HUMAN FACTORS, 1977, 19(5), 503—514

The Effects of Cockpit Environment on Long-Term Pilot Performance

ALLAN M. STAVE, *Sikorsky Aircraft*

A fixed-base helicopter simulator was used to examine pilot performance as influenced by noise, vibration, and fatigue. Subjects flew the simulator for periods ranging between three and eight hours while exposed to vibrations (at 17 Hz) ranging from 0.1 to 0.3 g, and noise stimuli varying between 74 (ambient) and 100 dB. Despite reports of extreme fatigue on these long flights, subject performance did not degrade. Within the limits of this study, performance tended to improve as environmental stress increased. However, subjects did suffer from lapses resulting in abnormally poor performance. These lapses are probably of short duration (seconds) and occur at unpredictable times. If such lapses occur in actual flight, they could provide an explanation for many so-called "pilot error" accidents.

INTRODUCTION

This report describes the results of a three year study of the effects of noise and vibration on pilot performance. The goal was to explore the effects of each parameter so that realistic limits could be placed on acceptable cockpit environments.

Pilots presently flying helicopter commercial routes spend about eight hours each day flying passengers between airports. During this time, the crew usually carries out about 30 landings and takeoffs. Future plans call for continuation of this type of operation even under severe instrument flying conditions.

At the present time helicopter airlines fly only in VFR conditions. The workload involved has not degraded crew performance, probably because pilots have been able to meet the demands of the tasks by increasing their personal effort. The concern that motivated the present study was that instrument conditions would increase workload and require almost maximum effort on the part of the crew. With this increased workload, any loss in efficiency through environmental stress (noise and vibration) would exceed the ability of pilots to compensate and thereby could cause accidents.

The study examined the hypothesis that the increased stress of environmental stimuli, such as noise and vibration, would degrade pilot performance on complex instrument flight patterns.

METHOD

Data on the effects of noise, vibration, and fatigue were collected while pilots accomplished tasks typically required of commercial helicopter aviators. Performance was measured in terms of deviations from desired flight path and altitude values. The simulator used for this study is a fixed-base device which makes use of both analog and digital computers. No external cues such as motion were presented.

Control/Navigation Task

The airline route shown in Figure 1 was set up to represent an IFR route structure within the New York metropolitan area. Four-minute rest periods were given at every stop except Kennedy, where an eight-minute rest period was permitted. Subjects were not allowed out of the simulator except at the 8-minute Kennedy stop. A lap around the route required one hour to accomplish with this schedule. At the start of data collection, pilots flew this route for periods of four hours. Later in the study this was changed first to a three-hour period and then to six-and eight-hour periods.

In the three-hour periods, subjects performed the same tasks as in the four-hour periods previously described. The rest periods, however, were not permitted. Thus, the pilot did not pause after landing but continued with the next takeoff. Subjects did not leave the simulator for the entire three hours.

The six-hour data runs consisted of two three-hour periods back to back. That is, subjects had no rest pauses and did not leave the simulator seat for the entire six-hour period.

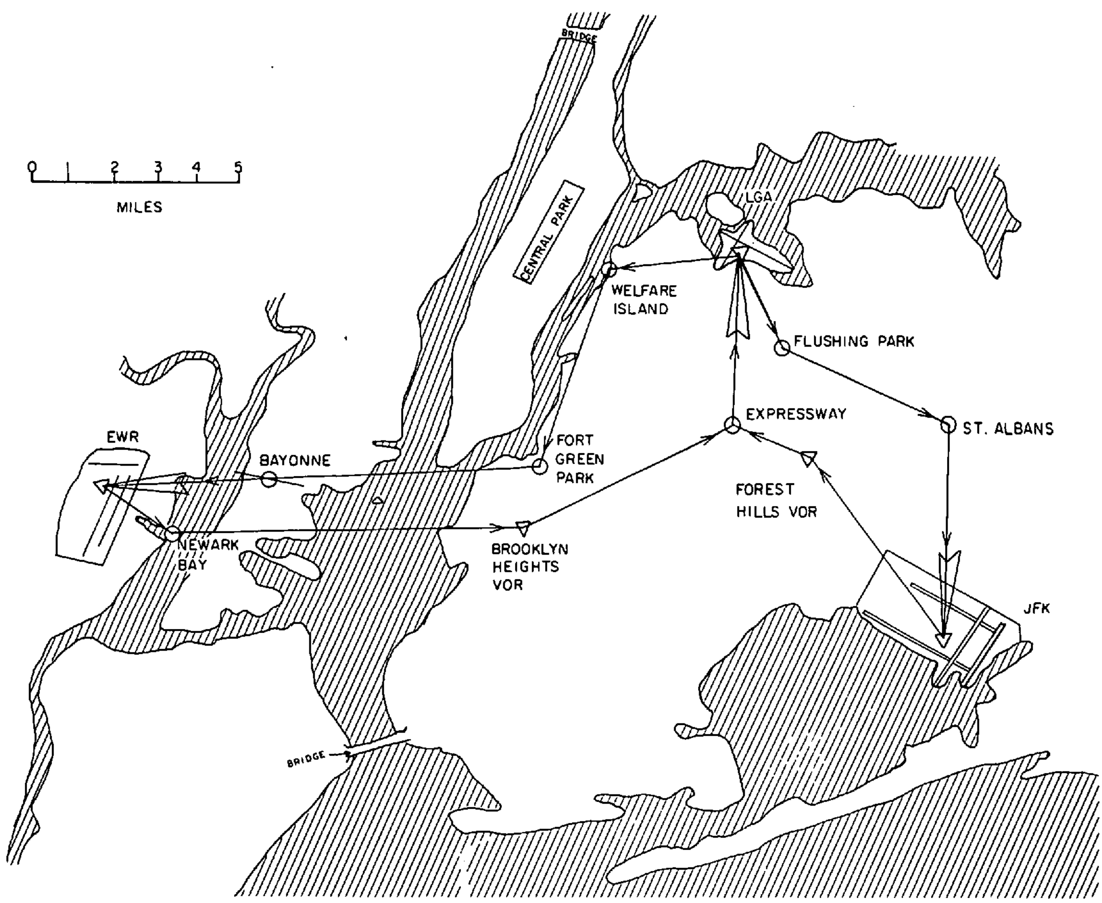

Figure 1. *Airline route used in simulation.*

The eight-hour run consisted of two, four-hour runs. The subjects were given four-minute rest periods after each landing and were permitted to leave the simulator once each hour for eight minutes.

The flight task performed by the subjects involved controlling the simulator which was programmed with a six-degree-of-freedom tracking task closely simulating a real helicopter. Pilots accomplished two basic tasks, airway navigation and an instrument (ILS) approach. The airway navigation task was simpler than the approach task and involved level flight while tracking radio and navigational aides using standard instrument flight practices. The approach task involved tracking an ILS type landing beam that brought the aircraft down from 305 to 23 meters where the problem was automatically terminated. This task was very demanding because it was difficult to maintain the vehicle precisely on the glide slope. The approach angle was steeper and the final altitude lower than the subjects had experienced before. A typical approach required about one minute to complete.

The flight instruments provided in the simulator corresponded exactly with those in a standard helicopter cockpit. In addition, four navigational aides were provided; VHF omnirange bearing information (VOR), distance from the VOR station (DME), automatic direction finding bearing from tuned stations (ADF) and the instrument landing system beam (ILS).

Data runs were conducted as though the subjects were making normal commercial transport flights. IFR procedures were simulated in which subjects were required to request and copy clearance, make position reports, and maintain specific climb, cruise, and approach airspeeds and altitudes.

Data runs were broken up into three units as follows:

Leg: That part of a data run lasting from takeoff at one heliport to landing at another (Kennedy to LaGuardia).

Round: That part of a data run consisting of four legs or one circuit of the route (*e.g.*, Kennedy to LaGuardia, to Newark, to Laguardia, to Kennedy).

Run: All of the data collected from one subject in one session. The data consisted of four or eight rounds.

Subjects

Five subjects were used for each data point. All were Sikorsky employees and former helicopter pilots. Each was instrument rated and had more than 1000 flight hours. Data runs took place in the evening (5 p.m. to 1 a.m.) after the subjects had completed a normal day's work in the Sikorsky engineering department.

Experimental Conditions

When the study began only the four-hour runs were planned. It was assumed that performance would slowly degrade due to noise, vibration, and fatigue effects. As data collection proceeded, however, it became apparent that performance was not degrading as expected. The experimenter began to suspect that the four-minute rest periods permitted subjects to recover from the effects of stress. This was supported by the results of Sussman (1970). In his study subjects operated an automobile simulator for four hours without rest. Performance degraded, but a four-minute rest period at the end of four hours restored proficiency. After collection of the data points for the conditions shown in Table 1, the effort to investigate the effects of rest periods on performance was modified.

The first step was to obtain a baseline condition of continuous performance. Accordingly, data for three hours of continuous performance were collected. This period was selected because it provided exactly the same amount of data (in terms of scores) as the original four-hour runs. It was assumed that performance would degrade, in the absence of

TABLE 1

Experimental Conditions

	17 Hz				12 Hz
Stress	4h	3h	6h	8h	3h
Noise	0	0	95 dB	95 dB	90 dB
Vibration	0	0	.2 g	.2 g	.1 g
Noise	100 dB	100 dB	90 dB		
Vibration	.2 g	.2 g	.1 g		
Noise	100 dB	95 dB	95 dB		
Vibration	.4 g	.1 g	.3 g		
Noise	0	0			
Vibration	.2 g	.2 g			
Noise	100 dB	100 dB			
Vibration	0	0			
Noise	95 dB				
Vibration	.1 g				

rest periods, as a function of noise and vibration stress. Once this occurred, it was planned to insert rest pauses of varying length for further study.

After collection of the five, three-hour conditions shown in Table 1, it was obvious that performance was not degrading as expected. It was then decided to experiment with longer data runs (eight hours with rest and six hours without rest). Completion of the first eight-hour data runs showed no obvious effects. In fact, performance seemed to improve. The six-hour data runs were set up as shown in Table 1.

Data Collection

While subjects were flying, both objective and subjective data were recorded. Objective data were recorded in a computer printout which provided both navigational score (RMS error off course) and approach score. An angular measure, the number of degrees off course, was used to score approach performance since it weights the final seconds of the approach more heavily than the initial part of the glide path. This corresponds to the situation in the real world because final position at break out is more important than positioning during the early part of an approach. Since the scoring system recorded errors, high scores mean poor performance.

Subjective data were obtained on a rating sheet devised for this study. This sheet was filled out at the end of each leg. It contained two types of self rating: flight performance and feelings of fatigue. The investigator also collected subjective data in terms of his own observations and an informal, open-end interview at the end of each data run or during rest periods. These were recorded in a log of each run and filed with the records for that run.

Vibration and Noise Stimuli

Vibration was imparted to the subjects by means of a hydraulic actuator attached to the pilot's seat. This device provided vertical sinusoidal motion at selected frequencies and g levels.

The literature on human reactions to vibration reveals conflicting results. In some studies, subjects were seated on hard surfaces such as metal plates, boards, or uncushioned seats. In others, subjects were seated on standard pilot or passenger seats. Since seats and seat cushions attenuate vibration, it seems evident that subjects exposed to equal floor vibration are not experiencing equal vibration at the base of the spine. This may explain the discrepancies between studies. Accordingly, it was decided to measure seat vibration attenuation (transmissibility) in the simulator. Knowledge of seat transmissibility permits accurate comparison of these results with those of other investigators. Table 2 is a comparison between floor and seat levels for each vibration condition used.

The initial goal of this study was to provide information to be used in real-world situations. Because of this, values were selected based on conditions which could be expected in actual helicopters. The vibration frequency

TABLE 2

Vibration: Floor versus Spine Level Comparison

Location	Frequency			
	17 Hz			12 Hz
Floor	0.3 g	0.2 g	0.1 g	0.1 g
Spine	0.11 g	0.06 g	0.02 g	0.08 g

of 17 Hz is the dominant frequency of the S-61 helicopter transport. The vibration amplitudes (0.1 to 0.3 g) were selected as a range extending from the actual level in the S-61 to a level which will probably never be exceeded in a commercial helicopter. The single data point at 12 Hz was used as a calibration point for a future study that investigates effects of variations in frequency as well as amplitude.

Noise stimuli provided to subjects consisted of a tape recording of S-61 cockpit noise played through high quality headphones at levels varying from 90 to 100 dB.

RESULTS

Lapsed Runs

In the early stages of the program, it was observed that a subject would suddenly perform quite poorly in the midst of otherwise normally distributed data. A gross examination of the data showed that each pilot's scores tended to cluster about some mean value except for a few extremely poor values. For example, a data run of six hours might have seven ILS scores clustering closely around an error value of 1.74. The eighth score might have a value of three or four times the average of the others, *i.e.*, about 6.0. A run with a score such as this was called a "lapsed" run.

It was decided that scores obtained in lapsed runs would not be included in the normal data analysis, since a single run could have a large effect on the average scores for that round. Accordingly, any run that exceeded three standard deviations from the mean of the rest of the scores for that round was classified as a lapse. These scores were separated from the data for separate analysis.

Examination of the lapse data shows that there is an inverse relationship between length of simulator run and the occurrence of lapses, *i.e.*, the longer the run the less lapse behavior is manifested. However, when the occurrence of lapses within individual runs is considered, it appears that there is no clear cut pattern of incidence in the first or last part of the six-and eight-hour data runs. One would think that if lapses are inversely related to time stress that they would also occur more frequently in the early parts of long data runs when all of the runs are considered. This did not happen.

Influence of Short Rest Periods

The suspicion that short rest periods were effective in maintaining performance was dramatically demonstrated during one of the early three-hour no-rest runs. On this run the subject began with poor scores which became all but unsatisfactory by the end of the second hour. At this point, the subject indicated that he was falling asleep and wanted a few moments rest to wake himself up. The experimentor was curious to see the effect of a short rest period so he permitted a four-minute rest. The result is shown in Figure 2. The average score before the rest period was 4.19 while the average score after the rest period was 1.29.

Stress Rating

In order to relate all of the variables mathematically, it was decided to order the conditions in terms of stress. To do this, stress was broken down into three categories: noise, vibration, and time (fatigue). Table 3 shows the categorization and the stress rating assigned to each. Noise was evaluated by con-

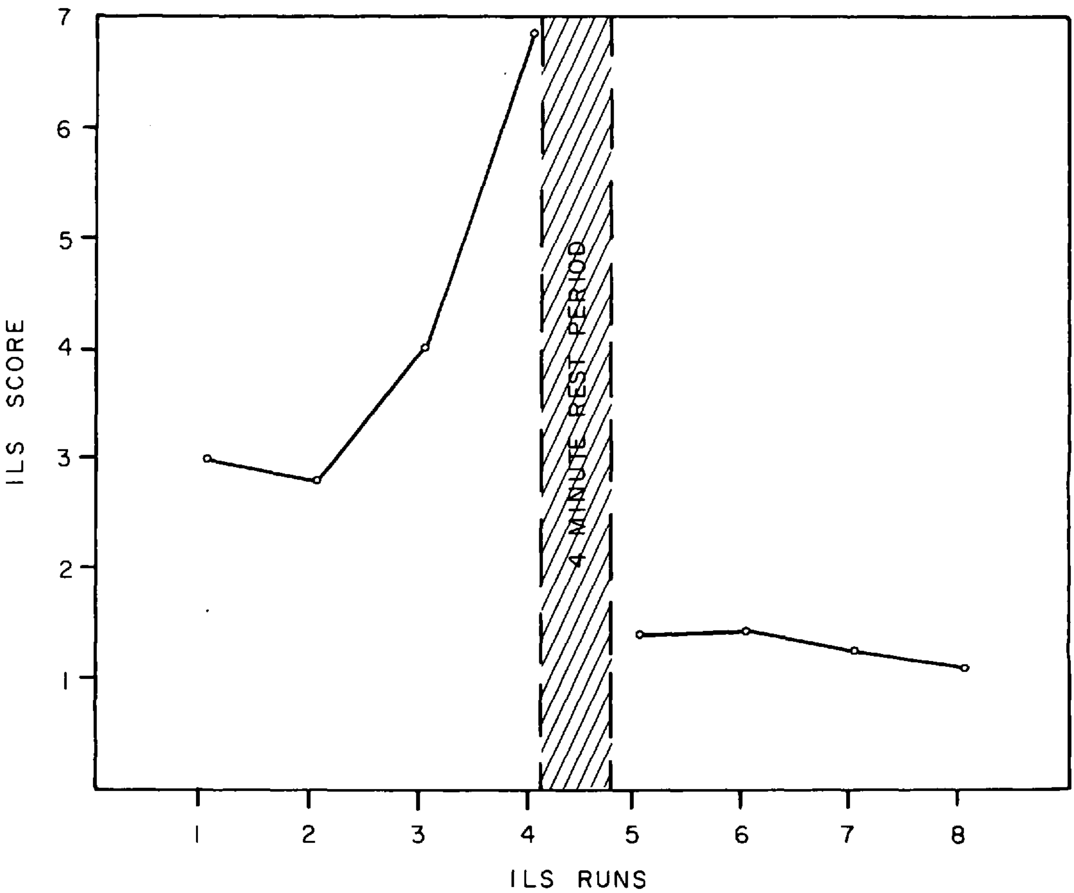

Figure 2. *Effects of short rest periods.*

verting dB levels into Sones. Fatigue was evaluated by having subjects rate their reactions to the time intervals in Table 3. Vibration exposures were quantified through use of Shoenberger and Harris' (1971) subjective equal intensity curves. Individual stresses were combined by simple addition.

Statistical Analysis

Three types of statistical analysis were used on the data. The first compared scores, lapses, and fatigue ratings with stress. The second compared fatigue ratings with scores. The final analysis compared scores with time in the simulator (fatigue).

Stress Analysis. Table 4 shows a matrix of Pearson product-moment correlations on the data. As shown in the table, almost all of the relationships are inverse, *i.e.*, the higher the stress, the better (lower) the score or the smaller the percentage of lapses. In observing the subjects, it was noted that the navigation task seemed easier and less demanding than the ILS task. This observation is supported by the correlations in Table 4, *i.e.*, none of the relationships between stress and navigation score are significant.

The noise stimuli used had no influence on performance. This is not unexpected since research by others has shown that steady-state

TABLE 3

Stress Rating Scheme

(a) Vibration Stress

Vibration Level (g)	Subjective Intensity	Rating
0	0	1
.1	8	1.8
.1	9	1.9
.2	20	3
.3	30	4

(b) Time Stress

Period	Rating
4 hour	1
3 hour	2
8 hour (1st half)	3.5
6 hour (1st half)	3.5
8 hour (2nd half)	5
6 hour (2nd half)	6

(c) Noise Stress

dB	Sones	Rating
74	10.6	1
90	32	3.2
95	45.3	4.5
100	64	6.4

TABLE 4

Pearson Product-Moment Correlations

	Stress			
	Time	Noise	Vibration	Total
Score				
ILS	−.466*	.038	−.603**	−.506*
Navigation	−.117	−.411	.299	−.178
Total	−.605***	−.252	−.422	−.517*
Lapse				
ILS	−.512*	−.359	−.491*	−.611***
Navigation	−.460*	.054	.069	−.180
Total	−.613***	−.169	−.230	−.474*

*$p < .05$.
**$p < .02$.
***$p < .01$.

noise does not seem to affect performance (Finkleman and Glass, 1971). Conversations with subjects indicated that noise added a note of realism rather than stress.

Fatigue analysis. Subject ratings of fatigue were compared with time in the simulator (Figure 3) and, as one would suspect, they increased with exposure to the simulator environment. Pilots seemed to adapt their feelings of fatigue as a function of time to go in the mission. The highest ratings in the three-, four-, and eight-hour missions were essentially the same. The results in the six-hour mission, however, are different. Subjects fatigued initially at the same rate as in the three-hour mission, but the curve levels off at three hours and then rises sharply to a rating in excess of four at the end of six hours.

These data are presented to illustrate the large subjective element in feelings of fatigue. In effect, perception of fatigue was influenced by the length of time a subject had left in the simulator. In many individual cases, the fatigue ratings seemed to reach a plateau at the three rating (tired but OK), until the subjects realized that the session was almost over. At that time, the ratings would climb rapidly to the four or five level. The normal rate of increase for fatigue ratings was on the order of a rating change every hour or so (four approaches). In the six-hour no-rest runs, it was not unusual for the rating to change from three to five within the space of a few legs (30 minutes).

ILS score vs. fatigue rating. Comparison of ILS scores with the subjects' ratings of their feelings of fatigue showed an inverse relationship, *i.e.*, as subjects became tired, their ILS scores tended to improve. Statistical analysis of the data yielded a significant correlation of −.56.

When lapse occurrence was correlated with

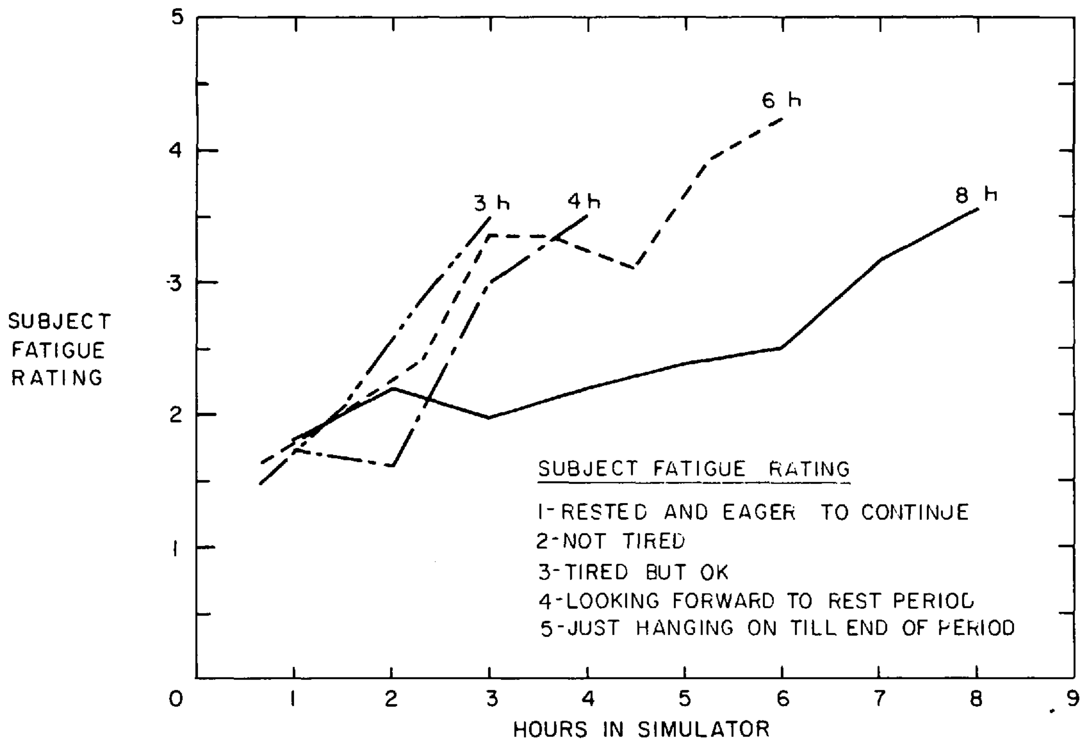

Figure 3. *Fatigue rating versus exposure time.*

fatigue ratings, the result was a correlation coefficient of .87. This means that whatever is causing "lapses" is directly related to an increase in fatigue.

Relationships of scores to time. Figure 4 shows performance score changes over time in the simulator. The scores are a combination of both ILS and navigation scores. (Navigation scores were divided by 200 so that both navigation and ILS performance would contribute evenly to the final result.)

The combined scores show a low point during the third hour and a rise beginning in the fourth hour. These data are similar to the findings of Jones (1960). In his study, radio operator performance was measured during 16-hour missions in actual aircraft. He found a slow improvement in performance for three hours and then a degradation. His data were taken over a series of five-hour watches. Thus, two independent studies both show that tasks requiring constant alertness result in slow performance improvement for a period of two to four hours and then degraded performance afterwards.

DISCUSSION

This study has many complex implications. Perhaps the most important is that performance seems to depend on motivation rather than fatigue. This conclusion is obvious especially when one watches subjects coping with the problem of working continuously for six hours. They start a session talking with the experimenter, making long involved position reports, singing to themselves, whistling under their breath, *etc.* As time goes on, this slowly changes to occasional muttered com-

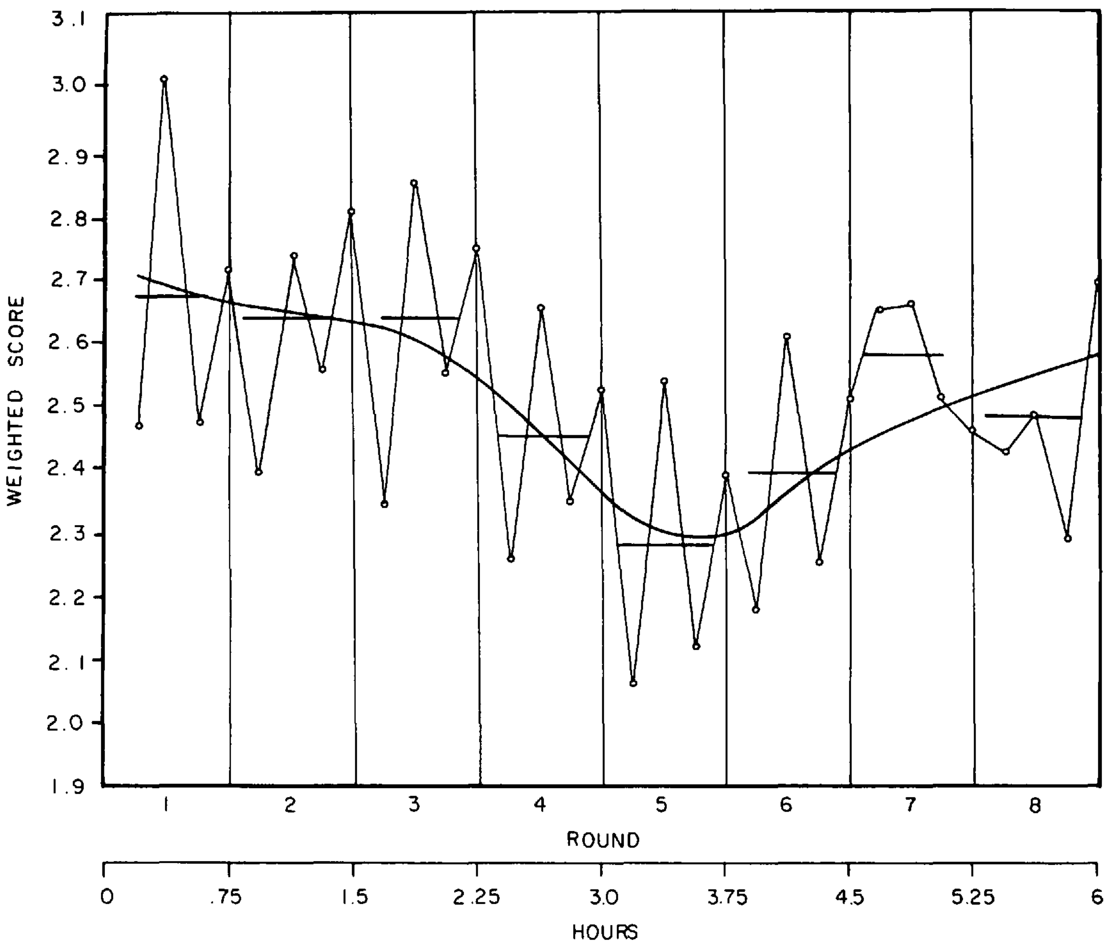

Figure 4. *Effects of simulator exposure on combined scores.*

plaints during approaches, shorter and shorter position reports, and finally into a grim silence interrupted only by brief position reports. This pattern is essentially the same for all subjects, differing only in its rate of onset. Despite the onset of fatigue, performance not only did not degrade, but it actually improved with time. This was apparent to the investigator during data collection and was amply backed up by the statistical analyses performed afterward.

The writer agrees with Pierson (1963) who, after measuring reaction time and movement time on a simple stimulus response task, reached three conclusions as follows:

(1) Subjective experience of fatigue is not a valid criterion of ability to perform speed or endurance type (muscular) work.
(2) Fatigue and endurance cannot be measured by work decrement.
(3) Fatigue, endurance, and work decrement are independent variables.

Although the tasks performed in the present study were a much more complex mixture of psychomotor elements, the findings can be summarized in almost the same way.

One of the more interesting observations of the present study was the effect of short rest periods. Results reported by others agree with this finding. Bergum and Lehr (1962) found that vigilance performance could be maintained effectively if subjects were granted short rest pauses (10 minutes). Coloquhoun (1959) in examining inspection tasks, found that men given short rest pauses of five minutes maintained undegraded effort for periods of one hour. Subjects working without rest could not maintain performance. Sussman (1970), as described earlier, found that a four-minute rest at the end of four hours of work resulted in almost complete performance recovery. Kraft (1965) found that the optimum rest pause for physical work on a dynamometer was one minute. Wilkinson (1959), in testing subjects after 30 hours of sleep deprivation, found that rest pauses of 30 seconds every five minutes could not prevent performance degradation.

The above studies are cited in descending order of rest period to illustrate that periods ranging from ten minutes down to only one minute have all been effective in preventing performance degradation. Wilkinson's study shows that there is a lower limit to the effectiveness of short rest pauses. Thus, the effectiveness of the four-minute rest in preventing performance degradation, as inadvertently demonstrated in this study, is a characteristic that should be taken into account in all performance studies. Explanation of why such short rest pauses should be effective must necessarily be tentative, but the experimentor feels that the brief pause improves motivation by allowing the subject to break his concentration and perhaps rid himself of some of his boredom.

Perhaps the key parameter in performance is subject motivation. Where motivation is high, performance does not degrade. Since motivation was not measured directly in this study, the only evidence for this is the daily observations of subject behavior while data were being collected.

One of the indicators of the high motivation level was the fact that only one of the six-hour no-rest data runs was aborted due to subject's inability to continue. The run was aborted in the fourth hour after the subject had exhibited very poor performance. During the run, the subject showed a great deal of irritability, impatience, and anger. In discussing the situation with him immediately after the abort, it was learned that he had had a particularly trying day and was facing a complex and demanding schedule of activities in the next few weeks. About three weeks later, the subject completed another six-hour data run without problems. The abort could be better explained by mental rather than by physical fatigue.

At the start of the study, it was assumed that performance would slowly degrade as the subjects became tired. This did not happen. A subsequent look at the literature showed that most reported research in this area has also failed to find any convincing relationship. Pierson (1963) points out that laboratory tasks of longer duration (20 or 30 hours) may be needed to demonstrate greater agreement between performance degradation and fatigue.

The question remains, however, as to why performance improved as stress increased. The author feels that the hypothetical curve of performance shown in Figure 5 might be used to explain the initial improvement in performance. As pilots feel the onset of fatigue, they put forth increased effort to compensate. This increased effort results initially in improved performance, which then degrades as the products of fatigue build up and cause a reduction in motivation. The plot of scores for eight rounds provides factual evidence for the curve in Figure 5. As a practical matter, however, subjects indicated that they would not fly a real vehicle if they considered themselves to be as tired as indicated by a fatigue rating of 4. One might assume, therefore, that no pilot would fly an actual aircraft beyond the point where motivation can sustain normal performance.

The focus of this study was to test performance under stress conditions comparable to those in a real-world environment. The results show that pilot performance is not degraded by noise, vibration, and duration conditions tested.

If one views the results of this study broadly, probably the most important finding has to do with the phenomenon of lapses. The phenomenon, as manifested herein, was not a function of training since lapses continued to occur throughout the data collection period (about one year). Approximately five percent of the ILS runs resulted in lapses.

Casual observation of pilot's reactions during data runs leads to the inference that lapses are caused by a gap in attention lasting for a short period of time (seconds in length). Such gaps would cause the subject not to correct position errors and thus allow error to build up. The ILS pathway was quite difficult to fly, and the larger the error build up, the more difficult it was to recover.

A gap of short duration would not, however, explain navigation lapses. One possibility is that several gaps might occur during a short period of time. This would explain the presence of unusually poor scores during the navigation part of the data runs.

Bills (1931; 1935a; 1935b) describes a phenomenon similar to lapses which he calls "blocking." This manifests itself in extra long reaction times of subjects working on a long series of relatively simple tasks. As a result of his extensive studies, Bills (1937) concluded that:

> Blocks are the tendency, shown by practically all persons, on continuous mental work, to show

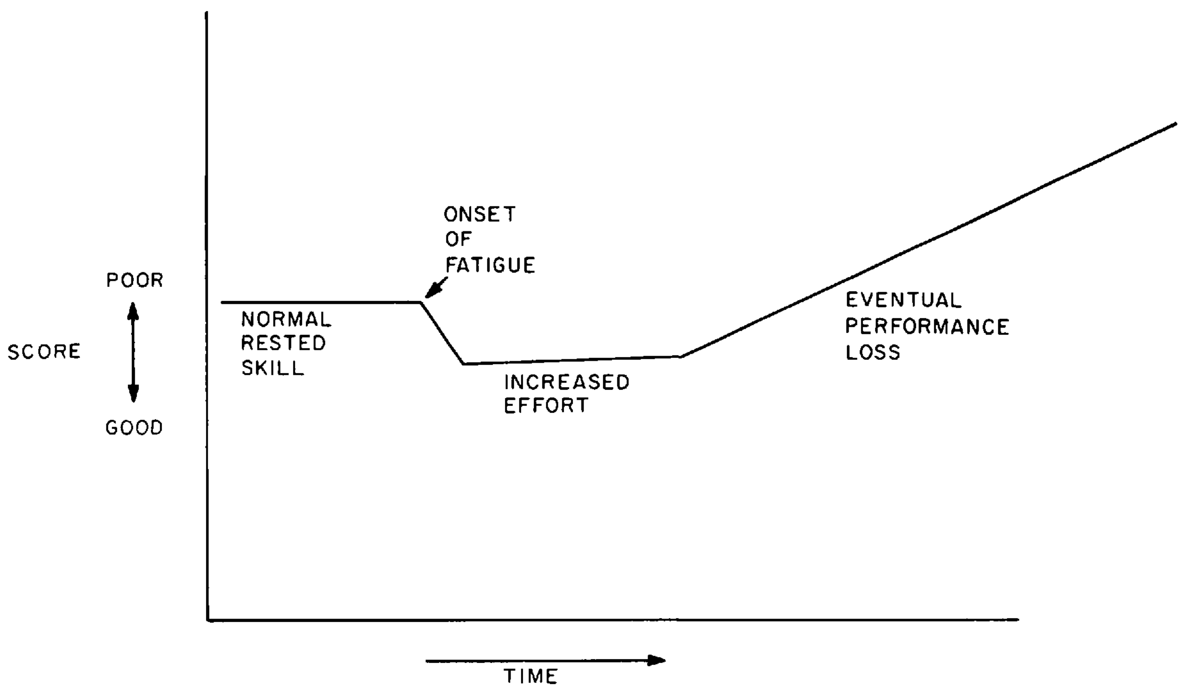

Figure 5. *Hypothetical curve of performance.*

> periodic gaps or pauses in their responses, which they are unable to prevent, no matter how hard they try. These breaks were found to occur about three times a minute on the average and to have a duration of from two to six average response times though practice decreases and fatigue increases their length and frequency, and individuals differ from one another widely.

Bills felt that blocking is the body's reaction to fatigue products in the nervous system. The effect of the block is to rest the brain cells and permit the continuance of mental work without degradation in performance. In a later study Bills (1937) found that a reduction in the amount of oxygen in the air breathed by subjects both increased the number of blocks and prolonged their duration. He reasoned that a momentary lack of oxygen in the brain will cause a block.

It is felt that blocks and lapses are manifestations of the same phenomenon and that lapse behavior is not peculiar to the conditions of this study.

If it can be shown that this phenomenon occurs in actual flight, then a possible causative mechanism for pilot-error accidents has been identified. Further investigation is required for clarification and possible corrective procedures.

ACKNOWLEDGMENTS

This study was sponsored by the National Aeronautics and Space Administration under contract NAS1-11222 Langley Research Center, Hampton, Virginia. The present article is based on material in the final report for this contract (Stave, 1973).

REFERENCES

Bergum and Lehr. Vigilance performance as a function of interpolated rest. *Journal of Applied Psychology*, 1962, *46*, 425-430.

Bills, A. G. Blocking: A new principle of mental fatigue. *Journal of Psychology*, 1931, *43*, 230-245.

Bills, A. G. Some causal factors in mental blocking. *Journal of Experimental Psychology*, 1935, *18*, 172-185. (a)

Bills, A. G. Fatigue, oscillation, and blocks. *Journal of Experimental Psychology*, 1935, *18*, 562-573. (b)

Bills, A. G. Blocking in mental fatigue and anoxemia compared. *Journal of Experimental Psychology*, 1937, *20*, 437-452.

Colquhoun, W. P. The effect of a short rest pause on inspection efficiency. *Ergonomics*, 1959, *2*, 367-376.

Finkleman, J. M. and Glass, A. A. An information processing evaluation of the effect of environmental noise on human performance. Paper presented at the annual meeting of the Human Factors Society, New York, October, 1971.

Jones, G. M. Fatigue effects on radio operators during a program of high intensity long duration flying. *Aerospace Medicine*, 1960, *31*, 478-484.

Kraft, R. Influence of alternating rest periods in static work with a dynamometer. *Travail Humain*, 1965.

Pierson, W. R. Work decrement and endurance in a simple repetitive task. *British Journal of Medical Psychology*, 1963, *36*, 279-289.

Shoenberger, R. W. and Harris, C. S. Psychophysical assessment of whole body vibration. *Human Factors*, 1971, *13*, 41-50.

Stave, A. M. Effects of helicopter noise and vibration on pilot performance. Langley, Va.: National Aeronautics and Space Administration, NASA CR-132347, December, 1973.

Sussman, D. An investigation of factors affecting driver alertness. Buffalo: Cornell Aeronautical Laboratories, Technical Report FH-11-7313, August, 1970.

Wilkinson, R. T. Rest pauses in a task affected by lack of sleep. *Ergonomics*, 1959, 2, (4).