limit are included in Appendix A.

Experimental Procedure

The research plan called for each of the thirty subjects, ten in each group, to "fly" one one hour sortie in ASPT. During each sortie each subject performed two APM's for each of the four maneuvers selected. The maneuvers were performed in order of ascending difficulty and this order was fixed for all subjects.

The sortie content was identical for groups two and three but differed for group 1. For groups two and three a sortie consisted of approximately 5 minutes of free flight warm up followed by two consecutive APMs on straight and level, two on turn to heading, two on vertical S "D", and two on formation flight. On these sorties the subjects "flew" alone.

The novice subjects were always accompanied by an instructor pilot (IP). Each sortie began with a preprogrammed demonstration of straight and level flight, a practice trial and then, two APM's. The IP provided instruction and guidance to the student pilot during the demonstration and practice trial but not during the APM's. This procedure was repeated for turns to heading and the vertical S "D". Because of technical difficulties the formation flight demonstration was flown "line" by the I.P. The sortie contents are outlined in Appendix A and the general design in Figure 2.

GROUP	SUBJECT	STRAIGHT AND LEVEL TRIAL 1	STRAIGHT AND LEVEL TRIAL 2	TURN TO HEADING TRIAL 1	TURN TO HEADING TRIAL 2	VERTICAL S DELTA TRIAL 1	VERTICAL S DELTA TRIAL 2	FORMATION FLIGHT TRIAL 1	FORMATION FLIGHT TRIAL 2
NOVICE	S1 S10								
INTERMEDIATE	S11 S20								
EXPERIENCED	S21 S30								

EXPERIMENTAL DESIGN

Figure 2

III. RESULTS

Analysis Approach

The analysis of the results was performed in two stages. In Stage I the relationship between the EFT scores and flying skill as measured by the ASPT APM total score was examined. In the second stage the effect of pilot skill level on the control movement measures, including the digital filter outputs, and the ability of these measures to discriminate pilot experience level was examined.

EFT Scores

Due to a programming error, the digital filter data for the first five subjects in Group 1 was not valid. This necessitated running five additional subjects. A sixth was then added in order to balance another experiment resulting in a total of sixteen subjects for group one. Although the digital filter data for five of these subjects was not valid, their EFT scores and APM total scores were accurate. Therefore all sixteen subjects in group one were included in the preliminary analysis.

Initially the correlation between the EFT scores and the APM total scores were examined. Since the total scores for formation flight were essentially all zero for all experience levels, this maneuver was omitted from the analysis.

The EFT scores correlation with the total score for each trial, the total of the two trials for each maneuver, and the total for all six trials was then estimated, with the results shown in Table I.

The results indicate a high negative correlation between EFT scores and pilot skill, as measured by the APM total score, for the novice group. All the correlation coefficients were negative, and as the maneuver difficulty increased the magnitude of these correlations increased, with the coefficients

TABLE I. EFT APM TOTAL SCORE CORRELATION RESULTS

GROUP (Number) (Mean EFT)	STRAIGHT AND LEVEL			TURN TO HEADING			VERTICAL S "D"			
	T1	T2	Total	T1	T2	Total	T1	T2	Total	Total For All Trials
NOVICE(16) (29.14)	-.20 (17.19)	-.38 (22.99)	-.32 (40.19)	-.17 (45.42)	-.40 (47.15	-.39 (92.57)	-.51* (22.22)	-.71** (26.75)	-.67** (52.97)	-.62* (185.74)
INTERMEDIATE(10) (29.05)	-0.26 (29.14)	-0.76** (29.09)	-0.55 (58.23)	+.35 (68.68)	-.04 (57.28)	+.11 (125.96)	-.22 (24.92)	-.36 (33.70)	-.33 (58.62)	-.46 (242.82)
EXPERIENCED(10) (20.59)	-.28 (38.21)	-.32 (51.78)	-.35 (89.99)	-.32 (51.01)	+.06 (65.49)	-.25 (116.59)	+.43 (41.13)	+.19 (43.46)	+.38 (84.59)	-.15 (291.18)

* significant at $\alpha = .05$
** significant at $\alpha = .01$

TABLE II. ANOVA TABLE FOR TOTAL SCORES

SOURCE	df	MS	F	Sig.
Groups (EXP. Level)	2	2,330.94	2.2956	n.s. $\alpha = .1$
Subj. w. groups	27	1,015.38		
MANEUVER	2	11,379.2	20.2743	$\alpha < .01$
Groups x Maneuver	4	687.42	1.2247	n.s. $\alpha = .1$
Maneuver x S. within G.	54	561.26		
TRIALS	1	1,057.88	5.3474	$\alpha < .05$
Groups x Trials	2	455.45		
Trials x s. within G.	27	197.83	2.3022	n.s. $\alpha = .1$
Trials x Maneuvers	2	41.838	.1556	n.s.
Groups x Trials x Man.	4	40.123	1.4924	n.s. $\alpha = .1$
B C x Sub w. groups	54	268.84		

for both of the vertical S "D" trials and total of the six trials significant at the .05 level.

This relationship is not as strong for the intermediate group, achieving statistically significance only for the second straight and level flight trial. The relationship diminished considerably for the experienced group as expected.

An analysis of variance was performed on the total scores using maneuvers, trials and groups or experience level as fixed independent factors. In order to maintain balance, only ten group one subjects were included in the analysis, which assumed subjects were nested within groups and crossed with trials and maneuvers. Again the formation flight scores were excluded.

The analysis, as shown in Table II, indicated the apparent systematic increase in mean total scores from 36.26 to 40.47 and to 48.53 for the novice, intermediate and advanced groups was not statistically significant. This lack of significance may be attributed to the large variation of subjects within groups, since both the change in means from 39.33 to 44.18 from trial one to trial two and of 33.39, 57.64 and 34.22 for straight and level flight, turns to heading, and vertical S "D" were significant at the .05 level.

Control Movements

The analysis of the control movement data was limited by the availability of only thirty subjects with which to access the contribution of thirty-six possible control variables. This limitation essentially determined the course of the data analysis. The primary objective of the analysis was to identify the variables with the greatest potential for discriminating pilot skill level while screening out those variables which showed little potential.

This screening was performed in two steps. The first step involved a univariate analysis of variance for each of the thirty-six potential dependent

TABLE III. SIGNIFICANT AILERON MEASURES

MANEUVER	STRAIGHT AND LEVEL			TURN TO HEADING			VERTICAL S "D"			FORMATION FLIGHT		
Trial	1	2	Total	1	2	Total	1	2	Total	1	2	Total
Minimum							0	0	0		0	*
Maximum				0		*	0	0	0		0	*
1st Moment												
2nd Moment			*				0	0	0		0	*
3rd Moment				*		*	0	*	0			
4th Moment				*			0	*	0		*	*
Mid S-Omega				*		*				0	0	0
1/8 Hz Filter				0		0	0	0	0	0	0	0
1/4 Hz Filter				0		*	0	0	0	0	0	0
1/2 Hz Filter				*			*	0	0	0	0	0
1 Hz Filter				*				0	*	0	0	0
2 Hz Filter				*				*		0	0	0

* = significant at p = .05
0 = significant at p = .01

TABLE IV. SIGNIFICANT ELEVATOR MEASURES

MANEUVER	STRAIGHT AND LEVEL			TURN TO HEADING			VERTICAL S "D"			FORMATION FLIGHT		
Trial	1	2	Total	1	2	Total	1	2	Total	1	2	Total
Minimum							*		*	*	0	0
Maximum							0	*	0	*	0	0
1st Moment					*	*						*
2nd Moment					*	*					*	0
3rd Moment					*	*			*			*
4th Moment					*			*	*			*
Mid S-Omega										*	0	0
1/8 Hz Filter										0	0	0
1/4 Hz Filter										0	0	0
1/2 Hz Filter										0	0	0
1 Hz Filter										0	0	0
2 Hz Filter										0	*	0

* = significant at p = .05
0 = significant at p = .01

TABLE V. SIGNIFICANT THROTTLE MEASURES

MANEUVER	STRAIGHT AND LEVEL			TURN TO HEADING			VERTICAL S "D"			FORMATION FLIGHT		
Trial	1	2	Total	1	2	Total	1	2	Total	1	2	Total
Minimum		*			*	*						*
Maximum							*					
1st Moment							*					
2nd Moment									*			*
3rd Moment							0	*	0			*
4th Moment								*	*			*
Mid S-Omega										*	*	0
1/8 Hz Filter										*	*	0
1/4 Hz Filter											*	*
1/2 Hz Filter												
1 Hz Filter												
2 Hz Filter												

* = significant at p = .05
0 = significant at p = .01

measures. The rational for this univariate analysis was that the variables exhibiting the greatest variation between groups would be the most useful in discriminating pilot experience level.

In an effort to reduce the noise inherent in the data the trial one and trial two values were added forming a total score. The univariate analysis was then performed on the individual trial data as well as the totals. The results of the analysis, which were performed individually for each maneuver are shown in Tables III through V for the aileron, elevator and throttle respectively.

Straight and Level: The most striking result of the results for straight and level is the lack of any significant differences between groups. Only the aileron second moment for the two trial total and the minimum throttle position for trial two significantly differed, and then only at the .05 level. None of the fifteen filter measures differed significantly for the maneuver.

Turn to Heading: There were fourteen significant F ratio's for the aileron, seven for the elevator and two for the throttle for this maneuver. Of these the most significant and their associated F, with two and twenty-seven degrees of freedom, values for trial one, trial two and the total are respectively, the aileron 1/8 Hz filter output ($F = 7.869$), the first moment of elevator position ($F = 5.049$), and the aileron 1/8 Hz filter output ($F = 5.49$).

Vertical S "D": There were twenty-seven, eight and eight variables statistically significant for the vertical S "D", for the aileron, elevator and throttle respectively. Of these the aileron second moment was most significant for both trials and the total with F's of 23.57, 18.11 and 24.69 respectively.

Formation Flight: The greatest difference between groups occurred for the formation flight group with twenty-six significant aileron, twenty-nine significant elevator and twelve significant throttle variables. The most significant variables were the aileron 1/4 Hz filter (F = 16.96) for trial one, the aileron 1/8 Hz filter (F = 17.18) for trial two, and aileron 1/8 Hz filter (F = 25.42) for the total. All the filter outputs were statistically significant at the .01 level for both trial for the elevator and aileron, but only the 1/8 and 1/4 Hz filters were significant for the throttle and then not for all trials.

Unfortunately the univariable analysis of variance only provides an indication of potential discriminating variables and the single most attractive variable. Since many of the dependent measures are highly correlated their relative discriminating power is a function of what other variables are included. To overcome this problem the univariable analysis was followed by a step-wise discriminate analysis, using Wilk's λ as the criterion for selecting the next variable to enter. The step-wise discriminate analysis was performed in two steps for each maneuver. First the results for trial one and two were combined to reduce the noise inherent in the data. Then a stepwise discriminate analysis was run on the combined data using a maximum of twenty steps using all thirty-six variables. The results of this analysis was then used to identify the first set of variables with entering F's in excess of a certain criteria level. The analysis was then repeated for the individual trials and the combined data l..king the analysis to this initial set of variables.

Straight and Level: The most striking feature of straight and level flight is the lack of any significant differences between the groups upon which skilled discrimination may be based. Based on the results of the first step of the discriminate analysis only two variables, aileron second moment and throttle minimum position, were identified as a potential second stage of the discriminating variables. The results of the discriminate analysis are shown in Table VI. The small percent of the two accounted for by the second discriminate function would suggest the three groups are close to colinear for

TABLE VI. STEPWISE DISCRIMINATE ANALYSIS RESULTS

MANEUVER	TRIAL 1 VARIABLE ENTERING	F	TRIAL 2 VARIABLE ENTERING	F	TOTAL VARIABLE ENTERING	F
STRAIGHT AND LEVEL	Aileron 2nd Moment	2.61	Throttle Minimum	3.57	Aileron 2nd Moment	3.77
	Throttle Minimum	5.79	Aileron 2nd Moment	1.55	Throttle Minimum	4.78
TURN TO HEADING	Aileron 1/8Hz Filter	7.87	Elevator 4th Moment	4.52	Aileron 1/8Hz Filter	5.49
	Elevator Maximum	4.43	Elevator Mod. S-Omega	3.12	Elevator Maximum	7.92
	Elevator 4th Moment	2.28	Elevator Maximum	1.44	Aileron 1 Hz Filter	2.77
	Aileron 1 Hz Filter	1.72	Aileron 3rd Moment	2.35	Elevator Mod. S-Omega	1.91
	Aileron 3rd Moment	.79	Aileron 1 Hz Filter	5.16	Aileron 3rd Moment	3.23
	Elevator Mod. S-Omega	.35	Aileron 1/8 Hz Filter	.66	Elevator 4th Moment	3.122
VERTICAL S DELTA	Aileron 2nd Moment	23.57	Aileron 2nd Moment	18.11	Aileron 2nd Moment	24.69
	Throttle 3rd Moment	8.32	Throttle 3rd Moment	14.77	Throttle 3rd Moment	15.29
	Throttle 4th Moment	3.07	Aileron Mod. S-Omega	2.31	Aileron Mod. S-Omega	2.31
	Elevator Mod. S-Omega	3.30	Throttle 4th Moment	1.76	Throttle 4th Moment	3.68
	Aileron Mod. S-Omega	1.11	Elevator Mod. S-Omega	1.18	Elevator Mod. S-Omega	2.93
FORMATION FLIGHT	Aileron 1/8 Hz Filter	13.03	Aileron 1/8 Hz Filter	17.18	Aileron 1/8 Hz Filter	25.42
	Throttle Minimum	2.21	Throttle Minimum	3.32	Throttle Minimum	3.04
	Throttle 1/8 Hz Filter	1.56	Elevator Minimum	1.26	Elevator Minimum	1.94
	Throttle 1 Hz Filter	3.91	Throttle 1/8 Hz Filter	.74	Elevator 1/8 Hz Filter	4.92
	Elevator Minimum	1.07	Throttle 1 Hz Filter	2.02	Throttle 1/8 Hz Filter	4.02
	Elevator 1/8 Hz Filter	2.67	Elevator 1/8 Hz Filter	1.38	Throttle 1 Hz Filter	3.06

this maneuver. The classification function coefficients suggest both variables were contributing approximately equally in classifying the pilots.

Using only these two variables, the thirty subjects were classified as shown in Table VII. From this table it may be observed the greatest misclassification occurred in discriminating between the novice and intermediate groups. Only two novices, one on trial one and one on trial two were misclassified as advanced while only one advanced pilot on trial one was classified as a novice.

Turn to Heading: Six variables showed some promise as possible discriminators for pilot skill level in this maneuver. However, the results were not consistent across trials. The two most important variables on trial one would manage no better than third and six place on trial two. Similarly the first two variables to enter for trial two were third and six on trial one. Moreover, the highest entering F value for either trial or the total was only 7.87. Although this represents an improvement over straight and level flight, it suggests the magnitude of the differences between groups is not large in comparison to the variation within groups, thus making discrimination of skill level difficult.

Using the order of entry on the two trial total as the measure of importance, it appears that of the six variables the Aileron 1/8 Hz filter output and the elevator maximum are the most promising discriminator variables for this maneuver. Although neither was significant on trial two, there entering F ratios of 5.49, and 7.92 for the totals were both statistically significant at the 0.01 level.

The classification of the thirty subjects on this maneuver using all six variables, is much better than for straight and level flight. The misclassification rates are slightly over 20% for the two trials and approximately 13% for the two trial total.

TABLE VII. SUBJECT CLASSIFICATION RESULTS

		CLASSIFIED AS TRIAL 1			CLASSIFIED AS TRIAL 2			TOTAL		
MANEUVER		Novice	Intermediate	Experienced	Novice	Intermediate	Experienced	Novice	Intermediate	Experienced
STRAIGHT AND LEVEL	NOVICE	6	3	1	6	3	1	6	3	1
	INTERMEDIATE	4	4	2	4	4	2	3	6	1
	EXPERIENCED	1	2	7	0	3	7	1	2	7
TURN TO HEADING	NOVICE	8	0	2	8	1	1	10	0	0
	INTERMEDIATE	2	8	0	3	7	0	2	8	0
	EXPERIENCED	0	2	8	1	1	8	1	1	8
VERTICAL S DELTA	NOVICE	8	0	2	10	0	0	10	0	0
	INTERMEDIATE	0	7	3	1	7	2	0	8	2
	EXPERIENCED	0	1	9	0	2	8	0	1	9
	NOVICE	8	1	1	9	1	0	9	1	0
	INTERMEDIATE	0	8	2	1	8	1	0	8	2
	EXPERIENCED	0	2	8	0	2	8	0	2	8

Vertical S Delta: The potential for discriminating pilot skill level is significantly enhanced for this maneuver. The aileron second moment and throttle third moment were the first and second variables to enter on both trials and the total, and there entering F ratio's ranging from 8.32 to 24.69 were substantially higher than for the previous two maneuvers. Although their F ratio's weren't particularly impressive, three additional variables were included in the second step of the discriminate analysis.

Using all five variables, the misclassification rate for the thirty subjects was slightly less than 20% for the two trials and as low as 10% for the two trial totals. From Table VII it may be observed that all the novices were correctly classified on trial two and for the two trial total and only two were misclassified on trial one.

Formation Flight: Six variables appear to have some potential for discriminating pilot experience level for formation flight, but these were dominated by the aileron 1/8 Hz filter. This variable had entering F ratios of 13.03, 17.18, and 25.42 for the two trials and the two trial total. Although the throttle minimum position was consistently the second variable to enter, its best entering F ratio was only 3.32 for trial two.

The misclassification rates for the thirty subjects were less than twenty percent for the two trials and about seventeen percent for the totals. Only one novice on trial one was classified in the advance category. None of the advance pilots were classivied as novice for formation flight.

Digital Filter Outputs

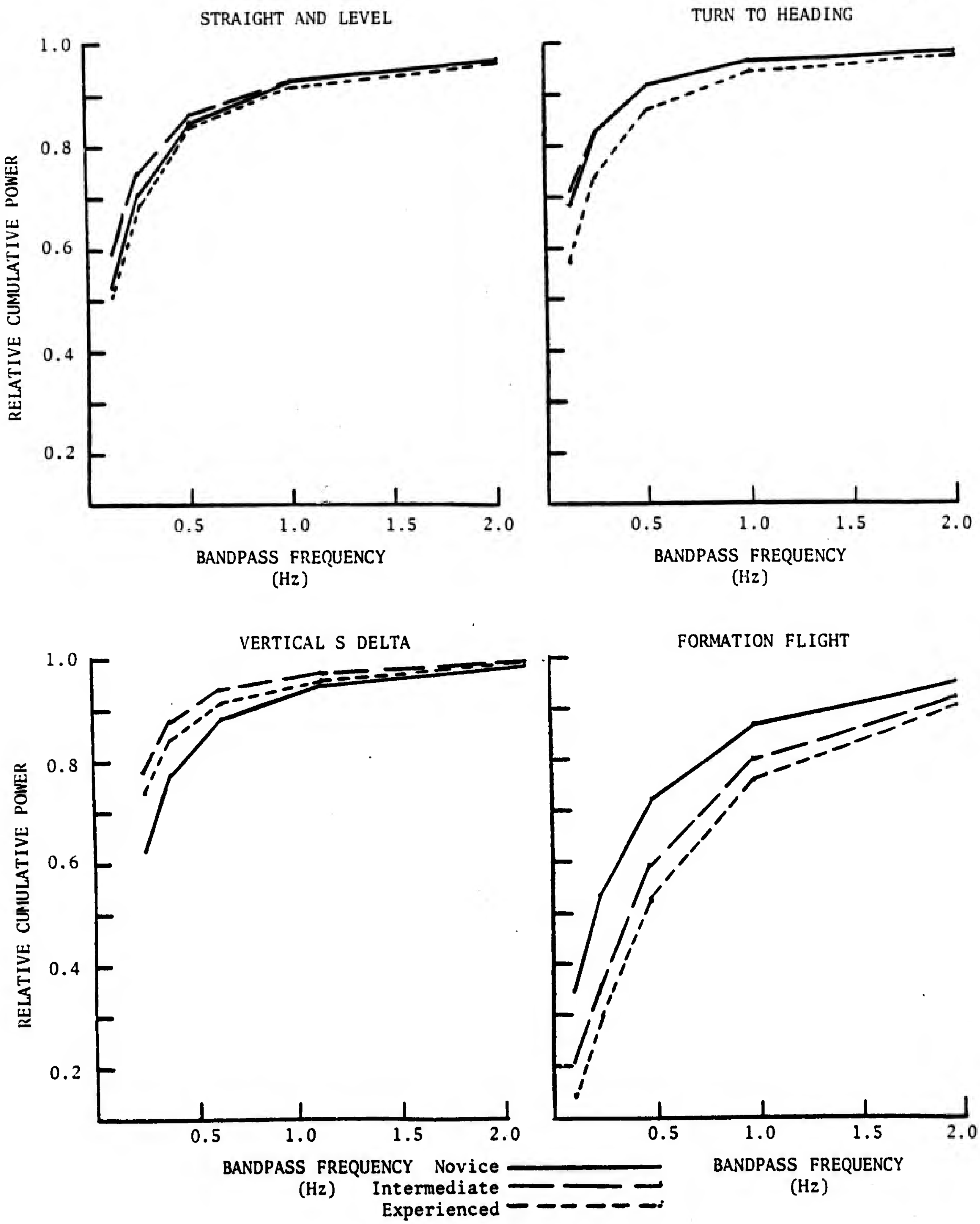
The previous analysis suggests that some of the filter outputs significantly differed between groups. The aileron position accounted for most of these differences, with both the throttle and elevator between group differences being statistically significant only for formation flight.

The aileron filter outputs for the average of the two trial totals are plotted for each group and maneuver in Figure 3. As plotted these output provide a rough estimate of the average cumulative relative power spectra across the pilots within each group.

The curves for formation flight and the turn to heading maneuvers closely conform to the hypothesized result, with the experienced pilots power spectra shifted toward the higher frequencies as expected. This relationship is, however, reversed for the vertical S delta maneuver, with the novice pilots power spectra actually shifted to higher frequencies than either the intermediate or experienced pilots. The large and highly significant increase in the aileron second moment may be partially responsible for this reversal. Recall that for the vertical S "D" the univariable F ratio was very high for the aileron second moment and that it was the first variable to enter during the discriminate analysis. The mean aileron second moments were .090, .051, and .053 respectively for the novice, intermediate and experienced groups. Thus the novice groups aileron second moment was approximately 80% higher than the intermediate and experienced groups values which were approximately equal. This was the only maneuver for which such a significant increase in the aileron second moment was observed. Apparently the novice group was injecting a high noise component into their aileron moments during the vertical S "D" maneuver.

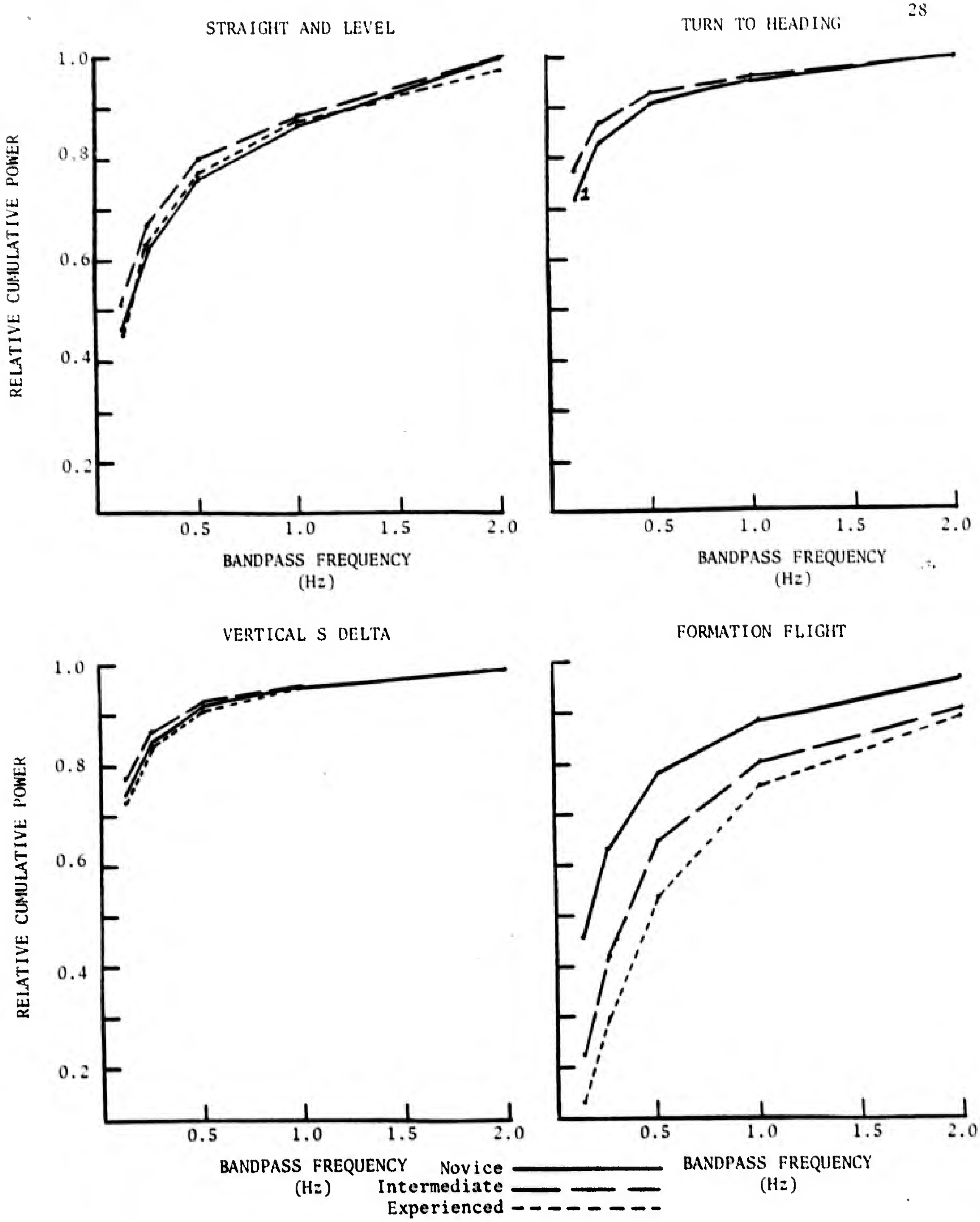
The elevator filter outputs, as shown in Figure 4 only differed significantly for the formation flight maneuver. The results due indicate that the experienced pilot's relative power spectra was shifted toward the higher frequencies.

Formation flight was the only maneuver for which there was sufficient throttle movement to obtain reliable filter output estimates. The output's, as plotted in Figure 5, generally conform to the successive organization of



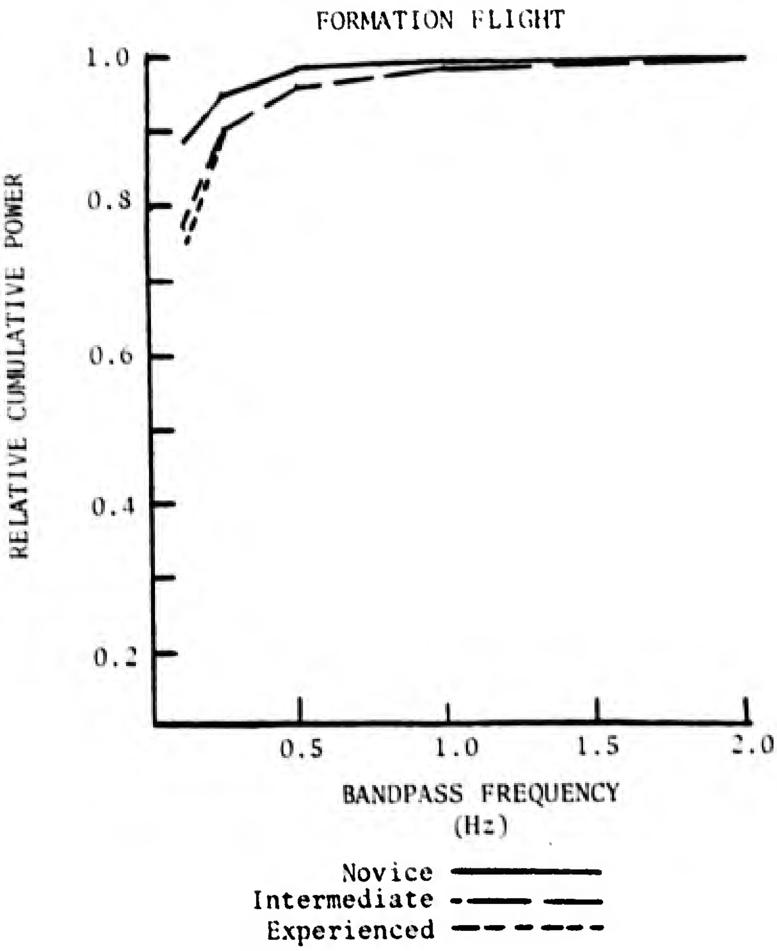
AILERON FILTER OUTPUTS

Figure 3



ELEVATOR FILTER OUTPUTS

Figure 4



THROTTLE FILTER OUTPUTS

Figure 5

perception hypothesis with the experienced pilots estimated spectra shifted toward the higher frequencies and the novice pilots toward the lower frequencies as expected.

IV. SUMMARY AND CONCLUSIONS

Certain insight may be derived by examining the subject's EFT scores. The means observed were considerably below averages for a male population of their approximate age range. If the pilots had represented a random sample, we would have expected an EFT average in excess of 45 seconds per item. The actual means of 29.14 and 29.05 for the novice and intermediate groups would suggest both groups were very field independent. The 20.59 sec/item average for the experienced group is considerably below the expected value. This would immediately suggest the subject populations were highly select groups, and that the selection mechanism, either implicitly or explicitly, selects high field independent individuals. Because of the lack of statistical significance for many of the novice groups correlation coefficient, no definitive conclusions may be drawn regarding the relationship between field dependence-independence and initial flying skill. The all negative correlation coefficients between the subjects EFT socres and the ASPT APM total scores for the novice group on each maneuver would imply field independent subject initially demonstrate greater skill levels. The systematic decline in the magnitude of the correlation coefficients as experience increases would imply that as skill levels increase, field dependence-independence becomes a less important factor. The lack of more statistically significant correlation coefficient might in part be attributed to the modest sample series and in part to the relatively homogeneous composition of the subject population with respect to field dependence/independence. Presumably a more heterogeneous population would produce higher correlations if a relationship does indeed exist.

From the analysis of the pilot's control movement data several conslusions may be drawn. First, the ability to discriminate pilot skill level from his .

control movements increases as the complexity or difficulty of the maneuver increases. There were few if any control measures which significantly differed between groups for straight and level flight. The converse was true for the vertical S delta and formation flight. In both cases there was a multiplicity of significant performance measures. Moreover the misclassification rates were generally below twenty percent for these maneuvers. It would appear that for simple undemanding maneuvers, novice pilots behave generally like more experienced pilots. Only on the more demanding maneuvers do novice and experienced pilot's behavior diverge.

Of the three control movements observed, aileron, elevator and throttle, the aileron was most significant in terms of the number of measures which significantly differed as a function of experience level, the level of significance of these measures and their discriminating ability. Of these measures the aileron second moment and 1/8 Hz filter output were the two dominate variables, each being the first to enter on the first step of the step-wise discriminate analysis for two of the four maneuvers. Which of the two is most important appears to be a function of the particular maneuver being performed.

After the two aileron measures the elevator maximum and throttle minimum position appear to be the most important discriminator variables. The throttle minimum position was the second variable to enter in the stepwise discriminate analysis on two of the four maneuvers. On a third maneuver the throttle third moment was the second variable to enter, which was highly correlated with throttle minimum position. At that stage the throttle minimum actually had the second highest F ratio and would have entered if the third moment had not been present. It would then appear reasonable to presume the throttle minimum position would be substituted for the third moment with little loss in discriminating power. Then, the throttle maximum position would have been the second variable entering on the turn to heading maneuver. Although it did not play an important role in any other maneuver its entering F of 7.92 for the two trial total would justify its being identified as an important discriminating variable.

The important skill level discriminators could then be summarized as aileron second moment, aileron 1/8 hz filter output throttle minimum position, and elevator maximum position.

The Butterworth filter results generally confirm the Successive Organization of Perception Hypothesis. The filter outputs for the more complex maneuvers did vary as a function of subject experience and generally in the direction anticipated. Moreover the magnitude of the change appears attractive in terms of devising measures sensitive to subject skill level, as evidenced by the importance of the 1/8 hz aileron filter output in determining subject experience level.

TABLE VIII. SUMMARY OF CONCLUSIONS

1) EFT scores are related to performance for novice pilots but the effect diminishes with experience.

2) Discriminability becomes better as maneuver difficulty increases.

3) The aileron dominated the elevator and throttle in terms of differences due to experience level.

4) The two most important variables appear to be aileron 2nd moment and 1/8 Hz filter output.

5) After the two aileron variables the throttle minimum and elevator maximum appear to be the most useful.

6) Pilot control movement based measures may be effectively used to discriminate pilot skill/experience level.

7) There are changes in the pilot's control moment power spectra as a function of skill level.

BIBLIOGRAPHY

Brown, J. E., W. L. Waag, and E. E. Eddowes, USAF Evaluation of an Automated Adaptive Flight Training System, FFF HRL-TR-75-55. Williams AFB, Arizona: Flying Training Division, Air Force Human Resources Laboratory, 1975.

Cream, B. W., F. T. Eggemeier and G. A. Klein, Behavioral Criteria in the Design of Air Crew Training Devices, Proceedings of the Human Factors Society, 19th Annual Meeting, 1975.

Goldstein, I. I., Training: Program Development and Evaluation, Brooks/Cole Publishing Co., 1974.

Knoop, P. A., Advanced Instructional Provisions and Automated Performance Measurement, Human Factors, Vol. 15, 1973.

Knoop, P. A. and W. L. Welde, Automated Pilot Performance Assessment in the T-37: A Feasibility Study. Air Force Human Resources Laboratory Technical Report TR-72-6, 1973.

Magdaleno, R. E., H. R. Jex and W. A. Johnson, Tracking Quasi-Predictable Displays, Proc. 5th Annual NASA-University Conf. on Manual Control, NASA SP-215.

Ping, J. A. and S. Gill, Evaluation of an Automated Flight Training System: Ground Controlled Approach Module. NAVTRAEQUIPCEN, IH-264, 1976.

Waag, W. L., E. E. Eddowes, J. H. Fuller, and R. R. Fuller, ASUPT Automated Objective Performance Measurement System, AF HRL-TR-75-3, Williams AFB, Arizona, 1975.

APPENDIX A

EQUATIONS USED FOR PERFORMANCE DATA

$V_1 = \Sigma P/N$

$V_2 = \Sigma P^2/N$

$V_3 = \Sigma P^3/N$

$V_4 = \Sigma P^4/N$

$S - Omega = \Sigma\ (\Delta P/h)^2/\Sigma P^2$

$P_1 = \Sigma X_1^2\ /\Sigma P^2$

$P_2 = \Sigma X_2^2\ /\Sigma P^2$

$P_3 = \Sigma X_3^2\ /\Sigma P^2$

$P_4 = \Sigma X_4^2\ /\Sigma P^2$

$P_5 = \Sigma X_5^2\ /\Sigma P^2$

$M\ (S - Omega) = (S - Omega) * \sqrt{V_2\ /\ (V_2 - V_1^2\ )}$

$Q_1 = (\Sigma X_1^2\ /N - (\Sigma X_1\ /N)^2)\ /(V_2 - V_1^2)$

$Q_2 = (\Sigma X_2^2\ /N - (\Sigma X_2\ /N)^2)\ /(V_2 - V_1^2)$

$Q_3 = (\Sigma X_3^2\ /N - (\Sigma X_3\ /N)^2)\ /(V_2 - V_1^2)$

$Q_4 = (\Sigma X_4^2\ /N - (\Sigma X_4\ /N)^2)\ /(V_2 - V_1^2)$

$Q_5 = (\Sigma X_5^2\ /N - (\Sigma X_5\ /N)^2)\ /(V_2 - V_1^2)$

BUTTERWORTH FILTERS

$$x_i(t) = \alpha_i * x_i(t-1) + \beta * (P(t) + P(t-1))$$

Filter No.	α_i	β_i
1	.9489	.0255
2	.9804	.0498
3	.8097	.0951
4	.6494	.1753
5	.3838	.3081

TOLERANCE VALUES FOR FREQUENCY DOMAIN STUDY

	BASE	LOWER	UPPER	BOUNDS	UNITS
050					
STRAIGHT & LEVEL					
ALTITUDE	15000.0	-31.4	31 4	3000.0	FEET
AIRSPEED	160.0	- 1.87	1.87	60.0	KNOTS
HEADING	180.0	- 1.92	1.92	60.0	DEGREES
051					
TURN TO HEADING					
ALTITUDE	15000.0	-43.1	43.1	3000.0	FEET
AIRSPEED	160.0	- 2.02	2.02	60.0	KNOTS
BANK	30.0	- 3.23	3.23	30.0	DEGREES
HEADING	65.0	- 2.79	2.79	60.0	DEGREES
052					
FORMATION					
X.POS	0.0	- 2.0	2.0	100.0	FEET
Y.POS	0.0	- 3.0	3.0	100.0	FEET
Z.POS	0.0	- 1.0	1.0	100.0	FEET
053					
VERTICAL S DELTA					
AIRSPEED	160.0	- 2.03	2.03	60.0	KNOTS
BANK	30.0	- 2.68	2.68	30.0	DEGREES
VVI	1000.0	-163.0	163.0	1000.0	FT/MIN
HEADING	Determined	- 2.25	2.25	30.0	DEGREES
ALTITUDE	15000.0	-32.4	32.4	5000.0	FEET

FREQUENCY DOMAIN STUDY
STUDENT DATA SYSTEM PRINT OUT PROFILE

VALUE	DESCRIPTION
1	Sample Size (15/second)
2	Minimum
3	Maximum
4	V 1
5	V 2
6	V 3
7	V 4
8	S-OMEGA
9	P 1
10	P 2
11	P 3
12	P 4
13	P 5
14	Modified S-OMEGA
15	Q 1
16	Q 2
17	Q 3
18	Q 4
19	Q 5

THE FOLLOWING IS AN INDEX OF PERMFILE EXERDESO

001	FDS MISSION 001 FOR GROUP 1
002	FDS MISSION 001 FOR GROUP 1 - FORMATION
003	FDS MISSION 002 FOR GROUP 2 & 3

001 FDS MISSION 001 FOR GROUP 1
EXERTBLE 001

01	TASK, 054	STRAIGHT & LEVEL	DEMO
02	TASK, 058		PRACTICE
03	TASK, 050		APM
04	TASK, 050		APM
05	TASK, 055	TURN TO HEADING	DEMO
06	TASK, 059		PRACTICE
07	TASK, 051		APM
08	TASK, 051		APM
09	TASK, 056	VERTICAL S DELTA	DEMO
10	TASK, 060		PRACTICE
11	TASK, 053		APM
12	TASK, 053		APM

002 FDS MISSION 001 FOR GROUP 1
EXERTBLE, 002

01	TASK, 057	FORMATION	MANUAL IP DEMO
02	TASK, 061		PRACTICE
03	TASK, 052		APM
04	TASK, 052		APM

003 FDS MISSION 002 for GROUPS 2 & 3
EXERTBLE, 003

01	TASK, 062	PRACTICE-15 MIN MAX
02	TASK, 050	STRAIGHT AND LEVEL APM
03	TASK, 050	
04	TASK, 051	TURN TO HEADING 090 APM
05	TASK, 051	
06	TASK, 053	VERTICAL S DELTA APM
07	TASK, 053	
08	TASK, 052	FORMATION APM
09	TASK, 052	

SORTIE CONTENT