US. Department
of Transportation

**National Highway
Traffic Safety
Administration**

DOT HS 808 299

June 1995

Semi-Annual Report

Research on Vehicle-Based Driver Status/Performance Monitoring: Seventh Semi-Annual Research Report

Technical Report Documentation Page

1. Report No. DOT HS 808 299	2 Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle RESEARCH ON VEHICLE-BASED DRIVER STATUS/PERFORMANCE MONITORING: SEVENTH SEMI-ANNUAL RESEARCH REPORT		5. Report date June 15, 1995 6. Performing Organization Code VPISU ISE
7. Author(s) Rollin J. Fairbanks Sarah E. Fahey Walter W. Wierwille		8. Performing Organization Report No. VPISU ISE 95-06
9. Performing Organization Name and Address Vehicle Analysis and Simulation Laboratory Virginia Polytechnic Institute and State University Blacksburg, VA 24061-0118		10. Work Unit No. (TRAIS) 11. Contract or Grant No. DTNH-22-91-Y-07266
12. Sponsoring Agency Name end Address Office of Crash Avoidance Research National Highway Traffic Safety Administration 400 Seventh Street, SW Washington, DC 20590		13. Type of Report and period Covered Semi-Annual Report, December 24, 1994 to May 23 1995 14. Sponsoring Agency Code NHTSA/NRD-50
15. Supplementary Notes		

16. Abstract

This report describes new findings supplementing previous project reports on driving simulation studies to develop detection algorithms and effective advisory, alarm, and alerting stimuli for a vehicle-based drowsy driver detection/warning system. Part I describes attempts to further optimize previously-reported driver drowsiness detection algorithms by the use of time-on-task variables, assessment of the predictive power of algorithms, use of prior time segment data along with current data, and "cropping" of outliers to minimize their possible spurious effects. Part II reports an experimental study to identify effective advisory, alarm, and alertness-maintaining stimuli to be used in a drowsy driver detection system. A three-stage system is conceptualized: monitoring/detection, advisory/alarm, and alertness maintenance. Various advisory, alarm, and alertness-maintainmg stimuli were tested to determine optimal stimuli to be used in a vehicle-based drowsy driver detection/warning system.

17. Key Words Drowsy Driver Detection, Driver Performance Monitoring, Asleep-at-the-Wheel, Driver Fatigue, Crash Avoidance Warning Devices, Alarm Stimuli, Alertness Maintenance.		18. Distribution Statement This document is available to the public through the National Technical Information Service, Springfield, VA 22161	
19. Security Classif. (of this paper)	20. Security Classif. (of this paper)	21. No. of Pager 125	22. Price

Form DOT F 1700.7 m-72) Reproduction of completed page authorized

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

The authors wish to thank Dr. Ron Knipling of NHTSA, who served as the contract technical representative. Dr. Knipling contributed many suggestions throughout the project and also provided results from his accident data base studies for inclusion in this report as well as in experimental planning. In addition, Dr. Knipling handled contractual matters promptly, efficiently, and constructively.

The authors also wish to thank Michael Perel, Dr. Michael Goodman, Robert Clarke, and William Leasure of NHTSA for helpful suggestions throughout the project.

EXECUTIVE SUMMARY

Part I of this document reports four attempts at further optimization of the previously reported driver drowsiness detection algorithms developed and validated by Wienville, Wreggit, Kim, Ellsworth, and Fairbanks (1994). All optimization studies utilize two groups of existing data: development data and validation data. In all cases, the development data are used for algorithm development and the validation data are used to validate the newly created algorithms.

The first section examines the effect of time-on-task variables on algorithm accuracy, the second looks at the predictive power of the algorithms, the third examines the effect of including previous time segment data with current data, and the fourth investigates the effect of eliminating outlier data from the dependent measure PERCLOS. . In the first study, time-on-task measures were found to have some improved capability, but they did not consistently add to the accuracy of the algorithms. In the second study, the predictive ability of the algorithms was found to be good. Advance detection of drowsiness is important since this allows a system more time to alert the driver or to confii drowsiness using other methods (thus reducing the false alarm rate). In the third and fourth studies, neither the use of combined preceding interval and current-interval data nor the elimination of outliers from PERCLOS data were found to significantly or consistently improve algorithm accuracy.

Part II of the document reports an experimental study which was carried out to identify effective advisory and alarm stimuli to be used in a drowsy driver detection system. The envisioned system has three stages. In the first stage, previously developed detection algorithms would compute on-line drowsiness levels. If a driver’s drowsiness level exceeds a predetermined threshold the system would proceed to stage two. At this point an initial advisory tone and a voice message would be played. If the driver does not respond, he or she would experience a m-alerting alarm. The third stage of the system

would give the driver an option of using a drowsiness countermeasure to help maintain the m-alerted state.

The goal of the present research was to determine the effectiveness of possible stimuli to be used in the second and third stages of the envisioned system. Eight initial advisory tones, two voice messages, eight alarm sounds, and five peripheral stimuli were investigated as part of stage two. In addition, six drowsiness countermeasures to be used in stage three were investigated. Eight graduate students in the Human Factors Engineering program at Virginia Tech volunteered as subjects. Subjects drove the automobile simulator throughout the experimental session. Data were collected using paired comparisons and effectiveness ratings.

The results of the study determined the characteristics of the optimal stimuli to be used in a drowsy driver detection system. These characteristics are presented below.

Initial Advisory Tone

- Duration: 0.8 second.
- Wave Shape: Rectangular.
- Frequency: 700 Hz to 1000 Hz.
- Amplitude: 8 dBa below the ambient noise level of the vehicle.

Voice Message

- Gender: Male and female voices were found equally effective.
- Amplitude: Using the peak sound levels taken for the voice messages, the mean chosen amplitude for the voice message was 10 dBA (peak) above the average ambient noise level of the vehicle. Using the average sound levels, the mean chosen amplitude was approximately 5 dBA below the average ambient noise level of the vehicle.

Alarm Stimuli

- Alarm characteristics: 1000 Hz, square wave on for 167 msec, off for 167 msec. (Repetition frequency of 3 Hz.)
- Amplitude: 3.5 dBA above ambient noise level in vehicle.
- Peripheral cue: Simulated brake pulse applied for 0.5 sec resulting in a 6 mph drop from 60 mph. Or, seat back and seat pan vibration (with specifications as given in Appendix 2D).

Countermeasure

- Peppermint scent was rated the highest on the effectiveness scales, with the lane-minder as a close second. However, the ratings were only slightly above "Moderately Effective."

This study succeeded in answering many question regarding stimuli to be used in a drowsy driver detection, advising, and alerting system. As stated above, the results indicated very effective stimuli to be used in the advising and alerting stages of the envisioned system.

PART I

ADDITIONAL DETECTION ALGORITHM OPTIMIZATION STUDIES

by

Rollin J. Fairbanks

and

Walter W. Wierwille

OVERVIEW

This part reports on four attempts at optimization of the previously reported driver drowsiness detection algorithms (Wierwille, Wreggit, Kim, Ellsworth, and Fairbanks, 1994). These studies use the previously reported development data for algorithm development and validation data for algorithm validation. The algorithm development and validation procedures followed in these studies are exactly the same as for the original studies. This consistency allows the success of new algorithms to be directly compared to that of the original algorithms.

Study 1 reports the effect of time-on-task variables on algorithm accuracy, Study 2 reports the predictive power of the algorithms, Study 3 reports the effect of mixing previous time segment data with current data, and Study 4 reports the effect of eliminating outlier data from the dependent measure PERCLOS.

STUDY 1.

EFFECTS OF TIME-ON-TASK VARIABLES

Introduction

The purpose of this project was to examine the ability of time on driving task measures to increase predictability when added to driver drowsiness detection algorithms. It was hypothesized that level of drowsiness tends to increase with driving time. Intuitively, it follows that the introduction of linear- and square-time variables to the algorithms would bias the output towards the drowsiness threshold, thus increasing detection accuracy.

Procedure

Typical algorithms were selected from previous studies (Wierwille, et al., 1994) to allow direct comparisons of algorithm accuracy. Six variable groups were selected from data obtained in phase II of the study, and compared to a total of three dependent variables: PERCLOS, EYEMEAS, and AVEOBS. A linear measure of time on task and a second order (squared) measure of time on task were each added to these six groups yielding a total of 12 variable groups (Tables 1.1 and 1.2). Multiple linear regression was performed on each variable group, optimal algorithms were obtained, and these algorithms were applied to validation data from phase III of the study. Classification matrices were developed and R values were obtained for each new algorithm, and these data were compared to classification matrix and R value data previously obtained using variable groups which did not include time-on-task measures.

Results

Twelve cases were examined. Six of these cases included the linear-time variable, and six included the squared-time variable. The results of these analyses are detailed in Tables 1.3 and 1.4. For half of these cases, the addition of a time variable produced no increase in R values or APAR rates over the original algorithms. In two cases a small increase was produced by the addition of time variables (in the magnitude range of 0.006 to

0.007 for R value, and 0.004 to 0.026 for APAR value). In four cases a modest increase was produced (in the magnitude range of 0.02 to 0.12 for R value, and 0.003 to 0.03 APAR value).

Table 1.1. Explanation of Variable Groups.

Variable Group	**Dependent Variables Used**	**Independent Variable Categories Included**
Group D	1. PERCLOS 2. AVEOBS	Steering and Acceleration
Group F	3. PERCLOS 4. EYEMEAS	Steering, Acceleration, LANDEV, LANVAR, LNMNSQ, LANEX, LNERRSQ
Group J	5. PERCLOS	Steering, Acceleration, and A/O Task
Group L	6. PERCLOS	NO Task, LANDEV, LANVAR, LANEX, LNMNSQ, LNERRSQ

Table 1.2. Explanation of Time Variables.

Time Variable	**Definition**
LINTIME	(T) 1/60, where T is the number of minutes since the start of the run.
SQTIME	(T)^2 (1/3600), where T is the number of minutes since the start of the run.

Conclusions

LINTIME and SQTIME were found to have some predictive capability although they did not consistently add predictive value to the algorithms.

Since data systems have built in time keeping components, time variables would be straightforward additions to implement. In most cases the improvement in algorithm accuracy contributed by time variables can be obtained using other measures already in use. As a result, it is important to carefully examine the benefits balanced against the costs of time variable use.

All testing was performed on subjects who had been deliberately sleep deprived prior to data collection. As a result, these subjects would be more likely to show trends of drowsiness over time and more profound increases with time than would normal drivers. Tying algorithms to time may potentially reduce overall validity when applied to real road conditions. Generally, drowsiness is far less prevalent on the road than in laboratory conditions using sleep deprived drivers. This may result in an increase in the false alarm rate by biasing the algorithm output towards the drowsiness threshold, especially as time goes on. Time variables increase in magnitude as time on task progresses. This increase will tend to bias the algorithm output towards the drowsiness threshold, thus increasing the potential for false alarms, particularly in cases of prolonged driving times.

Close examination of Variable Croup L data in Table 1.3 shows a slight reduction in R value with inclusion of LINTIME during the development phase. Theoretically, LINTIME should have been eliminated as an independent variable during the multiple regression process. It is clear that an algorithm can be developed without LINTIME to yield a slightly higher R value. It should be noted that in practice, multiple regression has a subjective component, the effect of which can be reduced with exhaustive substitution and resubstitution at the end of the regression analyses. This reduction in R value did not appear when the algorithms were applied to the validation data.

In this analysis of a number of “typical” variable groups, inclusion of time on task variables increased the accuracy of the resulting algorithms in one half of the cases. These contributions were modest at best and not consistent. The advantages of using time variables needs to be carefully balanced against the disadvantages (as discussed).

Table 1.3. Summary of Regression and Classification Matrices Results from inclusion of Time Variables: PERCLOS Dependent Variable.

VARIABLE GROUP	TERMS USED IN REGRESSION	ACCURACY MEASURE TYPE	ORIGINAL DATA SET	VALIDATION DATA SET
VARIABLE *GROUP D:* Steering and Acceleration vs. *PERCLOS*	w/o time variables (Earlier D analyses)	R Value	0.789	0.800
		APAR (large errors)	0.987	0.965
		APAR (all errors)	0.757	0.771
	w/ LINTIME	R Value	0.796	0.792
		APAR (large errors)	0.987	0.969
		APAR (all errors)	0.783	0.767
	w/ SQTIME	R Value	0.795	0.791
		APAR (large errors)	0.987	0.969
		APAR (all errors)	0.780	0.764
VARIABLE *GROUP F:* Steering, Acceleration, LANDEV, LANVAR, LNMNSQ, LANEX, LNERRSQ, vs. *PERCLOS*	w/o time variables (Earlier F analyses)	R Value	0.872	0.862
		APAR (large errors)	0.980	0.979
		APAR (all errors)	0.790	0.830
	w/ LINTIME (LINTIME actually removed by regression)	R Value	0.872	0.863
		APAR (large errors)	0.980	0.979
		APAR (all errors)	0.790	0.830
	w/ SQTIME (SQTIME actually removed by regression)	R Value	0.872	0.863
		APAR (large errors)	0.980	0.979
		APAR (all errors)	0.790	0.830
VARIABLE *GROUP J:* Steering, Acceleration, and A/O Task vs. *PERCLOS*	w/o time variables (Earlier J analyses)	R Value	0.836	0.599
		APAR (large errors)	0.970	0.958
		APAR (all errors)	0.790	0.833
	w/ LINTIME	R Value	0.863	0.698
		APAR (large errors)	0.990	0.986
		APAR (all errors)	0.830	0.854
	w/ SQTIME	R Value	0.856	0.719
		APAR (large errors)	0.990	0.986
		APAR (all errors)	0.820	0.861
VARIABLE *GROUP L:* A/O Task LANDEV/VAR, LNMNSQ, LANEX, AND LNERRSQ vs. *PERCLOS*	w/o time variables (Earlier L analyses)	R Value	0.875	0.796
		APAR (large errors)	0.950	0.979
		APAR (all errors)	0.930	0.868
	w/ LINTIME	R Value	0.872	0.836
		APAR (large errors)	0.960	0.979
		APAR (all errors)	0.830	0.826
	w/ SQTIME	R Value	0.763	0.738
		APAR (large errors)	0.980	0.979
		APAR (all errors)	0.830	0.833

Table 1.4. Summary of Regression and Classification Matrices Results from inclusion of Time Variables: Other Dependent Variable.

VARIABLE GROUP	TERMS USED IN REGRESSION	ACCURACY MEASURE TYPE	ORIGINAL DATA SET	VALIDATION DATA SET
VARIABLE *GROUP D:* Steering and Acceleration vs. *AVEOBS*	w/o time variables (Earlier D analyses)	R Value	0.747	0.727
		APAR (large errors)	0.920	0.944
		APAR (all errors)	0.687	0.785
	w/ LINTIME	R Value	0.815	0.841
		APAR (large errors)	0.930	0.965
		APAR (all errors)	0.670	0.799
	w/ SQTIME	R Value	0.786	0.813
		APAR (large errors)	0.923	0.962
		APAR (all errors)	0.663	0.806
VARIABLE *GROUP F:* Steering, Acceleration, LANDEV, LANVAR, LNMNSQ, LANEX, LNERRSQ, vs. *EYEMEAS*	w/o time variables (Earlier F analyses)	R Value	0.837	0.838
		APAR (large errors)	0.963	0.976
		APAR (all errors)	0.780	0.903
	w/ LINTIME (LINTIME actually removed by regression)	R Value	0.843	0.837
		APAR (large errors)	0.930	0.976
		APAR (all errors)	0.767	0.892
	w/ SQTIME (SQTIME actually removed by regression)	R Value	0.839	0.838
		APAR (large errors)	0.937	0.972
		APAR (all errors)	0.777	0.903

STUDY 2.

THE PREDICTIVE POWER OF DRIVER DETECTION ALGORITHMS: COMPARING REAL TIME ACCURACY WITH TWO, FOUR, SIX, AND TWELVE MINUTE ADVANCE INTERVALS.

Introduction

Previous research conducted under the current contract has been directed at developing algorithms for detection of drowsy drivers. These algorithms used a number of independent measures in categories such as lane-keeping, steering, seat, angular accelerometer, display-yaw, secondary task, and heart-related measures. Dependent measures were selected based on small studies which were conducted to determine and validate operational definitions of drowsiness.

Research thus far has used these independent variables to determine the corresponding dependent measures (current driver drowsiness level) at time t. The algorithm development process was based on 6 minute average data, and the resulting algorithms have been validated with data collected using a second group of subjects. Because of the use of six minute averages, there was some degree of "smoothing" of the data. The possibility also exists that, during on the road implementation, immediate warning may not allow for fast enough response as the driver becomes drowsy. Although all previously developed algorithms have been validated on a new group of subjects, the success of these algorithms given their response times has not been tested. Thus, to be complete, we must look at the predictive ability of algorithms based on these variables.

The purpose of this adjunct study was to evaluate the ability of various independent measures of driver performance to ***predict future*** driver drowsiness level. Specifically, algorithms were developed and validated to predict driver drowsiness level two minutes, four minutes, six minutes, and twelve minutes in advance of the independent measure data collection.

Method

Preparation of data for this study was quite involved. All of the data obtained in the main study were prepared, manipulated, and analyzed using six minute averages for all measures. As a result, all refmed data were in this form. In preparation for this study, the raw data which had been collected during the algorithm development phase of the main study were obtained from archival data bases as one-minute averages.

Baselining of Raw Data. The raw data, which consisted of 150 data sets for each of twelve subjects, were baselined using exactly the same procedure that was followed in the main study. Starting with one-minute average data for each subject, the first two minutes of data were deleted. The average of the first ten remaining minutes was calculated for each separate variable, and this average was then subtracted from every value in the data column, including the first ten. To confirm accuracy and correctness of these data, six minute averages were calculated and compared to the six minute averages previously calculated in the original study.

Each row of information contained the average values of each measure for a given subject during a specific minute. The one minute data were modified in four different ways to create four new data bases: two minute, four minute, six minute, and twelve minute advance data.

Preparation of Advance Data. Two minute advance data were created by deleting the first two minutes of dependent measure data and the last two minutes of independent measure data. This change was made for each of the twelve subjects, resulting in data sets that contained dependent measures which were two minutes in advance of the corresponding independent measures. This process was similarly performed on baselined data to create data files containing independent measure data which were four, six, and twelve minutes ahead of dependent measure data.

In final preparation of the data six minute averages were calculated for all modified data, and for each group (two, four, six, and twelve-minute advance data) complete data sets were formed by combining all subject data into one large file (subject 1 data, followed by subject 2 data. etc.).

Multiple Repression. Multiple regression was performed on these data to produce algorithms with outputs estimating the dependent definitional measure of drowsiness "PERCLOS." Two different groups of independent measures were used. Designation of these groups, "D" and "F," correspond to group names in the main study. Procedures in the analysis were followed according to those used in the main study to minimize variation due to technique.

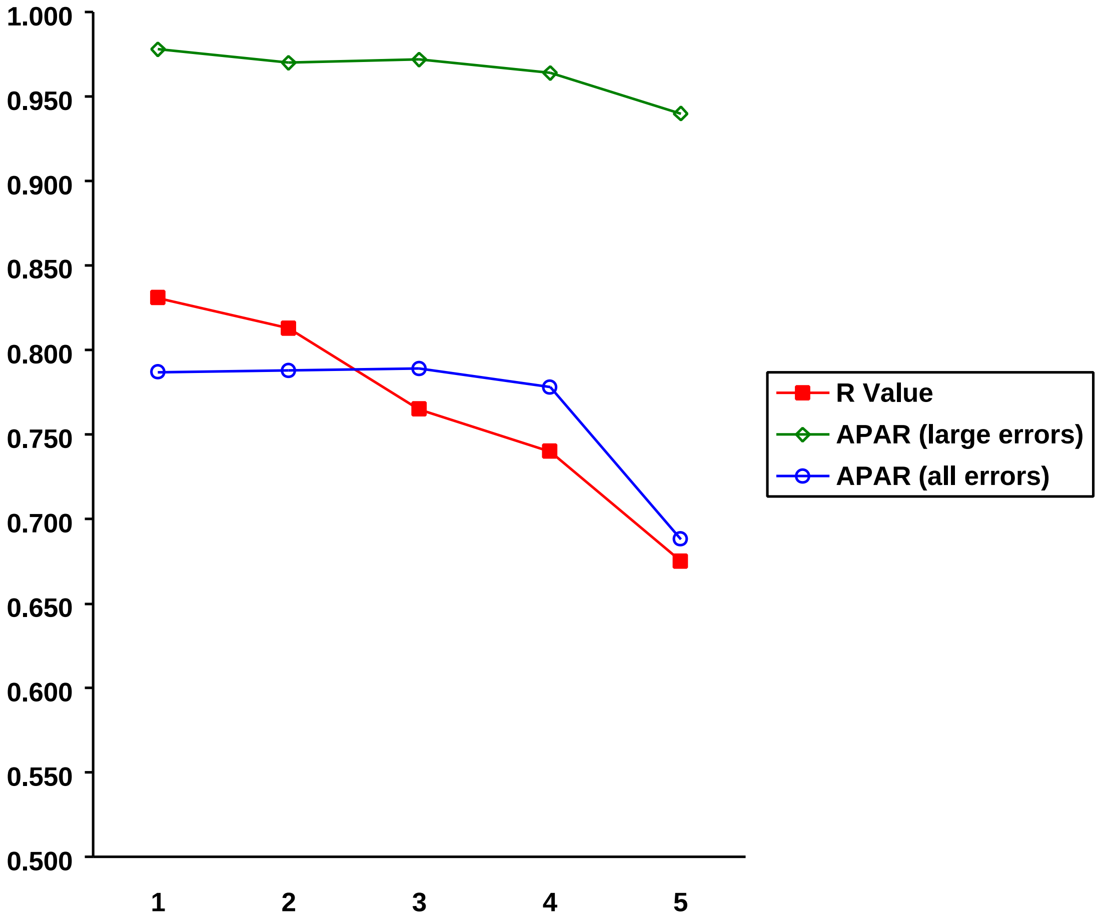
Validation of Resulting Algorithms. The developed algorithms were then applied to the validation data set gathered earlier. Data collected in the validation phase of the main study were prepared using the same procedure as previously described (in preparation of the algorithm development data).

Results

Multiple regression R values and classification matrices were determined for every data set in each variable group. Classification matrices were created using criteria which were defined in the main study. Apparent accuracy rate (APAR) with respect to large errors, and APAR with respect to all errors were calculated. The main study specified thresholds of PERCLOS values defining "drowsy," "questionable," and "awake." A large error was defined as a predicted classification of "drowsy" corresponding to a observed (measured) classification of "awake," or vice versa. All error is defined as any inconsistency between predicted and measured classification. R value and classification matrix results for all regressions and validations are shown in Table 1.5 along with the real time data as found and previously reported. Average values are included in this table as well, to allow the reader to get a general feeling for the trend of the data. Figures 1.1 through 1.5 graphically illustrate these trends. The algorithms developed for prediction appear in Tables 1.6 and 1.7.

Table 1.5. Summary of prediction algorithm accuracy results compared with corresponding real-time results.

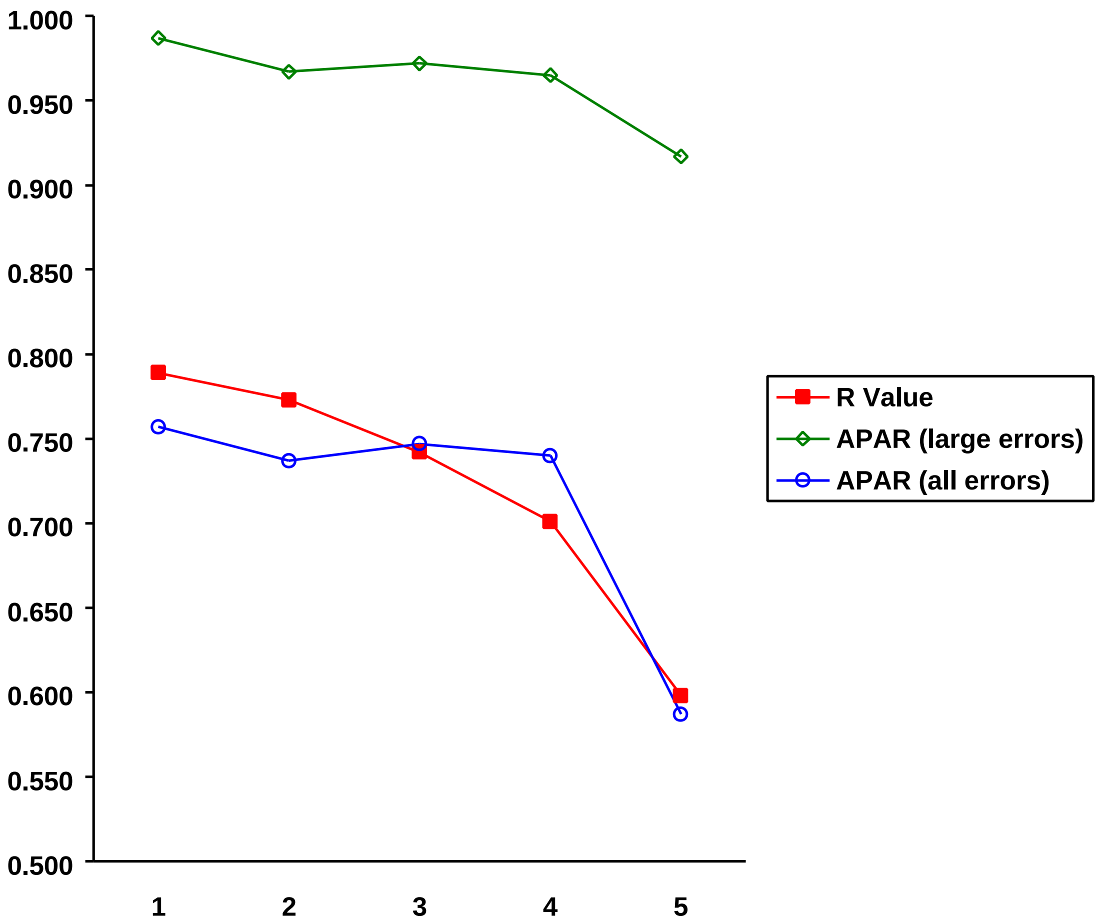
VARIABLE GROUP	DATA SET	ACCURACY MEASURE	REAL TIME	2 MIN ADV	4 MIN ADV	6 MIN ADV	12 MIN ADV
GROUP D: Steering and Acceleration vs. *PERCLOS*	DEV	R Value	0.789	0.773	0.742	0.701	0.598
		APAR (large errors)	0.987	0.967	0.972	0.965	0.917
		APAR (all errors)	0.757	0.737	0.747	0.740	0.587
	VAL	R Value	0.800	0.796	0.785	0.759	0.678
		APAR (large errors)	0.965	0.958	0.969	0.949	0.962
		APAR (all errors)	0.771	0.785	0.806	0.815	0.765
GROUP F: Steering, Acceleration, LANDEV, LANVAR, LNMNSQ, LANEX, LNERRSQ, vs. *PERCLOS*	DEV	R Value	0.872	0.815	0.771	0.733	0.650
		APAR (large errors)	0.980	0.977	0.969	0.972	0.902
		APAR (all errors)	0.790	0.773	0.750	0.736	0.620
	VAL	R Value	0.862	0.870	0.762	0.766	0.775
		APAR (large errors)	0.979	0.976	0.979	0.971	0.977
		APAR (all errors)	0.830	0.858	0.851	0.822	0.780
AVERAGES	DEV	R Value	0.831	0.794	0.757	0.717	0.624
		APAR (large errors)	0.984	0.972	0.971	0.969	0.910
		APAR (all errors)	0.774	0.755	0.749	0.738	0.604
	VAL	R Value	0.831	0.833	0.773	0.762	0.727
		APAR (large errors)	0.972	0.967	0.974	0.960	0.970
		APAR (all errors)	0.801	0.822	0.829	0.819	0.773
	ALL AVE	R Value	0.831	0.813	0.765	0.740	0.675
		APAR (large errors)	0.978	0.970	0.972	0.964	0.940
		APAR (all errors)	0.787	0.788	0.789	0.778	0.688



Key to x-axis labels

1- Real Time
2- 2 Minute Advance
3- 4 Minute Advance
4- 6 Minute Advance
5- 12 Minute Advance

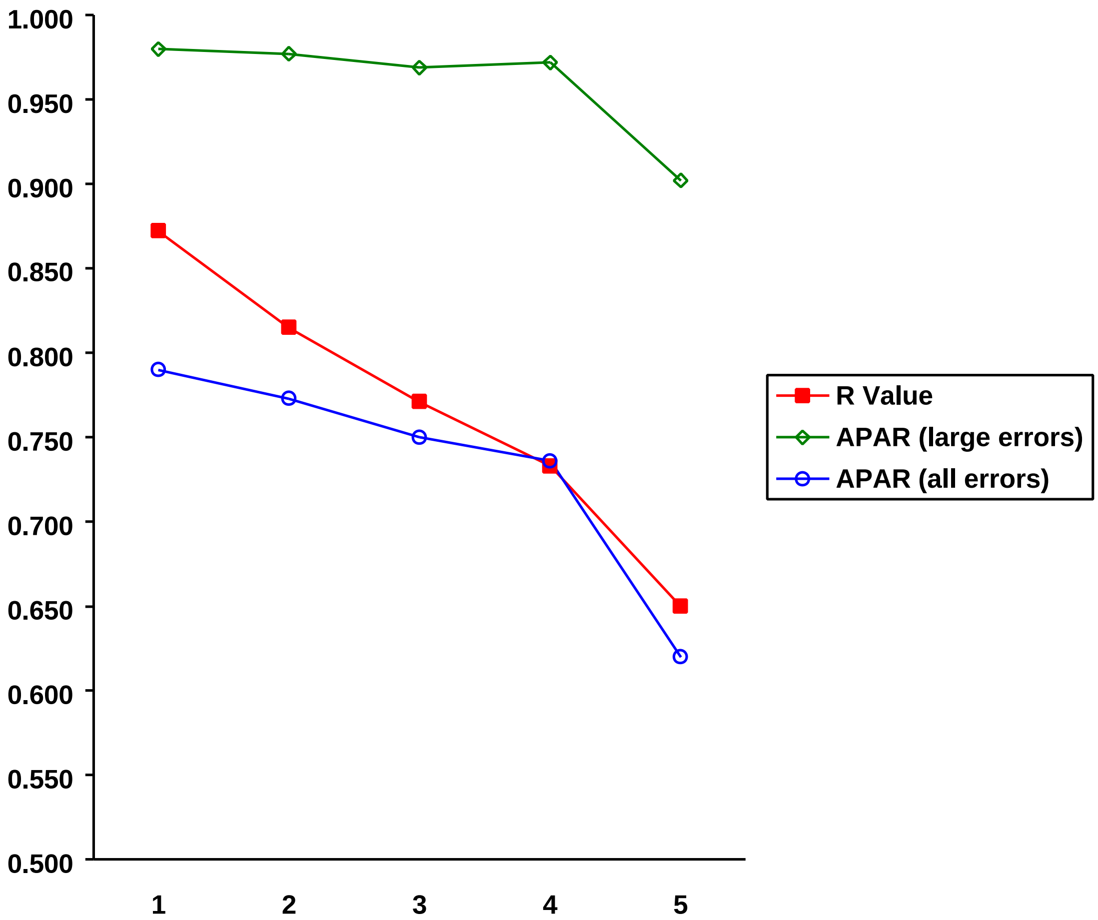
Figure 1.1. Overall average accuracy of algorithms as a function of advance time.



Key to x-axis labels

1- Real Time
2- 2 Minute Advance
3- 4 Minute Advance
4- 6 Minute Advance
5- 12 Minute Advance

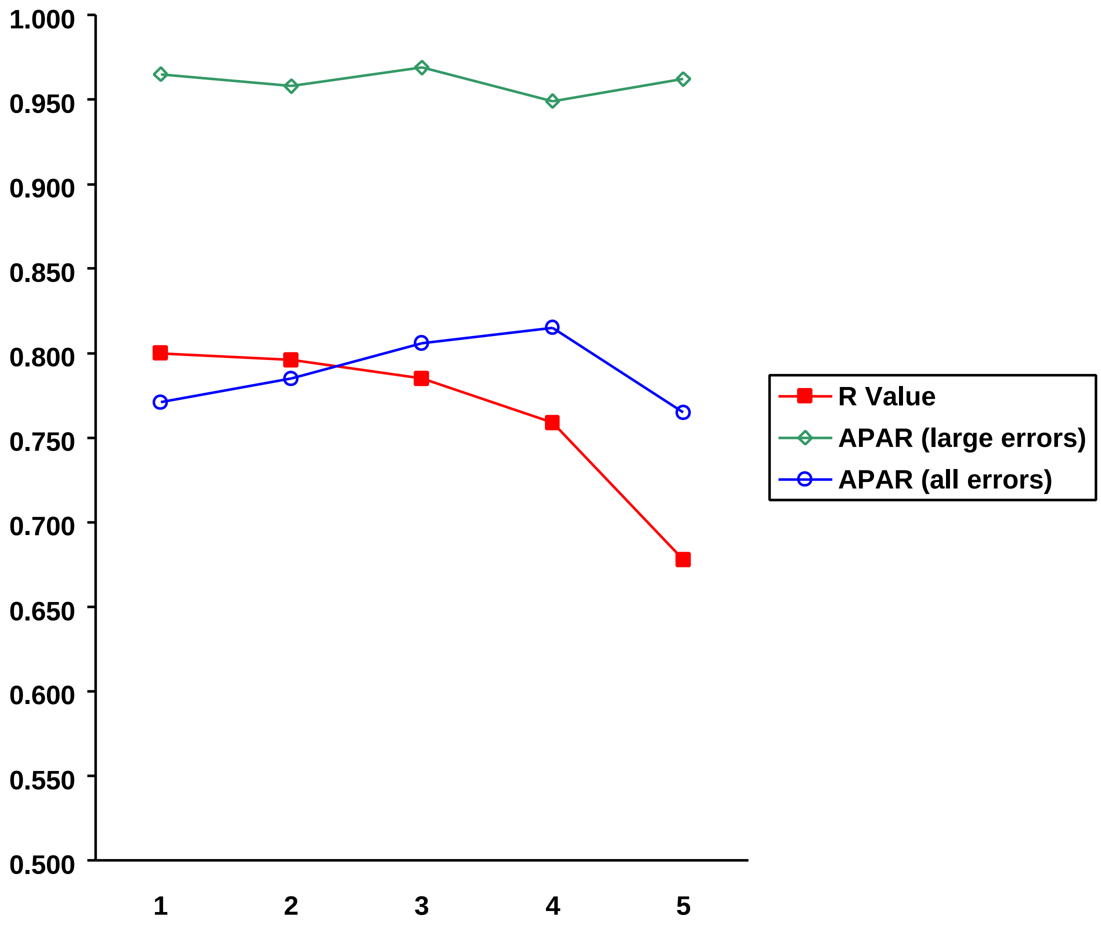
Figure 1.2. Accuracy of the Group D algorithms for the development data set.



Key to x-axis labels

1- Real Time
2- 2 Minute Advance
3- 4 Minute Advance
4- 6 Minute Advance
5- 12 Minute Advance

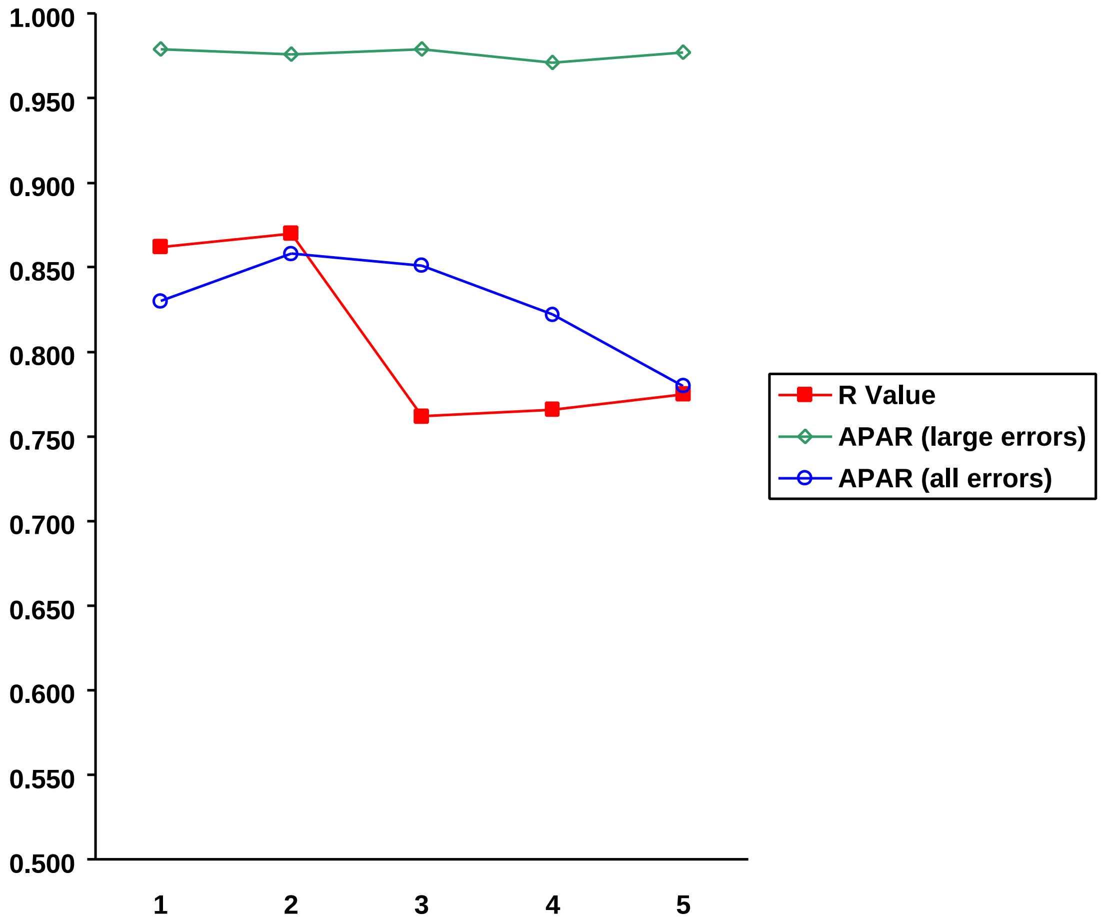
Figure 1.3. Accuracy of the Group F algorithms for the development data set.



Key to x-axis labels

1- Real Time
2- 2 Minute Advance
3- 4 Minute Advance
4- 6 Minute Advance
5- 12 Minute Advance

Figure 1.4. Accuracy of the Group D algorithms for the validation data set.



Key to x-axis labels

1- Real Time
2- 2 Minute Advance
3- 4 Minute Advance
4- 6 Minute Advance
5- 12 Minute Advance

Figure 1.5. Accuracy of the Group F algorithms for the validation data set.

Table 1.6. Multiple Regression Summaries for Group D with 2, 4, 6, and 12 Minute Advances. (Corresponding real-time regression summary appears in Wreggit, Kim, and Wierwille, 1993, p. 65.)

Regression Summary for Dependent Variable: PERCLOS, 2 Minute Advance Data. R = 0.77326541 R^2 = 0.59793940 Adjusted R^2= 0.58546166 F(9,290) =47.920 p<0.0000 Std.Error of estimate: 0.06343						
	BETA	St. Err. of BETA	B	St. Err. of B	t(290)	p-level
Intercpt			0.01578	0.00628	2.51214	0.01254
STVELV	0.25866	0.08563	0.00085	0.00028	3.02052	0.00275
LGREV	0.34659	0.07836	0.03495	0.00790	4.42306	0.00001
MDREV	-0.16731	0.06723	-0.00383	0.00154	-2.48850	0.01339
NMRHOLD	-0.29640	0.05884	-0.00515	0.00102	-5.03750	0.00000
THRSHLD	0.36236	0.05008	0.33794	0.04671	7.23540	0.00000
ACCVAR	-0.83456	0.20684	-0.03556	0.00906	-4.03478	0.00007
ACCDFV	0.84298	0.23227	0.13834	0.03812	3.62930	0.00034
INTACDE	-0.19723	0.07179	-0.12254	0.04460	-2.74719	0.00639
ACEXEED	0.21753	0.06240	2.03949	0.58505	3.48603	0.00057

Regression Summary for Dependent Variable: PERCLOS, 4 Minute Advance Data. R = 0.74237398 R^2= 0.55111912 Adjusted R^2= 0.54153448 F(6,281) = 57.500 p<0.0000 Std.Error of estimate: 0.06721						
	BETA	St. Err. of BETA	B	St. Err. of B	t(281)	p-level
Intercpt			0.0981	0.00542	1.80821	0.07164
LGREV	0.53794	0.05675	0.05320	0.00561	9.47967	0.00000
NMRHOLD	-0.26715	0.05917	-0.00464	0.00103	-4.51515	0.00001
THRSHLD	0.31107	0.05202	0.29358	0.04909	5.98016	0.00000
ACCVAR	-0.58659	0.20360	-0.02373	0.00824	-2.88104	0.00427
ACCDFV	0.45670	0.19513	0.07165	0.03061	2.34050	0.01996
ACEXEED	0.19855	0.05784	1.81877	0.52980	3.43295	0.00089

Regression Summary for Dependent Variable: PERCLOS, 6 Minute Advance Data. R = 0.70142431 R^2= 0.49199606 Adjusted R^2= 0. 48662982 F(3,284) = 91.684 p<0.0000 Std.Error of estimate: 0.07093						
	BETA	St. Err. of BETA	B	St. Err. of B	t(284)	p-level
Intercpt			0.01181	0.00560	2.10854	0.03585
LGREV	0.62232	0.05253	0.06465	0.00546	11.84687	0.00000
NMRHOLD	-0.14441	0.05278	-0.00253	0.00110	-2.30003	0.02217
THRSHLD	0.18626	0.05335	0.17518	0.05017	3.49150	0.00056

Regression Summary for Dependent Variable: PERCLOS, 12 Minute Advance Data. R = 0.73307939 R^2= 0.53740540 Adjusted R^2= 0.52752793 F(6,281) = 54.407 p<0.0000 Std.Error of estimate: 0.08805						
	BETA	St. Err. of BETA	B	St. Err. of B	t(290)	p-level
Intercpt			-0.00082	0.00600	-0.13708	0.89107
INTACDE	-0.11247	0.04706	-0.07098	0.02970	-2.39014	0.01750
ACEXEED	0.11012	0.05138	1.06862	0.49858	2.14333	0.03295
LANVAR	-0.33796	0.12795	-0.00245	0.00093	-2.64126	0.00872
LANDEV	0.42122	0.19122	0.03275	0.01487	2.20281	0.02842
LANEX	0.19928	0.08977	0.16370	0.07374	2.22001	0.02722
LGREV	0.38613	0.07220	0.04011	0.00750	5.34801	0.00000

Table 1.7. Multiple Regression Summaries for Group F with 2,4,6, and 12 Minute Advances. (Corresponding real time regression summary appears in Wreggit, Kim, and Wierwille, 1993, p. 87.)

Regression Summary for Dependent Variable: PERCLOS, 2 Minute Advance Data. R = 0.81495611 R^2 = 0.66415347 Adjusted R^2= 0.65610235 F(7,292) =82.492 p<0.0000 Std.Error of estimate: 0.05777						
	BETA	St. Err. of BETA	B	St. Err. of B	t(292)	p-level
Intrcpt			0.00390	0.00474	0.82227	0.41160
STVELV	0.14156	0.06942	0.00046	0.00023	2.03933	0.04232
LGREV	0.18249	0.07332	0.01840	0.00739	2.48875	0.01338
NMRHOLD	-0.13904	0.05489	-0.00241	0.00095	-2.53301	0.01183
THRSHLD	0.22236	0.04761	0.20737	0.04440	4.67060	0.00000
ACCVAR	-0.15717	0.04068	-0.00689	0.00178	-3.86380	0.00014
LANDEV	0.27408	0.07200	0.02079	0.00546	3.80684	0.00017
LANEX	0.25599	0.06468	0.20372	0.05147	3.95796	0.00009

Regression Summary for Dependent Variable: PERCLOS, 4 Minute Advance Data. R = 0.77068459 R^2= 0.59395474 Adjusted R^2= 0.58528474 F(6,281) = 68.507 p<0.0000 Std.Error of estimate: 0.06392						
	BETA	St. Err. of BETA	B	St. Err. of B	t(281)	p-level
Intrcpt			-0.00195	0.00529	-0.36967	0.71190
STVELV	0.49870	0.07954	0.00160	0.00025	6.27008	0.00000
STEXED	-0.28584	0.05895	-139.6402	28.79877	-4.84883	0.00000
THRSHLD	0.13312	0.04196	0.12564	0.03960	3.17230	0.00168
ACCVAR	-0.09591	0.04658	-0.00388	0.00188	-2.05913	0.04040
LANDEV	0.30771	0.08337	0.02400	0.00650	3.69089	0.00027
LANEX	0.26125	0.07397	0.21150	0.05988	3.53204	0.00048

Regression Summary for Dependent Variable: PERCLOS, 6 Minute Advance Data. R = 0.73307939 R^2= 0.53740540 Adjusted R^2= 0. 52752793 F(6,281) = 54.407 p<0.0000 Std.Error of estimate: 0.06805						
	BETA	St. Err. of BETA	B	St. Err. of B	t(281)	p-level
Intercpt			-0.00082	0.00600	-0.13708	0.89107
INTACDE	-0.11247	0.04706	-0.07098	0.02970	-2.39014	0.01750
ACEXEED	0.11012	0.05138	1.06862	0.49858	2.14333	0.03295
LANVAR	-0.33796	0.12795	-0.00245	0.00093	-2.64126	0.00872
LANDEV	0.42122	0.19122	0.03275	0.01487	2.20281	0.02847
LANEX	0.19928	0.08977	0.16370	0.07374	2.22001	0.02722
LGREV	0.38613	0.07220	0.04011	0.00750	5.34801	0.00000

Regression Summary for Dependent Variable: PERCLOS, 12 Minute Advance Data. R = 0.65029808 R^2= 0.42288759 Adjusted R^2= 0.41220032 F(5,270) = 39.589 p<0.0000 Std.Error of estimate: 0.07672						
	BETA	St. Err. of BETA	B	St. Err. of B	t(290)	p-level
Intercpt			0.00908	0.00696	1.30478	0.19308
ACCDEV	-0.24865	0.06264	-0.04140	0.01043	3.96939	0.00009
ACEXEED	0.13654	0.05373	1.37607	0.54152	2.54113	0.01161
LANDEV	0.40847	0.08065	0.03286	0.00649	5.06474	0.00000
LANEX	0.16853	0.08445	0.14595	0.07314	1.99552	0.04699
MDREV	0.23498	0.06064	0.00537	0.00139	3.87506	0.00013

Discussion

On the average, there was no appreciable degradation in APAR for large errors between algorithms developed using real time, two minute advance, four minute advance, and six minute advance (Figure 1.1). A slight decrease in accuracy occurred for twelve minute advance.

The APAR rates produced by the algorithms when they are applied to validation data stay fairly stable through the six minute advance intervals for both variable groups. The R values remain stable through the six minute advance interval when applied to group D algorithms (Figure 1.4), although they decrease significantly with four minute advance data when applied to group F algorithms (Figure 1.5).

Losses in algorithm accuracy are small over the range of zero to four minutes advance, and slight degradation is seen at six minute advance with a significant loss at twelve minute advance. In implementation of these algorithms for on-the-road use, four minutes should allow plenty of advance warning.

For both group F and group D, the development data R values experienced a steady decline as the tune interval between dependent and independent data increased. The APAR for large errors remained stable through the 2,4, and 6 minute intervals. As one might expect, there was a significant drop between the 6 and 12 minute intervals. In both groups, APAR values remain fairly stable through the six minute advance interval, and drop off at twelve.

These results suggest that there is minimal degradation in algorithm accuracy when the algorithms are predicting drowsiness up to six minutes in advance. These findings suggest that if earlier algorithms do not detect drowsiness soon enough, algorithms with prediction capability can be substituted which would not appreciably degrade detection performance. The structure of this study, which uses data collected from one group of subjects to develop algorithms and data collected from an independent group of subjects to validate them, contributes credibility to the results.

STUDY 3

TIME TREND ANALYSIS: THE EFFECT OF COMBINING PRECEDING-INTERVAL(S) AND CURRENT-INTERVAL DATA ON DRIVER DROWSINESS DETECTION ALGORITHMS

Introduction

All attempts at algorithm development and optimization thus far have used concurrently collected independent measure data. In this optimization study, independent measure data for the current six-minute interval, first previous six-minute interval, and second previous six-minute interval were used together to develop driver-drowsiness detection algorithms. The goal was to determine whether or not inclusion of previous interval data would improve detection accuracy.

Method

A previously developed algorithm, F4a (Wreggit, Kirn, and Wierwille, 1993), was selected to be used for comparison in this study, based on it's stability. This algorithm was developed starting with all steering and accelerometer measures, as well as LANDEV, LANVAR, LNMSQ, LANEX, and LNERRSQ. Similarly, these variables were used in this study, although in three forms:

a- current six minute interval

b- previous six minute interval

c- second previous six minute interval

Two studies were conducted. Study 1A involved the current and previous six minute intervals (a and b), and study 2A involved the current and two previous six minute intervals (a, b and c). In both cases, measures from each time segment were considered independent variables to be included in multiple regression, and the current six minute interval of the measure PERCLOS was the dependent measure.

Multiple regression was performed on both variable groups using the previously collected development data, and algorithms were obtained. R values were determined and classification matrices were created using criteria which are defined in the main study. Apparent accuracy rate (APAR) with respect to large errors, and APAR with respect to all errors were calculated. The algorithms were applied to previously collected validation data (Wierwille, et al., 1994), R values were determined, and classification matrices were calculated. These results were compared to the results previously obtained using current six minute interval data only (Wierwille, et al., 1994).

Results

The resulting data, including R value, APAR for large errors, and APAR for all errors, are presented in Table 1.8 along with the corresponding original data.

R values for the development data increase for group 1A and 2A by 0.009 and 0.011 respectively with respect to the development data. APAR (large errors) values increase by 0.003 and 0.006, and APAR (all errors) increase by 0.019 and 0.007. R values for the validation data decrease for group 1A and 2A by 0.006 and 0.021 respectively with respect to the original data. APAR (large errors) values increase by 0.006 and 0.001, and APAR (all errors) increase by 0.061 for study 1A and decrease by 0.03 1 in study 2A. These data appear in graphical form in Figure 1.6 for the purposes of comparison. The regression algorithms producing these results are presented in Table 1.9 and 1.10.

Discussion

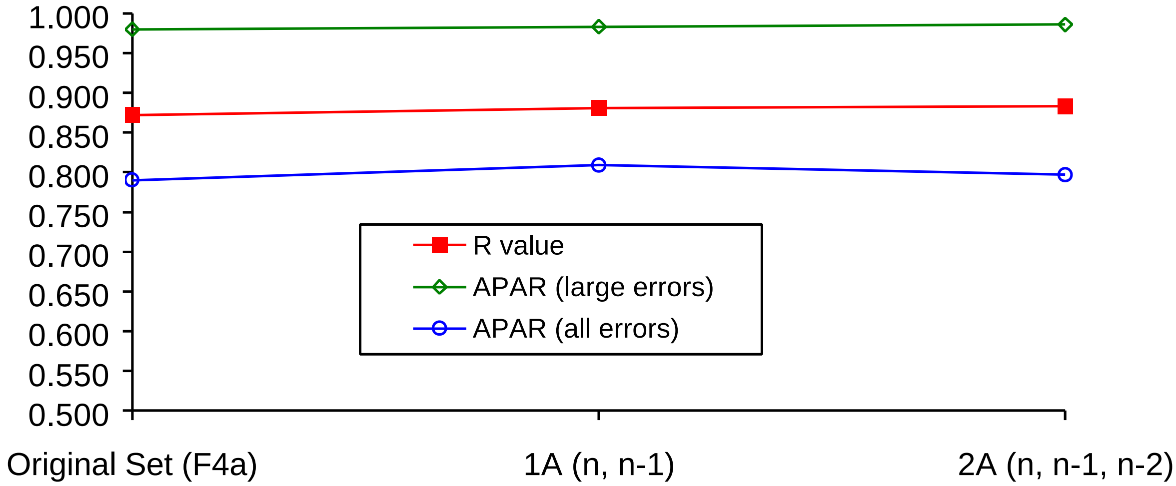
The use of previous time intervals in algorithm development produced a small improvement in algorithm accuracy for the development data, which did not remain stable when applied to validation data. This result occurred in algorithms produced using one previous segment as well as those produced using two previous segments.

These results suggest that the use of previous time intervals in driver-drowsiness detection algorithms does not improve algorithm accuracy. Additionally, while only the” two previous six-minute intervals were tested in this study, it seems unlikely that using data from further into the past would be of any benefit.

Table 1.8. A comparison of the original results with the results from studies 1A and 2A.

DATA SET	ACCURACY MEASURE	ORIGINAL DATA *(F4a)*	STUDY 1A (n, n-1) *(1A)*	STUDY 2A (n, n-1, n-2) *(2A)*
Development Data	R Value	0.872	0.881	0.883
	APAR (large errors)	0.980	0.983	0.986
	APAR (all errors)	0.790	0.809	0.797
Validation Data	R Value	0.862	0.856	0.841
	APAR (large errors)	0.976	0.982	0.979
	APAR (all errors)	0.830	0.891	0.799

Development Data



Validation Data

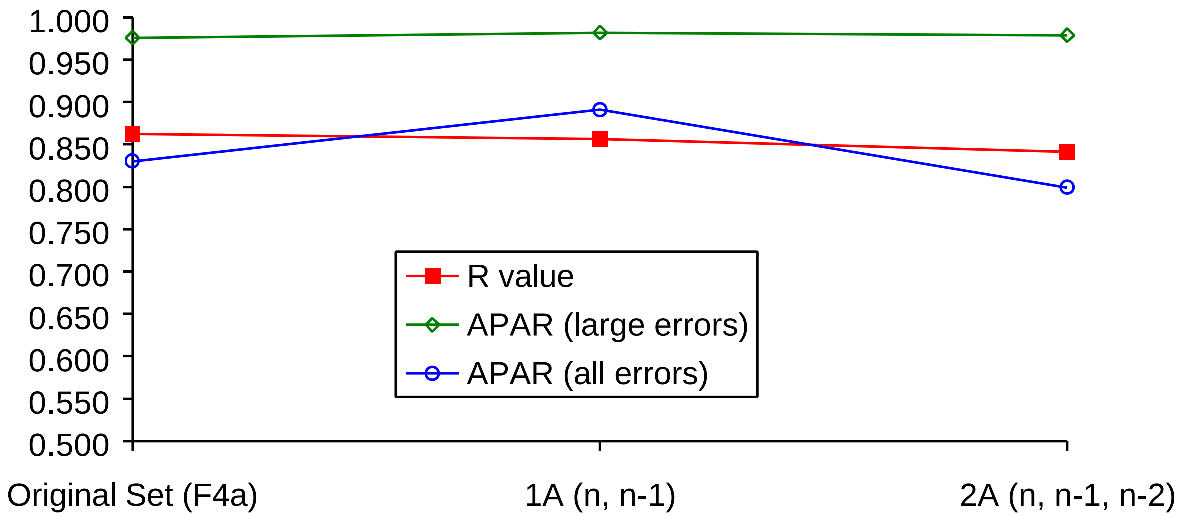


Table 1.9. Regression Summary for Algorithm 1A.
Dependent Variable: aPERCLOS
R=0.88079127, R2=0.77579327, Adjusted R2=0.77018810
F(7,280)=138.41 p<0.0000 Std.Error of estimate: 0.04746

	BETA	St. Err. of BETA	B	St. Err. of B	t(280)	p-level
Intercept			-0.007852	0.004	-1.824	0.069
aACCDEV	-0.117	0.034	-0.019121	0.006	-3.386	0.000
aLANDEV	0.909	0.072	0.068383	0.005	12.580	0.000
aLNERRS	-0.238	0.055	-0.001811	0.000	-4.300	0.000
aNMRHOL	-0.187	0.046	-0.003199	0.000	-4.033	0.000
aTHRSHL	0.240	0.041	0.221817	0.038	5.913	0.000
bLNMNSQ	-0.121	0.035	-0.000677	0.000	-3.442	0.000
bSTVELV	0.190	0.038	0.000619	0.000	5.006	0.000

Table 1.10. Regression Summary for Algorithm 2A.
Dependent Variable: aPERCLOS
R=0.88241777, R2=0.77866111, Adjusted R2=0.77202924
F(8,267)=117.41 p<0.0000 Std.Error of estimate: 0.04778

	BETA	St. Err. of BETA	B	St. Err. of B	t(267)	p-level
Intercept			-0.003214	0.004	-0.739	0.460
aLANDEV	0.714	0.052	0.054267	0.004	13.689	0.000
aNMRHOL	-0.147	0.052	-0.002517	0.000	-2.826	0.005
aTHRSHL	0.376	0.052	0.345770	0.048	7.202	0.000
bLNMNSQ	-0.124	0.037	-0.000687	0.000	-3.395	0.000
bSTVELV	0.250	0.038	0.000812	0.000	6.536	0.000
cLNMNSQ	0.806	0.188	0.004574	0.001	4.284	0.000
cLANVAR	-0.892	0.192	-0.006602	0.001	-4.653	0.000
cSMREV	0.137	0.057	0.001056	0.000	2.426	0.016

Measures preceded by “a” represent six-minute time segment n,
Measures preceded by “b” represent six-minute time segment n-1,
Measures preceded by “c” represent six-minute time segment n-2.

STUDY 4.

THE ELIMINATION OF OUTLIERS FROM PERCLOS DATA

Introduction

Multiple regression, which has been used to develop most of the driver drowsiness detection algorithms in this ongoing project, is a least-squares optimization technique. Consequently, dependent variable outliers can have a confounding effect on regression results. PERCLOS, which has been selected as a typical dependent measure and subsequently used in many algorithm optimization attempts, does contain some apparent outliers among the development and validation data.

The dependent measure PERCLOS can be clipped to a feasible range in an attempt to improve the regression output accuracy and thus use less of the optimization resource to absorb its effect. In this study, we investigate whether “clipping” PERCLOS to an operational range will reduce the effect of outliers.

Method

A PERCLOS value greater than 0.15 is classified as “drowsy.” In the examination of PERCLOS data, values greater than 0.25 appear to be outliers. The value 0.25 was selected as the maximum value, and all PERCLOS values greater than 0.25 were changed to the value of 0.25, that is, “clipped.”

The original set of independent measures from algorithm F4a were selected, and using “clipped” PERCLOS data as the dependent measure, multiple regression was performed in the same manner as previously reported. A second algorithm was developed using PERCLOS (unclipped) to allow for comparison. Both algorithms were applied to validation data. R values and APAR values from classification matrices were calculated.

Results

A summary of all results appears in Table 1.11. Multiple regression results are presented in Table 1.12. R values range from 0.817 to 0.868, and APAR values range from 0.787 to 0.983.

A reduction in R value occurred for the clipped PERCLOS data when applied to both the development and validation data sets. These results indicate that no improvement in algorithm accuracy is obtained from the clipping of PERCLOS prior to regression.

Table 1.11. A comparison of the regression and validation results using “clipped” PERCLOS and PERCLOS.

TRIAL	DATA SET	ACCURACY MEASURE	VALUE
Trial #1 Dependent Measure: “clipped” PERCLOS	DEV	R Value	0.864
		APAR (large errors)	0.983
		APAR (all errors)	0.800
	VAL	R Value	0.817
		APAR (large errors)	0.986
		APAR (all errors)	0.830
Trial #2 Dependent Measure: PERCLOS	DEV	R Value	0.868
		APAR (large errors)	0.983
		APAR (all errors)	0.787
	VAL	R Value	0.856
		APAR (large errors)	0.983
		APAR (all errors)	0.823

Table 1.12. Regression Summary for Algorithm 3C.
Dependent Variable: CLIPPED PERCLOS
R=0.86433999, R2=0.74708361, Adjusted R2=0.74278231
F(5,294)=173.69 p<0.0000 Std.Error of estimate: 0.03987

	BETA	St. Err. of BETA	B	St. Err. of B	t(280)	p-level
Intercpt			-0.002357	0.003	0.708	0.480
INTACDE	-0.125	0.031	-0.063314	0.016	-4.029	0.000
LANDEV	1.017	0.063	0.061510	0.004	16.085	0.000
LNERRSQ	-0.447	0.055	-0.002746	0.000	-8.160	0.000
NMRHOLD	-0.236	0.047	-0.003269	0.000	-5.069	0.000
THRSHLD	0.273	0.042	0.202633	0.031	6.518	0.000

Table 1.13. Regression Summary for Algorithm 3P.
Dependent Variable: PERCLOS
R=0.86810831, R2=0.75361204, Adjusted R2=0.74942176
F(5,294)=179.85 p<0.0000 Std.Error of estimate: 0.04901

	BETA	St. Err. of BETA	B	St. Err. of B	t(267)	p-level
Intercept			-0.003090	0.004	-0.755	0.451
INTACDE	-0.107	0.031	-0.067982	0.019	-3.519	0.000
LANDEV	0.910	0.062	0.068508	0.005	14.572	0.000
LNERRSQ	-0.250	0.054	-0.001911	0.000	-4.619	0.000
NMRHOLD	-0.205	0.046	-0.003536	0.000	-4.459	0.000
THRSHLD	0.242	0.041	0.223915	0.038	5.859	0.000

REFERENCES FOR PART I

Wierwille, W.W., Wreggit, S.S., Kirn, C.L., Ellsworth, L.E., and Fairbanks, R.J. (1994). Research on Vehicle-Based Driver Status/Performance Monitoring: Three Year Report. Vehicle Analysis and Simulation Laboratory, Virginia Polytechnic Institute and State University, VPISU Report No. ISE 94-04, NHTSA Report No. DOT HS 808 247, December, 1994.

Wreggit, S. S., Kim, C. L., and Wierwille W. W. (1993). Research on Vehicle-Based Driver Status/Performance Monitoring. (ISE Department Report Number 93-06). Vehicle Analysis and Simulation Laboratory, Virginia Polytechnic Institute and State University, Blacksburg, Virginia (October).

Wreggit, S. S., Kirn,C. L., and Wierwille W. W. (1994). Research on Vehicle-Based Driver Status/Performance Monitoring. (ISE Department Report Number 93-06). Vehicle Analysis and Simulation Laboratory, Virginia Polytechnic Institute and State University, Blacksburg, Virginia (April).

Ellsworth, L. A., Wreggit, S. S., and Wierwille, W. W. (1993). Research on Vehicle-Based Driver Status/ Performance Monitoring. (ISE Department Report Number 93-02). Vehicle Analysis and Simulation Laboratory, Virginia Polytechnic Institute and State University, Blacksburg, Virginia (April).

PART II

ADVISORY AND ALARM STIMULI OPTIMIZATION FOR A DROWSY DRIVER DETECTION SYSTEM

by

Sarah E. Fahey

and

Walter W. Wierwille

INTRODUCTION

Driver drowsiness is a major concern of transportation researchers today, and with good reason. Drowsy drivers are a major cause of motor vehicle accidents. Today's automobile designs concentrate on comfort and luxury. Reduced road noise and vibration are viewed as quality. However, these qualities in a vehicle can add to the problem of driver drowsiness. Operating a motor vehicle requires constant monitoring of the surrounding environment. When a driver becomes drowsy it becomes increasingly difficult to be attentive to the driving task.

There have been several reports on the severity of the drowsy driver problem. Over the course of one year 50% of all fatalities on the Ohio Turnpike were attributed to accidents involving drowsiness (Kearney, 1966). In a 1980 survey reported by Seko (1984), 75% of all drivers surveyed reported having experienced drowsiness while driving. Planque, Chaput, Petit, Tarriere, and Chabanon (1991) reported 26% of fatal accidents on the motorways in France were caused by drowsy drivers.

Data taken over the five year period of 1989-93 was recently summarized in a NHTSA Research Note (Knipling and Wang, 1994). These national statistics were taken from the NHTSA General Estimates System (GES) and the Fatal Accident Reporting System (FARS) data sources. A summary of the findings is as follows:

- There were 56,000 crashes annually in which driver fatigue/drowsiness involvement was cited in the police report (0.9 % of all crashes).
- There was an annual average of 40,000 non-fatal injuries due to drowsy drivers.
- There was an annual average of 1,357 fatal accidents involving a drowsy driver (3.6% of ail fatalities during the five year period),

- Actual involvement of driver fatigue/drowsiness is most likely much greater due to under reporting.

A 1973 study conducted at Duke University found that a significant number of drivers experienced drowsiness while driving and that drowsiness was a significant cause of automobile accidents (Tiley, Erwin, and Gianturco, 1973). This study found that 64% of 1500 drivers who were questioned were in agreement with the statement “I become drowsy while driving”. This study also found that 31.2% of those who had become drowsy while driving believed their drowsiness had occurred before they were aware of it. Also, 93.7% reported the drowsiness process was reversible; that is they felt they could take actions, while continuing to drive, to reduce their level of drowsiness (Tiley et al., 1973).

A study conducted by Itoi, Cilveti, Voth, Dantz, Hyde, Gupta, and Dement (1993) examined people’s ability to predict sleep onset. Subjects were asked to predict the likelihood (0% likelihood to 100% likelihood) of sleep for 30 consecutive 2-minute intervals. The results showed that subjects predicted significantly higher likelihood of sleep prior to intervals in which sleep occurred (78%) than prior to intervals in which no sleep occurred (42%). Results also showed that subjects predicted much lower sleep likelihoods before their first sleep event than before subsequent sleep events. In the Itoi et al.(1993) study, subjects were also asked to report any physiological indicators of sleepiness. It was found that those subjects who ignored the frequency of physiological indicators, as well as those whose physiological indicators were not regarded as strong indicators of sleep onset, were poor predictors.

Several points can be drawn from the findings of the two aforementioned studies. First, people often do not recognize they are drowsy until after they begin to exhibit symptoms. Second, some individuals may not exhibit any strong symptoms or indicators of sleep onset. Third, people’s ability to predict the likelihood of sleep varies widely.

Finally, the process of becoming drowsy is felt tobe reversible suggesting that countermeasures could be quite effective.

In light of the fact that drivers usually don't realize they are drowsy until after they have displayed symptoms, methods of operationally defining drowsiness have been investigated. Recent studies conducted at Virginia Polytechnic Institute and State University, funded by The National Highway Traffic and Safety Administration, have investigated both physiological and performance measures. Many physiological measures have been shown to be good predictors of driver drowsiness. However, problems arise when trying to implement these measures because they may be intrusive to the driving task or annoying to the driver. For these reasons performance measures have been the focus of more recent research. Drowsiness detection algorithms utilizing performance measures have been developed, refined, and validated (Wierwille, Wreggit, Kim, Ellsworth, and Fairbanks, 1994). The focus of research must now turn to the events which must take place after a drowsiness condition is detected. Some means of advising and alerting a driver and maintaining that m-alerted state are needed.

The goal of this research was to develop an appropriate warning system to produce a fast acting drowsy driver detection and alerting system for use with the detection algorithms described in Wierwille et al. (1994). The purpose of the warning signals is three-fold: first, to advise the driver that a drowsiness condition has been detected; second, to re-alert the drowsy driver; and third, to maintain his or her alertness for a period of time, making it possible to find a safe place to pull off the road and refresh. Because of the many factors involved in this study, a heuristic type evaluation process appeared to be the most efficient way to gather information necessary to design an optimal warning system.

LITERATURE REVIEW

Measures of Drowsiness

Several researchers have attempted to define drowsiness over the years. Torsvall and Akerstedt (1988) defined drowsiness as the "state during which sleep is perceived as difficult to resist, the individual struggles against sleep, performance lapses occur, and sleep eventually ensues." Much research has been conducted to operationally define drowsiness to aid in the detection of drowsy drivers.

Several physiological measures, performance measures, and subjective ratings have been shown to be good predictors of drowsiness. Much of this research has been conducted here at Virginia Tech and details of these studies can be found in previous technical reports. For the purpose of this report, the literature will focus on existing drowsiness detection systems, drowsiness countermeasures, and appropriate warning and alarming stimuli.

Countermeasures to Drowsiness

Because of the scope of the drowsy driver problem, many measures have been researched in hopes of developing a system that would detect a decrease in alertness before the driver is aware of his or her condition. Much of this research has shown promising results. However, problems still remain. Research is necessary to determine how to notify the driver of his or her condition and also how to counteract the drowsiness for a period of time so that the driver can reach a safe stopping area.

Simple countermeasures. Haworth and Heffeman (1989) showed that inadequate sleep or rest, prolonged hours of driving, and the intake of certain foods and drugs contribute to the onset of fatigue. Simple avoidance of these factors could be one way of counteracting drowsiness. However, this is not always possible.

Haworth and Vulcan (1990) classified countermeasures into three types. These are driver-oriented, vehicle-based, and environmental countermeasures. The driver-oriented countermeasures consist of educating drivers to recognize the signs and dangers of drowsiness while driving. Two ways to do this, suggested by Haworth et al. (1990), are driver training courses and mass media campaigns. Another fairly simple way to reduce the risk of drowsy driving is to avoid driving for extended periods of time after a long interval of work. However, this scenario is one that is not easily avoided in today's society. Many people commute an hour or more to and from work everyday. So, although this may seem like a simple way to reduce drowsy driving it is not very practical.

Vehicle-based countermeasures consist of listening to the radio and providing ventilation, possibly by opening the window. Mackie and O'Halon (1977) provide evidence of cool air having a preventive effect. However, there is no evidence of cool air alleviating already existing fatigue. Fatigue monitors are another type of vehicle-based countermeasure. Examples of these will be discussed in some detail later.

Environmental countermeasures include rest breaks and pavement treatments. There are conflicting results in research investigating the effectiveness of rest breaks. Harris and Mackie (1972) reported that rest breaks may have a refreshing effect before fatigue has developed but once fatigue has set in they have little beneficial value. Harris et al. (1972) also concluded that rest breaks of less than 20 minutes produced no recovery effect in driver performance while Lisper, Eriksson, Fagerstrom, and Lindholm (1979) found no difference in the effects of 15 and 60 minute rest breaks. Pavement treatments, specifically rumble strips, have been shown to be effective countermeasures to drowsiness (Haworth et al., 1990; Wood, 1994). Pavement treatments will also be discussed in some detail later.

Harris (1967) mentions several other simple methods of counteracting drowsiness. These include chewing gum, singing along with the radio, sitting on something hard, and taking off the right shoe. These methods may work for some, but unfortunately, they do not work consistently or for all people.

Chemical Countermeasures. Caffeine is considered by many as a favorable way to keep themselves awake while driving for extended periods of time. Haworth et al. (1990) cites many studies which examined the effects of caffeine on driving. Reaction times for a subsidiary task while driving were significantly shorter for a period of 1.5 hours for those drivers that had been given caffeine before a three hour drive (Lisper, Tornros, and Van Loon, 1981). Baker and Theologus (1972) found caffeine significantly reduced attention . lapses. Childs (1978) suggests the normal daily intake of caffeine influences the overall effect it will have on reaction times. It was found that subjects who did not normally consume three cups of coffee a day exhibited increased reaction times in short term visual scanning when given 400 mg of caffeine. In contrast, those subjects who commonly drank larger amounts exhibited a decreased reaction time to a visual stimulus. Childs (1978) concludes that a large dose of caffeine may disrupt visual attention in those who do not normally consume much caffeine, while enhancing visual attention in those who normally consume a large amount of caffeine.

Haworth et al. (1990) cites many studies that have shown stimulants, such as amphetamines, to improve tracking, concentration, and attention. The effect of these drugs is shown to be greater when the subject is sleep deprived. This suggests that they could be used to counteract fatigue. However, stimulants have also been found to increase risk taking behaviors. Because of this, they would probably not prove to be good countermeasures to use while performing a vigilance task such as driving (Haworth et al., 1990).

Nicotine has been supported in the literature as having a positive effect on the performance of vigilance tasks (Wesnes and Warburton, 1978). However, when considering any of the mentioned stimulants one must realize that they are drugs and might have serious health effects. For this reason, and because of individual differences in peoples' reactions to these drugs, they probably should not be considered viable sources of drowsiness countermeasures. A recently identified approach to combating drowsiness is the use of stimulating scents (Kaneda, Iizuka, Ueno, Hiramatsu, Taguchi, Tsukino, 1994). This research will be described further under the section entitled ***Recent Research.***

Mechanical and Electronic Countermeasures. There are several devices that have been developed as attempts to measure and in some cases counteract drowsiness in drivers. These devices include eyelid closure monitors, head nodding monitors, and reaction time monitors.

One device which monitors eyelid closures is the Onguard developed by an Israeli company, Xanadu Ltd.. This battery operated device, which consists of a small infrared sensor and an electronic processor, can be fitted to a standard pair of eyeglasses. The device directs a beam of infrared light at the eye and measures the light reflected back. When the eye is closed the amount of light reflected back is reduced. When the eye is closed for longer than 0.5 second an alarm is activated. Although this device has been shown to detect long eyelid closures, one flaw should be noted. Haworth et al. (1990) found that eyeglasses have a tendency to slip down the nose and as a result the alignment of the device may be in need of constant readjustment.

Head nod detectors have also been marketed to detect drowsiness in drivers. The Electronic Transistor Safety Alarm and Dozer’s Alarm are two examples of this type of detection device. These units, each of which consist of an angular rotational detector, are placed over the top of the ear and buzz loudly when the head nods forward past a certain angle. Dozer’s Alarm was studied by Haworth et al. (1990) and the results were less than satisfactory. Subjects found the device “Very Annoying.” Thorpy and Ledereich (1990) suggest that using a head nod detector gives drivers a false sense of security. Another problem, which Hulbert (1972) points out, is the possibility that an individual may become extremely drowsy and even fully asleep before his or her head falls forward.

Finally, there are reaction time monitors. Roadguard is one such device. Once installed, the device is activated when a car is put into high gear. A timer stops at random periods of 4-14 seconds and a small red light is illuminated on the dash board. This light must be deactivated by the driver within three seconds or an alarm will sound. Haworth et al. (1990) found this device to be the most reliable of the three types studied: Onguard, Dozer's Alarm, and Roadguard. However, there still exist limitations with the device. Lisper, Laurell, and Van Loon (1986) found that drivers could fall asleep behind the wheel for periods of less than two seconds. This suggests that the three second threshold of the Roadguard device may allow drivers to fall asleep and be involved in an accident before the alarm is activated (Haworth et al., 1990). Although the preceding devices have been shown to detect drowsiness in some cases, none has been shown to alleviate drowsiness for any significant length of time or lesson the decrease in driver performance (Haworth et al., 1990).

The Alert-0-Matic, developed by Frederik in 1966, is also a reaction time monitor (Hulbert, 1972). This device consists of a series of three alarms of increasing severity. A light is illuminated and the driver is able to turn it off by lightly tapping the horn. If the driver fails to do so within five seconds the device activates the car horn. If the driver still fails to tap the horn after three seconds of the horn sounding the device turns the ignition on and off for a period of five seconds. If there is still no response from the driver after an additional five seconds the ignition is shut off completely. Again the findings of Lisper et al. (1986) stated above suggest that a driver could fall asleep and crash before these alarms are activated. An additional problem with the Alert-0-Matic is that the light on the dashboard is illuminated at constant intervals of 60 seconds. Oswald (1962), cited in Harris (1967), noted that subjects exhibited a sleeping and waking pattern in which they woke up only to make a necessary response and then fell right back asleep. It seems likely that a driver could quickly learn the pattern of responding to the light every 60 seconds without being fully alerted (Harris, 1967).

The ALERTMASTER and Button Steering Wheel Alarm are similar devices. The ALERTMASTER requires the driver to apply constant pressure on a pedal located to the left of the clutch. When constant pressure is not maintained the horn sounds. The effectiveness of this device relies on the assumption that when a driver becomes drowsy the left foot will relax and not maintain the pressure needed (Hulbert, 1972).

The Button Steering Wheel Alarm operates on the same principle. In this device the button is located on the steering wheel. As with the ALERTMASTER, an alarm is activated when constant pressure is not maintained on the button. Fatigue may be a result of the constant pressure required by the drivers' finger or thumb. Also, steering maneuvers and control adjustments on the instrument panel may warrant removal of the finger or thumb from the steering wheel for short amounts of time (Hulbert, 1972).

Most of the drowsiness detection devices that have been discussed here are quite intrusive and distracting to the driver. All are driver based or require the driver to actively participate in the detection process in some way. These devices, while trying to combat the problem of drowsy drivers, may very well be causing another. Distraction alone is an important cause of vehicle accidents. Therefore, a drowsy driver detection system should be as non-intrusive as possible. A large part of the success or failure of a detection device lies in whether or not those who could benefit from it actually use it. Therefore, public acceptance is an important issue. Because of this, much of the recent research has focused on vehicle-based drowsy driver detection, which involves ways of detecting drowsiness without attaching any equipment to the driver or requiring the driver to do anything that is not part of the normal driving task.

Vehicle-Based Drowsy Driver Detection

Recent Research. As stated previously, many physiological measures have been shown to be very good predictors of drowsiness. The correlation of these measures with other non contact definitional measures has been investigated.

Kaneda et al. (1994) developed a detection system which uses image processing of the driver's face to detect diminished alertness. The image of the driver's face is processed to locate the eyes and then determine the degree to which they are open. This information was shown to have a moderately high correlation (0.77) with an alertness index which was based on a combination of brain wave measurements, blink ing rates, and facial expressions rated by observers.

Artaud, Planque, Lavergne, Cara, Lepine, Tarriere, and Gueguen (1994) also looked at facial image processing as a drowsiness indicator. Attempts to relate facial images to driving behavior, such as steering and respiratory signals, are underway.

Research at Virginia Tech. Many studies involving drowsiness detection have been conducted in the Vehicle Analysis and Simulation Laboratory at Virginia Tech. There are several physiological measures that have been proven time and time again to be good predictors of driver drowsiness (Erwin, 1976; Erwin, Hartwell, Volow, and Alberti, 1976; Wierwille and Muto, 1981; Hauri, 1982; Planque et al., 1991; Wierwille et al., 1994). However, there are problems with recording these measures because the necessary instrumentation is either intrusive upon the driving task or annoying to the driver. For these reasons, performance measures as predictors have been the topic of numerous research projects at Virginia Tech. Over several years algorithms for the detection of drowsy drivers have been developed, refined, and validated (Dingus, Hardee, and Wierwille,1985; Wierwille et al., 1994). These algorithms are ready to be implemented into a drowsy driver detection system. Now the focus of research must move to the sequence of events which will occur after drowsiness has been detected in order to re-alert the driver.

Usability

Usability is a very important consideration when designing systems for use by the general public . A user must be able to use the system correctly and with few to no problems. Usability tests can be conducted at all stages of development of a system. They are used to identify and correct problems throughout the iterative design process (Nielsen, 1993). There are several different ways the usability of a system can be tested. A brief overview of usability testing is presented below.

A common practice in usability testing is the use of verbal protocol. When using verbal protocol, a subject is asked to “think aloud” as he or she uses the system in question. This allows the experimenter to follow the subject’s train of thought and possibly obtain more useful information than if the subject were not speaking. There are two types of verbal protocol, concurrent and retrospective. During concurrent verbal protocol the subject verbalizes his or her thoughts while performing the tasks at hand. Retrospective verbal protocol requires the subject to verbalize his or her thoughts after the task has been completed, possibly while watching his or her performance of the task on video. In a recent study concurrent protocol was shown to facilitate task performance in software evaluation (Wright and Converse, 1992). Another study found subjects in the retrospective verbal protocol condition produced more valuable statements than those subjects in the concurrent verbal protocol condition (Ohnemus and Biers, 1993).

Three methods of usability testing were compared in a study conducted by Virzi, Sorce, and Herbert (1993). The three types were:

- Heuristic Evaluation - experts critique the user interface.
- Think-Aloud Evaluation - naive subjects comment on their thought processes as they use the system.
- Performance Test - error rates and task completion times are recorded as the subject interacts with the system.

The results of this study showed that the heuristic evaluation uncovered 81% of the problems present in a prototype of a voice mail system, while the performance test uncovered 46% and the think-aloud evaluation uncovered 69%. The heuristic evaluation also was found to be less expensive and less time consuming.

Heuristic evaluation is defined by Nielsen (1993) as a “systematic inspection of a user interface design for usability". The compliance of an interface with recognized usability principles is judged b y evaluators. These principles, or heuristics, include using simple dialogue, minimizing user memory load, maintaining consistency, and having proper feedback. Subjects in the present study were graduate students in Human Factors Engineering. This is to say that they possessed a more detailed knowledge of warning systems and human machine interfaces, such as that of a drowsy driver detection system, than the average person. In a sense they can be considered "experts".

It is possible that one evaluator could uncover the major problems in a design. However, it has been reported that only 35% of usability problems were discovered by a single evaluator and approximately 87% were discovered when ten evaluators were employed (Nielsen, 1993). Because different evaluators will uncover different problems it is best to have several evaluators and compile their findings (Nielsen, 1993).

A study conducted by Virzi (1992) investigated the appropriate number of subjects in usability testing. The results revealed that most severe problems are uncovered with the first few subjects and 80% of all usability problems are uncovered with 4 or 5 subjects. The results also showed there was little benefit to having many more than 5 subjects since the amount of new information revealed by each subject decreased with each additional one.

Lewis (1994) conducted a study which examined additional considerations concerning sample size for usability testing. Lewis (1994) refutes the results of the Virzi (1992) study which concluded that 80% of the problems associated with a system were uncovered by 4 or 5 subjects. Results showed that the likelihood of problem detection effects the percentage of problems found (Lewis, 1994). In the Virzi (1992) study the likelihood of problem detection was 0.32 to 0.42. However, Lewis (1994) found that with an average likelihood of problem detection of 0.16, a sample size of 10 subjects would be needed to uncover 80% of the problems. This finding suggests that it is important that usability evaluators are aware of the likelihood of problem detection for the particular system they are testing, so that they may more accurately estimate sample size needed.

METHODS OF ALERTING

The goal of a warning is to change a person's behavior. To be effective, warnings must be sensed, received, understood, and heeded (Sanders and McCormick, 1993). Thus, the warning must attract a person's attention, convey the correct message, and suggest the correct action to be taken.

Warning Systems for Motor Vehicles

When designing warning systems for use in motor vehicles, as opposed to aircraft or other systems, there are specific aspects which need to be examined carefully. When designing crash avoidance warnings for vehicles, the time available to react must be considered, and the alarm must be designed so that it conveys the appropriate level of urgency. Other aspects that should to be considered include variations in vehicles and drivers. The warning system must be compatible with many different types of vehicles. Another variation lies in the potential users of these warning systems. People of all ages drive. They possess different physical, sensory, and cognitive abilities. Some drive alone and some drive with passengers. These differences will impact the effectiveness of an alarm, and should be considered when designing a warning system.

General warning device design. Preliminary human factors guidelines for crash avoidance warnings have been developed (Lerner, Kotwal, Lyons, and Gardner-Bonneau, 1993). These guidelines suggest a general picture of an ideal warning system, and a summary of these guidelines is presented below.

An effective warning signal must be intrusive and convey a sense of urgency. However, warning systems in vehicles must not be so intrusive and urgent that they startle the driver, and possibly put him or her in more danger, or annoy the driver to the point that he or she will deactivate the system. At the same time, the warning must not be so conservative that it fails to result in the desired effect of alerting a driver of an approaching danger.

Multiple levels of warnings should be used to alleviate the activation of urgent false alarms. The most urgent alarms should be saved for imminent crash warning signals, with less urgent and less disturbing signals signifying earlier stages. The urgent alarms used for imminent crash warnings should also be unique. A certain frequency or pitch should be reserved for use only in a case of immediate danger.

Alarms signifying immediate danger should be presented through two sensory modalities to improve their likelihood of being received. Because of differences in drivers' perceptions and differences in driving environments, dual modality alarms are likely to be more effective. For example, an auditory tone combined with a haptic stimulus is likely to be more effective over a broad range of drivers than just a tone.

Warning systems should ordinarily be activated automatically every time the vehicle is started. However, in the case of motor vehicles, it is probably necessary for the driver to activate the detection/warning system for legal reasons. Nevertheless, drivers should be able to deactivate the system since it is possible that there will be unforseeable situations when the system is not necessary, and if left engaged, would produce false or nuisance alarms.

Built in diagnostic tests should be implemented with the detection/warning system. These tests should be engaged every time the system is activated. If the built in diagnostic test detects a system failure the driver should be notified immediately. However, this information should not be displayed in such a way that the driver could mistake the warning for one requiring a more immediate response.

Multiple settings should be available on warning alarms. These settings should be adjustable. When the vehicle is started the current settings should be obviously displayed or the settings should default to a predetermined setting. It is very likely that multiple drivers will drive one vehicle and possibly in different environments. For example, settings may be changed for night time driving or driving in noisy environments, such as driving in a convertible with the top down or with the radio at a high,volume. In situations like these it is important that the driver be aware of the alarm settings and assure that they are in the most effective mode for the specific driver and environment.

Driver alertness warnings. There are many possible ways to alert a driver in a drowsy condition. These include auditory displays such as tones, buzzers, rumble sound, and speech. Other possibilities include vibrations of the steering column or driver's seat. The maintenance of the driver's alertness is also necessary. Possible methods of maintaining a driver's alertness level are seat vibration, supply of fresh air, driver involvement in a secondary task, and the use of a stimulating scent, such as peppermint (Kaneda et al., 1994). Another possible method of maintaining a driver's alertness is the use of a lane-minder. This is a concept developed by Wierwille in 1989 which consists of sensors located in the vehicle which are able to sense the boundaries of the lane. If the vehicle exceeds those boundaries an alarm is activated.

Warning systems specifically designed for alerting drowsy drivers should follow the general guidelines presented above, as well as additional specifications which are presented below.

Auditory displays. Auditory displays are generally preferred for their effectiveness in alerting (Horowitz and Dingus, 1992). However, special precautions must be taken to . avoid startling or distracting the driver. Therefore, onset rates of 10 dB/msec and higher should be avoided. Also, it is suggested that sounds coming from a single area be avoided unless they are consistent with the direction of the hazard (Lerner et al., 1993).

The fundamental frequencies used in acoustic warnings should be in the range of 500-3000 Hz, and frequencies easily masked by the ambient noise of the environment should be avoided (Lerner et al., 1993). Edworthy, Loxely, and Dennis (1991) identified sound characteristics which increase the perceived urgency of a warning signal. These characteristics include a high repetition rate, high intensity, high fundamental frequency, and large frequency oscillations (warbling, for example).

When considering speech warnings there are several areas of concern. Voice displays are not capable of conveying a message as quickly as other modes of alerting. Also, care must be taken so that the warning voice is discernible from that of a passenger or radio messages. For these reasons, speech warnings may be most effectively used as part of a combination of two or more types of warnings.

Tactile displays. Subjects participating in driving simulator studies have shown a greater reduction in their level of alertness than subjects involved in road studies. Simulators, for the most part, have a smoother ride and less road vibration than an actual car. This suggests that vibration may have an alerting effect and for this reason, should be considered a viable warning signal for driver drowsiness. It is suggested that tactile displays, such as vibration, be located in the driver's seat or the steering column. The vibration frequencies of such warnings, according to Lemer et al. (1993), should be in the range of 100-300 Hz. However, these frequencies are much higher than vibration frequencies normally considered to affect operators. Therefore, lower frequencies may be appropriate.

Visual displays. A visual display is suggested to be used as an initial signal in a warning system. However, in a drowsy driver alerting system the use of visual displays is problematic. If a driver is drowsy and inattentive he or she may be less likely to perceive a visual warning in time to react appropriately. Also, most stimuli presented during driving are presented through the sensory channel of vision. Use of a visual display may overload this channel, and possibly distract the driver. If a visual signal is used as part of a warning, it should be presented within 15 degrees of the driver's normal line of sight of the roadway (Lemer et al., 1993).

Termination of warnings. Warnings which are automatically triggered by a specific condition should be presented for at least one second and until the triggering condition no longer exists (Lemer et al., 1993). This time period should give the driver enough time to recognize the purpose of the alarm. Immediately following the termination of the alarm the system should be reactivated in order to detect any quickly reoccurring decrease in alertness. Some warnings may require manual termination by the driver. The mode of termination in these cases should not be too easily accessed. It is possible that the termination of warning signals could become habit, just as in pressing a snooze button on an alarm clock. For this reason, the termination control should require some physical motion on the part of the driver.

<u>Related Research .</u>

Many accidents involving drowsy drivers occur when the driver allows the car to drift off the side of the road. These accidents have been labeled Drift-Off-Road @OR) accidents. A study of police accident reports in the 1980's showed this type of accident was the leading contributor to the total number of accidents on the Pennsylvania Turnpike. Fifty-seven percent of all accidents on the Turnpike in 1986 were classified as DOR accidents (Wood, 1994). Because of the frequency of DOR accidents, alerting drivers who are drifting became an area of interest.

A strip of patterned pavement along the shoulder of a road called a rumble strip has been effective in reducing the number of DOR accidents. If a driver drifts off the road, onto a shoulder having a rumble strip, the tires on the grooved or uneven pavement produce a loud sound and vibration. The California Department of Transportation tested a rumble strip pattern on a monotonous road between Las Vegas and Los Angeles. A 45% reduction in DOR accidents was reported after the installation of the rumble strip (Chaudoin and Nelson, 1985). The Pennsylvania Turnpike began installing their version of rumble strips, SNAPS (Sonic Nap Alert Patterns), in 1987. Data collected between one year and three and one-half years after the installation of the SNAPS showed a 70% reduction in DOR accidents (Wood, 1994).

Although these results indicate that rumble strips are quite effective, there are some negative issues which need to be addressed. The installation of rumble strips in every shoulder of every road would be very time consuming and expensive. Also, there is the problem of only one side of the road being equipped. Because of intentional lane changes (i.e. passing), the inside boundary of the lane is more difficult to equip with rumble strips. However, a simulated rumble strip effect (a combination of vibration and a rumble sound) produced from within the vehicle when a drowsy driver has been detected may prove to be a very effective countermeasure to drowsiness.

This idea has been investigated in recent research. Daimler Benz (1994) has developed an image processing system for lane position recognition. This process takes into consideration a vehicle's initial position, steering angle, and speed to determine the path the vehicle is predicted to follow if no adjustments are made. The Time to Line Crossing (TLC) is then calculated. If this time is one second or less a warning signal is activated.

Three types of warnings were investigated: acoustic, haptic, and corrective haptic. The acoustic warning used was a simulated rumble strip sound emanating from either side of the vehicle depending on which side of the lane the vehicle was close to exceeding. The haptic warning was steering wheel oscillation, and the corrective haptic warning was steering wheel oscillation combined with a corrective pull of the steering wheel. In other words, if the vehicle was moving toward the right side of the lane the steering wheel would be pulled slightly to the left to initiate correction.

The Daimler Benz (1994) study resulted in an overall positive judgment concerning the acoustic warning by subjects. Subjects also found the acoustic warning "enlivening". This suggests an acoustic warning is not only effective at immediately alerting a drowsy driver but it may have a sustained alerting effect as well.

The use of scents as a countermeasure to drowsiness was investigated in a study conducted by Kaneda et al. (1994). The refreshing effect of four scents, lavender, lemon, jasmine, and peppermint, were compared. The results showed that peppermint had the greatest refreshing effect. Kaneda et al. (1994) attributed this finding to the menthol found in peppermint. The stage of alertness at which the scent was introduced had an effect on the lasting refreshing effect of the scent. When the scent was introduced to a subject who was already experiencing a low level of alertness, its effect was brief in comparison to when the scent was introduced immediately following the first signs of a decrease in alertness. When a buzzer was sounded immediately before the release of the scent, the refreshing effect was extended from approximately 3 minutes to approximately 11-16 minutes (Kaneda et al., 1994). Although the Kaneda et al. (1994) results are impressive, there was no statistical analysis of the data, and the experimental method used was not clearly described. Therefore, further investigation into the use of scents as a method of alerting drowsy drivers is needed.

PRESENT STUDY

Research Objectives

This study was directed at evolving the potentially best configuration of advising/alarming stimuli to be used in a drowsy driver detection, advising, and alarming system. The proposed system consists of three stages. This research focused on optimizing the second and third stages of the system. A detailed diagram of the system is shown in Figure 2.1.

Stage-one – Initial detection of reduced alertness level (drowsiness). The first stage involves the use of performance algorithms to detect a decrease in a driver's alertness level. The algorithms used would be those developed and validated by Wierwille et al. (1994). "Step-up/step-down" procedures would be used, thereby allowing the detection system to continue to function during intervals when not all parameters are available. For example, two algorithms could be used, one employing steering and lateral accelerometer measures, and the other employing steering, lateral accelerometer, and lane-related measures. When lane-related measures are available the algorithm using all three types of measures would be used. If lane-related measures were not available for some reason, such as lane markers not being on all sections of road, the detection system would "step-down" to the algorithm that does not utilize the lane related measures and would thereby remain effective.

Once the system is engaged the algorithms would compute an alertness level using six-minute averages updated every minute. The first twelve minutes of driving would be used for establishing a baseline. Thereafter, alertness would be computed and evaluated each minute (using six-minute moving averages). As long as there is no reduction in alertness level of the driver the system would remain in stage one. Once the system detects that the driver's alertness level has fallen below a predetermined threshold, the system will progress to stage two. In the mean time, however, stage one processing would continue.

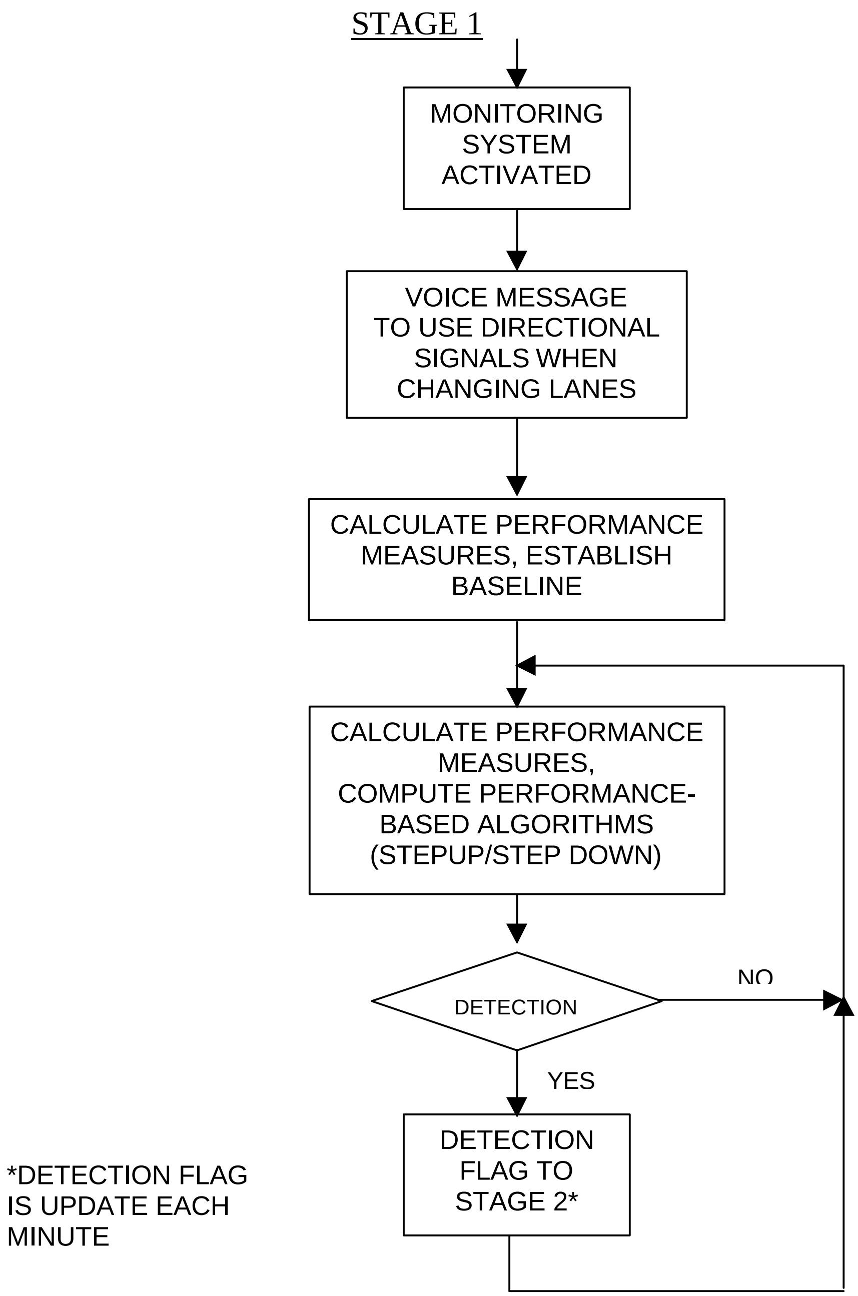


Figure 2.1A: System Flow Diagram for Initial Detection of Drowsiness (Stage 1)

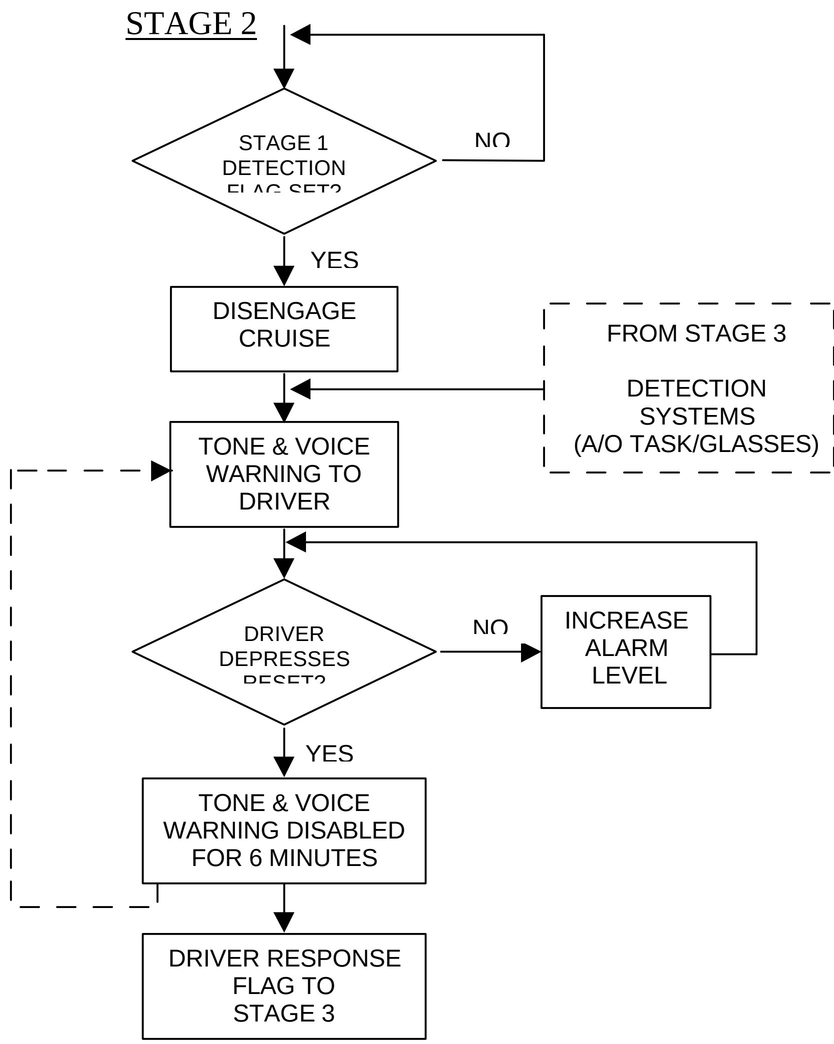


Figure 2.1B: System Flow Diagram for Re-alerting Driver (Stage 2)

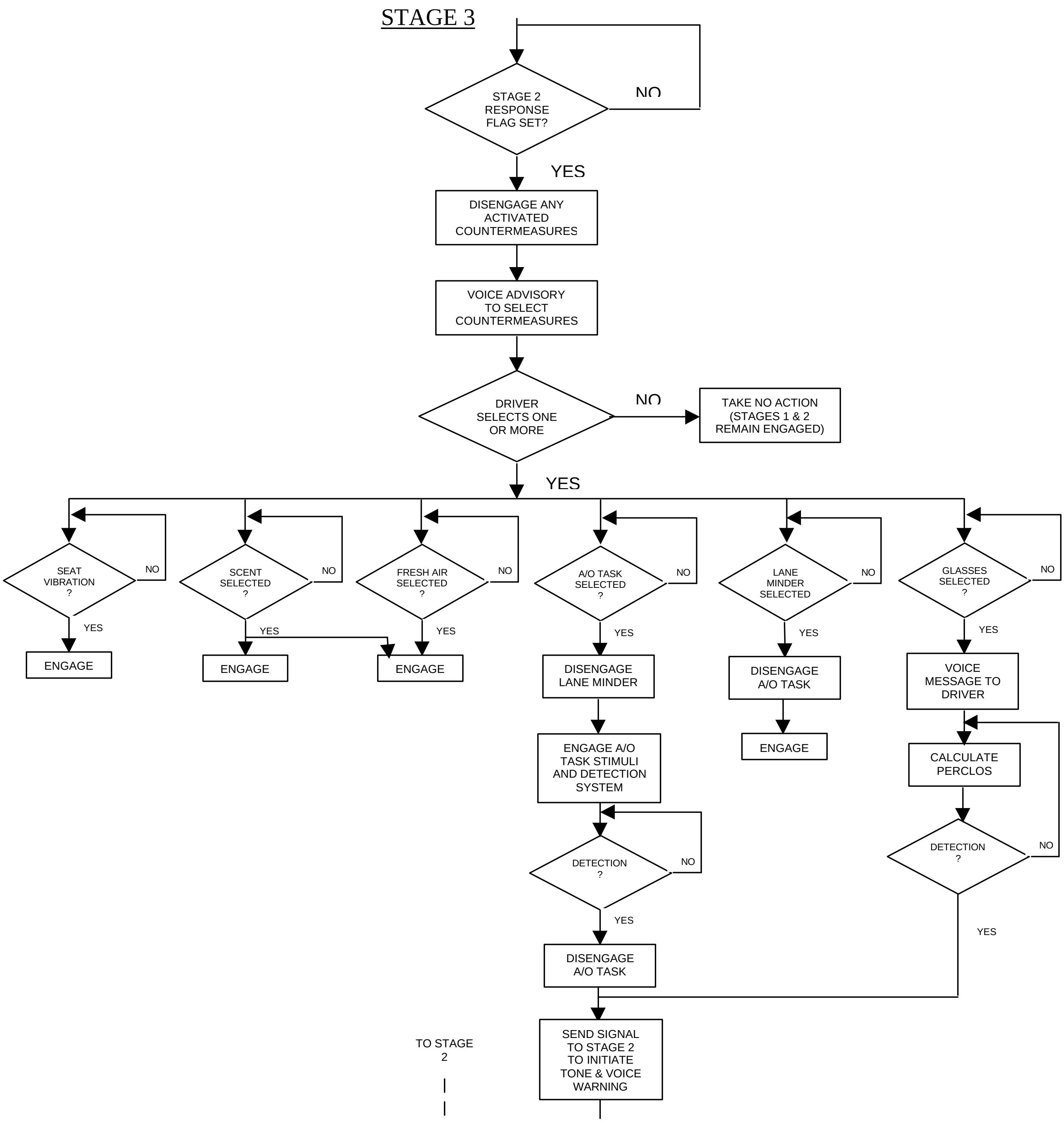


Figure 2.1C: System Flow Diagram for Maintaining Alertness (Stage 3)

Stage two - Re-alerting the driver. The second stage would begin with an auditory stimulus informing the driver that a drowsiness condition has been detected. A full alarm would then be activated unless the driver manually depresses the reset button. The advisory stimulus would consist of an audible tone followed by a voice message. If cruise control is engaged at this point it would be disengaged by the detection system. One of the objectives of the proposed research was to determine the optimal tone and voice message to be used for this advisory stimulus. The option to reset the system would give the driver the opportunity to avoid unnecessary exposure to full alarms. However, the reset button should be placed in an area that would not allow deactivation to occur too easily. When the driver depresses the reset button the initial alert would be disengaged for six minutes. After the reset button is depressed the algorithms will most likely still be detecting a noticeable reduction in alertness level. Therefore, the six-minute delay in the system is intended to avoid an immediate reactivating of the initial alerting tone and voice message after the driver has depressed the reset button.

If drowsiness is detected in the first stage and the driver does not reset the system after the initial advisory tone and voice message (at the beginning of the second stage), a subsequent full alarm would be sounded. There are many possible configurations for this alarm. Therefore, another objective of this research was to determine the optimal alarm to be used at this stage of the advising/alarming system. Characteristics of auditory displays, specifically sounds with a wide variety of wave shapes, frequencies, and modulations, were investigated. The effects of vibration of the steering wheel and the driver's seat, and a simulated braking effect were also investigated.

The alarm would continue until it is manually deactivated by the driver. Again; deactivation would be effected through the depression of the reset button. When the reset button is depressed the system would again delay r-e-activating of the initial advisory tone and voice message at the beginning of stage two for six minutes. Whenever the driver depresses the reset button, stage two would end and stage three would begin.

Stage three - Maintaining alertness. At the beginning of this stage, a voice message would advise the driver to engage an alertness aid to help maintain the driver's re-alerted state. If the driver feels that an alertness aid is needed to help remain alert while looking for a safe rest area, there may be several methods from which to choose. Possibilities would include seat vibration, introduction of a secondary task, activation of a lane-minder, use of eye glasses which measure slow eye closures, introduction of a stimulating scent, and directing fresh air toward the driver's face. The driver would select countermeasures by means of a control panel. Appendix 2A contains a description of its functions.

Seat vibration. When the seat vibration button is depressed momentarily the button would be illuminated and the seat would vibrate. When the button is depressed again the vibration would cease. It is possible to have vibration in the seat back and seat pan. The current study was directed at determining the most effective combination.

Scent. When the scent button is depressed the fresh air blower would be activated and a scent would be discharged into the air in front of the driver. This button would be illuminated along with the fresh air button. The blower would continue to operate for a predetermined length of time to disperse the scent. The subjective effectiveness of the use of a scent was tested in the current study.

Lane-minder. When the lane-minder button is depressed it would be illuminated and the lane-minder would be activated. If the driver allows the vehicle to exceed the lane boundaries an alarm would be activated. The lane-minder would only be able to be activated when the A/O task is deactivated and vice versa. The lane-minder would be able to be deactivated by depressing the lane-minder button again. Lane-minder effectiveness was also tested in the current study.

A/O task. When the A/O task button is depressed the button would be illuminated and the A/O task would begin. This task involves presentation of recorded words presented aurally to the driver. The driver responds using yes/no push buttons mounted on the steering wheel spokes. If the presented word contains the letter "a" or "o", the yes button is to be depressed. Otherwise the no button is to be depressed. During this task the monitoring system would employ the A/O task algorithms developed by Wreggit , Kim, and Wierwille (1993). If a decrease in the driver's alertness level is detected the A/O task would cease and the system will revert back to the beginning of stage two and the initial advisory tone and voice message will be sounded. The subjective effectiveness of the A/O Task was also tested in the current study.

Although the bulk of this research concentrates on the investigation of these aforementioned countermeasures, there are other possible ways to help maintain a driver's alertness level, as described below.

Fresh air. The fresh air button would activate the fresh air blower.

Glasses. An ordinary pair of eyeglass frames could potentially be equipped with a source which produces a beam of infrared light and sensors which detect when the beam of light is interrupted by a slow eyelid closure would be used. The measure of slow eyelid closure, PERCLOS, would be calculated, and if a high degree of slow eyelid closure is detected the system would revert back to the initial advisory tone and voice message at the beginning of the second stage. PERCLOS would continue to be calculated until detection is made or the countermeasure is deactivated by the depression of the glasses button on the driver's control panel.

The purpose of these alertness aids is to maintain the increased alertness level of the driver, achieved in stage two, until the driver can safely pull off the road and refresh.

Method

There are many unanswered questions regarding the interface of such a system. The purpose of this study was to answer as many of these questions as possible. Because of the many variables involved in a study of this kind, subjective usability testing, specifically a type of heuristic evaluation, utilizing rating scales and the method of paired comparisons appeated to be the best approach. As stated in the literature review, heuristic evaluation has been shown to be the most effective in uncovering major problems. Heuristic evaluation is also the least time consuming and least expensive with respect to other types of usability testing (Virzi et al., 1993).

The method of paired comparisons was used to allow the subjects to rate the different alarm stimuli. This method was chosen for several reasons. The subjects were presented with auditory stimuli which are likely to remain in memory for only a short period of time. There are statistical manipulations that may be performed on data obtained through the method of paired comparisons which allow the data to be analyzed as scaled values. Also, it was of some interest as to whether paired comparisons would result in the same rank order of the eight alarm sounds as straight effectiveness ratings.

Because subjects were driving the simulator while they were asked to rate the different stimuli, they gave their responses orally. A copy of the effectiveness scale can be seen in Figure 2.2. The scale was illuminated and located on the lower right side of the simulator's dash throughout the experiment. Subjects were asked to rate each stimulus as "Not Effective", "Slightly Effective", " Moderately Effective", "Very Effective", or "Extremely Effective". They were also permitted to rate stimuli as falling between these descriptors.

Subjects.Eight graduate students in the Industrial and Systems Engineering Department at Virginia Tech were used as subjects. Each had completed at least one semester of graduate work in Human Factors Engineering. Subjects were selected from this group because of their familiarity with behavioral methods.

Apparatus

Simulator. The moving-base automobile simulator at Virginia Tech, validated by Leonard and Wierwille (1975), was used in the study. The simulator handles like a mid-sized rear wheel-drive vehicle. It is computer controlled and is hydraulically powered. It has four degrees of physical motion (roll, yaw, lateral translation, and longitudinal translation). The ambient noise level in the simulator is adjustable. However, during the

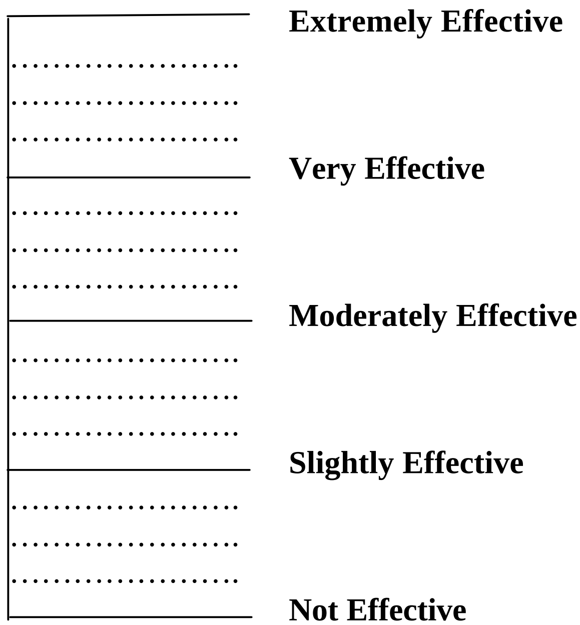


Figure 2.2: The Effectiveness Rating Scale Used in the Experiment.

experiment it was set at approximately 75.5 dBA with the sound level meter 5 inches from the right ear of the driver (73.1 dBA with the sound level meter 3 inches from the right ear of the driver). This level was based on a small road study which determined the ambient noise level of five different vehicles. The average noise level at 65 MPH was found to be 76.2 dBA, while the average at 55MPH was found to be 74.6 dBA (Table 2.1). The instructed speed in the experiment was 60 MPH. Further explanation of this road study can be found in Appendix 2B.

Table 2.1: Summary of Average Sound Pressure Level Measurements

		dBA
Speed	45 MPH	72.53
	55 MPH	74.61
	65MPH	76.22

Initial Advisory Tone. The initial advisory tone was produced from a function generator. The length of this tone was controlled by the experimenter through an interval timing device which allowed the adjustment of the length of the tone from 0.3 to 2.75 seconds.

Voice Message. The voice messages were recorded on a custom designed digital voice recording/playback system. The system is composed of four channels, each capable of recording and playing 10 seconds of audio information. Each of the channels can be played back on command and can be repeated indefinitely without delay.

Alarm Sounds. The audible sounds of the full alarm were produced from various function generators. These alarm sounds were recorded on and played back from a Sound Blaster 16 sound card installed on an IBM 433 DX/S PC.

The audio signals from the initial advisory tone, the voice message system, and the sound card were combined in a stereo audio mixing console. The outputs of the console were passed through dual power amplifiers and applied to small, high quality speaker enclosures containing 4-inch woofers and l-inch tweeters with a frequency response range of 100 Hz - 20 kHz. These speakers were located at mid dash height, approximately 35” to the right and left of the straight-ahead position.

Vibration. Vibration was produced in the driver’s seat and the steering wheel. Eccentrics (unbalanced rotational masses) located in the seat pan and the seat back were driven by high-quality servomotors to produce the vibration of the driver’s seat. The frequency of these vibrations was adjustable. The steering vibration was produced by a periodic signal applied to the active servo system that provides steering feel. A more detailed description of the vibration mechanisms as well as the other peripheral stimuli is presented in Appendix 2C.

Simulated Brake Pulse. A simulated braking effect was achieved by activating a switch at the investigator’s station. When the switch was activated, the simulator momentarily lurched backward to produce the feeling of braking. The speed also decreased somewhat as would be the case in an actual vehicle when the brake pedal is momentarily depressed.

Lane Minder. The simulator was equipped with a lane sensing device which was capable of activating an alarm when the simulated vehicle exceeded the lane boundaries. The alarm consisted of a chirping/beeping sound produced by piezo buzzers.

Scent. The peppermint scent was supplied through an aerosol spray can. A small amount of peppermint oil was placed in a pressurized, reusable (repressurizable) can. Vapors from the can were then be dispensed in front of the driver.

Experimental Design. The study was a within subject design. All subjects were exposed to all conditions. Stimuli were tested in three areas. In the second stage of the advising/alarming system (first phase of this research), the presence of a stimulus is necessary to indicate to the driver that he or she has been detected as exhibiting drowsiness. This stimulus was a tone followed by a voice message. The voice message was “Possible drowsiness has been detected; press reset now”. General features of the initial advisory tone and voice message that were tested were as follows:

Tone:

- Frequency (high, mid-high, mid-low, and low)
- Wave shape (rectangular and sinusoidal)
- Amplitude
- Duration

Voice message:

- Gender
- Amplitude

Eight candidate alarm sounds were also investigated. A detailed description of each alarm sound used is presented in Appendix 2D. These alarms varied in several characteristics including:

- Center frequency
- Wave shape
- Maximum amplitude
- Modulation characteristics

In addition to the alarm sounds, various tactile (haptic) stimuli were investigated. A detailed description of these stimuli is presented in Appendix 2C. They included the following:

- Steering wheel vibration

Seat back vibration

Seat pan vibration

Brake pulse

In the second phase of the study, the effectiveness of four drowsiness countermeasures was investigated. These countermeasures were:

Seat vibration

Lane minder

A/O Task

Scent

(It should be noted that the levels and frequencies of seat vibration were lower for the countermeasures phase than for the alarm phase. See Appendix 2D.)

Procedure. Each subject was instructed to awaken at 7 am the morning of the experiment. He or she was asked to go about normal daily activities with the exception of taking naps. At 6 pm the subject was picked up at home by an experimenter and taken out to dinner. The subject was allowed to eat whatever he or she wanted. However, the subject was not allowed to drink caffeinated or sugared beverages with his or her meal.

After dinner the subject was brought to the Vehicle Analysis and Simulation Laboratory and asked to read and sign an informed consent form. The subject was also asked to read the instructions to the study (Appendix 2E). A simple auditory test was then given, consisting of an experimenter and subject standing 3 feet apart with their backs facing each other. The experimenter spoke five or six words in a normal speaking voice and the subject was asked to repeat the words back to the experimenter. This test ruled out any severe hearing loss. Any questions or concerns the subject might have were addressed at this point. The subject was asked to remain in the lab until the experiment began. The subject could study or watch television.

At 11:45 pm the subject was asked to get into the simulator. The subject was given some final instructions and then asked to drive the simulator for approximately 5 minutes to become familiar with the simulator. After this period the subject was asked to exit the simulator and walk around in order to become acclimated to the simulator. The subject then re-entered the simulator and was asked to drive while the possible advising and alarming stimuli were explained in more detail. Each step of the advising/alarming system was explained to the subject, including where it fit into the envisioned system and the desired effect of each particular stimulus.

The initial advisory tone was played first. The subject was asked to optimize and rate eight different tones. The eight tones consisted of combinations of four different frequencies and two wave shapes. The wave shapes used were a rectangular wave and a sinusoidal wave. The eight tones were counterbalanced across subjects (Appendix 2F). Each subject was asked to set the duration of the tones, based on the first tone. Because of counterbalancing this was a different tone for each subject. Each of the eight tones was played and the subject was asked to adjust the amplitude of each to a preferred level. The subject was permitted to adjust the duration and the amplitudes a second time in a repeat presentation. The tones were then played at the subject's preferred amplitude and duration and the subject was asked to rate their effectiveness in advising a driver of a subsequent alarm.

The voice message was played next. There was a male and a female voice. Four of the subjects received the male voice first, and the other four received the female voice first. The appropriate amplitude for each voice was first chosen by the subject. He or she was then asked to rate the effectiveness with which each voice conveyed the message.

Once the initial advisory tone and voice message were set, the experiment focused on the full alarm. There were eight alarm sounds. The subject was first asked to set the amplitude for each alarm sound. The first alarm sound was played and the subject was asked to adjust the amplitude. The next alarm sound was played and again the subject was asked to adjust the amplitude. The amplitudes for all eight alarm sounds were set in this fashion. This process was repeated a second time to give the subject a chance to readjust the final chosen amplitude for each alarm sound. Once the amplitudes were set the method of paired comparisons was used to determine the preferred alarm sound. There were 28 possible pairs. The alarm sound pairs were presented in a different order for each subject so as to minimize possible order effects. Each alarm sound was played at its preferred amplitude during the comparisons. Once the paired comparisons were completed the study concentrated on possible peripheral stimuli. (Peripheral stimuli are those that could be combined with an alarm sound to produce a more effective alerting or awakening stimulus.)

Peripheral stimuli included steering vibration, seat back vibration, seat pan vibration, and a simulated brake pulse. The subject was exposed to each stimulus individually and asked to rate the effectiveness of the stimulus at re-alerting the driver. Next, the subject was reexposed to the alarm sounds and he or she was asked to rate the effectiveness of these sounds (by themselves) at re-alerting a drowsy driver. Once the subject had been exposed to both the peripheral stimuli and the alarm sounds, the alarm sounds were played in combination with each individual stimulus. The subject was asked to rate the effectiveness of each alarm sound combined with each peripheral stimulus. It should be mentioned that each particular subject was presented with the same order of alarm sounds each time they were played during his or her experimental session. However, the order of presentation was different for each of the 8 subjects. The order in which the peripheral stimuli were produced also differed for each of the eight subjects in order to correct for any systematic effects.

In the next phase of the experiment, the subject was asked to focus attention on possible drowsiness countermeasures. These countermeasures included possible combinations of seat back vibration and seat pan vibration as well as the use of a lane-minder, the introduction of a secondary task (A/O task), and exposure to a peppermint scent. The subject was exposed to each of the five countermeasures individually and then asked to rate the effectiveness of each countermeasure in maintaining the driver's un-alerted state. It should be noted that the seat vibration countermeasure cues used lower amplitudes and frequencies than those used for the peripheral alarm stimuli (Appendix 2D).

Data Analysis Overview

Four types of data were collected during the experiment. These included data obtained from the effectiveness rating scales and data collected from the paired comparisons, as well as volume and duration data. All subjects' responses were recorded manually by the experimenter. All volume data were originally recorded as values from 0 10 on a volume control. Sound level measurements of these values were taken after subject participation and the corresponding d BA levels were used in the analyses of the volume data. A summary of the sound pressure levels for all stimuli used in this study is presented in Appendix 2G.

T-test.T-tests were used to determine if there was a significant difference between the first and final chosen durations of the initial advisory tones. A t-test was also used to test the significance of the volume level and the effectiveness ratings of the two voice messages.

Analysis of variance. Analyses of variance (ANOVAs) were used to test for significant differences with respect to the volumes of the tones and alarms as well as for all of the effectiveness ratings. The effectiveness rating, excluding those of the voice messages, were taken using the scale shown in Figure 2.2. Subjects were permitted to select points in between descriptors if they wished. All variables were within subject, and a Greenhouse-Geisser correction was used when calculating the ANOVAs. Violations of the assumption of normality are present in some of the data groups. However, ANOVAs are known to be robust with regard to violations of these assumptions. A study of the effects of violating assumptions which underlie the analysis of variance was conducted by Glass, Peckham, and Sanders (1972). They found that non-normality skewdness had very little effect on the level of significance of the F-test. All reported p-values ate Greenhouse-Geisser corrected.

Paired Comparisons. The frequency of preference for each of the eight alarm sounds was recorded for each subject (Appendix 2H). The frequency data were converted into a proportion matrix. Maxwell's scaling technique was then used to transform the proportion data into an interval scale (Maxwell, 1974). This was achieved by converting the proportion data into z -deviates of the normal distribution. The mean z -values were then transferred to a scale on which the value of zero corresponds to the lowest mean z deviate. These data were then analyzed using a one-way analysis of variance with alarm being the single factor with eight different levels (Maxwell, 1974).

RESULTS

Initial Advisors Tone

Three aspects of the initial advisory tones were investigated in this study. These included the preferred length and volume of the tone as well as how effectively it attracted the attention of a driver and directed it to a forthcoming voice message. Figure 2.3 graphically depicts the average ratings for each of the eight tones. Table 2.2 shows each tone's wave shape and frequency as well as the average chosen dBA level and effectiveness rating (0 being "Not Effective" and 4 being "Extremely Effective"). The results of the Newman-Kuels post-hoc tests for amplitudes and effectiveness ratings are shown in the columns directly to the right of those values. Significant differences are reported between tones having different letters within these columns.

Duration. A t - test was performed on the two groups of duration data (Subjects' first chosen duration compared to their final chosen duration). No significant differences (a = 0.05) were found between the two groups . The final average tone duration chosen by the subjects was 0.80 second.

Amplitude.A one-way analysis of variance was performed on the preferred sound levels of the initial advisory tones. Significant results were found ($p = 0.0001$). Newman-Kuels post-hoc tests were conducted to identify specific differences. The results show that there were many cases where there were significant differences in the chosen amplitudes of specific advisory tones.

Patines. A one-way analysis of variance was performed on the numerical rating scale data, with a significant result ($p = 0.003$). The results of the Newman-Kuels post-hoc tests show several significant differences between the ratings of the eight different tones.

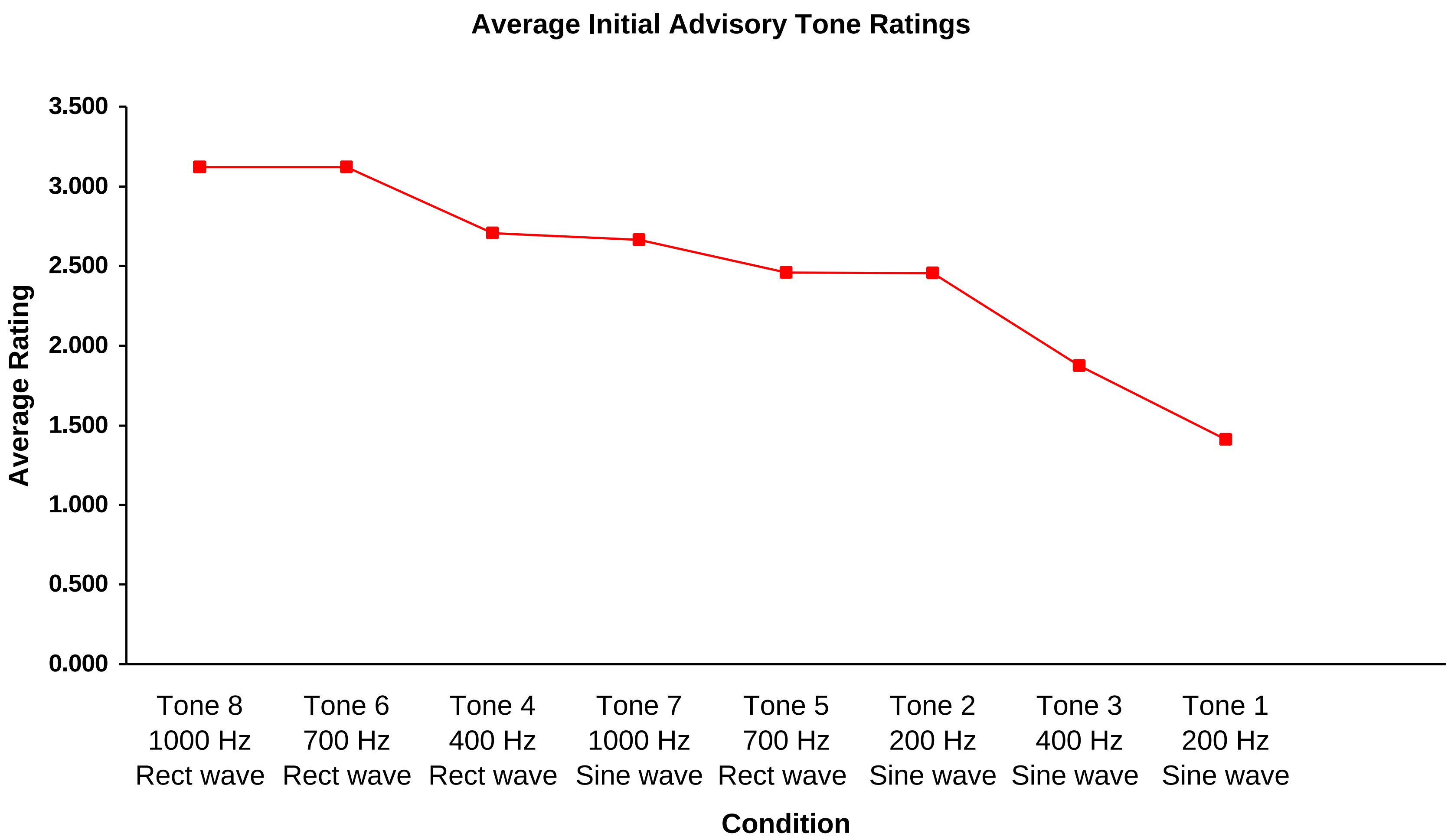


Figure 2.3: Mean Effectiveness Ratings for Initial Advisory Tones

Table 2.2: Summary Table of Initial Advisory Tone Data. Tones with same letter are ***not*** significantly different.

Condition	Wave Shape	Frequency (Hz)	Ave. dBA*	N-K Test Amp.	Avg. Effect. Rating (0 - 4)	N-K Test Ratings
Tone 8	rectangular	1000	65.150	A	3.121	A
Tone 6	rectangular	700	64.512	A	3.120	A
Tone 4	rectangular	400	65.800	A	2.706	B
Tone 7	sinusoidal	1000	71.975	B	2.665	BC
Tone 5	sinusoidal	700	71.700	B	2.460	BCD
Tone 2	rectangular	200	71.088	B	2.455	D
Tone 3	sinusoidal	400	73.288	B	1.874	E
Tone 1	sinusoidal	200	79.750	C	1.414	F

* It should be noted that, although several of these sound levels are below the ambient noise level of the simulated vehicle, they are audible because this energy is concentrated at specific frequencies.

Voice Message

Two types of data were collected on the voice messages. These were preferred amplitude levels and effectiveness ratings. The results of the two analyses are reported below. A table of the average amplitudes and effectiveness ratings is presented in Appendix 21.

Amplitude. A r-test was conducted to determine if there was a significant difference between the preferred volume level of the male and female voice messages. The result was not significant ***(a = 0.05).***

Ratings.A r-test was also conducted on the effectiveness ratings. Again, the result was not significant ***(a = 0.05).***

Alarm Stimuli

Several groups of data were collected in this segment of the experiment. The preferred volume levels were collected for each of the eight alarm sounds. Data on the alarm sounds by themselves were collected through the method of paired comparisons. Effectiveness ratings were collected for the peripheral stimuli by themselves, the eight alarm sounds by themselves, and the alarm sounds coupled with each of the peripheral stimuli. The results of these analyses are presented in Tables 2.3,2.4,2.5,2.6,2.7, and 2.8.

Amplitudes. A one-way analysis of variance was conducted to determine if any significant differences existed between preferred dBA levels of the alarm sounds. Significant results were found ($p = 0.0001$). The average dBA level chosen for each alarm and the results of the Newman-Kuels post-hoc tests are presented in Table 2.7.

Paired comparison. A one-way analysis of variance was conducted on the paired comparison data after it was converted into scaled values. These scaled values along with results of the Newman-Kuels post hoc tests are shown in Table 2.3. The scaled values were checked for additivity by comparing the statistic, X2 = fns2/4 (f = 1/2(k-l)(k-2) degrees of freedom), to the &i-squared distribution. The results of this test gave a value of 29.946 which when compared to the cm-squared value for 21 df is not significant (a = 0.05). This result indicates that the scale possesses additivity and the usual variance ratio test can be used to test for significance. The F-test did produce significant results (a = 0.05) indicating that subjects had significant differences in preference among the eight alarms. The ANOVA summary table is presented in Table 2.4.

Ratings. A one-way analysis of variance was conducted on the ratings of the peripheral stimuli. No significant differences ***(a = 0.05)*** were found. The mean effectiveness ratings for the five peripheral stimuli presented by themselves are shown in Appendix 21.

An 8 x 6 ANOVA was conducted to test the main effects and interactions of alarm type and peripheral stimulus condition (alone, with a brake pulse, with lower seat vibration, with upper seat vibration, with combination seat vibration, and with steering vibration). Table 2.5 shows the summary table for this analysis. Significant results were found for the main effects of Alarm Type and Stimulus Condition as well as the interaction

Table 2.3: Results of the Maxwell Scaling Procedure on the Frequency of Preference Data for the Alarm Sounds. Alarms with the same letter are not significantly different. Preference is from high (top) to low (bottom).

Condition	Short name	Frequency of Preference	Scaled value	N - K Test
Alarm 2	On-Off Tone	50	4.0073	A
Alarm 1	Dual Tone	42	3.2242	AB
Alarm 4	Freq. Swept Tone	38	3.0244	ABC
Alarm 7	Rapid Amp.-Mod. Tone	31	2.4316	BCD
Alarm 8	Gapped Freq-Swept Tone	27	2.0404	CD
Alarm 6	Overmod. Mid-Freq. Tone	22	1.7343	D
Alarm 5	Overmod. Low-Freq. Tone	11	0.7083	E
Alarm 3	Continuous Tone	3	0.0000	E

Table 2.4: ANOVA Summary Table For Scaled Values of Alarm Sound Preferences

ANOVA Summary Table

Source	**df**	**ss**	**MS**	**F**
Between Alarms	7	98.665	14.095	19.769*
Residual	21	14.9679	0.713	
Total	28	113.6329		

*Significant test results (a = 0.05)

between the two. Newman-Kuels post hoc tests were used to isolate significant differences. Only logical comparisons are reported. (Unimportant comparisons such as a comparison between Alarm 3 with a brake pulse and Alarm 6 alone are not reported.)

Table 2.6 shows the break down of the significant main effect of Alarm Type. Significant differences are shown by different letters in the right hand column of the table. Table 2.7 shows the comparisons between each of the alarm sounds by themselves and combined with each of the peripheral stimuli. These comparisons were made to determine if the peripheral stimuli added to the effectiveness of the alarm sounds. As can be seen in the table, there are increases in several cases. The effects can also be seen in Figure 2.4. Comparisons were also made within each stimulus condition. These results are reported in Table 2.8.

Table 2.5: ANOVA Summary Table of Stimulus Condition (SC) x Alarm Type (AT)

ANOVA Summary Table

Source	**df**	**ss**	**MS**	**F-value**	**P-value**	**G-G**
Subject (s)	7	51.852	7.407			
SC	5	26.197	5.239	6.118	0.0004	0.0048
SCxS	35	29.974	0.856			
AT	7	40.325	5.761	5.532	0.0001	0.0074
ATxS	49	51.024	1.041			
SCxAT	35	5.274	0.151	3.415	0.0001	0.0104
SCxATxS	245	10.811	0.044			

Table 2.6: Summary of the Analysis of the Effectiveness Ratings of Alarm Sounds Alone. Alarm sounds with the same letter are not significantly different.

Condition	Short name	Avg. dBA	N-K Test Amp.	Avg. Rating (0 - 4)	N-K Test Ratings
Alarm 2	On-Off Tone	74.575	A	3.462	A
Alarm 1	Dual Tone	83.387	C	3.219	AB
Alarm 4	Freq. Swept Tone	81.000	B	3.191	BC
Alarm 7	Rapid Amp.-Mod. Tone	80.775	B	3.039	CD
Alarm 8	Gapped Freq-Swept Tone	81.413	B	2.913	DE
Alarm 6	Overmod. Mid-Freq. Tone	77.725	A	2.850	EF
Alarm 5	Overmod. Low-Freq. Tone	77.0121	A	2.608	F
Alarm 3	Continuous Tone	88.900	D	2.392	G

Table 2.7: Average Effectiveness Ratings and Results of Post-hoc Comparisons Between Each Alarm Type Alone and that Alarm Type with Different Stimulus Conditions. Conditions with the same letter within one alarm type are not significantly different.

Alarm Type	Stimulus Condition	Avg. Rating (0-4)	N-K Test
1	Alone	3.080	A
	With Brake Pulse	3.539	B
	With Lower Seat Vibration	3.121	A
	With Upper Seat Vibration	3.122	A
	With Combination Seat Vibration	3.330	A
	With Steering Vibration	3.121	A
2	Alone	3.415	A
	With Brake Pulse	3.749	B
	With Lower Seat Vibration	3.331	A
	With Upper Seat Vibration	3.330	A
	With Combination Seat Vibration	3.663	A
	With Steering Vibration	3.287	A
3	Alone	1.537	A
	With Brake Pulse	2.914	B
	With Lower Seat Vibration	2.287	B
	With Upper Seat Vibration	2.414	A
	With Combination Seat Vibration	2.787	B
	With Steering Vibration	2.412	A
4	Alone	2.829	A
	With Brake Pulse	3.497	B
	With Lower Seat Vibration	3.081	A
	With Upper Seat Vibration	3.164	A
	With Combination Seat Vibration	3.414	B
	With Steering Vibration	3.162	A
5	Alone	2.039	A
	With Brake Pulse	3.038	B
	With Lower Seat Vibration	2.537	B
	With Upper Seat Vibration	2.579	B
	With Combination Seat Vibration	2.997	B
	With Steering Vibration	2.456	A
6	Alone	2.205	A
	With Brake Pulse	3.287	B
	With Lower Seat Vibration	2.787	B
	With Upper Seat Vibration	2.829	B
	With Combination Seat Vibration	3.122	B
	With Steering Vibration	2.870	B
7	Alone	2.622	A
	With Brake Pulse	3.374	B
	With Lower Seat Vibration	2.956	A
	With Upper Seat Vibration	2.997	B
	With Combination Seat Vibration	3.329	B
	With Steering Vibration	2.954	A
8	Alone	2.496	A
	With Brake Pulse	3.207	B
	With Lower Seat Vibration	2.871	B
	With Upper Seat Vibration	2.912	B
	With Combination Seat Vibration	3.204	B
	With Steering Vibration	2.789	A

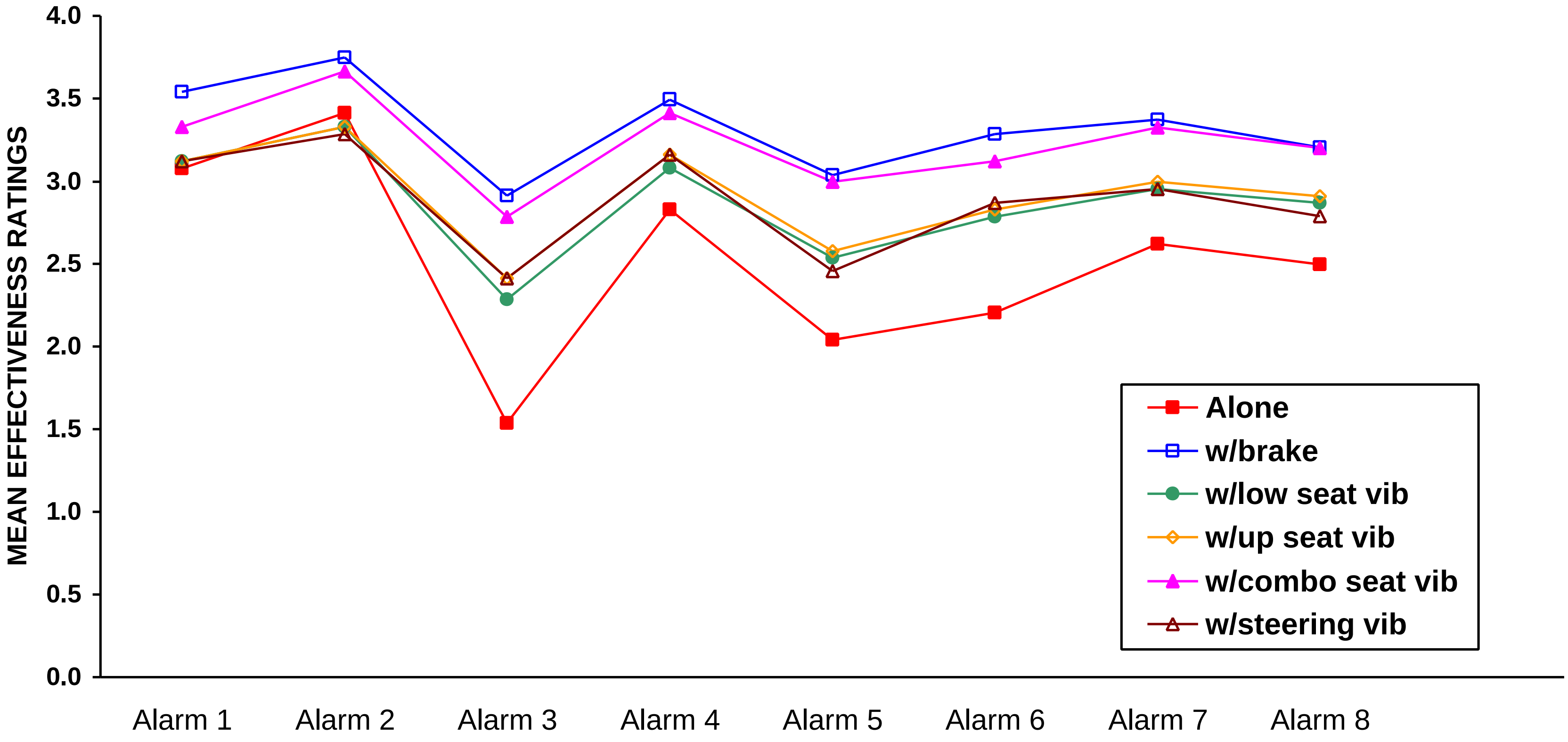


Figure 2.4: Mean Effectiveness Ratings for Alarm Sounds Alone and With Peripheral Stimuli

Table 2.8: Average Effectiveness Ratings and Results of Post-Hoc Comparisons Within Each Peripheral Stimulus Condition.

Alarms with Brake Pulse								
Alarm Type	3	5	8	6	7	4	1	2
Avg. Rating (0 - 4)	2.914	3.038	3.207	3.287	3.374	3.497	3.539	3.749
N – K Test	A	AB	ABCD	BCD	BCD	CDE	DE	E
Alarms with Lower Seat Vibration								
Alarm Type	3	5	6	8	7	4	1	2
Avg. Rating (0 - 4)	2.287	2.537	2.787	2.871	2.956	3.081	3.121	3.331
N – K Test	A	AB	BC	CD	CDE	CDEF	CDEF	F
Alarms with Upper Seat Vibration								
Alarm Type	3	5	6	8	7	4	1	2
Avg. Rating (0 - 4)	2.287	2.537	2.787	2.871	2.956	3.081	3.121	3.331
N – K Test	A	AB	BC	C	CD	CD	CD	D
Alarms with Combination Seat Vibration								
Alarm Type	3	5	6	8	7	1	4	2
Avg. Rating (0 - 4)	2.787	2.997	3.122	3.204	3.329	3.330	3.414	3.663
N – K Test	A	AB	ABC	BC	BC	BC	CD	D
Alarms with Steering Vibration								
Alarm Type	3	5	8	6	7	1	4	2
Avg. Rating (0 - 4)	2.412	2.456	2.789	2.870	2.954	3.121	3.162	3.287
N – K Test	A	A	B	BC	BCD	BCD	CD	D

Countermeasures

A one-way ANOVA was used to look for significant differences among the effectiveness ratings of the six countermeasures (lower seat vibration, upper seat vibration, combination seat vibration, A/O task, lane-minder, and the peppermint scent). No significant differences ***(a = 0.05)*** were found. The mean effectiveness ratings of the six countermeasures are shown in Appendix 2K.

DISCUSSION

Initial Advisory Tone

The results of the analyses of the initial advisory tone data demonstrate that particular characteristics are more effective at capturing the attention of an alert driver. As can be seen in Table 2.2, the higher frequency, rectangular wave tones were placed at lower preferred volume levels and rated higher on the effectiveness scale. The 1000 Hz and 700 Hz rectangular wave tones were placed in the “Very Effective” to “Extremely Effective” range on the rating scales. The sinusoidal waves were placed at much higher sound levels and still were only rated in the “Slightly Effective” to “Moderately Effective” range on the rating scales. Thus the most effective tones are not necessarily the ones with the highest sound level. The characteristics of the tone such as wave shape and frequency have a larger impact on the effectiveness of a certain tone.

Voice Message

The analyses of the voice message data resulted on no significant differences. The average preferred peak amplitude for the voice messages was found to be 83.4 dBA and 82.5 dBA for the male and female voices, respectively. The male voice on average was rated slightly higher (2.75) than the female voice (2.58). However, this difference is not statistically significant. Both the male or female voice seem to be quite effective at conveying the necessary information to the driver.

Alarm Stimuli

The effectiveness of the alarm sounds was investigated in two different ways. The ANOVA performed on the paired comparison data and the ANOVA performed on the rating scale data both produced significant results. Not only were these results significant, but they placed the eight alarm sounds in the same ranked order. The order of this rank can be seen in Tables 2.3 and 2.6. In both cases alarms #I2 (on-off tone) and #l (dual tone) are considered the most effective alarms. Both were placed well into the "Very Effective" to "Extremely Effective" range on the rating scales. In the paired comparisons, alarms #2 and #l were preferred over the other alarms 50 and 42, respectively, out of a possible 56 times. Alarms #2 and #1 share several characteristics. They are high-frequency, square waves with alternating periods of no sound or a sound at a lower frequency. Both have a repetition frequency of 3 Hz.

The 2-way ANOVA of Alarm Type and Stimulus Condition also produced significant results. Post-hoc tests revealed that in almost all cases the addition of a peripheral stimulus increased the effectiveness of a particular alarm. The brake pulse added to the effectiveness of every alarm sound. This can be seen clearly in Figure 2.4. The combination seat vibration also added to the effectiveness of most of the alarm sounds. An increased effectiveness was observed for all alarms with the combination seat vibration over the alarms by themselves. However, this increase was not significant in alarms #l and #2. This lack of significance may be due in part to the high ratings given to these two alarms by themselves. Subjects may have run out of room on the scales. Another interesting finding involving alarm #2 is that three of the peripheral stimuli; lower seat vibration, upper seat vibration, and steering vibration were reported to have detracted slightly from the effectiveness of this alarm (Figure 2.4). This again may be due to the initial high rating of this alarm by itself.

Countermeasures

As stated above, no significant differences were found among the effectiveness ratings of the countermeasures. For the most part they were only rated between Slightly Effective and Moderately Effective. The lane minder and the scent were rated slightly higher than Moderately Effective (2.12 1 and 2.165, respectively). A possible reason for the low rating of the seat vibration is that many subjects expressed that they found it relaxing. Many subjects also found the A/O task monotonous and something which they could perform in a very relaxed and inattentive state.

Conclusions

This study succeeded in answering a wide variety of questions regarding alerting/alarming stimuli to be used in a drowsy driver detection system. Conclusions drawn from the different areas of analyses in this study are presented below.

Initial Advisorv Tone. The purpose of the initial advisory tone is to capture the attention of a driver and direct his or her attention to a subsequent voice message. The results of this study indicate the preferred length of this tone to be 0.8 second. The recommended characteristics for the initial advisory tone are a rectangular shaped wave at a frequency of 700 Hz to 1000 Hz with a dBA level of approximately 8.1 dBA below the ambient noise level in the vehicle (The ambient noise level of the vehicle at 60 mph during this experiment was 73.1 dBA). Because of the concentration of high frequencies this tone is still audible even though it falls below the surrounding noise level.

Voice Message. As stated in the discussion section, there were no significant differences found between the male and female voice messages. However, the results do show that the voice message used in this study ("Possible drowsiness has been detected; press reset now"), presented at a peak amplitude of approximately 10 dBA above the ambient noise level in the vehicle (average amplitude of approximately 5 dBA below ambient noise level), was consistently rated from "Moderately Effective" to "Very Effective". Therefore, it may be concluded that either a male or a female voice may be used effectively as part of a drowsy driver detection system.

Alarming Stimuli. Both analyses of the alarm sounds showed alarms #2 and #1 as very effective. These alarms are both high-frequency, square waves with alternating periods of no sound or a sound at a lower frequency. Both have a repetition frequency of 3 Hz. The results of this study support increased effectiveness when dual modality is incorporated in an alarm. It can be concluded that the most effective way to re-alert a drowsy driver is to expose the driver to an alarm sound with characteristics similar to alarms #2 and #l at a amplitude of 3.5 dBA and 10 dBA (respectively) above the ambient noise level in the vehicle. This audible alarm should be accompanied by a simulated brake pulse to achieve maximum effectiveness. If brake pulses cannot be implemented, combined seat vibration can be used as an effective substitute.

Countermeasures. The results of this study found no countermeasures that were significantly different from one another. In addition, no investigated countermeasure rated much higher than "Moderately Effective". This leads us to conclude that further research may be needed. However, in the absence of further research, a lane-minder or peppermint scent system might be implemented. Both were rated as "Moderately Effective", and there is no quarantee that further research on other countermeasures will produce better results.

REFERENCES FOR PART II

Artaud, P., Planque, S., Lavergne, C., Cara, H., de Lepine, P., Tarriere, C., and Gueguen, B. (1994, May). An onboard system for detecting: lapses of alertness in car driving. Presented at the 14th International Technical Conference on the Enhanced Safety of Vehicles, Munich, Germany.

Baker, W. J. and Theologus, G. C. (1972). Effects of caffeine on visual monitoring. Journal of Applied Psychology, 56,422-427.

Chaudoin, J. H. and Nelson, G. (1985). Interstate routes 15 and 40 shoulder rumble strips. (Report Caltrans-08-85-l). California Department of Transportation.

Childs, J. M. (1978). Caffeine consumption and target scanning performance. Human Factors, 20, 9 l-96.

Daimler Benz. (1994, May). Driver drowsiness warning system. Slide presentation materials available upon request from Daimler Benz, Forschung und Technik, Stuttgart, Germany.

David, H. A. (1963). The method of paired comparisons. London: Griffin.

Dingus, T. A., Hardee, L. H., and Wierwille, W. W. (1985). Detection of drowsy and intoxicated drivers based on highway driving performance measures. (Tech. Report #8402). Blacksburg, VA: Virginia Polytechnic Institute and State University, IEOR Department.

Edwortby, J., Loxely, S., and Dennis, I. (1991). Improving warning design: Relationship between warning sound parameters and perceived urgency. Human Factors, 33(2), 205-23 1.

Erwin, C. W., (1976, February). Studies of drowsiness. (Final Report) Durham, NC: The National Driving Center.

Erwin, C. W., Hartwell, J. W., Volow, and Alberti, G. S. (1976). Electrodermal change as a predictor of sleep. Cited in Erwin, C. W. (1976, February) Studies of drowsiness (Final Report) Durham, NC: The National Driving Center.

Glass, G. V., Peckham, P. D., Sanders, R. J. (1972). Consequences to meet assumptions underlying the fixed effects analysis of variance and covariance. Review of educational research, 1972,42(3), 237-288.

Harris, W. H. and Mackie, R. R. (1972). A study of the relationships among: fatigue, hours of service, and safety of operations of truck and bus drivers (Report No. BMCS-RD-7 l-2). Washington, D. C.: Bureau of Motor Carrier Safety. (NTIS No. PB2 13-963).

Harris, W. V. (1967). The effect of ammonia and ammonium carbonate in the reduction of drowsiness in the human operator, Masters Thesis, Department of Systems Engineering, University of Arizona. Cited in Hulbert, S. F. (1972). Effects of driver fatigue. In T. W. Forbes (Ed.). (1972). Human Factors in Highway Traffic Safety Research (pp. 288-302). New York: Wiley and Sons.

Hauri, P. H. (1982). The sleep disorders (2nd ed.). Kalamazoo, MI: Upjohn.

Haworth, N. L., Heffeman, C. J., and Home, E. J. (1989). Fatigue in truck accidents. Melborne, Australia: Monash University, Accident Research Centte.

Haworth, N. L., Triggs, T. J., and Grey, E. M. (1988, June). Driver fatigue: Concepts, measurement and crash countermeasures (Report CR 72). Melbome, Australia: Monash University, Department of Psychology.

Haworth, N. L., Vulcan P,, Triggs, T. J., and Fildes, B. N. (1989). Driver fatigue research: Development of methodology. Melbome, Australia: Monash University, Accident Research Centre.

Haworth. N. L. and Vulcan, P. (1990). Testing of commercially available fatigue monitors (Draft Report). Melbome, Australia: Monash University, Accident Research Centre.

Horowitz, A. D., and Dingus, T. A. (1992). Warning signal design: A key human factors issue in an in-vehicle front-to-rear-end collision warning system. In Proceedings of the Human Factors Society 36th Annual Meeting (pp. 1011-1013). Santa Monica, CA: Human Factors Society.

Hulbert, S. F. (1963, October). Blood sugar level and fatigue effects on a simulated driving task. Engineering report 63-43, UCLA. Cited in Hulbert, S. (1972). Effects of driver fatigue. Human Factors in Highway Traffic Safety Research. New York: Wiley and Sons.

Hulbert, S. F. (1972). Effects of driver fatigue. In T. W. Forbes (Ed.). (1972). Human Factors in Highway Traffic Safety Research (pp. 288-302). New York: Wiley and Sons.

Itoi, A., Cilveti, R., Voth, M., Dantz, B., Hyde, P., Gupta, A., Dement, W. C. (1993). Can drivers avoid falling asleep at the wheel? Relationship between awareness of sleepiness and ability to predict sleep onset. Stanford University Sleep Research Center, February.

Kaneda, M., Iizuka, H., Ueno, H., Hiramatsu, M., Taguchi, M., and Tsukino, M. (1994, May). Development of a drowsiness warning system. Presented at the 14th Jnternational Technical Conference on the Enhanced Safety of Vehicles, Munich, Germany.

Keamey, P. W. (1966). Highway homicide. New York: Cromwell.

Khardi, S. and Vallet, M. (1994, May). Drowsiness of the driver: E.E.G. and vehicle parameters interaction. Presented at the 14th Jnternational Technical Conference on the Enhanced Safety of Vehicles, Munich, Germany.

Knipling, R. R. and Wang, J. S. (1994). Crashes and fatalities related to driver drowsiness/fatigue. NHTSA Research Note, November.

Leonard J. and Wierwille, W. W. (1975). Human performance validation of simulators: Theory and experimental verification. In Proceedings of the 19th Annual Meeting of the Human Factors Society (pp. 446-456). Santa Monica, CA: Human Factors Society.

Lemer, N. D., Kotwal, B. M., Lyons, R. D., and Gardner-Bonneau, D. J. (1993, November). Preliminary human factors guidelines for crash avoidance warning devices (Report No. DTNH22-9 l-C-07004). Washington, D.C.: NHTSA, OCAR.

Lewis, J. R. (1994). Sample sizes for usability studies: Additional consideration. Human Factors, 36(2), 368-378.

Lisper H. O., Eriksson, B., Fagerstrom, K. O., and Lindholm, J. (1979). Diurnal variation in subsidiary reaction time in a long-term driving task. Accident Analysis and Prevention, 11, 1-5.

Lisper, H. O., Laurell, H., and Van Loon, J. (1986). Relation between time to falling asleep behind the wheel on a closed track and changes in subsidiary RT during prolonged driving on a motorway. Ergonomics, 29, 445-453.

Lisper, H. O., Tomros, J., and Van Loon, J. (1981). Effects of caffeine or diazepam on subsidiary reaction time in a long-term driving task. Cited in Haworth, N. L. and Vulcan, P. (1990). Testing of commercially available fatigue monitors. (Draft Report). Melbome, Australia: Monash University, Accident Research Centre.

Mackie, R. R. and O'Halon, J. F. (1977). A study of the combined effects of extended driving and heat stress on driver arousal and performance. Cited in Haworth, N. L., Triggs, T. J., and Grey, E. M. (1988, June). Driver fatigue: Concepts, measurement and crash countermeasures (Department of Psychology Report #CR 72). Monash University, Melbome, Australia.

Maxwell, A. E. (1974). The logistic transformation in the analysis of paired-comparison data. British Journal of Psychology, 27, 62-71.

Nielsen, J. (1993). Usability Engineering. Boston, MA: Academic Press.

Office of Crash Avoidance Research (199 1, September). Drowsy/fatigued driver-crashes. IVHS/Crash Avoidance Countermeasure-Target Crash Problem Size Assessment and Statistical Description (Report No. 4). Washington, D.C.: NHTSA, OCAR.

Ohnemus, K. R. and Biers, D. W. (1993). Retrospective versus concurrent thinking-out-loud in usability testing. In Proceedings of the Human Factors Society and Ergonomics Society 37th Annual Meeting (pp. 1127-l 13 1). Santa Monica, CA: Human Factors Society.

Oswald, I. (1962). Sleeping and Waking. New York: Elsevier Press.

Plaque, S., Chaput, D., Petit, C., Tarierre, C., and Chabanon, C. (1991, November). Analysis of EGG and EEG signals to detect lapses of alertness in car simulation driving. Presented at the 13th ESV Conference, Paris, France.

Sanders, M. S. and McCormick, E. J. (1993). Human Factors in Engineering and Design. New York, NY: McGraw-Hill, Inc.

Seko, Y. (1984). Present technological status of detecting drowsy driving patterns. Jidosha Gijutsu, 30(5), 547-554.

Thorpy, M. J. and Ledereich, P. S. (1990). Medical treatment of obstructive sleep apnea. Cited in M. J. Thorpy (Ed.) (1990). Handbook of sleep disorders. Bronx, NY: Monte Fiore Medical Center and Albert Einstein College of Medicine.

Tiley, D. H., Erwin, C. W., and Gianturco, D. T. (1973, January). Drowsiness and driving: Preliminary report of a population survey (Report No. 730121). Detroit, MI: Society of Automotive Engineers.

Torsvall, L. and Akerstedt, T. (1988). Extreme sleepiness: Quantification of EOG and EEG parameters. International Journal of Neuos, 38, 435-441.

Virzi, R. A. (1992). Refining the test phase of usability evaluation: How many subjects is enough? Human Factors, 34(4), 457-468.

Virzi, R. A., Sorce, J. F., and Herbert, L. B. (1993). A comparison of three usability evaluation methods: Heuristic, think-aloud, and performance testing. In Proceedings of the Human Factors Society and Ergonomics Society 37th Annual Meeting (pp. 309-313). Santa Monica, CA: Human Factors Society.

Wesnes, K. and Warburton, D. M. (1978). The effect of cigarette smoking and nicotine tablets upon human attention. Cited in Haworth, N. L., Triggs, T. J., and Grey,

E. M. (1988, June). Driver fatigue: Concepts. measurement and crash countermeasures (Report CR 72). Melbome, Australia: Monash University, Department of Psychology.

Wierwille, W. W. and Muto, W. H. (1981, May). Significant changes in driver-vehicle response measures for extended simulated driving tasks. Presented at the First European Annual Conference on Human Decision Making and Manual Control, Delft, Netherlands.

Wierwille, W.W., Wreggit, S.S., Kim, C.L., Ellsworth, L.E., and Fairbanks, R.J. (1994). Research on Vehicle-Based Driver Status/Performance Monitoring: Three Vehicle Report Analysis and Simulation Laboratory, Virginia PolytechnicYear Institute and State University, VPISU Report No. ISE 94-04, NHTSA Report No. DOT HS 808 247, December, 1994.

Wood, N. E. (1994). Shoulder rumble strips: A method to alert “drifting’ drivers . Harrisburg, PA: Pennsylvania Turnpike Commission.

Wreggit, S. S., Kirn, C. L., and Wierwille, W. W. (1993). Research on vehicle-based driver status/performance monitoring (Report # 93-06). Blacksburg, VA: Virginia Polytechnic Institute and State University, ISE Department.

Wright, R. B. and Converse, S. A. (1992). Method bias and concurrent verbal protocol in software usability testing. In Proceedings of the Human Factors Society 36th Annual Meeting (pp. 1220-1224). Santa Monica, CA: Human Factors Society.

APPENDIX 2A

Descriptions of Functions in Driver's Control Panel

Lighted Push Buttons

"Monitoring System On" (Momentary: Push-on, Push-off)

Activates drowsiness detection system.

Begins performance monitoring.

Provides voice caution to driver to use directional signals when changing lanes.

Lights the "Monitoring System On" push button.

Initializes drowsiness detection system (new baseline established).

"Drowsiness Detected" (This is a light only)

Flashes yellow when performance monitoring system detects drowsiness and reset push button has not yet been depressed.

When reset is depressed momentarily, it goes to dull yellow for six minutes and then extinguishes.

Flashes yellow when A/O task monitoring system detects drowsiness and reset push button has not yet been depressed.

When reset is depressed momentarily, it goes to dull yellow for six minutes and then extinguishes.

Flashes yellow when lane-minder senses out of lane condition. Extinguishes when vehicle re-enters the lane.

"Reset" (Momentary: Push-on, Push-off)

Usually dull orange. Bright orange when momentarily depressed or when "drowsiness detected" light flashes.

Stops the flashing of the "drowsiness detected" light when depressed momentarily.

Prevents or stops the alarm system from sounding/vibrating for a period of six minutes.

Initiates message to select countermeasures when depressed momentarily.

Cancels all activated countermeasures when depressed momentarily.

"Seat Vibration" (Momentary: Push-on, Push-off)

Activates both the push button light and the seat vibration system.

Both the light and seat vibration are extinguished by depressing reset button.

"Scent" (Momentary: Push-on, Push-off)

Activates the scent push-button light and causes scent to be injected into air in front of the driver.

Activates the "Fresh Air" light and fresh air blower.

"Fresh Air" (Momentary: Push-on, Push-off)

Activates the fresh air push button and fresh air blower.

If scent light is lit, "Fresh Air" button has no effect when depressed.

" A/O Task" (Momentary: Push-on, Push-off)

Initiates A/O task and push-button light.

Causes monitoring system to activate A/O task algorithms.

When deactivated, it causes the monitoring system to rely on performance monitoring only.

Prevents lane-minder system from engaging.

"Lane-Minder"

Initiates lane-minder system and push-button light.

Prevents A/O task system from engaging.

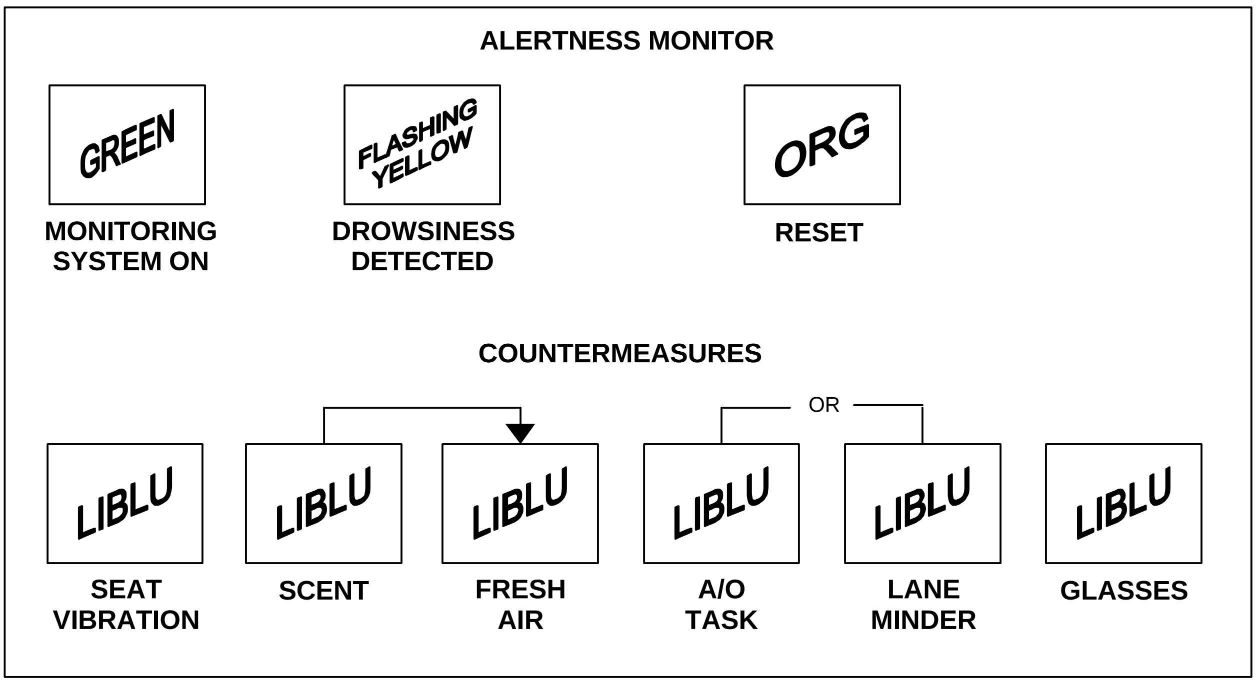


Figure 2A.1: Driver's Control Panel for Alertness Monitor

APPENDIX 2B
Average Sound Pressure Level Study

Measurement of Sound Pressure Levels

Method

An Extech Instruments digital sound level meter was used to measure the sound pressure levels in five different vehicles, The sound level meter was operated in its slow response mode. The measurements were taken on the A-scale at 45,55, and 65 MPH.

All measurements were taken while driving on the bypass section of Rt. 460 between North and South Main Streets in Blacksburg, Virginia. All readings were taken while in the highest gear.

All windows were closed. Any fan or air conditioning was turned off. Radios remained off. Neither driver nor experimenter spoke while measurements were taken.

For each run, the sound level meter was placed approximately 5 inches from the driver's right ear. Seven measurements were taken at each of the three speeds. These seven measurements were used to produce an average sound pressure level for each speed. The five vehicles used and the sound levels measured are shown in Table 2B1.

Table 2B. 1: Average Sound Levels for Individual Vehicles

	1988 Ford Range XLT			1994 Honda Accord LX			1995 VW Jetta GL			1992 Ford Explorer XLT			1992 Mazda Protégé DX		
	45 mph	55 mph	65 mph	45 mph	55 mph	65 mph	45 mph	55 mph	65 mph	45 mph	55 mph	65 mph	45 mph	55 mph	65 mph
	73.7	75.2	78.2	72.1	72.1	73.6	73.7	76.3	78.6	70.2	74.4	78.0	74.7	73.7	74.5
	72.6	76.1	77.9	71.8	72.5	73.1	74.9	78.6	78.3	70.8	72.4	77.9	75.5	74.4	75.7
	71.8	75.8	77.6	69.9	71.9	72.5	76.6	77.3	78.2	70.9	73.4	77.1	73.1	75.3	74.8
	73.0	76.3	77.5	70.1	70.9	72.6	74.4	77.1	78.4	69.8	73.9	76.8	73.6	75.6	74.1
	72.5	75.9	77.8	69.7	72.8	71.8	75.1	76.3	77.7	70.2	72.1	77.5	74.9	74.8	75.0
	72.9	76.0	78.0	69.1	72.2	72.8	74.9	76.5	77.8	69.9	71.7	77.2	74.1	75.0	75.3
	72.6	76.4	77.7	69.2	71.3	74.0	76.0	77.5	78.2	70.6	73.9	76.9	73.8	75.8	74.6
Average	72.7	76.0	77.8	70.3	72.0	72.9	75.1	77.1	78.2	70.3	73.1	77.3	74.2	74.9	74.9

APPENDIX 2X

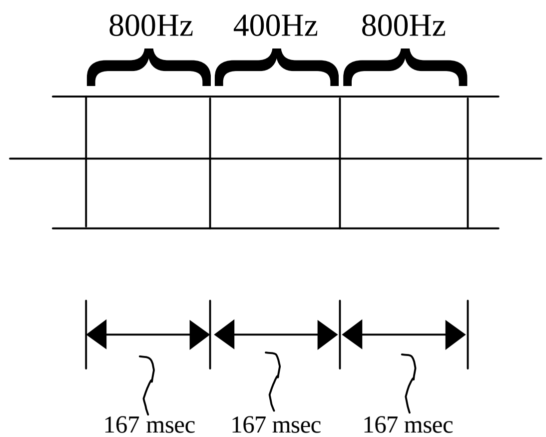
Technical Description of the Eight Alarm Sounds Used in the Experiment

Technical Description of the
Eight Alarm Sounds Used in the Experiment

All eight sounds are composed of rectangular waves with equal positive and negative going durations that are modulated in some way.* In other words, if the modulation were temporarily removed, the resulting wave would be a simple rectangular (square) wave.

Alarm 1. Short name: Dual Tone

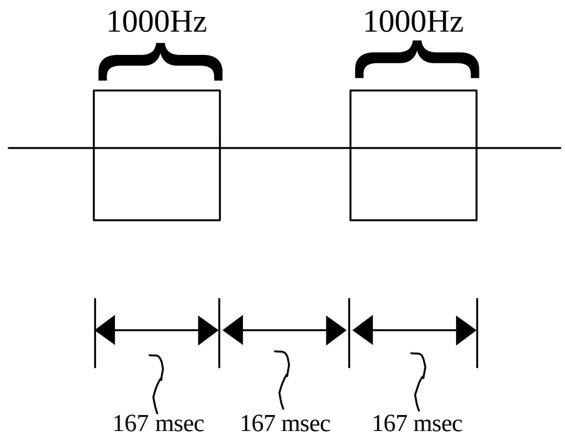
The waveform is at 800 Hz for 167 msec and then switches to 400 Hz for 167 msec. (Repetition frequency is 3 Hz).



*Alarm 3 is unmodulated

Alarm 2. Short name: On-off Tone

The waveform is at 1000 Hz for 167 msec and then is off for 167 msec. (Repetition frequency is 3 Hz.)

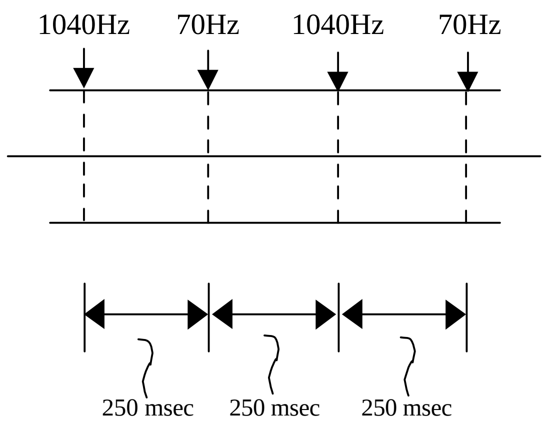


Alarm 3. Short name: Continuous Tone

The waveform is a continuous tone at 440 Hz. (There is no modulation.)

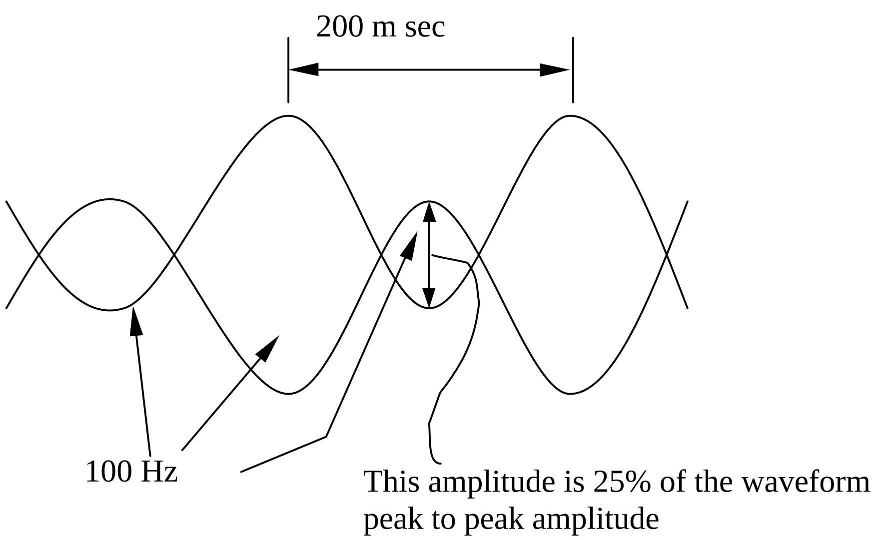
Alarm 4. Short name: Frequency Swept Tone

The waveform sweeps sinusoidally between 70 Hz and 1040 Hz. The sweep time from 70 to 1040 Hz is 250 msec, and the sweep time from 1040 to 70 Hz is 250 msec. (Repetition frequency is 2 Hz.)



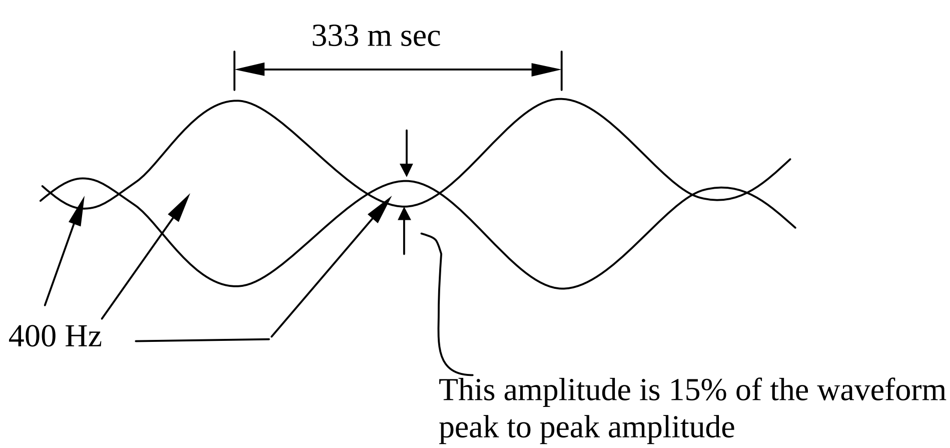
Alarm 5. Short name: Overmodulated low-frequency tone

The waveform is a sinusoidally amplitude-overmodulated 100 Hz tone. Peak amplitudes are 200 msec apart. (Repetition frequency is 5 Hz.)



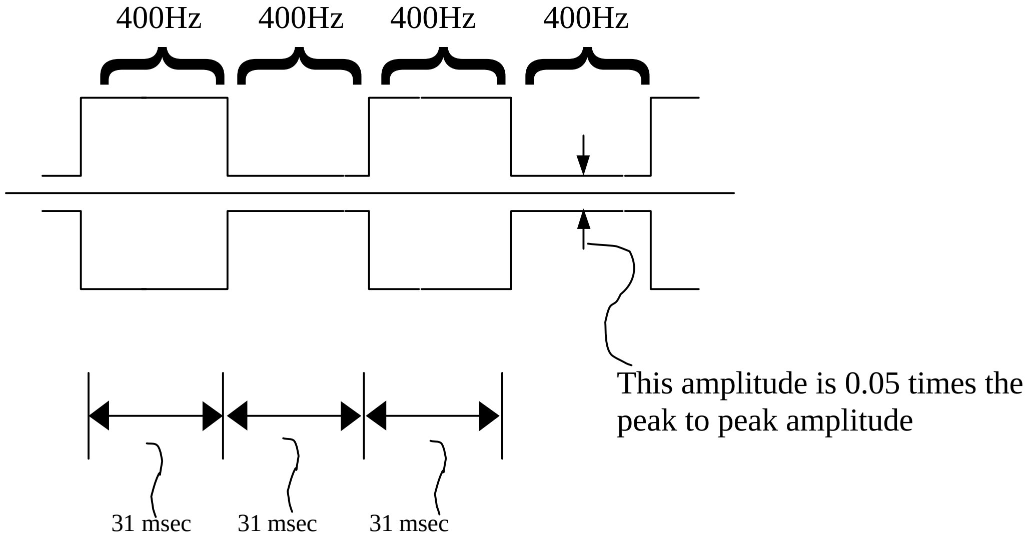
Alarm 6. Short name: Overmodulated mid-frequency tone

The waveform is a sinusoidally amplitude-overmodulated 400 Hz tone. Peak amplitudes are 333 msec apart. (Repetition frequency is 3 Hz.)



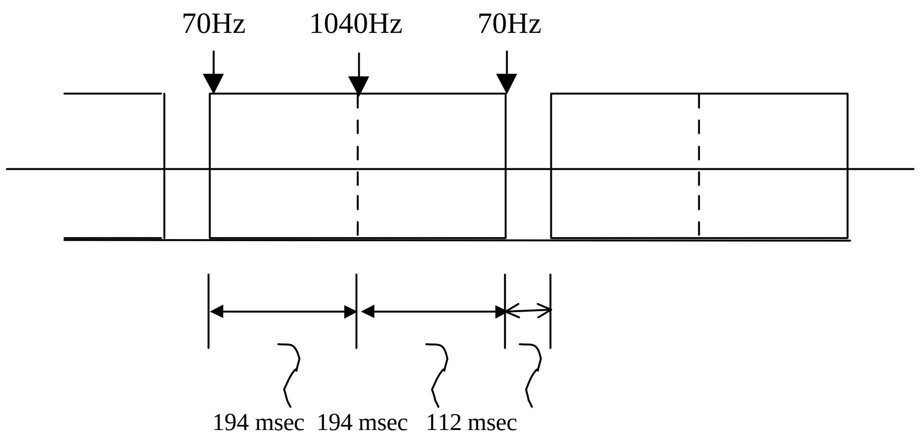
Alarm 7. Short name: Rapid, amplitude-modulated tone

The waveform is at 400 Hz with the amplitude at full amplitude for 3 1 msec followed by the amplitude at 5% of full amplitude for 31 msec. (Repetition frequency is 16 Hz.)



Alarm 8. Short name: Gapped frequency-swept tone

This waveform sweeps sinusoidally from 70 Hz and 1040 Hz and back again. Then there is a short gap before the process is repeated. The sweep time from 70 to 1040 Hz is 194 msec, and the sweep time from 1040 to 70 Hz is 194 msec. The gap before repeat is 112 msec. (Repetition frequency is 2 Hz.)



APPENDIX 2D
Specifications for Peripheral Alarm Cues and Countermeasures

Specifications for Peripheral Alarm Cues and Countermeasures

Steering Vibration

A 10 Hz square wave is applied to the servo system, whose output is torque applied to the steering wheel shaft.

For the alarm cue, the torque causes a 6.3 mm (0.25”) peak-to-peak vibration at the outside rim of the steering wheel. This measurement is made at 60 mph with the hands removed from the wheel. The wheel is assumed to be 38 cm (15”) in diameter (to the outside edge).

Seat Vibration

Vibrators are installed in the seat pan (bottom) and in the seat back (opposite the L-1 point of the spine for an average height male). The vibrators are composed of eccentric masses driven by small servomotors that are run open loop (Figure 2D1). The motors are mounted on brackets that are attached to flexible lucite sheets. The sheets are 7” by 7” by 3/16” thick and are lashed to the seat springs. For the seat pan installation, the lucite is between the springs and the foam padding of the seat (which is approximately 1 1/2” thick). For the seat back installation, the lucite is against the back side of the foam pad because the springs are embedded in the foam pad. (The foam pad thickness for the seat back is also 1 1/2” thick.)

For the alarm cues, the upper eccentric spins at 1380 RPM and the lower spins at 1440 RPM.

For the countermeasures cues, the upper eccentric spins at 875 RPM and the lower spins at 1030 RPM.

The eccentric for the upper vibrator is a rectangular solid made of mild steel. It is 1/2” by 1/2” by 2” long. The shaft hole for the drive motor and the set screw to hold the eccentric are 1 3/8” from one end of the eccentric.

The eccentric for the lower vibrator is also a rectangular solid made of mild steel. It is 1/2” by 1/2” by 2 3/16” long. The shaft hole for the drive motor and the set screw to hold the eccentric are 19/16” from one end of the eccentric.

Both motors are mounted on brackets with their shafts parallel to the lucite sheets. The shafts are 1 3/16” from the lucite sheets.

The upper vibrator is mounted in the seat back with the motor shaft vertical (actually, tilted rearward slightly) and the eccentric at the top. The lower vibrator is mounted with the shaft horizontal and aligned with the longitudinal axis of the vehicle. The eccentric is mounted at the front end.

Brake Pulse

The brake pulse alarm cue is estimated to be applied for 0.5 sec and results in a 6 mph drop in speed from a cruising speed of 60 mph.

> (Note that if the brake pulse alarm cue is implemented in an actual vehicle, the rear brake lights of the vehicle should begin flashing during the advisory tone

and voice message. When the brake pulse is applied and for a period of time thereafter, the brake lights should then be activated continuously.)

Scent

A repressurized (one pint) spray paint can (Jennican Products, Ltd., Guilford, Surry, U.K.) is used. The interior plastic hose is removed so that the can will spray a vapor and not the liquid from the bottom. One and one-half ounces of peppermint oil and an equal amount of water are poured into the can. After sealing, the can is pressurized to 85 psi.

When the scent is to be administered as a countermeasure, the experimenter shakes the can to mix the ingredients. The can is then handed to the subject who is instructed to dispense a burst of the vapor by directing the nozzle from right to left in front of the face. The subject may then inhale as much of the vapor from the single burst as he or she deems appropriate.

LaneMinder

The lane minder is a countermeasure that provides an audible warning that the (simulated) vehicle is partly out of lane. The alarm begins with a continuous tone as soon as the vehicle begins to edge out of lane. As the deviation continues, the sound level increases and then begins "beeping." Maximum sound level and beeping occur when the vehicle is three feet or more out of lane.

If the vehicle goes out of lane on the left, the sound emanates from a piezo-electric transducer to the left of the driver. If the vehicle goes out of lane on the right, the sound emanates from a piezo-electric transducer to the right of the driver. The transducers are mounted at mid-dash height, approximately 35" off to the left and off to the right of the driver's straight-ahead position.

The frequency of each transducer is 2800 Hz and the beep frequency is 3 Hz. Sound level at maximum excursion is 73 dBA.

A/O Task

This countermeasure is a subsidiary task presented auditorially to the driver. A recorded male voice presents a single word at each 15-second interval. The driver is to determine if the word contains the letter "A" or the letter "O", or both. If it does, the driver depresses the "yes" pushbutton mounted on the right spoke of the steering wheel at the usual right thumb position. If it does not, the driver depresses the "no" pushbutton mounted symmetrically on the left at the usual left thumb position.

Stimulus words are selected so that they do not contain both an A and an 0. Furthermore, they are selected so that presence or absence of an A or 0 is easy to determine.

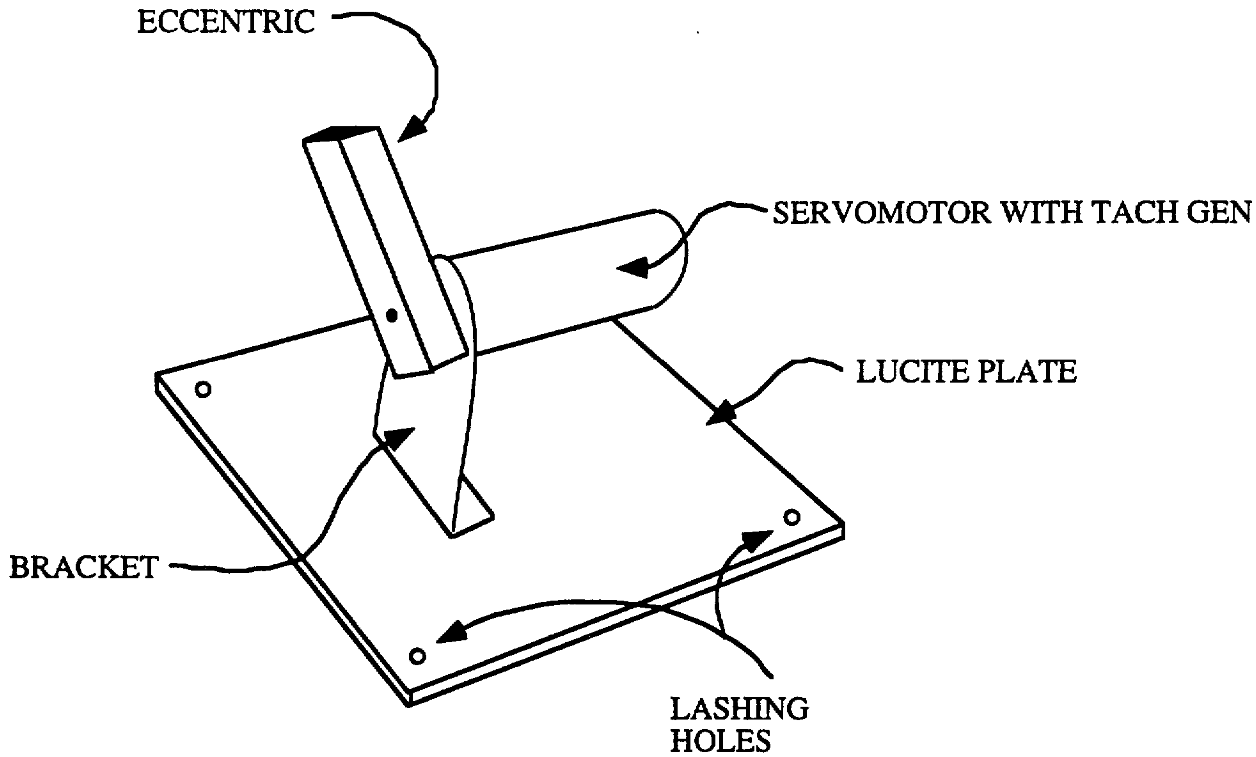


Figure 2D1. Sketch of vibrator assembly

APPENDIX 2E
Introduction to the Study and Informed Consent

Introduction to the Study

The purpose of this research is to investigate different characteristics of advising/alarming signals for use in a drowsy driver detection system in an automobile and to determine one or more appropriate configurations. The study is being conducted in the Vehicle Analysis and Simulation Laboratory, Department of Industrial and Systems Engineering, Virginia Polytechnic Institute and State University, Blacksburg. The research team consists of Sarah E. Fahey and Terry Fairbanks. The two researchers are graduate students in the Department of Industrial and Systems Engineering. Dr. Walter W. Wierwille is the Principle Investigator and Paul T. Norton Professor in the Department.

Your task will be to sit in an automobile simulator and drive as you would normally. The simulator will move so as to mimic the motions of an actual automobile. You will be asked to make preference choices and rate several different stages of the advising/alarming system. Phase one of the experiment will consist of three stages in which you will be asked to give your input. Phase two of the experiment will consist of only one stage.

The first stage of phase one will consist of advisory tones and voice messages. You will be asked to optimize and select parameters associated with each. For example, you will be asked to select the volume of each as well as other parameters such as the length of the advisory tones. You will also be asked to select which tones and voice messages are more effective at advising a driver of a subsequent alarm. It is important to understand that the intended purpose of the initial advisory tone and voice message is not to re-alert a drowsy driver but, rather, to inform a driver of a subsequent alarm. Therefore, you should keep in mind while setting the volumes that they do not need to be so loud as to get the attention of a driver who is not at least moderately alert. The volume should be set at a moderate level so as to capture the attention of a driver.

In some applications of the drowsiness detection system, drivers may not respond to the advisory tone and voice message. Therefore, the second stage of phase one will involve the selection and optimization of alarm sounds to re-alert the driver. Again, you will be asked to set the volume of the alarm sounds and other parameters. You will also be asked to select the alarms which you feel are most effective at alerting a drowsy driver. It is important to keep in mind that the intended purpose of these alarm sounds is to re-alert a drowsy driver. Therefore, in setting the volume you should be aware that the alarm sounds need to be loud enough to alert a drowsy driver but not so loud that they would startle the driver.

The third stage of phase one will deal with additional peripheral stimuli. You will be asked to rate these stimuli, such as steering and seat vibration, on how much they improve the effectiveness of the alarm stimulus complement.

The second phase of the study will focus on stimuli used to maintain a driver's alermess level. The purpose of this maintenance is so that the driver can stay alert for ten to fifteen minutes while driving to a safe place to stop and refresh. Stimuli which you will have experienced earlier, such as vibration, will be presented. You will, once again, be asked to rate the effectiveness of these stimuli. However, the intended purpose of these countermeasures is to maintain the driver's re-alerted state achieved by the previous alarm sounds. Therefore, you will be asked to rate the degree of effectiveness with which these countermeasures will help maintain a driver's re-alerted state for a period of approximately ten minutes.

If you decide to participate in this study, you must awake at 7 AM or before and go through your normal daytime activities without resting or napping. Then, at about 6 PM, a member of the experimental team will pick you up at your residence. This team member will buy you dinner at a fast-food restaurant. You may eat whatever you like, but you will not be permitted to drink caffinated or sugared beverages, such as coffee or coke. If you are a smoker, you will be permitted to smoke right after dinner, but not thereafter. You will then be taken to the laboratory where you will be allowed to read, study, watch TV (which will be provided), or listen to your own personal headset stereo. You will not be permitted to eat, smoke, or drink caffinated coffee, or caffmated soft drinks, since these may effect the outcome of the experiment. However, you will be permitted to drink water or non-caffmated, diet soft drinks.

Shortly after midnight the experimental session will begin. It will last roughly two (2) hours. You will be asked to drive the simulator and experience various stimuli. Your opinion will be sought in the volume adjustment of these stimuli and alarms. Your preferences between the various stimuli will also be sought.

After the completion of the experiment, you will be paid and any remaining questions will be answered. If you participate in this experiment you must agree to let one of the experimenters drive you home, since they will not be drowsy at this time.

Payment for the experiment will be $5 per hour between 6 PM and midnight, and $8 per hour after midnight. If you complete the experiment you will receive approximately $46. If you decide to withdraw during the experiment or simply cannot continue for whatever reason, you will be paid for the time actually spent. Since the simulator is a complex system, equipment failures do occasionally occur. If this happens it may be necessary for the experimenters to terminate the experiment, in which case you will similarly be paid for the time actually spent.

Once you are seated in the simulator, you must not attempt to leave the simulator until you have given the experimenters a chance to stop the simulator and guide you in exiting.

Initially, you will be asked to take a simple hearing test and fill out a brief questionnaire on your normal sleeping/waking patterns and your normal eating/drinking/smoking (if any) patterns. If you qualify, you will then be scheduled for the experiment.

There are some minor-risks and discomforts to which you will be exposed in this experiment. They are outlined in the attached informed consent form, which you should read carefully.

Participant's Informed Consent

1. You are being asked to volunteer to be a subject in a research project whose purpose and description are contained in the document **Introduction to the Study,** which you have already read.

2. There are some minor risks and discomforts to which you expose yourself in volunteering for this research. The risks are:

   The risk of possible interference with your next day's activities caused by less than a full night's sleep. This risk can be minimized by sleeping longer than usual the morning following your participation.

   The risk of injury if you attempt to leave the simulator without the help of the experimenters. Please inform one of the experimenters if you feel that you must leave the simulator. You will then be guided out of the simulator.

   The discomforts are:

   Possible discomfort associated with trying to drive while tired or drowsy.

   Possible minor motion sickness due to the movement of the simulator.

3. The data gathered in this experiment will be treated with anonymity. Shortly after you have participated, your name will be separated from your data.

4. While there are no direct benefits to you from this research (other than payment), you may find the experiment interesting. Your participation and that of other volunteers should aid in the determination of appropriate drowsiness waming and alerting alarms.

5. You should not volunteer for participation in this experiment if you have known hearing impairment, are under 18 years old, if you are pregnant, if you are not in good health, or if you have any other condition which would adversely affect you by being sleep deprived and staying up until approximately 2 AM.

6. You should know that the principle investigator of the research project (Dr. Wierwille) and the research team will answer any questions you may have about your participation, and you should not sign this consent form until you are satisfied that you understand all of the previous descriptions and conditions. (Dr. Wierwille's phone number is 23l-7952).

   You should further be aware that you may contact Dr. E. R. Stout, Chair of the University's Institutional Review Board, if you have questions or concerns about this experiment. Dr. Stout's phone number is 23l-9359.

7. You should know that at any time you are free to withdraw from participation in this research program without penalty for any reason.

8. You will be paid at a rate of $5.00 per hour before midnight and $8.00 per hour thereafter. If you complete your participation you will be paid $46.00. Payment will be made shortly after you have finished your participation.

9. You agree to allow one of the experimenters to drive you home following the experiment.

I have read and understand the scope of this research and I have no other questions. I hereby give my consent to participate, but I understand that I may stop participation at any time if I choose to do so.

Signature____

Date______________________

(A copy of this informed consent form shall be given to the participant)

APPENDIX 2F
Tone Characteristics and Presentation Order

Table 2F1: Characteristics of Initial Advisory Tones.

	1	2	3	4	5	6	7	8
Frequency (Hz)	200	200	400	400	700	700	1000	1000
Wave shape	sine	rect.	sine	rect.	sine	rect.	sine	rect.

Table 2F2: Presentation Order of Initial Advisory Tones

	Tones							
Subject 1	1	8	2	7	3	6	4	5
Subject 2	5	4	6	3	7	2	8	1
Subject 3	2	1	3	8	4	7	5	6
Subject 4	6	5	7	4	8	3	1	2
Subject 5	3	2	4	1	5	8	6	7
Subject 6	7	6	8	5	1	4	2	3
Subject 7	4	3	5	2	6	1	7	8
Subject 8	8	7	1	6	2	5	3	4

APPENDIX 2G

Sound Pressure Levels for Stimuli and Ambient Noise levels during experiment

Table 2G1: Sound Pressure Levels for Initial Advisory Tones. All measurements made with digital sound pressure level meter in slow response mode (Simulator not running).*

Tone 1 (200Sine)		Tone 2 (200Rect.)		Tone 3 (400Sine)		Tone 4 (400Rect.)	
level	dBA	level	dBA	level	dBA	level	dBA
6.5	76.2	2.5	65.0	4.0	68.4	2.0	60.0
7.0	77.8	3.0	69.2	5.0	70.7	2.5	63.8
7.5	79.6	3.5	70.9	5.5	72.0	3.0	67.3
8.0	80.7	4.0	72.3	6.0	73.4	3.5	69.6
8.5	81.1	4.5	73.4	6.5	74.5		
10.0	81.9	5.0	74.7	7.0	76.3		
				8.0	79.0		

Tone 5 (700Sine)		Tone 6 (700 Rect.)		Tone 7 (1000Sine)		Tone 8 (1000Rect.)	
level	dBA	level	dBA	level	dBA	level	dBA
3.5	69.3	2.0	60.2	2.0	61.4	2.0	62.0
4.0	70.5	2.5	64.6	2.5	65.6	2.5	66.4
4.5	71.6	3.0	67.9	3.0	69.2	3.0	70.2
5.0	72.2	3.5	69.4	3.5	71.5		
6.5	75.7			4.5	74.2		
				5.5	76.7		
				6.0	77.9		

Table 2G2: Sound Pressure Levels for Voice Messages. All measurements made with digital sound pressure level meter in slow response mode (Simulator not running).*

	MALE			FEMALE	
level	Peak values**	Slow response	level	Peak values**	Slow response
2.5	78.8	63.8	3.0	80.1	66.6
3.0	82.7	67.7	3.5	81.3	67.8
3.5	84.7	69.7	4.0	83.0	69.5
4.0	86.3	71.3	4.5	84.0	70.5
			5.0	85.3	71.8

* Measurements were made with the sound level meter 3 inches from the right ear of a typical driver.

** All analyses done with peak values.

Table 2G3: Sound Pressure Levels for Alarm Sounds. All measurements made with digital sound pressure level meter in slow response mode (Simulator not running)*.

ALARM 1		ALARM 2		ALARM 3		ALARM 4	
level	dB	level	dB	level	dB	level	dB
1.5	75.6	2.0	69.6	2.0	83.8	2.5	76.0
2.0	79.6	2.5	73.9	2.5	87.0	3.0	79.6
2.5	82.5	3.0	77.9	3.0	91.0	4.0	83.4
3.0	86.3	4.0	81.1	3.5	92.9	5.5	86.8
3.5	87.6	4.5	81.5	4.0	94.7		
4.0	89.6						

ALARM 5		ALARM 6		ALARM 7		ALARM 8	
level	dB	level	dB	level	dB	level	dB
3.0	71.0	2.5	71.4	2.5	74.3	2.5	76.7
3.5	73.3	3.5	76.8	4.5	82.6	3.5	82.0
5.0	77.0	4.0	78.2	5.0	83.9	4.0	83.4
5.5	78.3	4.5	79.4	6.0	86.4	4.5	84.4
6.0	79.6	6.0	83.2	6.5	87.8	6.0	88.0
8.0	86.3						

Table 2G4: Sound Pressure Levels for Ambient Noise in Laboratory and Simulator Running.*

Ambient noise in lab = 51 dBA	
Simulator up and running: dBA (digital, slow response)	
55 MPH	72.5
60 MPH	73.1
65 MPH	75.1

* Measurements were made with the sound level meter 3 inches from the right ear of a typical driver.

APPENDIX 2H
Preference Matrix and Maxwell's Scaling Procedure for Paired Comparison Data

Table 2H1: Preference Matrix for Paired Comparison Data.

When compared to:

		Alarm1	Alarm2	Alarm3	Alarm4	Alarm5	Alarm6	Alarm7	Alarm8	TOTAL
Prefer:	Alarm1		3	7	6	7	6	6	7	42
	Alarm2	5		8	6	8	7	8	8	50
	Alarm3	1	0		0	1	1	0	0	3
	Alarm4	2	2	8		8	6	5	7	38
	Alarm5	1	0	7	0		1	1	1	11
	Alarm6	2	1	7	2	7		1	2	22
	Alarm7	2	0	8	3	7	7		4	31
	Alarm8	1	0	8	1	7	6	4		27

Table 2H2: Proportion Matrix for Paired Comparison Data.

	Alarm1	Alarm2	Alarm3	Alarm4	Alarm5	Alarm6	Alarm7	Alarm8
Alarm1	X	0.375	0.875	0.75	0.875	0.75	0.75	0.875
Alarm2	0.625	X	1	0.75	1	0.875	1	1
Alarm3	0.125	0	X	0	0.125	0.125	0	0
Alarm4	0.25	0.25	1	X	1	0.75	0.625	0.875
Alarm5	0.125	0	0.875	0	X	0.125	0.125	0.125
Alarm6	0.25	0.125	0.875	0.25	0.875	X	0.125	0.25
Alarm7	0.25	0	1	0.375	0.875	0.875	X	0.5
Alarm8	0.125	0	1	0.125	0.875	0.75	0.5	X

Table 2H3: Matrix of z-deviates used in Maxwell's scaling technique.

$Z_{ij} = \log_e \{P_{ij}/(1-P_{ij})\}$ - OR - $Z_{ij}^* = \log_e\{(n_{ij}+1/2)/(n-n_{ij}+1/2)\}$ *if P's are 1 or 0

	Alarm1	Alarm2	Alarm3	Alarm4	Alarm5	Alarm6	Alarm7	Alarm8	Total(X)	X^2	Mean
Alarm1	0.000	-0.511	1.946	1.099	1.946	1.099	1.099	1.946	8.623	74.352	1.078
Alarm2	0.511	0.000	2.833	1.099	2.833	1.946	2.833	2.833	14.888	221.659	1.861
Alarm3	-1.946	-2.833	0.000	-2.833	-1.946	-1.946	-2.833	-2.833	-17.171	294.829	-2.146
Alarm4	-1.099	-1.099	2.833	0.000	2.833	1.099	0.511	1.946	7.025	49.344	0.878
Alarm5	-1.946	-2.833	1.946	-2.833	0.000	-1.946	-1.946	-1.946	-11.504	132.346	-1.438
Alarm6	-1.099	-1.946	1.946	-1.099	1.946	0.000	-1.946	-1.099	-3.296	10.863	-0.412
Alarm7	-1.099	-2.833	2.833	-0.511	1.946	1.946	0.000	0.000	2.282	5.209	0.285
Alarm8	-1.946	-2.833	2.833	-1.946	1.946	1.099	0.000	0.000	-0.847	0.718	-0.106
										789.319	

Table 2H4: Mean z-deviates and corresponding scaled values.

	Alarm3	Alarm5	Alarm6	Alarm8	Alarm7	Alarm4	Alarm1	Alarm2
mean x	-2.1463	-1.438	-0.412	-0.1059	0.2853	0.87807	1.07784	1.86103
scaled	0.0000	0.7083	1.7343	2.0404	2.4316	3.0244	3.2242	4.0073

APPENDIX 2I

Average Amplitude and Effectiveness Ratings for the Voice Messages

Table 2I1: Average Amplitude and Effectiveness Ratings for the Voice Messages.

Voice	Avg. dBA (peak values)	Avg. dBA (slow response)	Avg. Effect. Rating
Male	83.4	68.4	2.75
Female	82.5	67.5	2.58

APPENDIX 2J

Average Effectiveness Ratings for Peripheral Stimuli

Table 2J1: Average Effectiveness Ratings for Peripheral Stimuli

Peripheral Stimuli	Avg. Effect. Rating
Upper Seat Vibration	2.496
Lower Seat Vibration	2.789
Combination Seat Vibration	2.997
Steering Vibration	2.581
Brake Pulse	2.914

APPENDIX 2K
Average Effectiveness Ratings for Countermeasures

Table 2K1: Average Effectiveness Ratings for Countermeasures

Peripheral Stimuli	Avg. Effect. Rating
Upper Seat Vibration	1.540
Lower Seat Vibration	1.289
Combination Seat Vibration	1.662
A/O Task	1.580
Lane Minder	2.121
Peppermint Scent	2.165