admin

UMTRI-85-49

73041

HUMAN FACTORS TEST OF A DRIVER ALERTNESS DEVICE

Paul Green

DECEMBER 1985

UMTRI The University of Michigan Transportation Research Institute

Technical Report Documentation Page

1. Report No. UMTRI-85-49	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle HUMAN FACTORS TEST OF A DRIVER ALERTNESS DEVICE		5. Report Date December 1985
		6. Performing Organization Code 388641
7. Author(s) Paul Green		8. Performing Organization Report No. UMTRI-85-49
9. Performing Organization Name and Address The University of Michigan Transportation Research Institute Huron Parkway & Baxter Road Ann Arbor, MI 48109-2150 U.S.A.		10. Work Unit No.
		11. Contract or Grant No.
12. Sponsoring Agency Name and Address Amway Corporation Research & Development - Home Care Products 7575 E. Fulton Road, Mail Shop 50-2E Ada, MI 49355		13. Type of Report and Period Covered Final Report 9/1983 - 7/1985
		14. Sponsoring Agency Code
15. Supplementary Notes		

16. Abstract

This project assessed the effectiveness of a prototype driver-alertness device that measures driver response time to a tone. In previous studies response time has been found to be a indicator of fatigue. Fatigue is often a contributing factor to accidents, though the percentage varies greatly with the data base and the method of calculation.

Eight men drove a driving simulator for two-hour sessions very late at night. Across sessions drivers were exposed to four test conditions--device with a fixed interval tone (every 20 or 60 seconds), device with a random schedule (between 20 and 60 seconds), listening to the radio, or nothing (no device or radio). While driving, data on steering error, the interval between heart beats, response times to the device and intervals between tones, and experimenter ratings of the subject's alertness, were recorded. Participants were most alert while listening to the radio and least alert with nothing. Of the two versions of the device, the greatest levels of alertness were associated with the random mode. A post-test survey led to several suggestions regarding the button, tone, device placement, and other aspects that might be used to improve the prototype.

17. Key Words Human Factors, ergonomics, engineering psychology, human engineering, fatigue, driving, alertness devices, response time, simulators		18. Distribution Statement	
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 83	22. Price

ACKNOWLEDGMENTS

I would like to thank the University of Michigan Engineering students who contributed to this project: Sue Brian for her preliminary examination of the simulation architecture; Alan Olson for his development of the assembly language graphics routines for the simulator; Mike Scheller for his work on the simulation user interface and checkout of the prototype alertness device; Stacy Reifeis for her help in assembling the literature and accident statistics; and Kara Heinrichs and Donald Ottens for carrying out the test.

I thank the Amway Corporation for its support of this research. Because research involves exploring the unknown, it is often difficult to estimate how long something will take and what it will cost. This project had more than its share of twists and turns, and Amway's great patience in seeing it through to completion is appreciated.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Scope

This report describes a test of a prototype driver-alertness device. The purpose of the test was to establish whether it is effective.

To put the test in perspective, something needs to be said about what driving fatigue is, and what research has been done on it. The funded project work did not include a detailed literature review, and hence this report is by no means a complete review of accident statistics, surveys, or human performance studies on the subject.

Readers interested in a more comprehensive review should look at Olsen and Post (1979), Nagasuka and Ohtz (1979), Ryder, Melio, and Kinsley (1981), and McDonald (1979).

Definition of Fatigue

Driving fatigue is clearly an issue of major concern. According to Brown (1982), driver sleepiness or fatigue is the second leading causal factor of traffic accidents in California. (Alcohol is first.) Driver fatigue or drowsiness is a serious problem that will probably affect more than two-thirds of all drivers sometime in their life (Snook, 1976), often leading to near fatal or fatal accidents.

Many definitions of fatigue have been offered. Lisper, Laurell, and Stening (1973) define fatigue as "falling asleep" at the wheel; Attwood and Scott (1981) say fatigue is the "drowsiness" that is induced either through time at the wheel or by sleep loss; and Snook and Dolliver (1976) consider fatigue as "driver drowsiness." Van Der Nest (1978) defines fatigue in terms of symptoms: (1) decrease in attention; (2) decrease in motivation; (3) slowed and impaired perception; (4) impairment in thinking; (5) decrease in performance speed; (6) increase in errors; and (7) decrease of performance capability for physical and mental activation. Brown (1972) notes that fatigue can be "physical" or "psychological." He defines fatigue as the subject's increasing disinclination and unwillingness to continue performing the test at hand. Further, he notes that psychological fatigue differs from physical fatigue, in that its onset and recovery from it can be sudden. Hulbert (1972) states that there are two categories of fatigue, "task-induced" (driving) and fatigue due to non-task factors such as sleep deprivation. Ryder, Manlin, and Kinsley (1981) likewise draw distinctions between the various types of fatigue. Thus, as McFarland (1971, p. 1) puts it, "Definitions of the nature of fatigue are almost as numerous as the articles that have been written about it."

In this experiment, fatigue was defined as the inability to perform well due to a lack of sleep, an unnatural break in the circadian rhythm, or excessive "time on task," such as driving a car. Those forms of fatigue are distinguished from "whole body fatigue" (e.g., the weariness experienced at the end of a marathon), and "localized muscle fatigue" (e.g., the lack of strength in a pitcher's arm at the end of a game). Both of those types of fatigue were excluded from consideration.

Accident Statistics

Driver fatigue is often reported to be a contributing cause of traffic accidents. Table 1 shows some of the percentages. Even from a quick examination of this table, it is clear that the accident figures differ widely. Sometimes that happens because the figures are compiled in different ways. For example, the percentage of <u>all accidents</u> (including those for which the cause is unknown) in which fatigue was identified as a contributing factor varies from .2 to 1.8%. For the percentage of accidents <u>for which the cause was known</u>, .2 to 2.4% are fatigue-related. However, when one changes the baseline to accidents for which <u>some contributing cause is identified</u>, the range is from .2 to 54.5%.

Differences also arise due to the collection procedures that are unique to each data base. FARS (Fatal Accident Reporting System) is a data base maintained by the federal government. It contains information on every fatal motor vehicle accident in the United States and is based on police-collected information.

The Tri-Level statistics come from in-depth investigations of accidents performed by specially trained teams. Both the FARS and Tri-Level data are high quality.

The data from the states of Michigan, North Carolina, Texas, and Washington are all based on police-collected information for all accidents. Only the Texas data must be viewed with some caution. The Texas reporting scheme does not allow for an "unknown cause" (shown as 0% in the table). (Those accidents are sometimes classified as "apparently normal.")

Finally, BMCS contains occasionally suspect, self-reported data on interstate motor carrier accidents (mostly truck) that occur in the United States. Since most interstate trips are long, one would expect more fatigue-related accidents to occur. Thus, accident statistics must be cited with some care. While fatigue is commonly a cause, the estimated frequency depends very much upon how it was computed and where the numbers came from.

TABLE 1

FATIGUE AS A CAUSAL FACTOR IN ACCIDENTS (%)

	BMCS[1]	FARS[2]	Michigan[3]	North Carolina[4]	Texas[5]	Tri-Level[6]	Washington[7]
1. Unknown Cause	0.2	1.8	44.9	4.2	0.0	0.0	84.4
2. Known Cause	99.8	98.2	55.1	95.8	100.0	100.0	15.6
2a. No Contributing Cause	96.5	68.5	42.8	91.4	99.0	5.0	5.6
2b. Fatigue	1.8	0.7	1.3	1.3	0.5	0.2	0.2
2c. Other Cause (not none (2a) or fatigue (2b))	1.5	29.0	11.0	3.1	0.5	94.8	9.8
% of known due to fatigue (2b/2)	1.8	0.7	2.4	1.4	0.5	0.2	1.3
% of the "some cause" due to fatigue (2b/(2b + 2c))	54.5	2.4	10.6	29.5	50.0	0.2	2.0

Sources:

1. Bureau of Motor Carrier Safety (BMCS), 1980. HSRI Accident Data System Codebook, Number 81–10. The University of Michigan Transportation Research Institute, Ann Arbor, Michigan, November, 1982.
2. Fatal Accident Reporting System (FARS), 1975–1980. HSRI Accident Data System Codebook, Number 82–8. The University of Michigan Transportation Research Institute, Ann Arbor, Michigan, September, 1982.
3. Michigan, 1981. UMTRI Data System Codebook, Number 81–10. The University of Michigan Transportation Research Institute, Ann Arbor, Michigan, November, 1982.
4. North Carolina Traffic Accident Summary, January-December, 1982. North Carolina Department of Transportation, Division of Motor Vehicles. Raleigh, North Carolina, 1982, as published in Accident Facts, 1983 Edition, National Safety Council, Chicago, Illinois, 1983, 55.
5. Texas, 1982. UMTRI Data System Codebook, Number 83–10, The University of Michigan Transportation Research Institute, Ann Arbor, Michigan, May, 1983.
6. Treat, J.R., Tumbas, N.S., McDonald, S.T., Shinar, D., Hume, R.D., Mayer, R.E., Stansifer, R.L., and Castellan, N.J. Tri-Level Study of the Causes of Traffic Accidents, Executive Summary. Technical Report #DOT HS-805 099, Washington, D.C., 1979.

Several studies have examined the accident data in somewhat greater detail. Harris (1977), using BMCS data, found that the highest percentage of accidents occurred between 4 and 6 a.m. In addition, twice as many accidents occurred between midnight and 8 a.m. (66%), when one would expect drivers to be more tired than in the other 16 hours of the day (34%), particularly since most driving is done in the daytime. Further, 88% of the "dozing-driver" accidents were single-vehicle accidents. The remaining 12% were collisions with other vehicles, most of which were rear-end collisions, again suggesting fatigue as a factor. The data also suggest that heart rate might be a good predictor of fatigue, since accident rates and average deviation of the heart rate (from the daily mean) were negatively correlated. When the data were examined from the perspective of consecutive hours of driving, periods of three hours or more were clearly associated with greater-than-expected accident rates.

Kearney (1966) reported that in one year the Ohio Turnpike charged 50% of its fatalities to drowsiness-related accidents. He also noted that on virtually all of the major turnpikes, "driver asleep" ranked well ahead of speeding as a major cause of fatalities. The Oklahoma Turnpike Authority suggest falling asleep is a particular problem for young drivers: 50% of all the drivers who fell asleep at the wheel were less than 24 years of age, and 78% were below 34 years. Other evidence suggests that older drivers (over 45 years old) should be the focus of attention. They tend to be more susceptible to fatigue after driving for a shorter period of time than young drivers (Ryder, Manlin, Kinsley, 1981).

In a related context, Wilde and Stinson (1983) did an extensive investigation on the fatigue of locomotive engineers. Of the 222 fatal and injury-producing train collisions and derailments, as reported by Kruz in 1972, 75% were attributed to human performance shortcomings, mostly of the train crew. Among the descriptions of human performance shortcomings was "crew asleep."

The accident data provide ample evidence that drowsiness is commonly a factor in traffic (and other) accidents most often at night, hardly a surprising result. However, the data are inconsistent in terms of an overall estimate or which age group is affected.

Driver Surveys

Paralleling the accident data have been surveys asking drivers how and when they have experienced driving fatigue. Because of the long hours and miles traversed by long-haul truck drivers, it is often reported they have relatively more fatigue-related problems than car drivers. McDonald (1979) reviewed the literature on heavy truck driving in the UK. In examining the work of Prokop and Prokop (1955), he noted that of 569 drivers interviewed in motorway cafes, 18% admitted that they had fallen asleep at the wheel "sometime." In a study by Tilley (1973) McDonald noted that 64% of 1,500 driving license renewal applicants said that they "had become drowsy while driving"; 6.4% said that fatigue or sleepiness had caused near accidents; and .64% of them said that they had at least one accident under these circumstances. McDonald noted, among other things, that statistics underestimate the full contribution of fatigue-related accidents. In reference to his investigation relating heart rate to fatigue, he said that a decrease in heartbeat should be seen as test times wear on, especially after 11:00 p.m.

Linklater (1980) interviewed 615 truck drivers and 551 other motorists at eight locations on major New South Wales roadways. The number of hours spent behind the steering wheel of a vehicle in a typical week was used as the variable to indicate driver fatigue. He found that truck drivers had significantly more traffic crashes than other motorists because of their longer driving hours. When a compensatory factor was used for relative exposure rates, there was no significant difference between truck drivers and other motorists in terms of numbers of traffic crashes reported. Research indicated that truck drivers who drove 55–75 hrs/week were more likely to experience fatigue-related accidents, than those who drove less than that. Linklater noted that of the accidents he investigated, road design was a contributing factor, and that it might be possible to specify certain simple design features to reduce the occurrence of fatigue-related crashes, such as a different surface texture on the road pavement in the form of a rumble strip.

Storie (1984) describes a driver survey that was a follow-up to almost 1,000 accidents. About 1,500 responses were received from drivers. Of those responding, only 3% reported they were sleepy, a surprisingly low figure if drivers responded truthfully.

It is evident from these surveys that many drivers have fallen asleep while driving. Fortunately, not all of those events led to accidents. Those personal experiences can make drivers aware of the dangerous consequences of fatigue and the need to do something about it.

Previous Research on Driver Performance

A substantial amount of research has been completed on driver fatigue, of which seven studies are especially pertinent.

Dureman and Baden (1972) had eight subjects steer a crude driving simulator for several four-hour test sessions. At random times (.5 to 3.5 minutes, mean = 3) a .65 kHz tone (70 db) was presented. Half of the subjects received a mild electric shock when the steering error was large, the other half heard a click when that occurred. Performance measures included pulse rate, skin resistance, respiratory rate, EMG, and reaction time to the tone.

They reported that steering errors increased with time, and for the non-shock group, errors were significantly correlated with reaction time ($\underline{r}$ = .41). Also significant were the correlations of steering error with heart rate ($\underline{r}$ = −.52) and respiration rate ($\underline{r}$ = −.38). Only the correlation with heart rate was significant for all of the four non-shock subjects.

Yajima, Jkeda, Oshima, and Sugi (1976) report two extensive studies of driving performance. In the first, involving 25 drivers, five cars were driven in a caravan on a freeway. Tests lasted 7-10 hours, with breaks for rest and load. The five vehicles had somewhat different instrumentation in them, so not all performance measures were collected for all drivers. Typically control flicker frequency, blood pressure, reaction time (simple and choice), and fatigue ratings were collected. In a number of instances EEG, EOG, EKG, GSR, and inspiratory volume were collected. Details of how the data were collected—for example, of what the stimuli were in the reaction time test—are not given.

In a second series, two cars were driven for up to 24 hours around a test track. The number of subjects and performance measures were unspecified.

Yajima et al. report there were no significant changes in the mean heart rate with time. However, reaction time variance did increase with time, though it is unclear if the change was statistically significant. No mention is made of mean reaction time. Thus, the utility of the behavioral measures is difficult to assess because of the lack of statistical analysis with performance measures, such as steering/lane position error, with which these measures might be compared.

Snook and Dolliver (1976) had ten people drive a car simulator for three-hour test periods. Each person completed six experimental sessions, using one of four countermeasures (music, news programs, lateral position feedback, and speed information feedback), preceded by six sessions without them. Measures of performance included mean lateral position error, speed variation (lap time standard deviation), mean steering reversals, mean heart rate, and fatigue ratings.

An analysis of variance indicated that lateral position error (measured by partitioning the lane into eight error bands), speed variation, and subjective ratings of fatigue increased significantly with driving time, while heart rate decreased significantly. Lateral position error (staying centered in the lane), was found to be the most sensitive indicator of fatigue. With regard to the four countermeasures, subjects said that the music and speed feedback countermeasures made them feel less tired, but their views were not verified by the performance data.

Riemersma, Biesta, and Wildervanck (1977) performed two experiments using a Volvo 145 instrumented to record the driver's heart rate, base position, steering wheel reversals, speed, and longitudinal acceleration. In addition, the instrumentation provided a means to identify when driving was other than straight ahead—for example, when passing.

Only the first experiment examined fatigue in detail. In that experiment twelve subjects repeatedly drove a 6 km circuit of divided highways ten times. Subjects drove from 10 p.m. to 6 a.m. with a stop at 2 a.m. for fuel. While driving, subjects were asked to report each time they had driven an additional 20 km. In addition, drivers reacted to changes in color of a light mounted on the top of the dashboard by pressing the horn button. Changes occurred randomly every .5 to 4.0 minutes.

The results showed statistically significant increases in reaction time, and errors in reporting distance traveled, the standard deviation of lane position, and speed, with time. Also, both the standard deviation and mean interbeat intervals increased significantly with time.

Fagerstrom and Lisper (1977) had 12 people drive a Volvo 145 for just over three hours at high speeds over 370 km of Swedish roads. Participants were tested under three conditions—listening to radio talk shows, listening to radio music, and silence (no radio). About every 50 seconds while driving (range 10 seconds to 2 minutes), a 90db/1000 Hz tone was sounded, which the driver turned off with a foot switch. The results showed a steady decrease in heart rate over time (about 10 beats/minute over the test session). Reaction times (an indicator of alertness) increased linearly with time, though the increase was smaller while listening to the radio, for experienced drivers, and for introverts (as determined by an Eysenck Personality Inventory).

Egelund (1982) had eight inexperienced drivers drive a Volvo 145 estate car on a 340-km four-hour country route. He measured three physiological variables: heart rate variability (HRV) in the .05 – .15 Hz region, the standard deviation of HRV (S.D. HRV), and heart rate (HR). He found there was not a significant relationship between S.D. HRV and HR (heart rate being the most widely used measure in determining driver fatigue). He did, however, find filtered HRV to be correlated with distance driven and concluded it seems to be a sensitive indicator of driver fatigue.

Attwood and Scott (1981) had four people drive around a closed 7.2 km. oval track.

Each person drove at 80-kph for three hours on the first test day, stayed awake for the next 21 hours, and then drove for an additional 3 hours. At the end of two weeks they returned for the same test a second time. The drivers wore headphones that played music, except when drivers wandered out of their driving lane or were judged to be dozing off. Then a loud 1000-Hz. tone interrupted their music for approximately 3 seconds. The performance of each subject was monitored by an on-board computer. Driver drowsiness was judged by a multivariate decision rule that was based on 30-second averages of data from each trial. (See Eatock Demmery, Williams, and Attwood (1978) for a description of the rule.)

Together these experiments show there is a good correlation between mean heart rate and fatigue, and that both listening to the radio and responding to loud auditory tones may combat it.

Countermeasures

Large-scale campaigns against drinking and driving are being conducted now in the United States. But the most one hears about fatigue is the well-used expression "Stay Awake! Stay Alive." In the past, many countermeasures have been proposed. To improve railroad safety, Wilde and Stinson (1983) proposed the Device for Attention Monitoring and Excitation (DAME). The device, mounted in a locomotive cab, employs three lights (green, yellow, red). Passing a block signal on the side of the tracks turns on the green light. If a button is not pressed within a specified time, the yellow light is turned on. If there is no response to the yellow, then both an auditory alarm and the red light go on. If all three warnings are not turned off, then the brakes are applied and the train will come to a halt.

In addition, Wilde and Stinson have outlined some useful ideas regarding fatigue countermeasures.

1. In designing a fatigue countermeasure, one must be careful not to take the driver's attention away from the road while activating the device.

2. Any vigilance device defeats its purpose if it becomes a distraction to a driver, with a feeling of "working against them" often arising.

Hulbert (1972), in reviewing proposed countermeasures in his report, notes the difficulty of collecting reliable evidence to substantiate the general value of some methods. He states that there may in fact be no reliable method for most people. Some of the countermeasures he found in the literature include:

1. Singing, chewing a pack of gum, taking off the right shoe, or sitting on something hard (Harris, 1967).

2. Electronic Transistor Safety Alarm – a plastic device that curls around the driver's ear, and buzzes when the driver's head tilts (Traffic Safety, 1959). This is similar to current products ("Sleeper Beeper"/"Driver Alert") now on the market.

3. Button Steering Wheel Alarm – a device that plugs into the car radio. The driver cannot release the button without triggering alarm (Traffic Safety, 1959).

4. Alertmaster – a pedal positioned left of the clutch that, when activated, must be maintained with light pressure or else the horn will sound (Williams, 1966).

5. Alert-O-Matic – a mechanical device wired into the electrical system of a car. It produces a sequence of three alerting (first visual and then auditory) signals of increasing severity before it turns off the ignition (Frederik, 1966).

6. Highway-To-Car-Communications System – a roadside system that initiates a warning signal within the driver's car. The driver must then turn it off manually (Delco Radio, 1961).

Other anti-fatigue techniques, such as amphetamine drugs and caffeine, have been less vigorously investigated. Van Der Nest (1978) found that rest periods significantly decreased the influence of fatigue on driving. As is commonly known, the best way to avoid fatigue is to drive when one is well rested, take frequent rest stops, and drive during the daytime when one is normally awake.

TEST PLAN

Test Overview

As is evident from the previous discussion, fatigue is a significant traffic safety problem. It is commonly experienced by drivers. While not every instance of falling asleep results in a crash, fatigue is a frequent causal factor. A number of the studies have data on response time, but none contain the specific test data of interest to the sponsor of this research.

The purpose of this test was to evaluate a particular alertness device. During the test participants sat in a mockup of a car. They "steered" it in response to a dynamic nighttime road scene shown in front of them, similar to that of a simple video game. At various times in the experiment, the alertness device generated a tone. The participant turned off the tone by pressing a button. Two computers continuously monitored the participant and equipment, recording steering error, the time between tones, response times to it, the time between each heart beat, and periodic ratings of the participant's alertness.

Driver alertness was examined in four test conditions. The conditions were (1) responding to the alertness device with a fixed interval between tones (about every 60 seconds), (2) alertness device with a random interval (ranging from 20 to 60 seconds), (3) listening to a car radio, and (4) no device or radio (control). When the alertness device was used, the radio was not.

Each person completed one test session for one condition each evening. The order of conditions was counterbalanced across subjects, so that practice effects could be examined.

People Tested

Eight men ("morning people") in good health served as participants in this experiment. All were licensed drivers. Four were younger (ages 21 to 25) and four were older (54 to 74). The younger men, all students at the University of Michigan, were friends of an experimenter. The older men were reliable individuals who had participated in previous UMTRI studies and indicated an interest to continue to do so. Most of them lived in local retirement communities.

All participants were paid $7.00 for a one-hour daytime screening session and $25.00 for each of the four test sessions ($100.00 total). When equipment failures made it necessary to repeat sessions, subjects were paid at the same rate.

Test Equipment and Materials

The hardware used in this test included a mockup of a car, a rear-projection screen, a video projector, a Commodore 64 computer system, an IBM personal computer, a low-light level video system, the alertness device, and several miscellaneous items. The general arrangement of the equipment is shown in Figures 1 and 2.

All tests were conducted with the subject seated in a full-size mockup of a 1982/3 Ford Escort/Mercury Lynx. The car had a production steering wheel and the low-line 2 speaker AM/FM radio found in the 84 Escort. The steering wheel was linked by ropes to bungee cords, giving the system a spring-centered feel. The steering system's resistance was light, and it pulled slightly to the left.

Shown on a large screen in front of the vehicle were six pairs of rectangles simulating road edge markers for a single-lane road as it would appear at night. (Figure 3 shows what the driver saw.) To enhance the visual qualities of the simulation and avoid forcing the subjects to stay awake, tests were conducted in a windowless room with the lights off. There was some illumination from the experimenter's worklight scattered around the barrier, the glow from the car radio face, and scatter from the projection display.

The road scene was generated by a Commodore 64 computer, which was connected to a video projector. The Commodore was also responsible for storing the participant's steering error data. A color monitor used with the Commodore displayed a duplicate copy of the road scene to the experimenter.

A proprietary UMTRI-developed assembly language program loaded by a BASIC language IO program generated the road image. Deviations of the road center were based upon repeated use of 400 point sequences, computed by adding together four nonharmonic sinusoids. The sequence repeated about every 3.5 minutes. To the subject each road looked continuous and curved in an unpredictable manner.

For subject's warmup runs, function [1] ("Huron") generated the road. (The road scene reminded one of the programmers of Huron River Drive, a winding road in Ann Arbor.) For test sessions, roads were generated using function [2] ("Observ", short for Observatory Avenue). There were four variants, one for each test condition. They included the pattern itself, a pattern in which left and right were reversed, and the previous two patterns run backwards. Because of roundoff errors, these reversed patterns

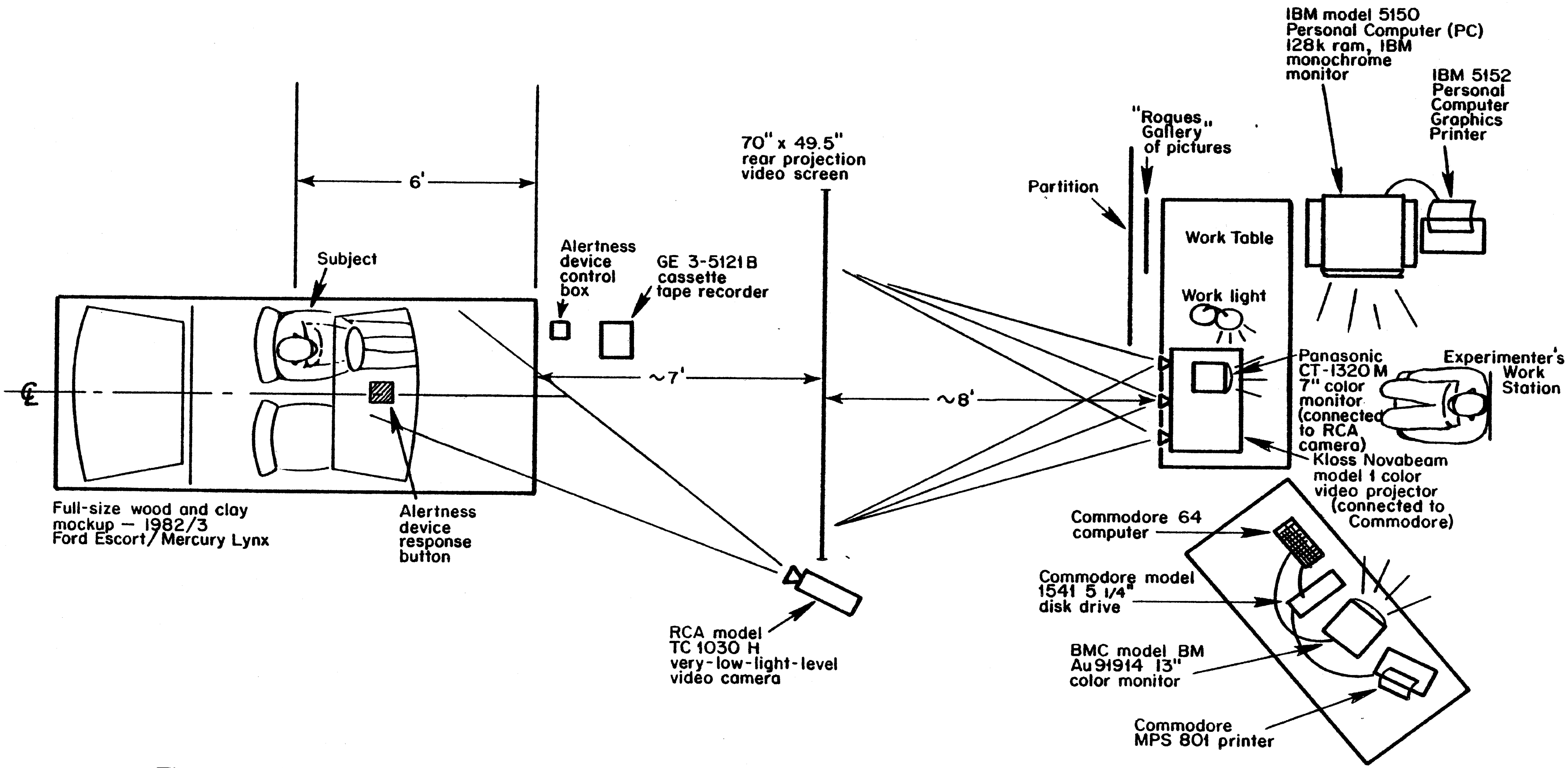


Figure 1. General arrangement of equipment.

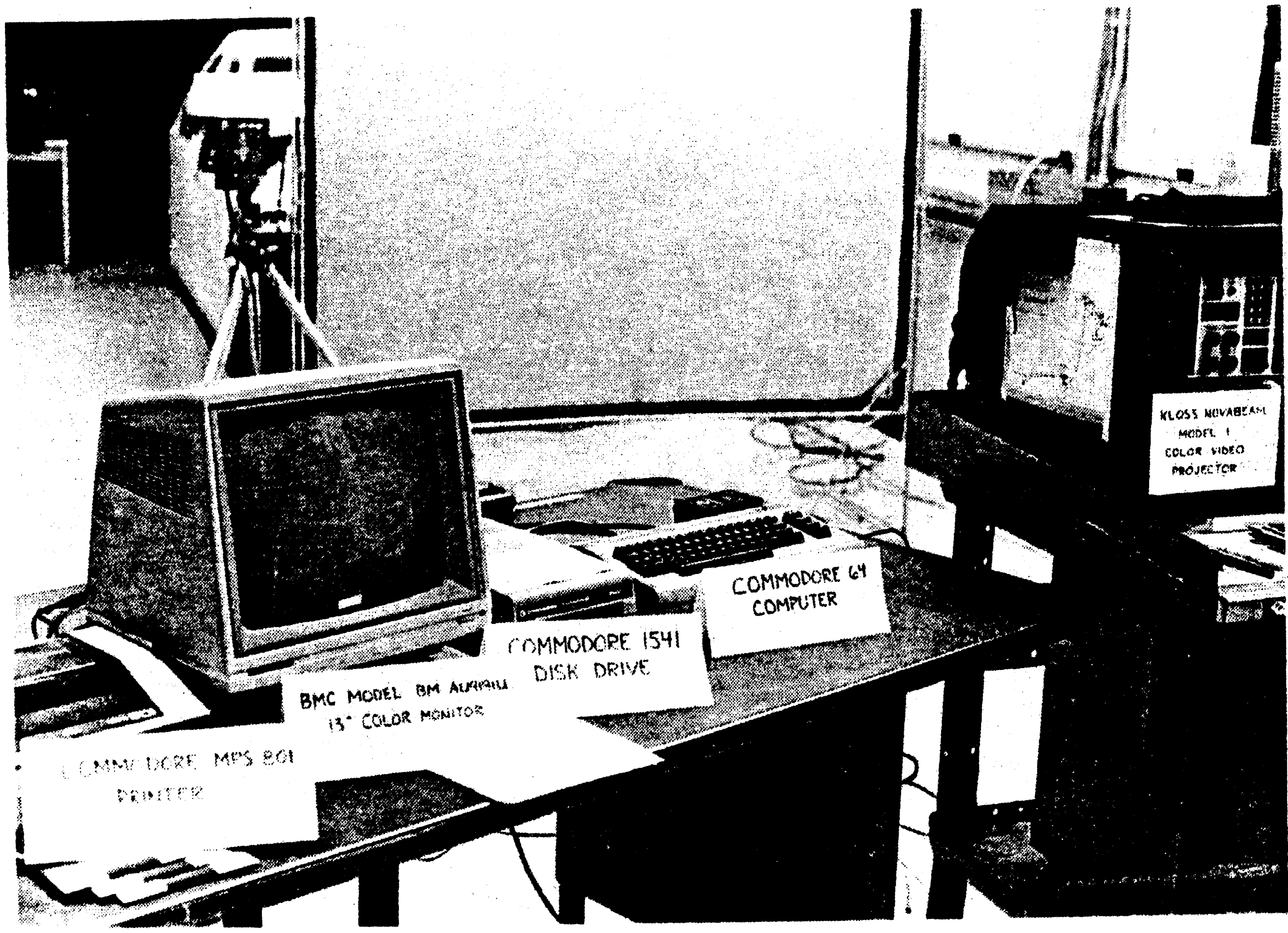


Figure 2. Experimenter's workstation.

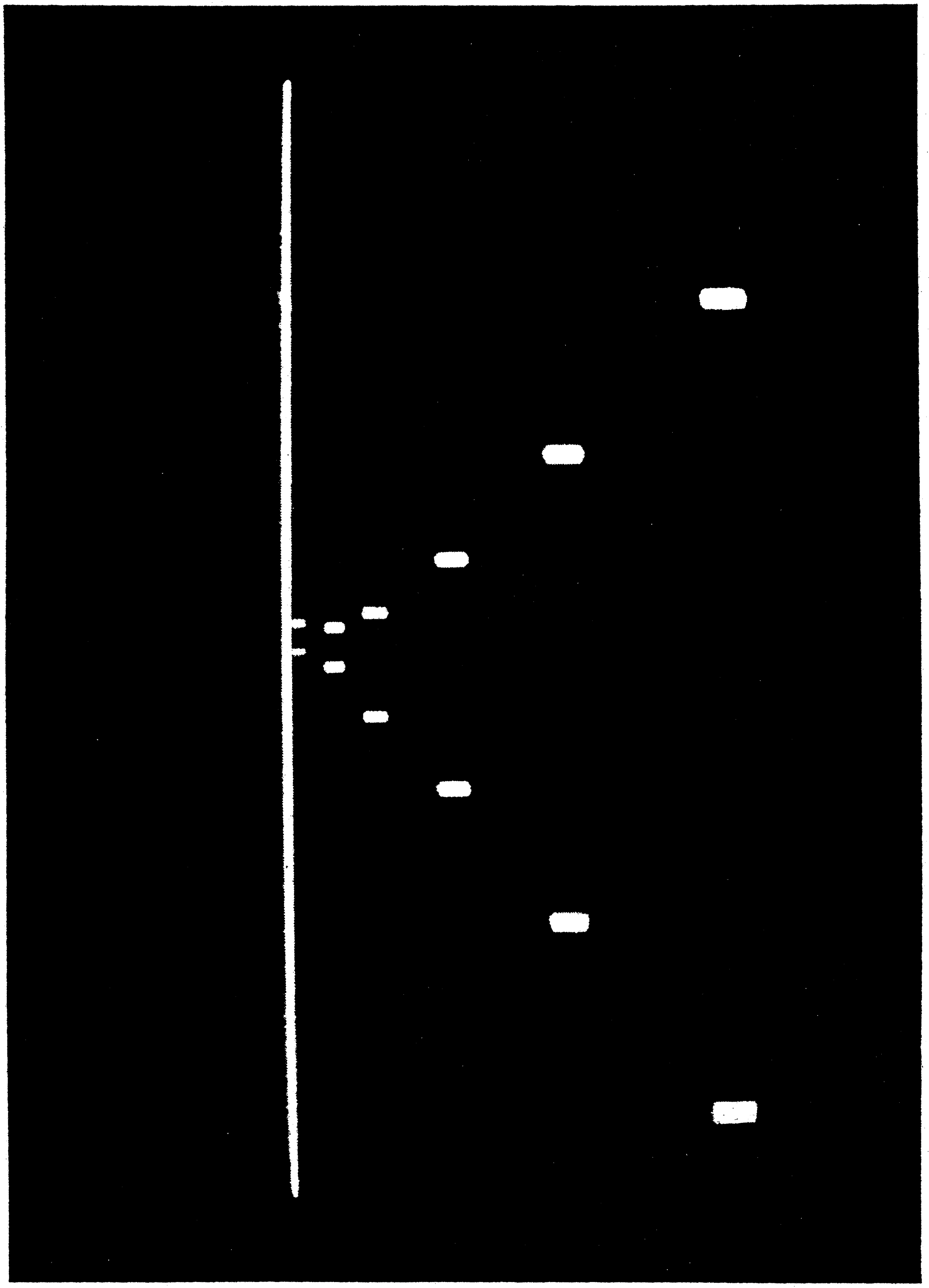


Figure 3. Road scene.

were not exact reversals. These manipulations led to groups of four different patterns of identical difficulty.

[1] $\text{deviation} = .9\sin(1.1t) - .7\sin(1.7t) + .9\sin(1.9t) - .4\sin(3.7t)$

(Huron)

[2] $\text{deviation} = .9\sin(1.1t) - .7\sin(1.7t) + .9\sin(1.9t) - .4\sin(2.9t)$

(Observ)

While steering the vehicle, people responded to a prototype alertness device developed by the Amway Corporation. The prototype consisted of a 6 x 6 x 3 inch control box (see Figure 4) and a 3-3/4 x 2-1/2 x 1-1/4 inch response box with a 13/16 x 1-7/16 inch short travel pushbutton on it. The response box also contained a tone generator (constant 5.1 kHz tone, 100 dBA at the driver's ear). The response pushbutton was mounted on the top of the instrument panel near the centerline of the vehicle. (See Figure 5.) It was about a 10-inch reach for the right hand from the 2 O'clock position on the steering wheel. Next to the button was a small LED that illuminated when the tone sounded.

The alertness device had three operating modes—no tone, fixed-interval, and random interval. Normally in the fixed-interval mode the tone would sound about every 60 seconds for three seconds and then cease. If the button was not pressed within six seconds of when the tone sounded, the tone would go on and stay on until the button was pressed. The next tone would sound 20 seconds later. The device would continue to sound at 20-second intervals until two consecutive response times were less than three seconds. In the random mode, the fixed 60-second interval was replaced by one that was randomly distributed between 20 and 60 seconds.

To enhance the simulation and mask extraneous noise, a cassette recorder played a 60-second continuous loop of "road noise" (65 dBA-slow) during the test. The recording was of a 1975 Honda Civic (with a defective muffler) revving at about 1500 RPM.

Input for measurement of the participant's heart rate came from three chest electrodes linked to a Respironics Exersentry personal heart rate monitor. The electrodes were mounted on a webbed belt worn beneath the participant's shirt. The Exersentry was in turn connected to an IBM Personal Computer (PC) with a custom-made digital IO card to handle input from the heart rate monitor and the alertness device, and take care of timing. Using a BASIC language program, the PC recorded the time between button presses to the nearest second. It also recorded the time from when the tone went on until

Figure 4. Prototype control box.

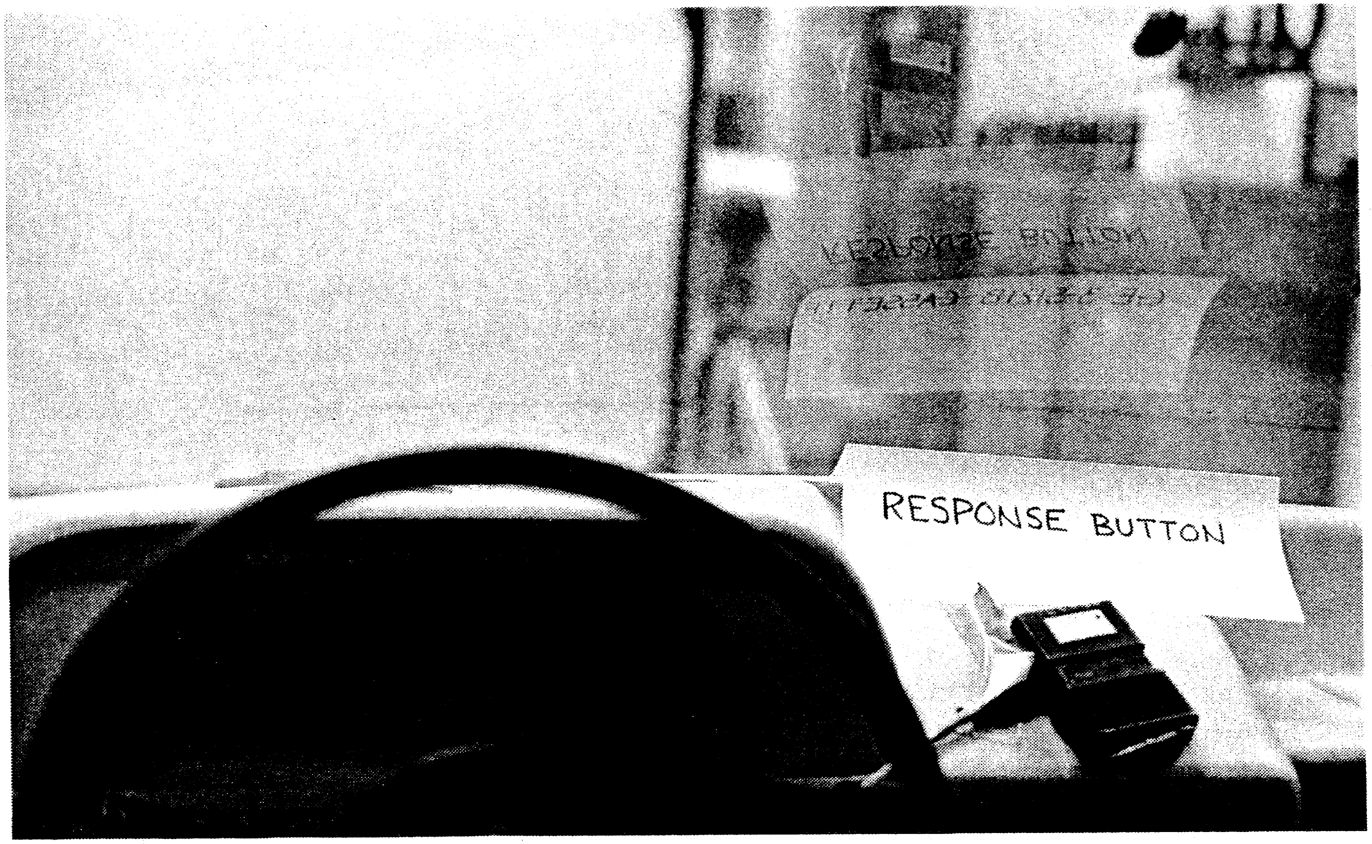


Figure 5. Response box installed on the instrument panel.

the subject turned it off (response time), and the interbeat intervals of the heart to the nearest millisecond.

In addition, the PC recorded experimenter comments, and every 10 minutes cued them to make the fatigue judgments. Judgments of how alert the participant appeared were made by viewing the subject on a small video monitor that was connected to a very-low-light-level video camera. Judgments were on a 1 to 5 scale (1=alert, 5=asleep). The video images were compared with pictures of four young men sitting in the test mockup in various states of alertness ("the Rogue's Gallery" – Appendix A). The pictures were mounted on a posterboard by the experimenter's station.

Test Activities and Their Sequence

Each subject was recruited using the instructions in Appendices B and C. Each participated in a one-hour screening session and four three-hour late night test sessions. The screening sessions verified that the participant's heartbeat could be reliably recorded by the computer and they could perform the steering task. See Appendix D for the screening session protocol. In that session the instructions were read and participants completed a biographical information (Appendix E). After fitting the electrode belt to the participant, the consistency of the heart beat signal was checked. To do this, the experimenter watched the LED on the heart beat interface box to cycle on and off for every heart beat, observed the heart rate on the Respironics digital display, or listened for the Respironics box to beep (by disconnecting it from the computer).

Subsequently, the subject was shown how to turn off the alertness device tone (by depressing the instrument panel mounted button). In addition, proper operation of the heart beat recording software was reconfirmed by watching the heart beat display near the computer blink and the screen display of the number of beats update.

With the preliminary steps completed, the experimenter explained and then ran the driving simulation program on the Commodore. The participant completed several 2–3 minute trials and received feedback as to how well he was steering. When participants had difficulty, the experimenter stood next to the car and told the participant which way to turn the wheel. If there were no problems, the participant was invited back for a test session.

Test sessions consisted of usually two 2–3 minute warmup blocks of driving followed by the main two-hour test session. Test sessions took place late at night to assure that the participant was fatigued. For the young participants, often two were tested per evening, one from 11:00 PM until 1:45 AM (driving from 11:30 until 1:30) and a second from 2:00 AM until 4:45 AM (driving from 2:15 until 4:15 in the morning). Older subjects were usually tested from 10 PM until midnight. Most reported they usually went to bed at 10, much earlier than the young subjects. Participants were tested once or twice a week, usually with at least two days rest between sessions to avoid a shift towards night of the person's daily cycle.

At the end of the final session the subject completed a questionnaire concerning the alertness device's features. (See Appendix E.)

RESULTS

Data Reduction

The data analysis was performed in four steps—screening out the bad data and generating minute-by-minute summaries, uploading the data files to the University's mainframe, examining correlations between variables to establish which were related to fatigue or steering performance, and finally, looking for statistically significant condition differences using analysis of variance. The by-minute means and standard deviations were computed using custom programs written in BASIC for the IBM PC. The correlations and related initial analysis was carried out using MIDAS (Michigan Interactive Data Analysis System, Fox and Guire, 1976). Most of the analysis of variance calculations were made using the UCLA BMDP package (BioMeDical statistical software – P series procedures, Dixon, Brown, Engelman, Frane, Hill, Jennrich, and Toporek, 1985).

There were far more than normal problems encountered in collecting this data. They seemed to occur at random. The average subject had to return for at least one extra two-hour session (block) because of hardware and software failures. For example, in one instance the potentiometer used to measure steering wheel position came loose and broke the connection to the computer. In another instance, the steering wheel was turned so hard that it went past the mechanical stop and destroyed the potentiometer.

Software-related problems were more common. At least four blocks were lost because of incorrect disk management procedures (re-using file names, not checking for sufficient space on a disk) and one was lost because of a programming error that caused the program to stop prematurely at midnight rather than continuing well past it. Finally, in one block the subject fell asleep, and when he awoke, he could not get back on the road. All of these events should have been trapped by the software so that the data weren't lost.

While this was intended to be only a preliminary study, nonetheless a considerable amount of data were collected. It is estimated that from the screening sessions, four two-hour test sessions per subject, and repeated sessions, 96 hours of data were accumulated for the eight test subjects. During each test, the steering error was recorded every three seconds, the time between each heart beat was measured to the nearest millisecond, the time between beeper actuations and associated response times was recorded, and how alert the each subject appeared was recorded every ten minutes. The raw data and data summaries almost filled 25 360K floppy disks, and additional large summary files were created on the University's mainframe.

Data reduction was carried out to make the analysis manageable and reliable. Only blocks for which there was complete data for all measures were used. Most of the initial analysis was based upon 3,840 (8 subjects x 4 blocks/subject x 120 minutes/block) cases for each variable. Had the raw data been used, for example for steering error, the number of cases would have increased by a factor of 20, with consequent increases in uploading time to the mainframe, analysis cost, and analysis time. While the summary process can obscure brief but severe changes in the data, for example, where a person falls asleep and stopped steering for a moment, those instances tend to be repeated over a short time period and should be apparent in the by-minute summaries.

Fortunately, the procedure for determining the mean and standard deviation for each minute of the steering error was straightforward, as there were no missing data.

Summary of the Intertrial Interval and Response Time (RT) data was not so simple because of several minor problems with the design of the experiment. First, because of a peculiar aspect of the IO card, the interrupt generated by a response allowed the previous and not the current response time to be retrieved. Second, response times were measured to the nearest millisecond, while intertrial interval times were only to the nearest second. Assuming that responses occurred on the average of once per minute (the actual rate tended to be higher), at the end of the experiment this could lead to a cumulative error as large as two minutes as to when an event occurred. Finally, the steering error and RT data were collected by different computers that were not fully synchronized, though they did start within 30 seconds of each other. This made it difficult to pin down exactly when a response occurred. In most cases the last button press in a block was lost because of the lags and lack of synchronization.

During pilot testing it became clear there would be problems with the heart rate data, and those arose during the test. Sometimes the electrodes broke contact with the skin, leading to interbeat intervals that were double or triple surrounding intervals. In other instances, shifts in the baseline voltage (due to a weak ground) and the nature of the heart's electrical cycle led to intervals that were one-half or one-third those of adjacent intervals. (After the experiment was completed, it was learned these problems could have been eliminated by the use of disposable electrodes. Unfortunately, they do leave a mark on the skin for short time after they are removed.)

For normal adults engaged in light physical work, the interbeat interval rarely varies by more than 10% within a minute, and the faulty data were quite obvious. It was originally planned to use a computer program to identify the faulty points and suggest corrections (splitting or pooling) to an operator. This procedure was used on only one block of data that had some unusual problems with it. This editing process proved to be extremely time-consuming. While including the correction rules in the software to make the process automatic was considered, the rules proved to be too complex, context dependent, and time-consuming to code quickly.

Instead, to eliminate the bad data, a listing of the data was inspected and a range of acceptable intervals was selected. Typically the range was 250 milliseconds above and below the mean interval for each block for each subject. In most instances, 250 milliseconds was more than three standard deviations from the mean. Assuming the interbeat intervals are normally distributed, which is a reasonable first approximation, less than 5 out of 1,000 points should fall outside of that range. In a typical minute of data there were two bad data points (out of 60–70 in the raw data). This procedure usually reduced the standard deviation for a given minute to 1/3 the uncensored value and shifted the mean interbeat interval by 50–100 milliseconds. In a few instances where the contact problems were acute, the mean rate was adjusted by averaging it with the means from the preceding and following minutes.

Correlations Between Measures

Correlations were used to determine how the measures related to each other. Of particular importance were the correlations between response time and the fatigue and steering performance measures. Shown in Table 2 are the correlations of several measures averaged across ten minute intervals, the limit of the fatigue ratings. There are 378 and not 384 cases (8 subjects x 4 blocks x 12 periods of 10 minutes) represented, because there were six missing cases of the standard deviation of the interbeat interval. Correlations with response time have not been included in this table, because they were not collected in two of the four conditions, and including them would have cut the number of cases in half.

The table shows an extremely strong correlation between the mean and standard deviation of the interbeat interval. Also shown are highly statistically significant (though not very large) correlations between several other variables. The fatigue rating and the mean interbeat interval were correlated. (People who look tired have longer interbeat intervals (slower heart rate).) The fatigue rating was also correlated with the standard deviation of the steering error. Because they weren't correlated with most of the variables of interest (though they were correlated with each other), these correlations suggest that the standard deviation of the interbeat interval and the mean steering error should not be viewed as predictive of alertness as the other variables.

TABLE 2

CORRELATIONS BETWEEN MEASURES, EXCLUDING RESPONSE TIME DATA

Variable	Fatigue Rating	Interbeat Interval		Steering Error	
		Mean	SD	Mean	SD
Fatigue Rating					
Mean Interbeat Interval	.21				
Standard Deviation Interbeat Interval	.07	.76			
Mean Steering Error	−.02	.08	.26		
Standard Deviation of Steering Error	.28	.05	.27	.20	

Note: p(.1)=.08, p(.05)=.10 p(.01)=.13, p(.001)=17

Table 3 contains the correlations with the Response Time data included. The sharply reduced number of complete cases (125) is due to the reduced number of conditions described before and other reasons. Following common statistical practice, increasing the number of cases tends to increase correlations as well as the values needed to reach particular significance levels.

The general pattern of correlations is the same as the previous table, except that the correlation between fatigue ratings and the standard deviation of the steering error is not significant nor is the one between it and the mean steering error. Of special interest are the correlations of the response time data. The only significant correlation with intertrial interval was with the mean steering error. Mean response time and the standard deviation of it were correlated with several variables (fatigue rating, mean interbeat interval, steering error) and, hence, deserve further examination.

Mean Steering Error

When one drives a car, one tries to stay in the middle of the lane. It therefore seems reasonable to look at mean steering error as a measure of how well people drove. In contrast, in the correlation analysis, mean error was correlated only with other variables for the response time conditions.

Using Analysis of Variance (ANOVA) the mean steering error data for each ten minutes were examined. As a reminder, eight men were tested, four young and four old. Each subject was tested four times with a different alertness condition (device, fixed interval; device, random interval; radio; nothing) occurring each time. The order of conditions was different for each person in each age group to counterbalance for practice effects. The statistics from that ANOVA are shown in Table 4.

None of the factors in this analysis was significant, though the effect of age almost was. Mean steering errors were −.8, −1.6, −3.9, and 12.2 for the younger subjects and −3.2, −15.3, −5.6, and −1.4 for the older subjects. For this simulation, a negative average steering error corresponds to steering to the right, and a value of +/−128 would correspond to driving on the road edge. Some of the bias may be due to a steering system hardware misalignment. It also appears that subjects, at times, drove as if they were on a two-lane road, even though they were instructed to drive as if they were on a single-lane expressway ramp. That is particularly true for subject six, whose mean steering error was always negative.

Shown in Figure 6 is a plot of mean steering error versus time by condition. In

TABLE 3

CORRELATIONS BETWEEN MEASURES, INCLUDING RESPONSE TIME DATA

Variable	Fatigue Rating	Interbeat Interval		Mean ITI	Response Time		Steering Error	
		Mean	SD		Mean	SD	Mean	SD
Fatigue Rating								
Mean Interbeat Interval	.26							
Standard Deviation Interbeat Interval	.01	.72						
Mean Intertrial Interval	.10	.10	−.02					
Mean Response Time	.16	.17	−.03	−.50				
Standard Deviation of Response Time	−.16	.29	.36	−.08	.33			
Mean Steering Error	−.12	.17	.18	−.30	.34	.30		
Standard Deviation of Steering Error	.11	.05	.36	.11	−.25	.11	−.03	

Note: $p(.1)=.15$, $p(.05)=.18$, $p(.01)=.23$, $p(.001)=.29$

TABLE 4

ANOVA OF MEAN STEERING ERROR

Source	SS	df	MS	F	p
Subject	9940.78	3	3313.59	3.11	.19
Age	5880.48	1	5880.48	5.52	.11
Error-1	3194.23	3	1064.74		.11
Condition	1261.84	3	420.61	1.24	.38
Block	575.23	3	191.74	<1	–
CB	4408.76	6	734.79	2.12	.19
CA	3042.84	3	1014.28	2.92	.12
BA	1139.79	3	379.93	1.09	.42
Error-2	2077.68	6	346.28		
Time	271.65	11	24.69	1.12	.38
TS	966.13	33	29.27	1.32	.21
TA	244.52	11	22.22	1.00	.46
Error-3	726.27	33	22.00		
CT	598.19	33	18.12	<1	–
BT	853.24	33	25.85	1.26	.21
CBT'	1310.18	66	19.85	<1	–
BTA	765.00	33	23.18	1.13	.32
CTA	588.15	33	17.82	<1	–
Error-4	1353.10	66	20.50		

general, people did best with the device set in the random-interval mode, somewhat worse in the fixed-interval mode, even worse with nothing, and worst with the radio on, though these differences were not statistically significant.

An alternative perspective is that the device may not have an overall effect on steering performance, since most of the time the driver is alert and steering well, but rather that it minimizes how often steering is extremely poor. Shown in Figure 7 are histograms of the mean steering error for each minute by condition. Using a mean error equal to or in excess of +/−35 as "extremely poor," there were 24 such minutes ("outliers") for the fixed interval device, 10 for the random interval, 11 for the radio, and 55 when neither the device or radio were present. Using +/−45 as the criterion, the values are 8, 0, 2, and 15 for the same conditions. These numbers suggest poorest performance with nothing, some improvement with the fixed-interval device, and no difference between the device in the random interval mode and the radio.

Also noteworthy was the gradual, though nonsignificant ($\underline{F}<1$) decrease in steering error across blocks. (See Figure 8.) The lack of significance was due in part to the accuracy of the error measurement (one screen location) and rules regarding significant figures. It is also for those reasons the lines in Figure 8 are so jagged.

It is clear this improvement was not due to subjects' learning the road pattern. While the same road-generating function was used in all four conditions, so they would be equally difficult, the sequence and direction of turns differed. It was thought that subjects' considerable on-the-road (in some cases 50 years) and simulator experience (typically 10 hours by the end of the experiment) would eliminate any suggestion of improvements with practice. It did not. It was therefore appropriate to control for practice effects, as was done in this experiment.

Standard Deviation of the Steering Error

Good driving is not only associated with being in the center of one's lane on the average but doing so consistently. Such variation is measured by the standard deviation of the steering error. In a previous analysis, it was found to be as well as correlated with other performance measures as well mean steering error. Table 4 contains the ANOVA of the standard deviation of steering error computed for each 10-minute period. Except for the effect of Time ($\underline{F}(11,33) = 7.88$, $\underline{p}<.001$), none of the main effects or interactions was statistically significant. The effect of Time is shown in Figure 9. Clearly, as time progresses within blocks, steering becomes much more variable, though across blocks, except for block three, the trend is for variability to decrease. Prior to the experiment it

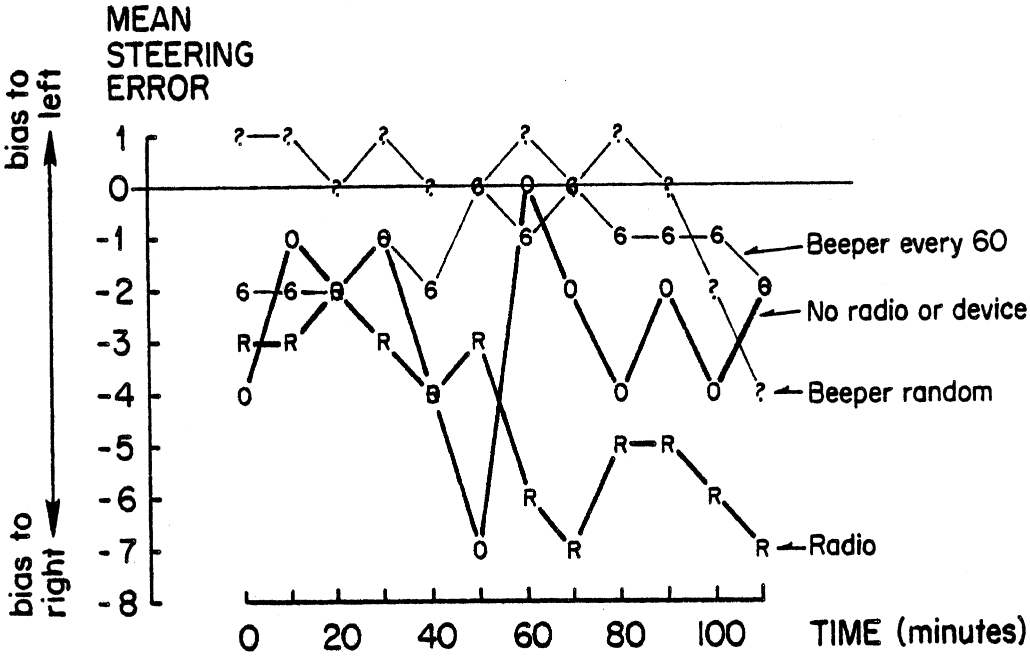


Figure 6. Mean steering error versus time by condition.

Device: Fixed Interval | Device: Random Interval

COUNT FOR MEAN STEERING ERROR (Each X = 10)		MIDPOINT	COUNT FOR MEAN STEERING ERROR (Each X = 10)	
+	0	-75	0	+
X+	1	-65	0	+
X+	1	-55	0	+
+	0	-45	0	+
X+	3	-35	2	+X
XXX+	27	-25	43	+XXXXX
XXXXXXXXXXXXXXXXXXXX+	198	-15	140	+XXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX+	331	- 5	302	+XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXX+	243	5	233	+XXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXX+	105	15	197	+XXXXXXXXXXXXXXXXXXXX
XXXX+	32	25	36	+XXXX
XX+	13	35	7	+X
X+	4	45	0	+
X+	2	55	0	+
+	0	65	0	+
+	0	75	0	+
+	0	85	0	+
	960	Total	960	

Radio | Nothing

COUNT FOR MEAN STEERING ERROR (Each X = 10)		MIDPOINT	COUNT FOR MEAN STEERING ERROR (Each X = 10)	
+	0	-75	0	+
+	0	-65	0	+
X+	1	-55	0	+
X+	1	-45	0	+
X+	6	-35	12	+XX
XXXXXX+	60	-25	46	+XXXXX
XXXXXXXXXXXXXXXXXXXXXXX+	223	-15	252	+XXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX+	347	- 5	345	+XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXX+	233	5	168	+XXXXXXXXXXXXXXXXX
XXXXXXXX+	73	15	57	+XXXXXX
XX+	13	25	37	+XXXX
X+	3	35	28	+XXX
+	0	45	7	+X
+	0	55	3	+X
+	0	65	3	+X
+	0	75	1	+X
+	0	85	1	+X
	960	Total	960	

Figure 7. Histograms of steering error by condition.

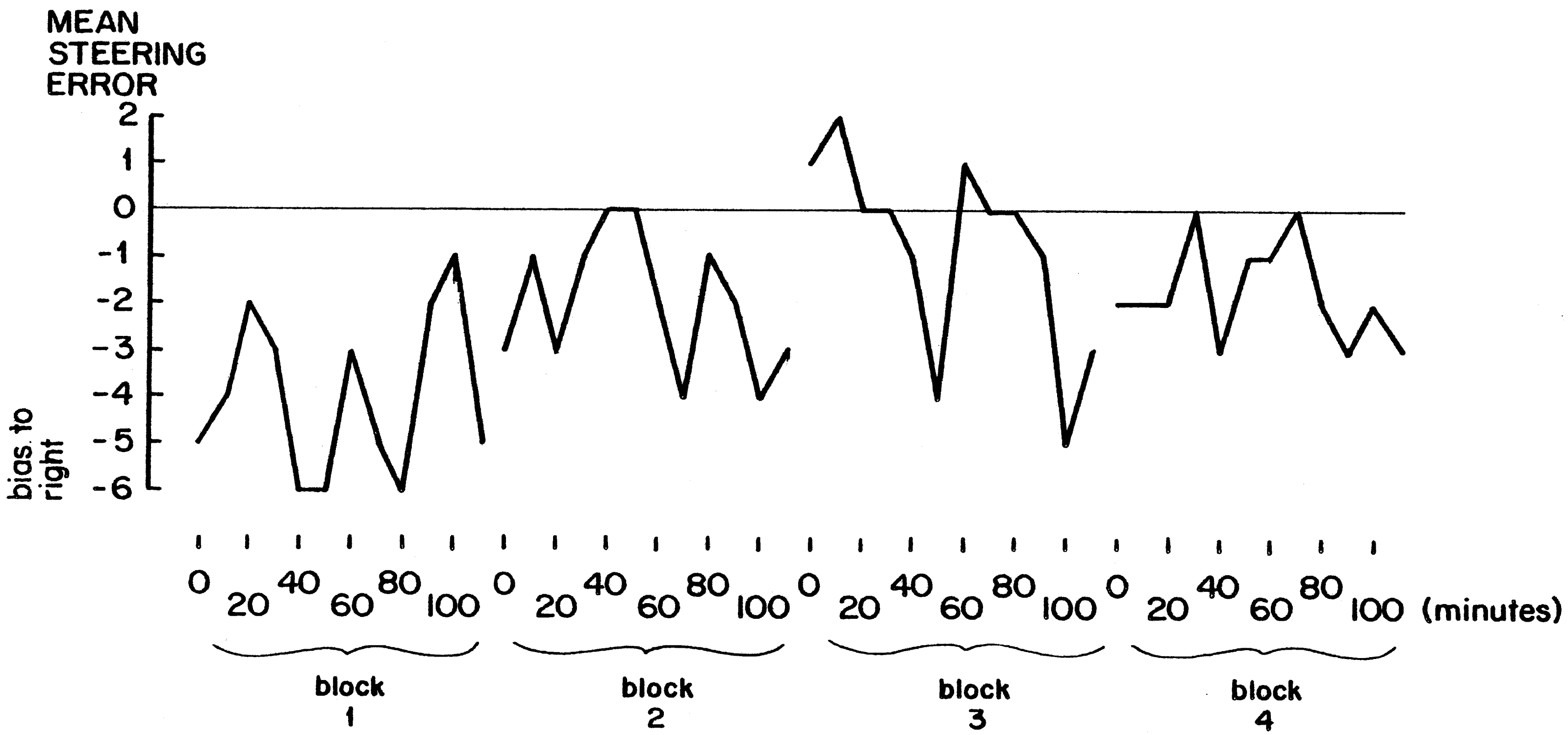


Figure 8. Mean steering error vs. time by block.

was expected that variability would increase in an exponential manner with time. That is, error variability would be low for most of each block and then rapidly increase at the end. Here, the opposite happened, with much of the degradation happening within the first 20 minutes of each two-hour session.

Shown in Figure 9 are the differences due to test condition. The figure shows that the standard deviation for the radio condition was lower (less variable) with time.

Thus, the mean error data suggest that people on the average are most likely to be in the center of the roadway while using the device, somewhat less likely while no external stimulation is present, and least likely while listening to the radio. On the other hand, people are less variable in steering when listening to the radio than when nothing or the device is present. From the perspective of the outlier data, the grouping is radio or random interval device, fixed interval device, and then nothing from best to worst performance. While none of these differences is statistically significant when viewed independently, the pattern of results makes sense. Apparently, the radio (and to some degree, the alertness device) have qualities of both keeping people alert but at the same time distracting them from steering. For them to work, they must demand the driver's attention. In the case of the radio, people stay awake but don't pay close attention to staying in the exact center of the lane. This strategy tends to allow for larger mean steering errors but tends to minimize extreme errors, thus leading to decreased standard deviation and fewer outliers. In terms of the two versions of the device, the random version is more effective, because subjects cannot predict when it will go off, and it goes off more often.

At an intuitive level this explanation fits with common observations. When people listen to and especially when they adjust the radio, they do pay slightly less attention to staying dead center in the lane. (Just watch the cars with only a driver as the driver sings and taps out the beat on the steering wheel.) On the other hand, when they are doing that, they are not going to doze off.

Intertrial Intervals

Since the intertrial interval is controlled by the duration of the subject's two previous response times, and response time should reflect the subject's alertness, intertrial interval could reflect the subject's state of alertness. But since the intertrial interval was uncorrelated with the other performance measures, it was not explored in detail.

However, the data summaries and the distribution of the mean intertrial intervals

TABLE 5

ANOVA OF STANDARD DEVIATION OF STEERING ERROR

Source	SS	df	MS	F	p
Subject	5572.14	3	1857.38	1.40	.39
Age	5582.13	1	482.13	0.44	.55
Error-1	3968.38	3	1322.79		
Condition	415.87	3	138.62	1.05	.43
Block	516.67	3	66.57	<1	–
CB	183.34	6	30.55	<1	–
CA	487.18	3	162.39	1.23	.37
BA	129.63	3	43.21	<1	–
Error-2	787.13	6	131.18		
Time	1453.21	11	132.11	7.88	<.001
TS	589.89	33	17.87	1.07	.42
TA	228.60	11	20.78	1.24	.30
Error-3	553.37	33	16.76		
CT	289.56	33	8.77	<1	–
BT	427.72	33	12.96	1.02	.46
CBT	570.07	66	8.63	<1	–
BTA	279.77	33	8.46	<1	–
CTA	378.73	33	11.47	<1	–
Error-4	832.57	66	12.61		

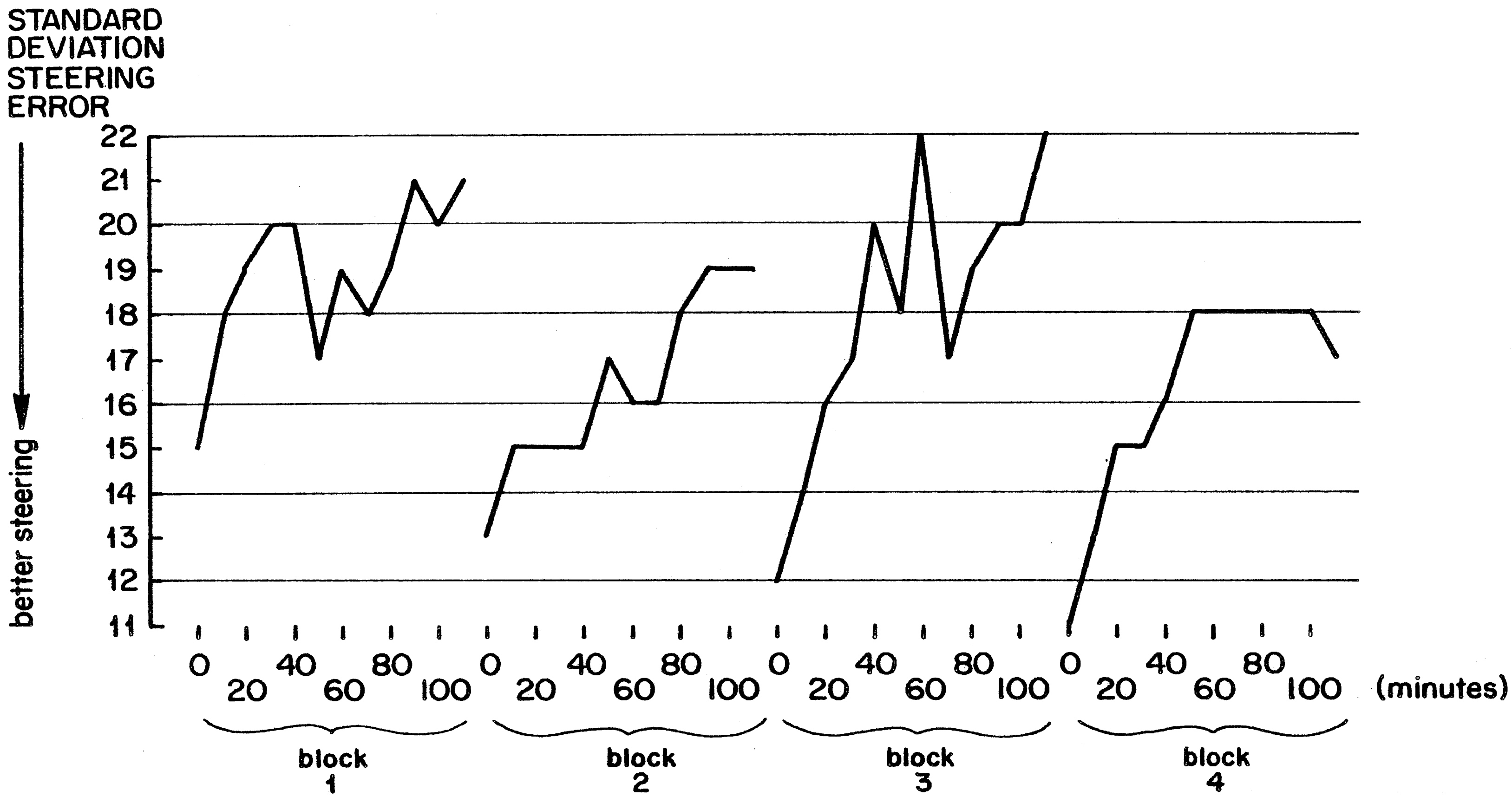


Figure 9. Standard deviation of steering error versus time by block.

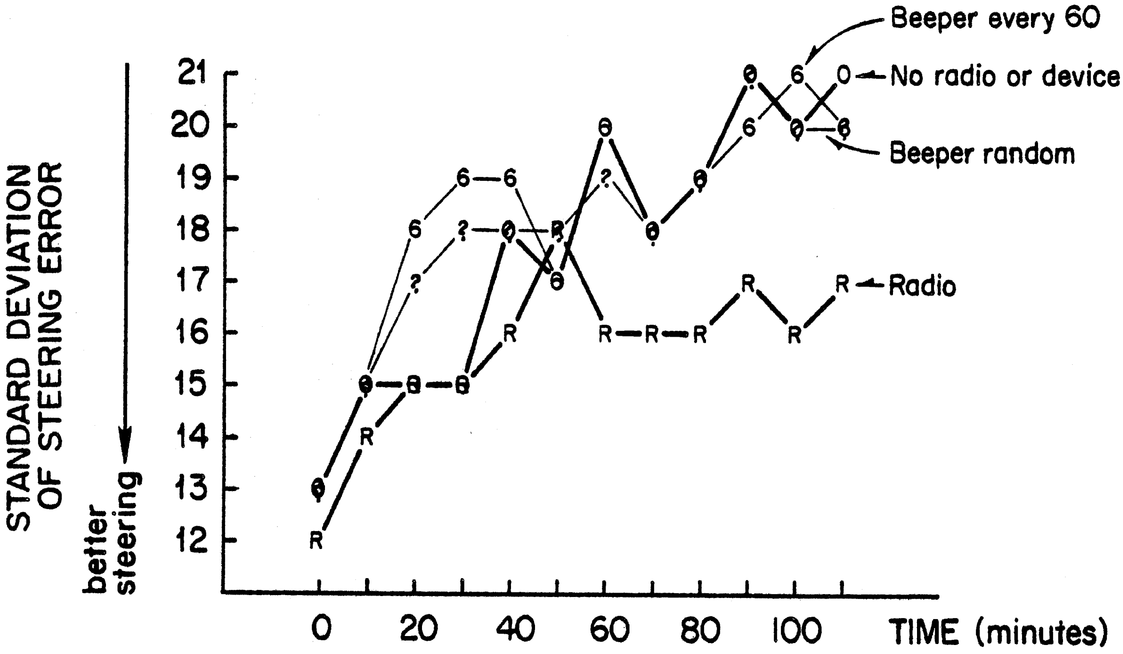


Figure 10. Standard deviation of steering error versus time by condition.

reveal some interesting findings. There were several scattered minutes in device-related conditions where no button presses occurred. This came about when the intertrial interval was 61 seconds and the interval happened to begin at just the right time to skip a minute. Also, there were two instances where the intertrial interval was two minutes (see Figure 11), one where it was three minutes and ten seconds, and one where it was 15 seconds. These were durations the device was not programmed by the sponsor to generate. Otherwise, the response time trial scheduling worked as designed, 20 or 60 seconds in the fixed mode, randomly varying between 20 and 60 in the random mode. The distribution of intervals in the fixed mode shows values other than 20 or 60 because of the one-minute averaging period.

Mean Response Times

Shown in Table 6 is the ANOVA of the mean response times. In that analysis the effect of block has been ignored because of incomplete balancing and the lack of a block effect in previous analyses. With regard to individual differences, neither subject nor age differences had a significant effect on response time, thought the age effect came close. Mean response times were 1628, 1345, 1688, and 1340 milliseconds for the younger subjects and 2304, 1337, 2661, and 1680 milliseconds for the older subjects.

Unlike several of the previous analyses, there was a significant difference due to the condition, with response times to the fixed interval device (1684 milliseconds) being briefer than those of the random interval version (1812 milliseconds). Histograms of the response time distributions by condition are shown in Figure 12.

Since brief response times are associated with greater alertness, one potential interpretation of these data is that people were more alert, or kept more alert, in the fixed-interval mode.

An alternative interpretation is more likely. Some subjects reported that in the fixed-interval mode they knew about when the next response would occur. They prepared for it by moving their hand near the button so as to shorten the duration of the annoying tone (and consequently their response times). In a few cases the response times are so short (e.g., 88, 181, 192 milliseconds), that subjects must have begun the motion to press the button before they heard the tone. There were also reports that sometimes the button was struck before a tone had been presented, though these responses were not picked up by the computer, because it scanned for a button press only after a tone was presented.

While subjects could have been instructed not to anticipate the tone, the purpose of

```
        COUNT FOR MEAN ITI (Each x = 20)      MIDPOINT    COUNT FOR MEAN ITI (Each x = 20)

Device:  Fixed Interval                  + 0      0       0 +                 Device:  Random Interval
                                         + 0     10       0 +
                                     XXXX+ 65    20      84 +XXXXX
                                        X+ 13    30     118 +XXXXXX
                                       XX+ 25    40     284 +XXXXXXXXXXXXXXX
                                         + 0     50     312 +XXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX+ 807   60     149 +XXXXXXXX
                                         + 0     70       0 +
                                         + 0     80       0 +
                                         + 0     90       0 +
                                         + 0    100       0 +
                                         + 0    110       0 +
                                        X+ 2    120       0 +
                                         + 0    130       0 +
                                         + 0    140       0 +
                                         + 0    150       0 +
                                         + 0    160       0 +
                                         + 0    170       0 +
                                         + 0    180       0 +
                                         + 0    190       0 +
                                         + 0    200       1 +X
                                           960  Total   960
```

Figure 11. Histograms of mean intertrial intervals by condition.

TABLE 6

ANOVA OF RESPONSE TIME

Source	SS	df	MS	F	p
Subject	21679014	3	7226388	3.37	.17
Age	11761200	1	117661200	5.48	.11
Error-1					
Condition	785664	1	785664	5.85	.09
CS	4611173	3	1537057	11.45	.04
CA	55624	1	55624	<1	—
Error-2	402886	3	134295		
Time	2882959	11	262087	4.07	<.001
TS	3437022	33	104152	1.62	.09
TA	657424	11	59765	<1	—
Error-3	2126245	33	64431		
CT	1722788	11	156617	1.69	.12
CTS	3895244	33	120764	1.30	.22
CTA	1645717	11	149610	1.62	.13
Error-4	3053910	33	92542		

```
COUNT FOR MEAN RESPONSE TIME  (Each X = 10)      MIDPOINT    COUNT FOR MEAN RESPONSE TIME  (Each X = 10)

 Fixed Interval                            X+ 3        0       0 +                         Random Interval
                                       XXXXX+ 41     500      61 +XXXXXXX
                         XXXXXXXXXXXXXXXXXXXX+ 199   1000     209 +XXXXXXXXXXXXXXXXXXXXX
 XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX+ 367         1500     314 +XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
                             XXXXXXXXXXXXXXXX+ 156   2000     164 +XXXXXXXXXXXXXXXXX
                                     XXXXXXX+ 63     2500      84 +XXXXXXXXX
                                       XXXXX+ 41     3000      43 +XXXXX
                                         XXX+ 22     3500      29 +XXX
                                           X+  5     4000      10 +X
                                           X+  2     4500       8 +X
                                           X+  3     5000       3 +X
                                           X+  1     5500       0 +
                                            +  0     6000       0 +
                                           X+  3     6500       6 +X
                                           X+  2     7000       8 +X
                                           X+  1     7500       2 +X
                                            +  0     8000       1 +X
                                            +  0     8500       1 +X
                                            +  0     9000       0 +X
                                            +  0     9500       0 +X
                                            +  0    10000       1 +X

                                              960    Total    960
```

Figure 12. Histograms of response times by condition.

this experiment was to simulate actual on-the-road use, and in that context people would behave as they did here. While some might argue that the product would come with instructions saying not to defeat the device, many people, because the product is so simple, will not read them (Wright, Creighton, and Threlfall, 1982). Based on related experience with automobile owner's manuals, it is quite likely that once the product is installed, the instructions will be lost, so that only the installer, and not most users, will even know there were instructions (Green, 1984).

If a device were to be produced, it would have to handle both the responses to no signal and the unrealistically short response times (say less than 250 milliseconds). The processor could check for every button press and compare the response time with a fast threshold time. If the time was unrealistically brief, either because the tone had not been sounded (a "negative" time) or had just been sounded, then the processor should schedule the tone to go off in the next few seconds. Thus trying to beat the clock would cause the tone to sound more often. This should reduce efforts to anticipate the tones.

Another strategy users might follow would be to tape down or continuously hold down the response button. If this were detected, the tone should go on and stay on until the button was released.

Also unusual about the response time data is the distribution of the long times. As a reminder, when the tone sounded, it was for three seconds and then went off. If the subject had not responded by six seconds, the tone went off again. This strategy was chosen so the tone could be made loud enough to be alerting but brief enough so that the annoyance level was reduced.

As shown in response time histograms (Figure 12), this pattern had a profound effect. The frequency of response times increases from 0 to a peak of 1500 milliseconds and then trails off to zero at 5000–5500 milliseconds. At the 6500 millisecond interval (6000 milliseconds plus some time to respond), the number of responses increases and then gradually trials off to 0 at longer time intervals. Some of the responses in excess of six seconds are due to the subject's not being alert. However, this pattern suggests that some might be due to the subject's forgetting to press the button or misunderstanding how the device works. (The driver might reason that if the tone goes off by itself, it's just a warning to stay alert, for which no response is needed.) To avoid the false indications of a lack of alertness suggested by long response times, the tone intensity should not drop to zero in the three-to-six second interval but rather to a reduced level.

Beyond the condition effects described previously, several other factors were statistically significant in the ANOVA, in particular the effect of time with the test block ($p<.001$). That effect is shown in Figure 13. In addition, there were significant interactions between conditions and subjects and time and subjects.

Standard Deviation of the Response Time

As was noted earlier, the standard deviation of response time was often as well correlated with the performance measures of interest as the mean response time. In this case the data have not been subjected to ANOVA because the measure was of secondary interest and so many data were missing that the exact analysis procedure was unclear. (For much of the experiment, there was only one button press in a given minute. For those minutes a standard deviation is not computable.)

Nonetheless, there was an interesting tendency for the standard deviation to be both less and less variable for the device in the random-interval mode than in the fixed interval mode. (See Figure 14.) This could be because the random device sounded more often and therefore more consistently assessed the subject's state of alertness.

Mean Interbeat Interval

Heart rate is an indicator of alertness. In brief, the more active one is, the greater the heart rate. Also, as was noted in the literature review in the introduction, some have suggested that heart rate variability is a measure of alertness.

Shown in Table 7 is the ANOVA of the mean interbeat interval. For the 8 participants the mean interbeat intervals were 881, 1039, 1193, 854, 1024, 741, 954, and 784 milliseconds corresponding to heart rates of 68, 58, 50, 70, 59, 81, 63, and 77 beats per minute. The differences between them were not significant. Also not significant were differences due to the test condition, though the effects of time and time interacting with age were significant. The condition and time effects are shown in Figure 15. Despite the lack of significance, the pattern of the results is similar to that of other measures. Subjects had the shortest interbeat intervals (most rapid heart rate – were most awake/ alert) while using the radio, with the device in the random mode, nothing, and the device in the fixed-interval mode following in that order.

Standard Deviation of the Interbeat Interval

As was noted in the correlation analysis, heart rate and heart rate variability are related with a more rapid heart rate (greater alertness) leading to less variability.

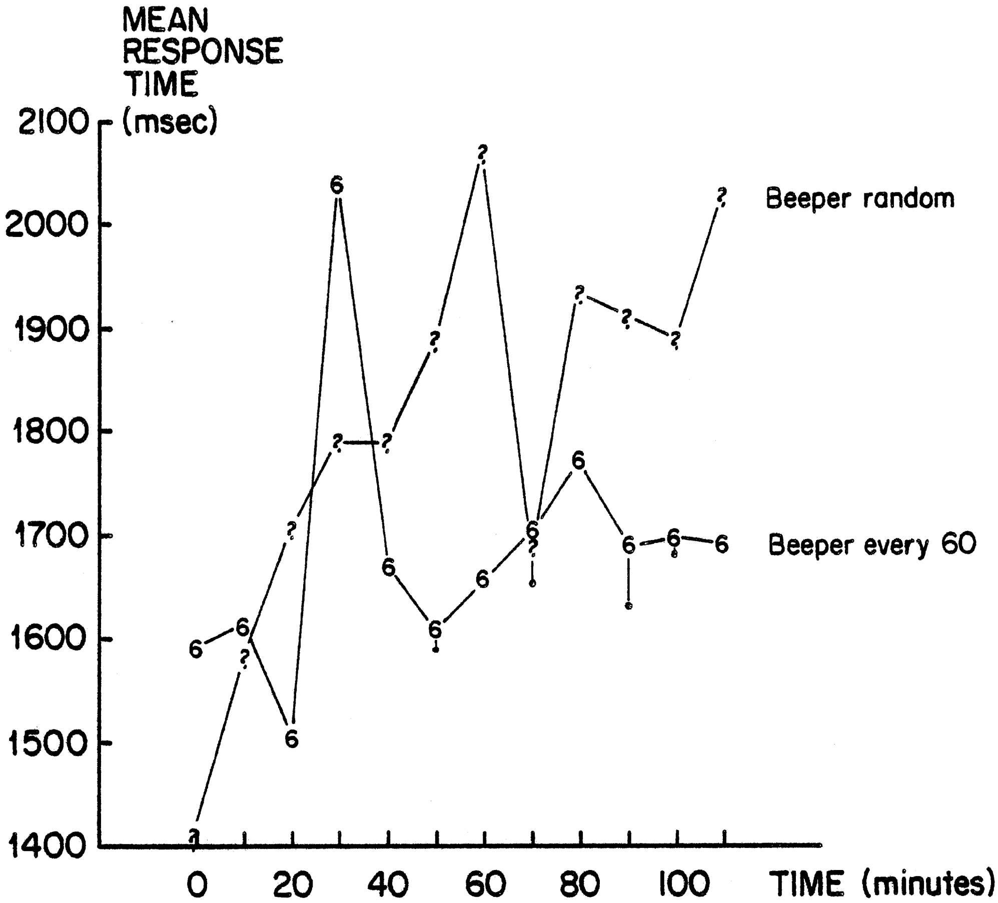


Figure 13. Mean response time by condition.

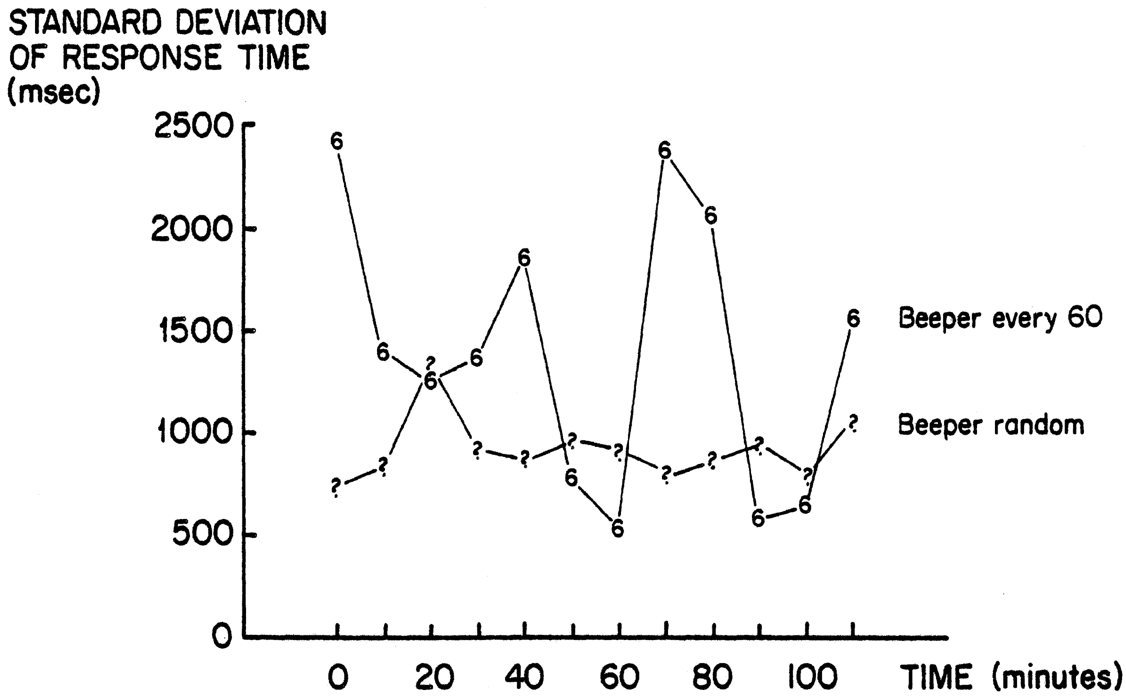


Figure 14. Standard deviation of response time by condition.

TABLE 7

MEAN INTERBEAT INTERVAL

Source	SS	df	MS	F	p
Subject	3376928	3	1125642	1.20	.44
Age	1304101	1	1340101	1.39	.32
Error-1	2807540	3	935846		
Condition	125382	3	41794	<1	—
Block	191640	3	63880	1.09	.42
CB	210442	6	35073	<1	—
CA	184820	3	61606	1.05	.43
BA	30290	3	10096	1.16	.33
Error-2	348978	6	58163		
Time	104937	11	9539	5.60	<.001
TS	65314	33	1979	1.16	.33
TA	44587	11	4053	2.38	<.01
Error-3	56180	33	1702		
CT	25170	33	762	<1	—
BT	19616	33	594	<1	—
CBT	51571	66	781	<1	—
BTA	21753	33	659	<1	—
CTA	51870	33	1571	<1	—
Error-4	70779	66	1072		

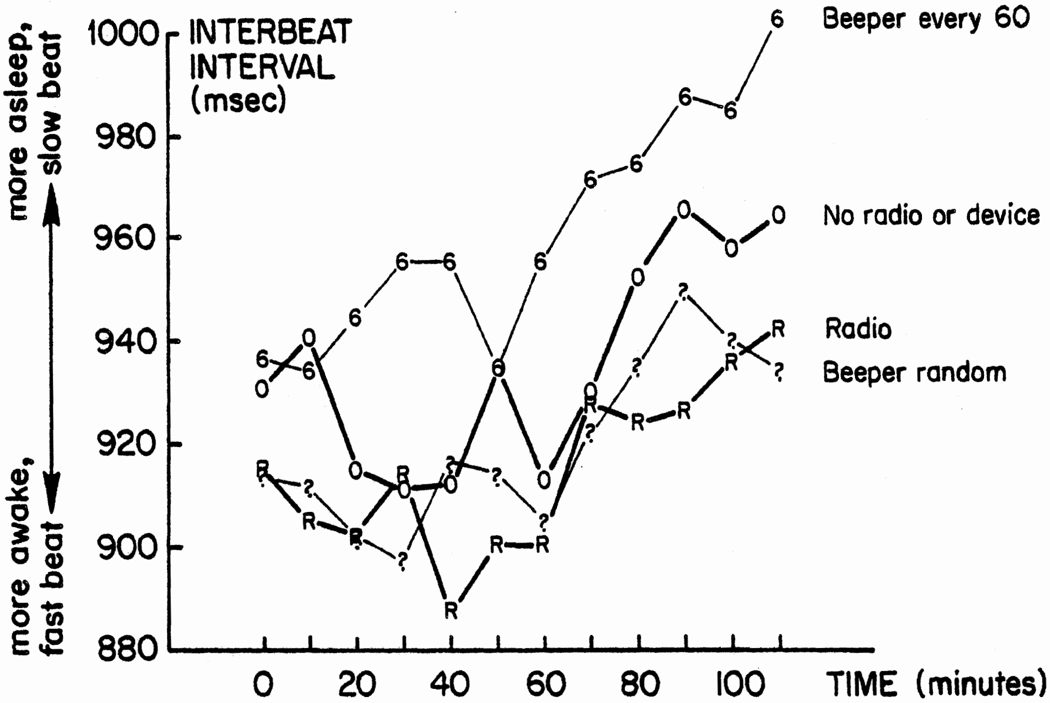


Figure 15. Mean interbeat interval versus time by condition.

Furthermore, changes in the heart rate (due to waking up or dozing off, being startled by the tone, etc.) could potentially make the standard deviation of the interbeat interval a better measure of alertness than the mean.

Shown in Table 8 is the ANOVA of the standard deviation of the interbeat interval. The pattern is similar to that for the mean interval. Differences between people were not significant, though they did span a large range. Values of 70, 113, 105, and 66 milliseconds were computed for younger subjects and 84, 15, 38, and 28 for the older subjects.

Differences due to condition were not significant, though the pattern was similar to that seen before. Standard deviations were lowest with the radio, somewhat greater for both modes of the device, and greatest for the control condition (nothing). These differences are shown in Figure 16.

Finally, as shown in Table 8, several other factors were significant, including the condition by age interaction, and the condition by block by time interaction.

Fatigue Ratings

Shown in Table 9 is the ANOVA of the fatigue ratings. Unlike several of the other analyses, many factors had significant effects. One should view this analysis with some caution, as the ratings are more like interval-scale data than the ratio-scale data the ANOVA technique assumes them to be.

In that ANOVA both the effect of condition and the interaction of condition with time were significant. People were rated as least fatigued while using the radio, more fatigued with the device in the random mode, and most fatigued with either the device in the fixed-interval mode or no device at all. Those differences tended to grow with time. They are shown in Figure 17. It is curious that there is so much overlap of conditions in that figure and the differences are statistically significant, whereas for other measures of condition differences, the overlap is much less and the differences are not significant. The reason is that the variability for the other measures is much larger.

Also significant in this analysis was the effect of subject, time, and the block by age, time by age, and condition by time by age interactions.

Clearly, the data show consistent trends that the radio is best as an alertness device, followed by the random-interval device, the fixed-interval device, and finally nothing. To varying degrees this pattern was found for the standard deviation of the steering error, the mean steering error outliers, the mean interbeat interval, and the

TABLE 8

ANOVA OF THE STANDARD DEVIATION OF INTERBEAT INTERVAL

Source	SS	df	MS	F	p
Subject	49324	3	16441	.30	.82
Age	216362	1	216362	3.94	.14
Error-1	164783	3	54927		
Condition	6783	3	2261	1.32	.35
Block	3211	3	1070	<1	–
CB	19379	6	3229	1.89	.22
CA	17132	3	5710	3.35	.09
BA	4730	3	1576	<1	–
Error-2	10220	6	1703		
Time	3473	11	315	<1	–
TS	10093	33	305	<1	–
TA	7436	11	667	1.06	.41
Error-3	20739	33	628		
CT	3146	33	95.3	<1	–
BT	4067	33	123.2	1.19	.26
CBT	10434	66	158.0	1.53	.04
BTA	4157	33	125.9	1.22	.24
CTA	2831	33	85.8	<1	–
Error-4	6799	66	103.0		

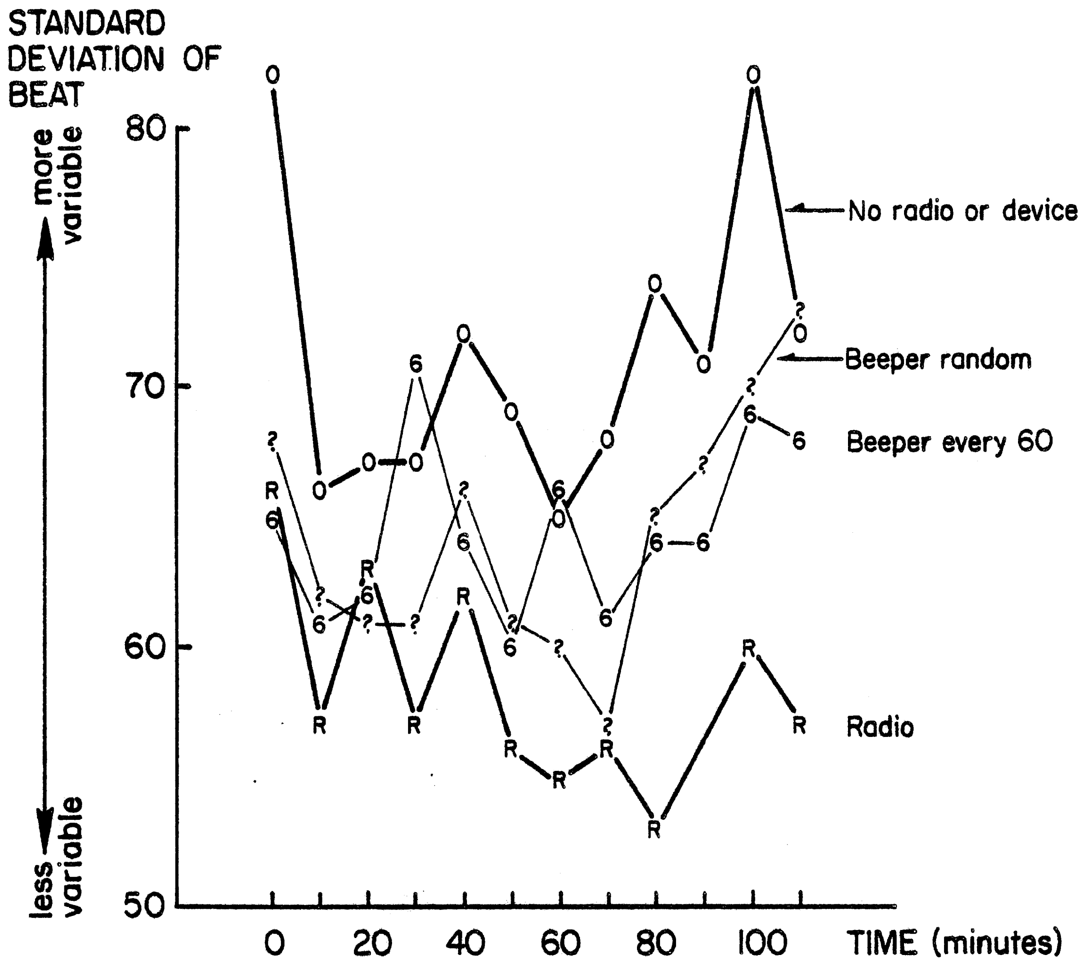


Figure 16. Standard deviation of interbeat interval versus time.

TABLE 9

ANOVA OF FATIGUE RATINGS

Source	SS	df	MS	F	p
Subject	11.521	3	3.840	7.58	.07
Age	1.500	1	1.500	2.96	.18
Error-1	1.521	3	.506		
Condition	6.145	3	2.048	5.36	<.05
Block	1.000	3	0.333	<1	–
CB	4.416	6	0.736	1.93	.22
CA	1.604	3	0.534	1.40	.33
BA	8.500	3	2.833	22.25	<.01
Error-2	2.291	6	0.381		
Time	149.771	11	13.615	67.4	<.001
TS	10.042	33	.304	1.51	.12
TS	11.313	11	1.028	5.09	<.001
Error-3	6.667	33	.202		
CT	8.666	33	.262	1.56	.06
BT	6.562	33	.198	1.18	.27
CBT	9.208	66	.139	<1	–
BTA	5.187	33	.157	<1	–
CTA	8.333	33	.252	1.50	.08
Error-4	11.083	66	.167		

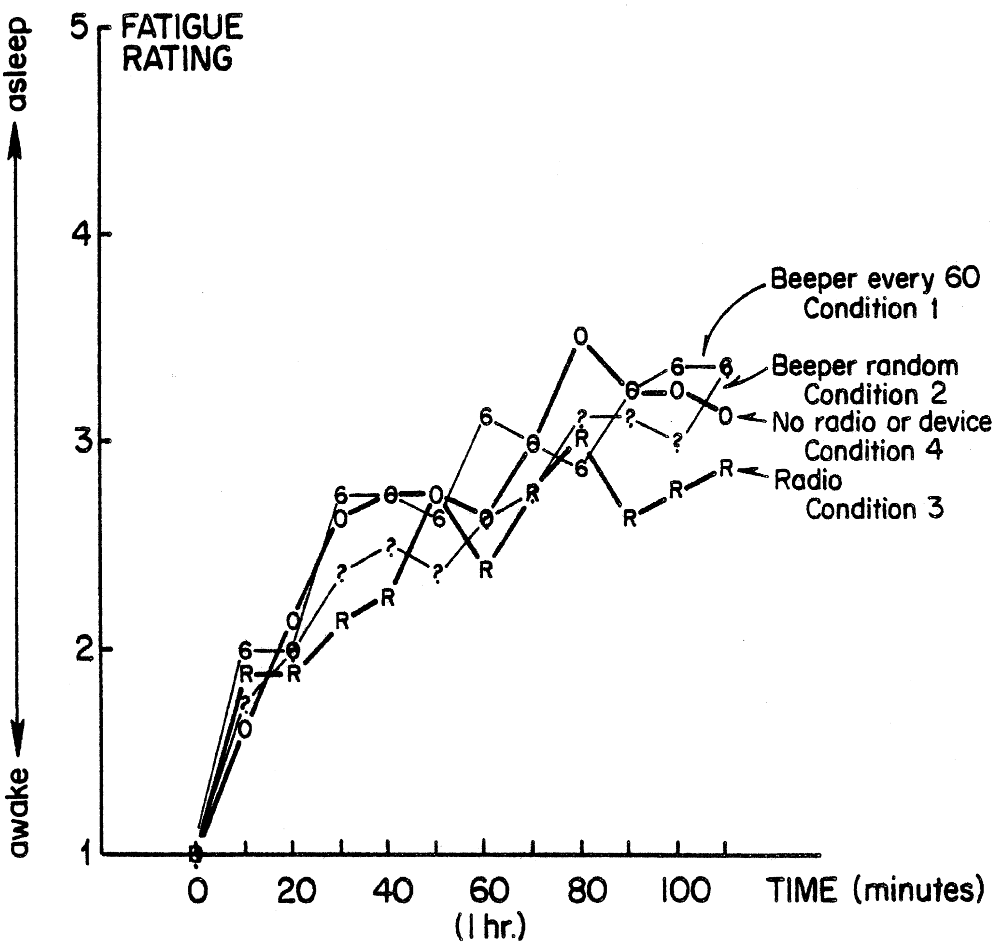


Figure 17. Mean fatigue ratings versus time by condition.

fatigue ratings. Except for the differences in fatigue ratings, these differences were not statistically significant when all factors were included in the analysis of variance. However, in several preliminary one-way ANOVAs, the differences were statistically significant.

Since many of the factors in the ANOVAs were not significant, one could argue that one should do some pooling before looking at the main effects (e.g., condition differences). While many scholarly journal articles could be written by statisticians debating the merits of pooling here, it is a moot point, since the differences were in the "opposite" direction, favoring the radio over the device. There are very few cars and trucks on American roads that do not have radios.

One might speculate from these data that if the radio helped keep people alert, and sometimes so did the alertness device, that together they would be very effective. There are suggestions here that this might not be so, and in fact having them together could be worse than just having one or the other (or even nothing at all). Both the radio and the device keep one alert by providing a distraction. It could be that having both distractions at the same time could degrade performance as alertness decreased, rather than enhance it.

Questionnaire Data

The questionnaire data provided some insight into how the alertness device should be designed and how it might be used. Shown in Table 10 is a summary of the questionnaire data. While eight people is a fairly small number for a survey, one must keep in mind that these people had considerable experience with the device under carefully controlled conditions and their collective views are a more accurate representation of it than collective views from a much larger sample who might only be shown the device.

Drivers had varied opinions about where the device should be mounted. This suggests that if the device were to become a real product, it would need brackets or supports so it could be mounted both on the instrument panel and the steering wheel structure. (Most likely placing it on one's lap or on the header, as participants suggested, would not prove to be workable.) Also having some impact on its mounting is its size, which is largely dependent upon the button size. Most drivers thought the current size was about right.

Drivers felt the current tone loudness was about right (6/8) but they still wanted to be able to adjust it (7/8). They had diverse opinions as to how often it should normally go

TABLE 10

POST-TEST QUESTIONNAIRE RESPONSES

Issue/Question	Response
Button	
Where should it go?	dashboard 2-left of center 1-center steering wheel 2-center 1-right side (thumb switch) 1-header/visor 1-lap of driver
Size	1-too big 5-about right 2-too small
Tone	
Loudness	1-too loud 6-about right 1-too soft
Should loudness be adjustable?	7-yes 1-no
How often should it normally sound?	every... 1–600 seconds 1–120 seconds 1–60 seconds 3–30 seconds 1-I don't know 1-missing (wants only random)
When should it sound next if the driver was slow?	in... 2–20 seconds 1–15 seconds 3–10 seconds 1–5 seconds 1-immediately
Should how often it sounds be adjustable?	6-yes 2-no
Driver Behavior	
Ever fallen asleep while driving?	5-yes 3-no

Issue/Question	Response
How often would you use it if a rented car had one?	1-always 5-sometimes 2-never
How would you use it?	3-learn when tired (and then stop) 2-not use 2-to stay awake longer (1 would use as last resort after radio) 1-sometimes to stay awake longer, sometimes to to learn when tired and stop
How effective was it in keeping you alert? (1=useless, 5=very effective)	4–2 3–3 1–4
How much would you pay for it?	4−$0 1−$5 1−$10 1−$35 1−$50

Note: In this table each answer consists of a pair of values separated by a dash. The entry after the dash indicates the response category. The value before it is the number of people choosing that category.

off, ranging from 10 minutes (10 times the current value) to 30 seconds. The general tendency was to choose durations for the normal mode that are briefer than the current setting. When response times were long, those responding thought the device should go off about every 10 seconds, half of the current setting. They felt the timing should be adjustable by the driver (6/8).

When asked if they had ever fallen asleep before while driving, five of the eight said they had fallen asleep/nodded off while driving. Of those five, three said the device would have prevented them from falling asleep. Most of the instances of falling asleep were associated with an unusual long-distance trip (e.g., driving home for a holiday) and driving at night. One subject said he ran into a snow bank when he fell asleep.

When asked about rental cars (where they don't pay for the device directly), five of the eight drivers said they sometimes would use the device, but only if they were very tired and had to push on. Drivers were mixed in their views as to whether they would use it as a warning device to stop, or as an "electronic amphetamine" to keep them awake when they are tired. Because this last application might encourage people to continue driving when they shouldn't, its extent should be investigated in detail. Although the emphasis in developing a product should be on good design to assure proper use, those intentions can be reinforced by clear instructions and supportive advertising. Some means of printing the instructions on the device or making them inseparable should be considered.

Finally, when asked how much they would pay for a device, half of the subjects indicated zero. There was a distinct difference due to age, with the young subjects offering figures of 0, 0, 5, and 10 dollars, and the old subjects reporting 0, 0, 35, and 50 dollars. It is likely that an alertness device is viewed as a highly discretionary expenditure, and price may not be as important as product features. The younger subjects were students with minimal income. The older subjects were more financially secure.

Beyond the extensive objective and subjective measurements of human performance that were collected using state-of-the-art computer techniques and were rigorously analyzed, there still remains the ultimate question, did people fall asleep. The answer is yes, twice; once in the no-radio-or-device condition and once with the device in the fixed-interval mode. In the first instance, the subject awoke quickly enough that he was able to get back on the road. In the second, the incident was long enough that the road disappeared from the screen and the subject was not able to recover. The block was stopped prematurely and the data could not be saved. Maybe more than any other result from this experiment, this incident of falling asleep suggests the device needs additional engineering work.

CONCLUSIONS

1. Based on the literature review, it is clear that fatigue is a major causal factor in road vehicle accidents and, as one would suspect, is usually associated with long-distance driving at night. A surprisingly large number of drivers admit they have fallen asleep while driving, though most of those incidents did not result in a crash.

2. It is also clear that there is considerable variability in the statistical estimates of the extent to which fatigue is associated with accidents. Estimates vary widely between accident data bases and with the baseline figure (all accidents, accidents for which the cause is known, and accidents for which some contributing cause is identified). In examining accident statistics, one must keep in mind that there is rarely a single well-identified cause, but rather a variety of factors contribute.

3. Use of the radio and response-time devices has been examined in several previous studies and shown to be potentially effective in assessing the level of driver fatigue.

4. While there were many small flaws in the experimental procedure, overall it went quite well. The road scene, full-size mockup, and road noise captured the essence of driving at night. The test times and durations assured that people were quite tired at the end of the test, and in two cases they fell asleep during the test. Except for occasional missing heart beats, the data set was complete.

5. In this experiment, low mean steering error, low standard deviation of the steering error, few extremely large steering errors, high heart rates, and low alertness scale ratings are all associated with greater levels of alertness. Except for mean steering error, the rank order of performance on all of these variables tended to be quite consistent, with the measurements indicating the lowest levels of alertness for no extra stimulation (no device or radio), greater levels for the device, and the highest levels for the radio. Of the two modes of the device, alertness, in general, was greatest with the device in the random mode. Except for the fatigue ratings, the differences between the four conditions were not statistically significant when each variable was considered independently in a complete ANOVA model. The differences were significant when one-way ANOVAs were used.

6. Subjects fell asleep twice during the test, once when neither the device nor the radio were present, and once while using the device in the fixed-interval mode. This evidence, combined with the other performance measures, indicate that the alertness device, in its current form, was much less effective in keeping drivers alert or assessing how alert drivers were than listening to the radio, though it was better than nothing at all.

7. One might speculate that if the radio helps keep people alert, and to some degree, so too does the alertness device, then the optimal solution would be to use both at the same time. That is not necessarily so. In fact, driving and using both could be much worse than just using one or nothing at all. Both the radio and the device keep the driver alert by providing a distraction, and together they could be extremely distracting. The distraction property is most evident for the radio. When listening to the radio and driving, people tend not to concentrate as much at keeping their vehicle in the exact center of their lane. That occurred in this experiment (where the highest mean steering error was in the radio condition) and is commonly observed in expressway driving.

8. Based on the data collected, there are a number of steps that can be taken to improve the effectiveness of the device. It is possible drivers might try to defeat the device by holding the button down continuously, as was reported earlier. If the software detects that condition, the tone should sound and continue to sound until the button is released. The natural response will be for drivers to pound the box, causing the button to be released.

9. The software should also trap extremely brief responses (about less than 250 milliseconds) or instances where the button is pressed but no tone has been presented. Several of the former were found in this experiment. They come about because the driver anticipates the onset of the tone and begins to move his or her hand to the button before the tone goes off. If this occurs, the processor should reschedule the tone to again sound soon thereafter. That time should be random rather than fixed.

10. The software for scheduling when the tone sounds needs to be made more reliable. While there were only four instances detected in which the intertrial interval was not what it should have been (usually they were too long), there should have been no deviations.

11. In the post-test questionnaire, the button size was reported to be about right. At other times subjects reported the device was difficult to find in the dark. Currently the box location is identified by a small red LED on it. Backlighting the entire button surface should be considered.

12. The tone intensity was clearly alerting, and subjects reported its loudness to be about right. However, there was clear sentiment for making its loudness adjustable. Further, the response time data shows that the tone should not go off until the button is pressed, though to reduce annoyance it may be desired to have its intensity reduced after a few seconds.

13. The tone schedule should be somewhat random. In this experiment subjects performed as if they were more alert with the random-interval device than with the fixed-interval version, though the differences were not statistically significant. A problem with the current device is that it is too insensitive to the subject's state of alertness. For example, in the fixed-interval mode, the processor only checks if the response time is less than six seconds, and based on that and the previous response, selects the delay for the next presentation of the tone. A more complex scheduling algorithm should have been considered. For example, it might partition the response time continuum into several intervals (say 350–2500, 2500–3000, 3000–4000, 4000–5000, 5000–7000, and >7000 milliseconds). Based on those intervals, the previous response, and possibly the response before that, a time for a subsequent presentation of the tone could be scheduled. There should be some variability in selecting that interval as the time of the tone onset needs to be somewhat unpredictable. It might also be desirable to include engine-on-time (a measure of trip duration) and time-of-day in that calculation. Those two parameters can be easily measured by the device.

14. Should additional testing be performed using the same simulation, a number of modifications to the software should be made. It is evident from the individual differences that something should be done to improve the quality of feedback during training sessions. At least one person consistently steered so that he was off center. Feedback might take the form of sounding a tone while the person was steering whose intensity or frequency (in Hertz) corresponded to the instantaneous steering error. Another possibility would be to show a plot on the screen of mean steering error for each block.

15. The software for both the driving simulation and heart beat programs needs

to be modified to verify there is enough space on the disk to save the data and to prevent the use of duplicate file names. Almost half of the data that was lost occurred for these reasons.

16. More heart beat data was lost than should have been because of poor contact between the skin and the electrodes. Standard disposable EKG electrodes should be used instead of the reusable ones provided with the Respironics device. In addition, the software should alert the experimenter by an on-screen message to instances where missed or extra beats seem to be occurring.

17. The button press software should record both the response time and the intertrial interval to the nearest millisecond to allow for a more accurate determination of when in a two-hour session a button press occurred. The hardware should be changed so that retrieved responses from the IO card don't lag actual button presses.

An important point to note about this study is that it was done at all. Many manufacturers have developed and sold products that were claimed to have safety benefits without having subjected them to testing, let alone the independent and thorough effort described in this study. All too often the product doesn't assume the quality of "making driving safer" until it reaches the advertising copywriter's desk.

In this case, the literature suggested that response-time measures could be used in driver-alertness devices. However, the prototype device examined in this experiment, while more effective than nothing at all, was less effective than a radio, an item found in most cars and trucks. The data reported here also identified many ways the prototype might be improved. Once such changes were made, the device could be more effective than a radio or other devices in assessing alertness.

Also needing further exploration is how various drivers might use the device. While the number of people who asked this was extremely small, it is troublesome to find a few saying they would use it as an electronic amphetamine, a way to stay awake when they are tired and shouldn't drive, rather than as a warning device telling them when to stop driving.

REFERENCES

Attwood, D.A. and Scott, P.L. The on-line use of performance measures to predict driving fatigue. Proceedings of the Human Factors Society – 25th Annual Meeting, 1981, 726–730.

Brown, I.D. Driving fatigue. Endeavor, 1982, 6(2), 83–90.

Brown, M.D., Engleman, L., Frane, J.W., Hull, M.A., Jennrich, R.I., and Toporck, J.D. BMDP Statistical Software, Berkeley, CA, University of California Press, 1985.

Bureau of Motor Carrier Safety (BMCS), 1980. HSRI Accident Data System Codebook, Number 82–10. The University of Michigan Transportation Research Institute, Ann Arbor, Michigan, November 1982.

Dureman, E.I. and Boden, C. Fatigue in simulated car driving. Ergonomics, 15(3), 299–308, 1972.

Eatock, B.C., Demmery, R.D., Williams, R.D., and Attwood, D.A. Portable interactive data acquisition and analysis system for driver behavior research. Transportation Research Record, Washington, D.C., 1978, 672, 44–49.

Egelund, N. Spectral analysis of heart rate variability as an indicator of driver fatigue. Ergonomics, 1982, 25(7), 663–672.

Fagerstrom, K.O. and Lisper, H.O. Effects of listening to car radio, experience, and personality of the driver on subsidiary reaction time and heart rate in a long term driving task. In R.R. Mackie (Ed.). Vigilance: Theory, Operational Performance and Physiological Correlates, New York: Plenum Press, 1977.

Fatal Accident Reporting System (FARS), 1975–1980. HSRI Accident Data System Codebook, Number 82–8. The University of Michigan Transportation Research Institute, Ann Arbor, Michigan, September 1982.

Fox, D.J. and Guire, K.E. Documentation for MIDAS (3rd ed.), The University of Michigan Statistical Research Laboratory, Ann Arbor, MI, September 1976.

Green, P. Driver understanding of fuel and engine gauges. Society of Automotive Engineers paper no. 840314, Warrendale, PA: 1984.

Harris, W. Fatigue, circadian rhythm, and truck accidents. Chapter 8 (pp. 133–146) in Vigilance: Theory, Operation Performance, and Physiological Correlates (R.R. Mackie, ed.) New York: Plenum Press, 1977.

Hulbert, S. Effects of driver fatigue. In Forbes, T.W. (Ed.). Human Factors in Highway Traffic Safety Research, Wiley, N.Y., 1972, 288–302.

Kearney, P.W. Highway Homicide. Thomas Y. Crowell Company, New York. 1966, 84–86.

Linklater, R. Fatigue and long distance truck drivers. Traffic Engineering, Proceedings, Part 4, 1980, 4, 192–201.

Lisper, H.O., Laurell, H., and Stening, G. Effects of experience of the driver on heart-rate, respiration-rate, and subsidiary reaction time in a three hours continuous driving task. Ergonomics, 1973, 16(4), 501-506.

McDonald, N. Fatigue, safety and heavy goods vehicle driver. Unpublished Ph.D. Dissertation, Trinity College, Dublin, Ireland, 1979.

McFarland, R.A. Understanding fatigue in modern life. Ergonomics, 1971, 14(2), 1-10.

Michigan, 1981. UMTRI Data System Codebook, Number 82-10. The University of Michigan Transportation Research Institute, Ann Arbor, Michigan, November 1982.

Nagasuka, Y. and Ohta, H. Changes in psychological functions in prolonged driving I, Tohoku Psychologica Folic, 1929, 38(1-4), 90-101.

North Carolina Traffic Accident Summary, January-December, 1982. North Carolina Department of Transportation, Division of Motor Vehicles. Raleigh, North Carolina, 1982, as published in Accident Facts, 1983 Edition. National Safety Council, Chicago, Illinois, 1983, 55.

Olsen, R.A. and Post, D.L. Fatigue and performance: A literature review. University Park, PA: Pennsylvania State University, Pennsylvania Transportation Institute, Report PTI 7804, February 1979.

Riemersma, J.B.J., Biesta, P.W., and Wildervanck, C. Fatigue and stress due to prolong driving and changing task demands. Warrendale, PA: Society of Automotive Engineering paper no. 770134.

Ryder, J.M., Manlin, S.A., and Kinsley, C.H. The effects of fatigue and alcohol on highway safety. Washington, D.C., Report No. DOT-HS-805-854, 1981.

Snook, S.H. and Dolliver, J.J. Driver fatigue: A study of two types of countermeasure. Proceedings of the 20th Annual Meeting of the Human Factors Society, 1976, 304-311.

State of Washington, 1983. UMTRI Data System codebook, Number 84-2. The University of Michigan Transportation Research Institute, Ann Arbor, Michigan, May 1984.

Texas, 1982. UMTRI Data System Codebook, Number 83-10. The University of Michigan Transportation Research Institute, Ann Arbor, Michigan, October 1983.

Treat, J.R., Tumbas, N.S., McDonald, S.T., Shinar, D., Hume, R.D., Mayer, R.E., Stansifer, R.L., and Castellan, N.J. Tri-level study of the causes of traffic accidents, executive summary. Technical Report No. DOT HS-805 099, Washington, D.C., 1979.

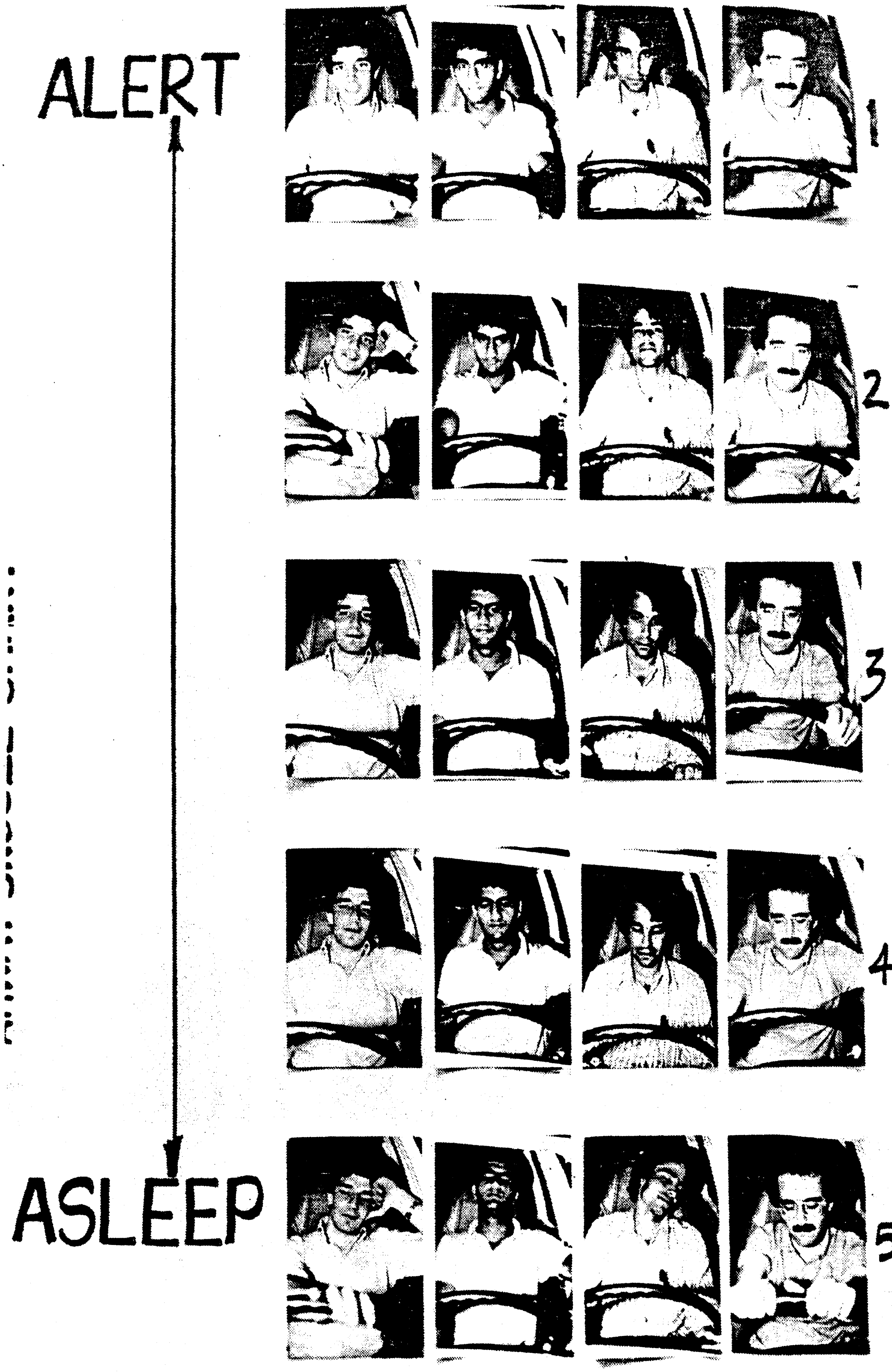
Van Der Nest, M. The role of fatigue in the driving situation. A survey of literature on psychophysiological factors. The National Road Safety Council, Johannesburg, Republic of South Africa, 1978.

Wilde, G.J.S. and Stinson, J.F. The monitoring of vigilance in locomotive engineers. Accident Analysis and Prediction, 1983, 15(2), 88-93.

Wright, P., Creighton, P., and Threlfall, S.M. Some factors determining when instructions will be read. Ergonomics, 1982, 25(3), 225-237.

Yajima, K., Ikeda, K., Oshima, M., and Sugi, T. Fatigue in automobile drivers due to long time driving. Warrendale, PA: Society of Automotive Engineering, paper no. 760050, 1976.

APPENDIX A: The "Rogues' Gallery"


ALERT
1
2
3
4
ASLEEP
5

APPENDIX B: General Instructions to Subjects

****

UNIVERSITY OF MICHIGAN TRANSPORTATION
RESEARCH INSTITUTE HUMAN FACTORS DIVISION

****

GENERAL INSTRUCTIONS TO SUBJECTS
(used during initial contact)

Notes:

Don't say it's for Amway.

Don't suggest it's a product test.

Don't say "helps keep you awake."

I WOULD LIKE TO KNOW IF YOU ARE INTERESTED IN PARTICIPATING IN AN EXPERIMENT. THE PURPOSE OF THE EXPERIMENT IS TO ASSESS THE EFFECTIVENESS OF A DEVICE FOR ASSESSING DRIVER ALERTNESS. WHILE SITTING IN A DRIVING SIMULATOR, WE WILL ASK YOU TO PRESS A BUTTON IN RESPONSE TO A TONE AND RECORD YOUR HEART RATE.

TO BE ABLE TO PARTICIPATE IN THIS EXPERIMENT, YOU NEED TO COME TO THE LAB SO WE CAN CHECK YOUR HEART RATE AND SHOW YOU THE DRIVING SIMULATOR. THE SCREENING WILL TAKE ABOUT AN HOUR AND WE'LL PAY YOU $7.00 DOLLARS FOR DOING THAT.

UNLESS THERE IS A PROBLEM, WE WILL THEN ASK YOU TO RETURN FOR FOUR TWO-HOUR SESSIONS IN THE DRIVING SIMULATOR. THOSE SESSIONS WILL HAPPEN VERY LATE AT NIGHT BECAUSE WE ARE LOOKING AT ALERTNESS PROBLEMS, AND THE BEST TIME TO TEST IS WHEN YOU ARE TIRED. FURTHERMORE, FOR THE SAKE OF CONSISTENCY, WE WILL TEST YOU THE SAME NIGHT FOUR WEEKS IN A ROW. FOR COMPLETING THE TEST YOU WILL BE PAID AN ADDITIONAL $100.

ARE YOU INTERESTED IN PARTICIPATING?

If the person says yes, then read the following:

BECAUSE WE ARE ONLY TESTING A FEW PEOPLE, IT IS ESSENTIAL THAT THEY BE IN GOOD HEALTH AND THEY NOT BE DOING ANYTHING THAT WILL ALTER THEIR ALERTNESS FROM WEEK TO WEEK. THERE ARE SEVERAL REASONS WHY WE CANNOT TEST A PERSON:

1) THEY ARE TAKING DRUGS (EITHER PRESCRIPTION ON NONPRESCRIPTION) THAT EITHER KEEP THEM AWAKE OR MAKE THEM SLEEPY

2) THEY ARE USING ILLEGAL DRUGS OF ANY KIND

3) THEY ARE WORKING ON A ROTATING OR NIGHT SHIFT. (WE WANT TO TEST "DAY" PEOPLE ONLY.)

4) THEY HAVE A HEART CONDITION

5) THEY ARE NOT A LICENSED DRIVER.

IF YOU DO NOT QUALIFY TO PARTICIPATE IN THIS EXPERIMENT, PLEASE SAY SO NOW. IT IS NOT NECESSARY TO SAY WHY.

If they are ok, then set up a date and time for the screening. Fill out and then give them the "Information for Participants ..." form. Make sure to point out the phone numbers and indicate the times are pickup and not start times.

APPENDIX C: Information for Participants in the Alertness Study

****

UNIVERSITY OF MICHIGAN TRANSPORTATION RESEARCH INSTITUTE
HUMAN FACTORS DIVISION
HURON PARKWAY & BAXTER ROAD
ANN ARBOR, MICHIGAN
48109-2150

****

Information for Participants in the Alertness Study

Scheduled sessions:

screening – ____________________

(we will NOT pick you up for this session)

test 1 – ______________________

(we will pick you up for this & all following sessions)

test 2 – ______________________

test 3 – ______________________

test 4 – ______________________

Questions:

If you have any questions about this research or there is a problem with your schedule, please call one of the people listed below at 764 4158:

Kara Heinrichs – Experimenter

Paul Green – Project Director

Flora Simon – Secretary

Do's and Don'ts

1. Try to keep your sleep schedule fixed. Even though the test time might be quite late for you, do not take a nap beforehand. It's ok if you are tired. In fact, it's best if you are. Since we want to compare performance across the four weeks, it is important that you come in the same mental and physical condition each week.

2. Each week try to get to bed the same time the night before a test.

3. Please avoid caffeine (coffee, Coke, etc.) or other stimulants or depressants (alcoholic beverages, etc.) at dinner or anytime thereafter on evenings when you are being tested.

4. Don't worry about falling asleep while driving back and forth to Transportation Research. We will pick you up at home and return you there.

5. Smoking is not permitted during the test.

6. You might want to bring a light-weight sweater with you because sometimes the test room is a bit cool.

APPENDIX D: Driver Alertness Study – Screening Session Protocol

****

UNIVERSITY OF MICHIGAN TRANSPORTATION RESEARCH INSTITUTE
HUMAN FACTORS DIVISION

****

Driver Alertness Study – Screening Session Protocol

1. Before the participant arrives turn the IBM PC on and load in (but do not run) the software. Also turn on the projection display (and check its alignment). Finally, turn on the Commodore and load the Road program.

2. Before leaving make sure the display monitors are turned down to avoid etching the screen and the room lights are off.

3. Pick up the subject.

4. Ask the subject if he wants to use the lavatory or get a drink of water. Remind them that we want to avoid "rest stops" during the two-hour test.

5. Put the electrode belt on the subject, and using the PC heart rate software, make sure it is working.

6. If it is a screening session, turn on the alertness box and demonstrate its use. Make sure the first two responses times are long (>1 or 2 seconds) so the device cycles properly.

7. Plug in the radio, turn it on, and if necessary, show how it works.

8. Flip over the sign on the door ("Experiment in progress. Do not enter.")

9. Finish loading in the driving simulation software. (runs? 1st for 2 minutes, 2nd for five? for test session there will be 1 2-hour run)

10. Turn off the room lights. Only the desk lamp by the experimenter's stations should be on.

11. Turn on the monitor and then the low-light-level TV camera.

12. When the subject is ready, begin.

13. For the screening session only, show the participant a listing or summary of the steering error data.

14. If it is a test session and the last subject for the evening, begin saving the data on disk. Start the Commodore first, since it may take a while. If it is

the last subject for a test session, turn the equipment off. To check everything is off, stand by the experimenter's station and turn off the room lights. If anything is glowing, you forgot to turn something off. Don't forget to flip the "experiment in progress" sign over.

15. If it is the end of the screening session, pay the subject $7.00. Make sure they sign the subject voucher form. If the subject's data are ok, select times for the test sessions.

16. If it is the last test session, have the subject complete the end of test survey. Then pay the subject $100.00. Have the subject sign the voucher form, and thank him for his cooperation.

17. If it is a test session, drive the subject home.

APPENDIX E: Alertness Device Questions

****

UNIVERSITY OF MICHIGAN TRANSPORTATION RESEARCH INSTITUTE
HUMAN FACTORS DIVISION

****

ALERTNESS DEVICE QUESTIONS

The biographical section of this form should be completed by the experimenter while the participant is being screened. At that time the participant should not see this form. The alerting device questions should be asked only after all four test conditions have been completed.

BIOGRAPHICAL DATA:

Name ____________

Age ____________

Sex M F (circle one)

Handedness: left right ambidextrous (circle one)

Field of Study/Occupation (be specific, e.g., mechanical engineering, European history, etc.)

Do you wear glasses or contacts when driving?

Y N (circle one)

Vision (corrected if you wear glasses or contacts, e.g., 20/20, 20/50)

ALERTING DEVICE:

Hardware:

1. Place an "X" in the drawing below, showing where you would want this device installed in your car.

2. The pushbutton is

   too small about right too big. (circle one)

3. The tone is

   too loud about right too soft (circle one)

4. If you had this device in your car, would you want to be able to adjust the loudness of the tone?

   yes no (circle one)

5. Normally the tone goes off about every 60 seconds. How often should it sound?

   every _____ seconds

6. When you don't respond fast enough, the tone now goes off every 20 seconds. How often should it sound?

   every _____ seconds

7. If you had this device in your car, would you want to be able to adjust how often it goes off?

   yes no (circle one)

APPLICATION:

1. Whether or not you like the device, how effective was it in keeping you alert? (circle one)

   1 2 3 4 5

   useless very effective

2. Have you ever fallen asleep while driving?

   yes no (circle one)

   if yes, when? ______________________________

If yes, would this device have prevented you from falling asleep?

   yes no (circle one)

3. If you rented a car that had this device installed, how often would you use it? (circle one)

   never

sometimes (if so, when?) ______________________

always

4. If you used this device, how would you use it? (circle one)

   A. would not use it

   B. to help stay alert longer (continue driving) when you are tired

   C. learn when you are tired and when warned by the device, stop and rest

   D. sometimes B, sometimes C

   E. ignore it

   F. other, please explain ______________________

5. How much would you be willing to pay for this device?

   $_______