

SIMULATOR TEST AND EVALUATION OF A DROWSY DRIVER DETECTION
SYSTEM AND REVISIONS TO DROWSINESS DETECTION ALGORITHMS

by

Mark Gustav Lewin

Thesis submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of
Master of Science
in
Industrial and Systems Engineering

APPROVED



Dr. Walter W. Wierwille, Chair



Dr. Robert J. Beaton



Dr. C. Patrick Koelling

September, 1996
Blacksburg, Virginia

Keywords: Driver Drowsiness, Driver Performance, Detection, Alarms, Test and
Evaluation, Algorithm Development

LD
5655
V855
1996
L494
c.2

SIMULATOR TEST AND EVALUATION OF A DROWSY DRIVER DETECTION SYSTEM AND REVISIONS TO DROWSINESS DETECTION ALGORITHMS

by

Mark Gustav Lewin

Committee Chairman: Dr. Walter W. Wierwille

Industrial and Systems Engineering

(ABSTRACT)

This study was undertaken to simulator test and evaluate a complete drowsy driver detection system. The goal of the study was to recommend optimal specifications for a system to be further studied in an actual vehicle. The system used a set of algorithms developed from previously collected data and a set of previously optimized advisory tones, advisory messages, alarm stimuli, and drowsiness countermeasures. Detection occurred if eye closure or lane excursion exceeded predetermined thresholds. Data were obtained from six sleep-deprived subjects who drove a motion base automobile simulator late at night. Each subject was trained in carefully observing lane boundaries, using a device which sounded an alarm if lane boundaries were exceeded.

The performance aspect of the system dominated the detection process. None of the algorithms tracked well with the measures they were designed to estimate; correlations were much lower than expected. The algorithms relied heavily on the positioning of the vehicle relative to the lane.

In a subsequent experiment, large lane exceedances were curtailed using a lane departure warning system with seat vibration. Twelve sleep-deprived subjects drove the simulator late at night, carefully observing lane boundaries. A variety of measures were collected and revised algorithms were developed. Correlations were low, but classification accuracy was higher than expected and comparable to previous algorithms.

The drowsiness phase of a system to be tested in full scale should use new three-minute average algorithms, a “drowsiness or performance” criterion for detection, and a supplemental lane departure warning system.

ACKNOWLEDGMENTS

I would like to begin by thanking my committee chair, Dr. Walter W. Wierwille. His advice, guidance, and insight in matters related to (and unrelated to) this project have been absolutely invaluable. It has a pleasure and an honor to work with such a knowledgeable, experienced individual, and I am confident that the lessons I have learned from Dr. Wierwille will serve me well both on a professional level and a personal level; I don't think I could have asked for a better major advisor. I would also like to thank the other members of my committee, Dr. Robert J. Beaton and Dr. C. Patrick Koelling.

I would like to thank my lab colleagues Terry Fairbanks, Vicki Lewis Neale, and David Martin for the wealth of support they have given me throughout this research. Terry was with me through over 20 late-night experimental sessions performing the rather menial task of eye-tracking for almost three hours at a time; I don't know how he survived and that deserves to be recognized. In addition, Terry provided a great deal of help with data analysis. He has also been a great friend and source of advice and support in a variety of situations. Vicki and Dave both spent many hours of their time sleep-depriving subjects, and provided a great deal of advice and support.

Special thanks go to my parents, Gus and Ann Lewin, my brother, Michael Lewin, and my grandmothers, Hope Tiemeyer and Vera Lewin. Without their love, support, and encouragement not only during my years in graduate school but all my life, much of what I have accomplished (and hope to accomplish) would not have been possible. I would also like to thank the many good friends not already mentioned that I have made during my six-plus years at Virginia Tech for their support and companionship. As I prepare to (finally!) leave the collegiate life (for now), I can look back on an experience that would not have been nearly as enjoyable (or bearable) without you.

This research was sponsored under a cooperative agreement (DTNH 22-91-Y-07266) funded by the National Highway Traffic Safety Administration. Ron Knipling and Jeff Woods served as contract technical representatives.

TABLE OF CONTENTS

ABSTRACT.....	ii
ACKNOWLEDGMENTS	iv
TABLE OF CONTENTS.....	vi
LIST OF TABLES.....	xi
LIST OF FIGURES	xiv
INTRODUCTION	1
LITERATURE REVIEW	4
Magnitude of the Problem	4
Existing Drowsiness Detection Systems.....	6
Physiological Measures of Drowsiness	9
Eyelid Closure.....	9
Eye Movement.....	10
Muscle Activity.....	10
Brain Wave Activity	10
Skin Potential	11
Heart Rate Variability	11
Subjective Rating of Drowsiness	12
Development of Operational Definition	12
Performance Measures as Indicators of Drowsiness.....	14
Lateral Control Measures.....	14
Driver Response Measures.....	18
Longitudinal Control Measures	19
Algorithm Development	20
Algorithm Validation	23
Effects of Cruise Control, Secondary Task, and Velocity-Related Measures	24
Further Algorithm Refinement and Optimization Studies.....	26
Effects Of Higher-Order Algorithms	26
Elimination Of Outliers From A/O Based Algorithms	26
False Alarm Rates During Alert-Driver Segments	27
Effects Of Time-On-Task Variables	27
Comparison of Real-Time Accuracy With Advance Intervals	28

Time Trend Analysis.....	29
Elimination Of Outliers From PERCLOS Data.....	29
Appropriate Drowsiness Warnings for Motor Vehicles	30
Auditory Displays	31
Tactile Displays	32
Visual Displays	32
Termination of Warnings.....	33
Countermeasures to Drowsiness.....	33
Simple Countermeasures	33
Chemical Countermeasures	34
Mechanical and Electronic Countermeasures	35
Optimization of Advisory and Alarm Stimuli	37
Initial Advisory Tones	40
Voice Messages	41
Auditory Alarm Stimuli	41
Peripheral Stimuli	42
Countermeasures.....	42
PHASE ONE -- REMOVAL OF BASELINING TECHNIQUES FROM DROWSY DRIVER DETECTION ALGORITHMS.....	44
Overview.....	44
METHOD	54
RESULTS AND DISCUSSION	57
PHASE TWO -- SIMULATOR TEST AND EVALUATION OF A COMPLETE DRIVER STATUS/PERFORMANCE MONITORING ADVISORY-ALARM-COUNTERMEASURES SYSTEM.....	64
Research Objectives.....	64
System Description	65
Stage One: Initial Detection of Drowsiness Status or Reduced Performance Level	65
Stage Two: Re-Alerting the Driver.....	69
Stage Three: Maintaining Alertness.....	70
METHOD	72
Subjects.....	72
Apparatus	73
Simulator.....	73

Eye Closure Monitoring Equipment	74
Drowsiness Detection (Stage 1) Equipment	75
Driver Re-Alerting (Stage 2) Equipment	77
Alertness Maintenance (Stage 3) Equipment	78
Experimental Design	78
Procedure	82
Data Analysis Overview	85
RESULTS	89
Qualitative Observations	89
Detection	89
Advisory, Alarm, and Countermeasures	91
Quantitative Analyses	91
Algorithm Performance	91
Consistency of Data	91
Effects of Alarms and Countermeasures on Status and Performance	92
DISCUSSION	95
Qualitative Observations	95
Detection	95
Advisory, Alarm, and Countermeasures	95
Quantitative Analyses	96
Algorithm Performance	96
Consistency of Data	97
Effects of Alarms and Countermeasures on Status and Performance	97
Conclusions	98
Recommendations	99
PHASE THREE – DEVELOPMENT OF REVISED ALGORITHMS FOR DETECTION OF DRIVER DROWSINESS	101
Research Objectives	101
METHOD	103
Subjects	103
Apparatus	104
Simulator	104
Eye Closure Monitoring Equipment	105
Subjective Drowsiness Rating Equipment	105

Lane Enforcement Equipment	106
Data Collection Equipment.....	107
Data Analysis Equipment	108
Experimental Design.....	108
Procedure	113
Data Analysis Overview	116
RESULTS AND DISCUSSION	122
Correlation Analyses.....	122
Distribution Analyses.....	130
Multiple Regression.....	131
Driver Status (Drowsiness) Algorithms.....	131
Driver Performance Algorithms.....	134
Classification Matrices.....	136
Driver Status (Drowsiness).....	136
Driver Status <u>or</u> Performance.....	138
Conclusions.....	142
Recommendations.....	144
REFERENCES	148
APPENDIX A: Regression Summaries and Classification Matrices for Unbaselined LANEX Algorithms Used In Test and Evaluation Phase (Fairbanks, Lewin, and Wierwille, 1995)	157
APPENDIX B: Logic Diagram Representation of Test and Evaluation System Specifications.....	160
APPENDIX C: Potential Subject Screening Questionnaire	169
APPENDIX D: Subject and Experimenter Rating Scales Used In Test and Evaluation Experiment.....	172
APPENDIX E: Information Sheet Regarding Procedures for Experiment and Participant Informed Consent Form.....	175
APPENDIX F: Line Charts Comparing Measured PERCLOS with ePERCLOS Algorithm Outputs	182
APPENDIX G: Line Charts Comparing Measured LANEX with eLANEX Algorithm Outputs	198
APPENDIX H: Analyses of Variance for Evaluation of Effects of Alarms and Countermeasures on Driver Status and Performance	206

APPENDIX I: Distribution Analyses for EYEMEAS, PERCLOS, DRVDROW, LANDEV, LANEX, and LNMNSQ in One-Minute Interval Data Set	212
APPENDIX J: Distribution Analyses for EYEMEAS, PERCLOS, DRVDROW, LANDEV, LANEX, and LNMNSQ in Three-Minute Interval Data Set.....	219
APPENDIX K: Descriptive Statistics for Individual Measures in Three-Minute and Six-Minute Interval Data Sets	226
APPENDIX L: Regression Summaries, Unconditional Classification Matrices, and Conditional “OR” Classification Matrices for Driver Status Algorithms Developed Using Three-Minute Interval Data Set	229
APPENDIX M: Regression Summaries, Unconditional Classification Matrices, and Conditional “OR” Classification Matrices for Driver Status Algorithms Developed Using Six-Minute Interval Data Set	251
APPENDIX N: Regression Summaries and Classification Matrices for Driver Performance Algorithms Developed Using Three-Minute and Six-Minute Interval Data Sets.	271
VITA	276

LIST OF TABLES

1: History of Detections with Subjective Ratings and Experimenter Ratings	90
2: Summary of r Values Obtained in Correlation Analyses	92
3: Summary of Data Consistency Analyses Between Algorithm Development Experiment and Test and Evaluation Experiment.....	93
4: History of Vibration Activations (data truncated for each subject after fourth vibration).....	117
5: Sets of Measures Used in Multiple Regression Analyses for Each Dependent (Definitional) Variable.....	120
6a: Correlation Coefficients Between Measures for 1-Minute Interval Data Set	123
6b: Correlation Coefficients Between Measures for 2-Minute Interval Data Set.....	124
6c: Correlation Coefficients Between Measures for 3-Minute Interval Data Set	125
6d: Correlation Coefficients Between Measures for 6-Minute Interval Data Set.....	126
7a: Correlation Coefficients Between SLEEPER1, SLEEPER2 and SLEEPER3, and All Other Measures for 3-Minute Interval Data Set.....	127
7b: Correlation Coefficients Between SLEEPER1, SLEEPER2 and SLEEPER3, and All Other Measures for 6-Minute Interval Data Set.....	128
8a: Summary Table of Multiple Regression Analyses Showing R Values for 3- Minute Interval Data Set and Dependent Driver Status Measures	133
8b: Summary Table of Multiple Regression Analyses Showing R Values for 6- Minute Interval Data Set and Dependent Driver Status Measures	133
9a: Summary Table of Multiple Regression Analyses Showing R Values for 3- Minute Interval Data Set and Dependent Driver Performance Measures.....	135
9b: Summary Table of Multiple Regression Analyses Showing R Values for 6- Minute Interval Data Set and Dependent Driver Performance Measures.....	135
10a: Summary Table of Unconditional Classification Matrix Analyses Showing Apparent Accuracy Rates for Driver Status Algorithms developed from 3-Minute Interval Data Set	137
10b: Summary Table of Unconditional Classification Matrix Analyses Showing Apparent Accuracy Rates for Driver Status Algorithms developed from 6-Minute Interval Data Set	138

11a: Summary Table of Conditional Classification Matrix Analyses Showing Apparent Accuracy Rates for Driver Status Algorithms developed from 3-Minute Interval Data Set -OR- LANEX/eLANEX	140
11b: Summary Table of Conditional Classification Matrix Analyses Showing Apparent Accuracy Rates for Driver Status Algorithms developed from 3-Minute Interval Data Set -OR- LNMNSQ/eLNMNSQ.....	140
11c: Summary Table of Conditional Classification Matrix Analyses Showing Apparent Accuracy Rates for Driver Status Algorithms developed from 6-Minute Interval Data Set -OR- LANEX/eLANEX	141
11d: Summary Table of Conditional Classification Matrix Analyses Showing Apparent Accuracy Rates for Driver Status Algorithms developed from 6-Minute Interval Data Set -OR- LNMNSQ/eLNMNSQ.....	141
H1: ANOVA Summary Table for Countermeasure Type After Detection (Dependent Variable: PERCLOS)	207
H2: ANOVA Summary Table for Countermeasure Type After Detection (Dependent Variable: LANEX)	207
H3: ANOVA Summary Table for Countermeasure Type After Detection (Dependent Variable: LANDEV)	208
H4: ANOVA Summary Table for Test and Evaluation Experiment vs. Development Experiment, Drowsy Condition, No Countermeasure Operating (Dependent Variable: PERCLOS)	208
H5: ANOVA Summary Table for Test and Evaluation Experiment vs. Development Experiment, Drowsy Condition, No Countermeasure Operating (Dependent Variable: LANEX)	209
H6: ANOVA Summary Table for Test and Evaluation Experiment vs. Development Experiment, Drowsy Condition, No Countermeasure Operating (Dependent Variable: LANDEV)	209
H7: ANOVA Summary Table for Test and Evaluation Experiment vs. Development Experiment, Drowsy Condition, A/O Task Operating (Dependent Variable: PERCLOS).....	210
H8: ANOVA Summary Table for Test and Evaluation Experiment vs. Development Experiment, Drowsy Condition, A/O Task Operating (Dependent Variable: LANEX).....	210
H9: ANOVA Summary Table for Test and Evaluation Experiment vs. Development Experiment, Drowsy Condition, A/O Task Operating (Dependent Variable: LANDEV).....	211

K1: Descriptive Statistics for Individual Measures for 3-Minute Interval Data Set
(474 Observations).....227

K2: Descriptive Statistics for Individual Measures for 6-Minute Interval Data Set
(238 Observations).....228

LIST OF FIGURES

1: Baselined Regression Summary and Classification Matrices Showing Accuracy of Algorithm D4a When Applied to Original Data and New Data (Wierwille et al., 1994)	48
2: Baselined Regression Summary and Classification Matrices Showing Accuracy of Algorithm F4a When Applied to Original Data and New Data (Wierwille et al., 1994)	49
3: Baselined Regression Summary and Classification Matrices Showing Accuracy of Algorithm J4a When Applied to Original Data and New Data (Wierwille et al., 1994)	50
4: Baselined Regression Summary and Classification Matrices Showing Accuracy of Algorithm L3a When Applied to Original Data and New Data (Wierwille et al., 1994)	51
5: Unbaselined Regression Summary and Classification Matrices Showing Accuracy of Algorithm D4a-N When Applied to Original Data and New Data	59
6: Unbaselined Regression Summary and Classification Matrices Showing Accuracy of Algorithm F4a-N When Applied to Original Data and New Data	60
7: Unbaselined Regression Summary and Classification Matrices Showing Accuracy of Algorithm J4a-N When Applied to Original Data and New Data	61
8: Unbaselined Regression Summary and Classification Matrices Showing Accuracy of Algorithm L3a-N When Applied to Original Data and New Data	62
9: Comparison of R Values and Apparent Accuracy Rates For Baselined and Unbaselined Algorithms When Applied to Original Data and New Data	63
A1: Regression Summary and Classification Matrix Showing Accuracy of Algorithm LDV-1 When Applied to Combined Data Set (24 Subjects)	158
A2: Regression Summary and Classification Matrix Showing Accuracy of Algorithm LDV-4 When Applied to Combined Data Set (24 Subjects)	159
B1: Logic Diagram for Detection System Computational Procedure	162
B2: Logic Diagram for Detection System Computational Procedure	163
B3: Logic Diagram for Detection System Computational Procedure	164
B4: Procedure for Calculation of Moving Averages	166
D1: Subjective Rating Scale Used By Subject At Each Detection During Test and Evaluation Experiment	173

D2: Subjective Rating Scale Used By Experimenter At Each Detection During Test and Evaluation Experiment	174
F1: All Subjects: PERCLOS vs. ePERCLOS (straight 6-minute averages, minutes following detection removed)	183
F2: All Subjects: Measured PERCLOS vs. Algorithm F4a-N (straight 6-minute averages, minutes following detection removed)	184
F3: All Subjects: Measured PERCLOS vs. Algorithm D4a-N (straight 6-minute averages, minutes following detection removed)	185
F4: Subject #4: PERCLOS vs. ePERCLOS (straight 6-minute averages, minutes following detection removed)	186
F5: Subject #4: Measured PERCLOS vs. Algorithms F4a-N and D4a-N (straight 6-minute averages, minutes following detection removed)	187
F6: Subject #5: PERCLOS vs. ePERCLOS (straight 6-minute averages, minutes following detection removed)	188
F7: Subject #5: Measured PERCLOS vs. Algorithms F4a-N and D4a-N (straight 6-minute averages, minutes following detection removed)	189
F8: Subject #6: PERCLOS vs. ePERCLOS (straight 6-minute averages, minutes following detection removed)	190
F9: Subject #6: Measured PERCLOS vs. Algorithms F4a-N and D4a-N (straight 6-minute averages, minutes following detection removed)	191
F10: Subject #7: PERCLOS vs. ePERCLOS (straight 6-minute averages, minutes following detection removed)	192
F11: Subject #7: Measured PERCLOS vs. Algorithms F4a-N and D4a-N (straight 6-minute averages, minutes following detection removed)	193
F12: Subject #8: PERCLOS vs. ePERCLOS (straight 6-minute averages, minutes following detection removed)	194
F13: Subject #8: Measured PERCLOS vs. Algorithms F4a-N and D4a-N (straight 6-minute averages, minutes following detection removed)	195
F14: Subject #9: PERCLOS vs. ePERCLOS (straight 6-minute averages, minutes following detection removed)	196
F15: Subject #9: Measured PERCLOS vs. Algorithms F4a-N and D4a-N (straight 6-minute averages, minutes following detection removed)	197
G1: All Subjects: LANEX vs. eLANEX (straight 6-minute averages, minutes following detection removed)	199

G2: Subject #4: LANEX vs. eLANEX (straight 6-minute averages, minutes following detection removed)	200
G3: Subject #5: LANEX vs. eLANEX (straight 6-minute averages, minutes following detection removed)	201
G4: Subject #6: LANEX vs. eLANEX (straight 6-minute averages, minutes following detection removed)	202
G5: Subject #7: LANEX vs. eLANEX (straight 6-minute averages, minutes following detection removed)	203
G6: Subject #8: LANEX vs. eLANEX (straight 6-minute averages, minutes following detection removed)	204
G7: Subject #9: LANEX vs. eLANEX (straight 6-minute averages, minutes following detection removed)	205
I1: EYEMEAS 1-Minute Histograms	213
I2: PERCLOS 1-Minute Histograms	214
I3: DRVDROW 1-Minute Histograms	215
I4: LANDEV 1-Minute Histograms	216
I5: LANEX 1-Minute Histograms	217
I6: LNMNSQ 1-Minute Histograms	218
J1: EYEMEAS 3-Minute Histograms	220
J2: PERCLOS 3-Minute Histograms	221
J3: DRVDROW 3-Minute Histograms	222
J4: LANDEV 3-Minute Histograms	223
J5: LANEX 3-Minute Histograms	224
J6: LNMNSQ 3-Minute Histograms	225
L1: Regression Summary for Algorithm D4e-3	230
L2: Regression Summary for Algorithm D2e-3	231
L3: Regression Summary and Classification Matrices for Algorithm D6e-3	232
L4: Regression Summary for Algorithm SLEEPER1-D-3	233
L5: Regression Summary for Algorithm SLEEPER2-D-3	234
L6: Regression Summary for Algorithm SLEEPER3-D-3	235
L7: Regression Summary for Algorithm E4e-3	236
L8: Regression Summary for Algorithm E2e-3	237

L9: Regression Summary and Classification Matrices for Algorithm E6e-3	238
L10: Regression Summary for Algorithm SLEEPER1-E-3	239
L11: Regression Summary for Algorithm SLEEPER2-E-3	240
L12: Regression Summary and Classification Matrices for Algorithm SLEEPER3-E-3.....	241
L13: Regression Summary and Classification Matrices for Algorithm F4e-3	242
L14: Regression Summary and Classification Matrices for Algorithm F2e-3	243
L15: Regression Summary and Classification Matrices for Algorithm F6e-3	244
L16: Regression Summary and Classification Matrices for Algorithm SLEEPER1-F-3.....	245
L17: Regression Summary and Classification Matrices for Algorithm SLEEPER2-F-3.....	246
L18: Regression Summary and Classification Matrices for Algorithm SLEEPER3-F-3.....	247
L19: Regression Summary and Classification Matrices for Algorithm G4e-3.....	248
L20: Regression Summary and Classification Matrices for Algorithm G2e-3.....	249
L21: Regression Summary and Classification Matrices for Algorithm SLEEPER3-G-3	250
M1: Regression Summary for Algorithm D4e-6	252
M2: Regression Summary for Algorithm D2e-6	253
M3: Regression Summary and Classification Matrices for Algorithm D6e-6.....	254
M4: Regression Summary for Algorithm SLEEPER1-D-6.....	255
M5: Regression Summary for Algorithm SLEEPER2-D-6.....	256
M6: Regression Summary for Algorithm SLEEPER3-D-6.....	257
M7: Regression Summary and Classification Matrices for Algorithm E4e-6	258
M8: Regression Summary for Algorithm E2e-6.....	259
M9: Regression Summary and Classification Matrices for Algorithm E6e-6	260
M10: Regression Summary for Algorithm SLEEPER1-E-6	261
M11: Regression Summary for Algorithm SLEEPER2-E-6	262
M12: Regression Summary and Classification Matrices for Algorithm	

SLEEPER3-E-6.....263

M13: Regression Summary and Classification Matrices for Algorithm F4e-6 264

M14: Regression Summary and Classification Matrices for Algorithm F2e-6 265

M15: Regression Summary and Classification Matrices for Algorithm F6e-6 266

M16: Regression Summary and Classification Matrices for Algorithm

SLEEPER1-F-6.....267

M17: Regression Summary and Classification Matrices for Algorithm

SLEEPER2-F-6.....268

M18: Regression Summary and Classification Matrices for Algorithm

SLEEPER3-F-6.....269

M19: Regression Summary and Classification Matrices for Algorithm

SLEEPER2-G-6 270

N1: Regression Summary for Algorithm eLANEX-D-3 272

N2: Regression Summary for Algorithm eLNMNSQ-D-3.....272

N3: Regression Summary for Algorithm eLANEX-E-3.....273

N4: Regression Summary for Algorithm eLNMNSQ-D-3.....273

N5: Regression Summary for Algorithm eLANEX-D-6 274

N6: Regression Summary for Algorithm eLNMNSQ-D-6.....274

N7: Regression Summary for Algorithm eLANEX-E-6.....275

N8: Regression Summary for Algorithm eLNMNSQ-E-6 275

INTRODUCTION

A number of reports exist that document the contribution of driver drowsiness to highway accidents and fatalities. It has often been difficult to pinpoint drowsiness as the cause of an accident. However, continuing research concerning drowsy drivers is leaving little doubt that driver drowsiness is a major safety hazard on highways around the world.

Possibly adding to the problem is that cars are increasingly being designed with luxury and comfort in mind. Soundproofing and reduced vibration are viewed as increases in quality. Also, devices intended to reduce the workload on the driver, such as cruise control, have been and will continue to be increasingly implemented in cars. These factors may cause drivers to become mentally complacent and underloaded, decreasing the need or the desire to constantly monitor the driving environment and increasing the likelihood of drowsiness behind the wheel.

In recent years, many advances have been made in the implementation of devices in automobiles that increase the level of safety of their occupants. These devices are designed to mitigate the effects of accidents, if accidents happen. Few automotive safety devices are designed to help the driver avoid accidents, and virtually none addresses the hazards inherent in the driving task that cause the driver to have accidents. There is a definite need to recognize these hazards and design devices that address them. Conceivably, an effective device that monitors driver performance, detects driver drowsiness, and alerts the driver as to his or her impaired status could be constructed.

There are currently many devices on the market that claim to prevent drivers from becoming impaired due to drowsiness. However, none of these devices has been consistently supported by scientific research. In recognition of the need for an adequate detection device, research has been conducted at Virginia Polytechnic Institute and State

University through a cooperative agreement with the National Highway Traffic Safety Administration (NHTSA). The goal of the research has been to develop reliable algorithms for the detection of driver impairment due to drowsiness. This was done by developing operational definitions of drowsiness which were then correlated with driver performance measures that would be obtainable on-line in an actual vehicle. The resulting algorithms were refined and validated (Wierwille, Wreggit, Kirn, Ellsworth, and Fairbanks, 1994).

Further research has resulted in the development of a warning system for use with the algorithms. This system was designed to advise the driver of detection of a drowsiness condition, to re-alert the drowsy driver, and to maintain driver alertness long enough for the driver to find an acceptable place to pull off the road and refresh. An initial advisory tone, a voice message, various alarming stimuli, and various countermeasures were optimized in the study (Fairbanks, Fahey, and Wierwille, 1995).

The original goal of the present study was to integrate and simulator test the detection algorithms described in Wierwille et al. (1994) with the warning system described in Fairbanks, Fahey, and Wierwille (1995) to produce a fast-acting, on-line drowsy driver detection and alerting system. In the first phase of the study, some revision of the original algorithms was performed to remove baselining techniques, greatly simplifying on-line computation of the algorithm outputs. In the second phase of the study, the integrated detection system with the new, unbaselined algorithms underwent simulator test and evaluation. In addition, the effectiveness of a lane minder device, proposed in Fairbanks, Fahey, and Wierwille (1995), was further investigated.

The detection system effectively controlled lane-keeping behavior and helped to keep drivers awake. However, it did not perform well in detection of drowsiness. Therefore, a third phase of the study became necessary. The goal of the third phase was

to develop revised drowsiness detection algorithms in a simulator study. The study was performed under conditions in which drivers were required to perform the lane-keeping task in a manner more closely reflecting actual driving in an actual vehicle. The ultimate goal of the present study, taken as a whole, was to specify a drowsy driver detection, alarm, and countermeasure system (DDDACS) with a maximal probability of success in field tests.

LITERATURE REVIEW

Magnitude of the Problem

Driver drowsiness has been under investigation as a highway safety concern for a number of years. In 1966, Kearney reported that drowsy drivers were responsible for 50% of all fatalities on the Ohio Turnpike. In a 1973 study conducted by Tilley, Erwin, and Gianturco at Duke University, questionnaires were administered to 1500 licensed drivers at a North Carolina Department of Motor Vehicles office. It was found that 64% of these drivers had, at some point, become drowsy while driving. In addition, 7% of the drivers surveyed reported that they had gone to sleep for short periods while driving. Of those drivers who reported becoming drowsy while driving, 31.2% indicated that the drowsiness occurred before they were aware of it. Also of the drivers who reported becoming drowsy, 10% had been in one or more accidents and another 10% had nearly been in an accident due to drowsiness or sleeping at the wheel.

The drowsy driver problem is not limited to the United States; it is a worldwide problem. In 1980, the Kanagawa Prefectural Police conducted a survey near the Tokyo-Nagoya Expressway in Japan. The results of the survey showed that 75% of the drivers surveyed admitted having been sleepy while driving (Seko, 1984). Seko (1984) also reported a 1973 survey that cited driver drowsiness as the cause of 45% of all traffic accident fatalities on Japanese highways. In 1991, Planque, Chaput, Petit, Tarriere, and Chabanon reported that 26% of fatal accidents on motorways in France were caused by drowsy drivers. An analysis of 1992 crash data in New South Wales, Australia implicated fatigue in 6% of all crashes, 15% of fatal crashes, and 30% of rural fatal crashes (Fell, 1994).

A recent summary of automotive crash statistics (Knipling and Wang, 1995) strongly supports driver drowsiness as a major cause of single- and multiple-car accidents in the U.S. today. The summary was based on an analysis of General Estimates System (GES) data for the years 1989-1993. The following results were reported:

- Police report approximately 56,000 crashes per year (0.9% of all crashes) in which driver fatigue/drowsiness involvement is cited.
- Driver fatigue can be discerned as a causal factor in 79,000 to 103,000 crashes per year (1.2% to 1.6% of all crashes) based on further analysis of the GES data.
- Each year, 2.5% of all crashes in which incapacitating injuries occur are attributed to drowsy drivers.
- Each year, 3.2% of all fatal crashes are attributed to drowsy drivers.
- Actual involvement of driver drowsiness/fatigue is probably far greater due to under reporting.

An earlier summary of crash statistics compiled by the NHTSA Office of Crash Avoidance Research (OCAR) summarized crash characteristics (1991):

- The peak time for crashes is between 1:00 A.M. and 7:00 A.M., when over 40% of all crashes occur. A secondary peak occurs between 3:00 P.M. and 5:00 P.M. The lowest crash rate occurs between 9:00 A.M. and noon. Fatal crashes exhibit a similar time distribution.
- About 70% of all crashes occur outside urban areas.
- Over 50% of all crashes occur on non-divided highways.
- The first harmful events of crashes are distributed as follows:
 - 64%: Collision with fixed objects (trees, guardrails, signposts, culverts, etc.)
 - 17%: Collision with another motor vehicle in transport
 - 7%: Rollover

6%: Collision with a parked vehicle

- Most crashes occur on roadways where the speed limit is 55 to 65 miles per hour.

Existing Drowsiness Detection Systems

There have been a number of commercially available devices that have been developed to measure drowsiness and, in some cases, warn the drowsy driver of his or her impaired status. These devices generally fall into three categories: head nod detectors, eye closure monitors, and reaction time devices.

The typical head nod detector is an ear piece that contains a battery, an angular motion detector, and an alarm. When the wearer's head nods forward beyond a given angle, the angular motion detector activates the alarm. Head nod detectors have been shown to have numerous problems in several studies. In a 1990 study evaluating a number of fatigue detection devices, Haworth and Vulcan showed that Dozer's Alarm, a head nod detector marketed in Australia, was not a reliable detection device. It is possible that before a driver's head nods past the predetermined angle, the driver may have lost a great deal of alertness or even consciousness, or the vehicle may already have crashed (Hulbert, 1972). It is also possible that a head nod detector attached to the ear can give a driver a false sense of security against sleepiness, effectively creating a hazard (Thorpy and Ledereich, 1990). It is difficult to manufacture a head nod detector that all drivers can wear; Haworth and Vulcan found that Dozer's Alarm slipped off subjects who either had small ears or wore eyeglasses (1990).

An eye closure monitor known as Onguard has been developed in Israel. Mounted on a standard eyeglass frame, this battery-operated device consists of an infrared sensing unit, an electronic processor, and an alarm buzzer. It detects reductions in the amount of

light reflected when the eyelid covers the cornea of the eye and sounds an audible alarm when the eye is closed for more than one half second. This device seems to detect fatigue; however, Haworth and Vulcan (1990) noted that the infrared monitors easily deviated from their proper alignment due to eyeglass slippage on the nose, severely limiting their reliability.

A Danish company has developed a reaction time device called Roadguard. When the vehicle is put into high gear, an electronic circuit in the device is activated, engaging a timer that randomly stops at periods ranging from 4 to 14 seconds. When this timer stops, a red light on the instrument panel immediately comes on. If one of two touch contacts on the steering wheel is not pressed within three seconds to turn the light off, an alarm sounds. Haworth and Vulcan (1990) found this device to be the most reliable one they tested, but one with many drawbacks. The red light on the instrument panel often was hidden behind the steering wheel and had to be readjusted for different subjects based on height. They noted that drivers could easily ignore the light and press the buttons automatically, preventing the alarm from sounding even though the driver's eyes could be completely closed. Also, they questioned the three-second threshold for resetting the alarm; Lisper, Laurell, and Van Loon (1986) found that drivers falling asleep at the wheel did so for periods of 0.5 to 1.5 seconds.

Another reaction time monitor, the Alert-O-Matic, consisted of a series of three alarms of increasing severity (Hulbert, 1972). The driver must turn off a light on the dashboard that is illuminated at constant intervals of 60 seconds by lightly tapping the horn. If the driver does not tap the horn within five seconds, the device will activate the horn. If the driver still fails to tap the horn after three seconds, the device turns the ignition on and off for five seconds; if there is still no response from the driver after five seconds, the ignition shuts off completely. Along with the problems inherent in the

Roadguard device, the Alert-O-Matic has the disadvantage that a driver could quickly learn the pattern of responding to the light every minute without being fully alerted.

Two other drowsiness warning devices are the ALERTMASTER and the Button Steering Wheel Alarm. Both devices require the driver to exert constant pressure on a foot pedal or a button; when constant pressure is not maintained, an auditory alarm sounds. There are several flaws inherent in these types of devices. First, they rely on the assumption that drowsiness will result in a relaxation in pressure (Hulbert, 1972). They also ignore that fatigue can result from exerting constant pressure, or that body movements required in the driving task (e.g., steering maneuvers, adjustments on the instrument panel) may require removal of pressure from the pedal or button (Hulbert, 1972).

Though all the above commercially available devices claim to prevent drowsiness, the fact is that none of them has been proven by experimental research. They may detect some aspects of drowsiness, but not one has been shown to effectively alleviate drowsiness. Also, most of these devices and other existing devices overlook the body of existing research on the physical and psychological components of driver drowsiness, thereby overlooking much of the complexity of the design problem (Skipper, Wierwille, and Hardee, 1984). In addition, most of the devices are very intrusive to the driver. All are driver-based or require the driver to actively participate in the detection process in some manner. Public acceptance is of great importance in the implementation of drowsiness alerting devices; no device is beneficial if those who could benefit from it refuse to use it.

Physiological Measures of Drowsiness

Over the years, a large amount of research has been conducted toward developing an operational definition of drowsiness. There are six physiological measures that have been shown to be related to drowsiness: eyelid closure, eye movement, muscle activity, brain wave activity, skin potential, and heart rate variability. Wierwille, Wreggit, and Mitchell (1992) conducted a literature review examining these six measures. An abbreviated summary of their findings is presented here.

Eyelid Closure

Previous research by Erwin (1976) and Skipper, Wierwille, and Hardee (1984) has shown this measure to be a very reliable indicator of drowsiness. Erwin (1976) studied several physiological measures and found slow eyelid closure to be the best predictor of sleep onset and degraded task performance. Eye closures can be unintrusively monitored by a low-light level camera. An experimenter views the image on a video display and operates a potentiometer indicating the degree to which the driver's eyes are open. Two slow eye closure measures used by Ellsworth, Wreggit, and Wierwille (1993) and supported by research conducted by Wreggit, Kirn, and Wierwille (1993) are defined here:

- PERCLOS: The proportion of time the eyes are 80% to 100% closed over a one-minute interval.
- EYEMEAS: The mean square of the percentage eyelid closure signal sampled over a one-minute interval. (EYEMEAS is more heavily weighted as eye closure increases.)

These measures have been shown to be excellent indicators of drowsiness, but they ignore the possibility that drowsiness could occur without eyelid closures in some individuals. Therefore, other physiological measures of drowsiness have been examined.

Eye Movement

With electrooculography (EOG), which measures electrical potentials through electrodes in contact with the skin around the eyes, it is possible to track the deterioration of alertness (Planque et al., 1991). As a person becomes very drowsy, he or she begins to exhibit slow, rolling lateral eye movements as opposed to quick, voluntary eye movements. This change in movements can be detected by the EOG. However, a person must be very drowsy before these rolling eye movements occur; therefore, this indicator may be a poor predictor.

Muscle Activity

As a subject becomes drowsy, activity in the muscles decreases, especially in those muscles controlling facial expressions (Yabuta et al., 1985). Through electromyography (EMG), this change of muscle potential may be monitored. A study conducted in 1982 by Hauri showed a decrease in muscle potential of the chin as an individual neared the first stage of sleep. However, Erwin (1976) argued that muscle activity has poor predictive value because a period of sleep is possible before any significant change in muscle tone occurs.

Brain Wave Activity

Using the electroencephalogram (EEG), brain wave activity can be measured. However, it is not clear whether brain wave activity can be used as a predictive measure

of drowsiness. Erwin (1976) found a rapid shift in brain wave patterns upon eyelid closure, but found no reliable alteration in brain wave activity before eyelid closure. However, Planque et al. (1991) found evidence that sharp changes in the frequency of brain wave activity occurred as an individual progressed from alertness to extreme drowsiness and sleep.

Skin Potential

Skin potential level (SPL) may be measured by attaching electrodes to the skin that can detect the potential difference between the outermost layer of skin and the layer immediately beneath it. In a 1976 study, Erwin, Hartwell, Volow, and Alberti found a correlation between SPL and stages of arousal. The findings further suggest that a reduced SPL must occur for a person to fall asleep, but sleep onset does not necessarily follow a decrease in SPL. Also, baseline values of SPL exhibit considerable variation between and within subjects.

Heart Rate Variability

Heart rate may be measured in a relatively unintrusive manner through use of a plethysmograph, which attaches to the ear. A number of studies have found a correlation between heart rate interval variability and driver fatigue level (Wierwille and Muto, 1981; Sugarman and Cozad, 1972; Riemersma, Sanders, Wildervanck, and Gaillard, 1977). However, Riemersma et al. (1977) suggest that the subjects' adaptation to experimental conditions could have caused the changes in heart rate variability.

Subjective Rating of Drowsiness

Most studies of drowsiness involving subjective ratings have been based on ratings from the subjects themselves. However, Carroll, Blisewise, and Dement (1989) studied subjective rating of drowsiness by observers in several nursing home settings. The results of their study suggest that the use of observer ratings is a valid approach to studying drowsiness.

Ellsworth, Wreggit, and Wierwille (1993) further researched this finding by showing videotapes of sleep deprived subjects to observers. These observers were briefed on the Description of Drowsiness Continuum, a rating scale basing drowsiness on eyelid closures, mannerisms, and facial tone. The accuracy, bias, and consistency of the various observers were compared when rating drowsiness levels. The results showed good interrater reliability ($r = 0.81$) as well as good intrarater reliability ($r = 0.88$) and test-retest reliability ($r = 0.81$). This demonstrated consistency both between and within raters shows great promise for the average observer rating (AVEOBS) as a reliable indicator of drowsiness.

Development of Operational Definition

Development of an operational definition of drowsiness based solely upon a review of physiological indicators of drowsiness in the literature is nearly impossible. For nearly every study supporting a particular physiological measure to be included in such a definition, there is a study that disagrees with use of the same measure. Ellsworth, Wreggit, and Wierwille (1993) studied promising physiological indicators of drowsiness with the objective of resolving these conflicts and developing one or more new operational measures of drowsiness to predict potentially hazardous levels of impairment.

Four male subjects and four female subjects from the Blacksburg, Virginia community volunteered to participate in the study. Each participant was required to awaken by 7:00 A.M. and perform normal daily activities without napping. Each subject was picked up at 6:00 P.M., taken to dinner, and then brought to the laboratory. An experimenter was present in the laboratory to ensure that the subject stayed awake and did not ingest any stimulants. Each session began at midnight and lasted slightly over two hours.

Subjects were asked to complete two alternating cognitive tasks: a mathematical task (considered a high-level task) and a letter search task (considered a low-level task). During the experiment, a number of physiological measures were recorded, as well as performance measures for the cognitive tasks (response time, number of correct responses, number of incorrect responses, number of no responses).

The data were analyzed in two parts. First, correlation analyses were performed to determine which of the physiological measures varied reliably with impairment due to drowsiness. Then, linear multiple regression analyses were performed to find optimized combinations of variables that would best predict impairment resulting from drowsiness.

It was shown that performance impairment due to drowsiness was reliably detected with linear combinations of detection measures. The best definition of drowsiness obtained in the analyses included the physiological measures of eyelid closure, simple EEG, and simple heart rate. Eyelid closure received the heaviest weighting, followed by EEG and then heart rate. The final regression equation obtained is:

$$Y = 0.940 (\text{PERCLOS}) - 0.660 (\text{MNALPHA}) + 0.345 (\text{MNTHETA}) \\ - 0.257 (\text{MNBETA}) + 0.129 (\text{MNSQHRT})$$

- PERCLOS is the percentage of time the eyes were 80% closed to 100% closed.

- MNALPHA, MNTHETA, and MNBETA are the mean alpha, theta, and beta EEG amplitudes.
- MNSQHRT is the mean square of the heart rate signal.

The constants shown in the above equation are standardized regression beta weights.

Performance Measures as Indicators of Drowsiness

Although several physiological measures have been studied as indicators of drowsiness, the implementation of equipment to monitor any of them can be highly intrusive in the context of a driving task. As a result, performance measures have been under investigation as a usually unintrusive alternative; the measures are activities that occur normally while driving. Three main groups of performance measures that can be used for driver impairment prediction are lateral control measures, driver response measures, and longitudinal control measures. Wierwille et al. (1994) have included measures from each of these groups in development and validation of drowsy driver detection algorithms.

Lateral Control Measures

These measures are subdivided into lane-related measures, steering-related measures, and heading and lateral acceleration measures. Many components of the driving task are over-learned and easily taken for granted by an impaired driver; maintaining lane position is one of these components (Dingus, Hardee, and Wierwille, 1985). In addition, lane position errors have been shown to increase over extended periods of time (Mast, Jones, and Heimstra, 1966; Sussman, Sugarman, and Knight, 1971).

Wierwille et al. (1994) considered several lane-related measures in development of algorithms for driver drowsiness detection. All five are feasible for implementation in a simulator as well as in an actual vehicle. The measures are presented here:

- LANDEV: The standard deviation of lateral position relative to the lane, where lane position is measured in feet.
- LANVAR: The variance of the lateral position relative to the lane (equal to the square of LANDEV).
- LNMNSQ: The mean square of lane position in feet, where the “zero” position is defined as that position occurring when the vehicle is centered in the lane.
- LNRTDEV: The standard deviation of the time derivative of lane position (relative to the lane) in feet per second.
- LNRTVAR: The variance of the time derivative of lane position (equal to the square of LNRTDEV).
- LANEX: The proportion of time that any part of the vehicle exceeds either lane boundary.
- LNERRSQ: The mean square of the horizontal difference (in feet) between the outside edge of a vehicle and the lane edge when the vehicle exceeds the lane. When the vehicle does not exceed the lane, the contribution to the measure is zero.

Similar to lane position, steering behavior deteriorates as a driver becomes impaired due to drowsiness (Wierwille et al., 1992). Hulbert (1963) found that sleep-deprived subjects make fewer small steering reversals than rested drivers. Lane deviations begin to increase. When a driver detects a deviation, he or she attempts to correct the error with a large steering reversal. Thus, large steering reversals increase with drowsiness, and small

steering reversals decrease with driving time (Wierwille and Muto, 1981; Khardi and Vallet, 1994).

Several steering-related measures have been considered in development of drowsy driver detection algorithms (Wierwille et al., 1994). Again, these measures are suitable for implementation in either a simulator or an actual vehicle. The measures are presented here:

- NMRHOLD: The number of times the hold circuit output on the steering wheel exceeds a threshold value. The threshold value corresponds to holding the steering wheel still for 0.4 seconds or longer.
- THRSHLD: The proportion of time that the hold circuit on the steering wheel exceeds a threshold value. This proportion begins to increase after 0.4 seconds of hold and continues until the steering wheel is moved.
- STVELV: The variance of steering velocity, where velocity is measured in degrees per second.
- LGREV: The number of times that steering excursions exceed 15° after steering velocity passes through zero.
- MDREV: The number of times that steering excursions exceed 5° but do not exceed 15° after steering velocity passes through zero.
- SMREV: The number of times that steering excursions exceed 1° but do not exceed 5° after steering velocity passes through zero.
- STEXED: The proportion of time that steering velocity exceeds 125° per second.

In high-speed driving, heading must be monitored very frequently to ensure safe driving. At highway speeds, a deviation of only one degree can cause a car to cross a

lane boundary in less than two seconds. Therefore, a heading measuring apparatus could be a great benefit to a drowsiness detection system. Several heading-related measures have been considered in development of drowsy driver detection algorithms (Wierwille et al., 1994). They are described here.

- HPHDGDEV: The standard deviation of the high-pass heading signal, in degrees. The heading signal is passed through a high-pass single-pole filter with a corner frequency of 0.016 Hz.
- HPHDGVAR: The variance of the high-pass heading signal (equal to the square of HPHDGDEV).
- DSYAWDEV: The standard deviation of the display yaw signal in degrees. The display yaw signal is the angular difference between vehicle heading and instantaneous roadway tangent.
- DSYAWVAR: The variance of the display yaw signal in degrees (equal to the square of DSYAWDEV).

In conjunction with heading-related measures, measures related to lateral acceleration have also been considered in algorithm development (Wierwille et al., 1994). These measures are described here:

- ACCDEV: The standard deviation of the smoothed output of a simulated lateral accelerometer, where the output is first converted to feet per second-squared. (Smoothing is accomplished with a single-pole low-pass filter with a corner frequency of 7.25 Hz.)
- ACCVAR: The variance of the smoothed output of the accelerometer (equal to the square of ACCDEV).

- INTACDEV: The standard deviation of the lateral velocity of the vehicle. (The signal is obtained by passing the smoothed accelerometer signal through an additional single-pole low-pass filter with a corner frequency of 0.004 Hz.)
- INTACVAR: The variance of the lateral velocity of the vehicle (equal to the square of INTACDEV).
- ACXEED: The proportion of time that the magnitude of lateral acceleration exceeded a threshold of 0.3 g (9.66 ft/second²).

Driver Response Measures

Two types of driver response measures that have been studied are driving task response measures and subsidiary task response measures. Driving task response measures are not likely to interfere with normal driving performance because the events for which responses are measured are normally part of the driving task. Research has shown reaction time to increase with sleep loss and time on task (Ryder, Malin, and Kinsley, 1981). Thus, driving task response measures could provide reliable monitoring of performance decrement due to drowsiness (Hardee, Dingus, and Wierwille, 1985). However, those stimuli associated with driving task response measures, such as lead cars and roadside obstacles, are not always present in the driving environment (Wierwille et al., 1992).

Hardee et al. (1985) studied subsidiary task response measures by exposing sleep-deprived subjects to various types of stimuli while they were driving a simulator. The number of correct responses on this subsidiary task in a three-minute period (NUMHIT) was recorded. NUMHIT was found to have potential as a second-level drowsiness detector.

In development of driver drowsiness detection algorithms, Wierwille et al. (1994) investigated a subsidiary task known as the A/O task. Every fifteen seconds, subjects were presented with a previously recorded spoken word and asked to indicate whether the word contained the letters A or O. Three response measures related to this task are presented here:

- AOTIME: Mean response time to a correct response. Incorrect responses and no-responses are specified as 12 seconds.
- NMWRONG: Mean number of incorrect responses. Instances in which there is no response are not included in this measure.
- NUMNR: Mean number of stimuli for which there is no response.

Longitudinal Control Measures

Longitudinal measures include vehicle following-related measures and measures related to speed, acceleration, and braking behavior. These measures have given mixed results at best in the literature; furthermore, they would be difficult to obtain in an actual vehicle.

A study by Safford and Rockwell (1967) found that accelerator pedal reversals were correlated with time over a 24-hour driving task. However, a literature review conducted by Hardee et al. (1985) found no evidence supporting a correlation between accelerator-related measures and drowsiness. Measures related to braking or deceleration have not been found to change significantly as a function of eight or twelve hours on the driving task (Brown, 1965; 1966; Brown, Simmons, and Tickner, 1967) or as a function of 29 hours of sleep deprivation (Huntley and Centybear, 1974).

During night driving, speed variability has been found to increase significantly (Riemersma et al., 1977). However, several studies found no significant variability in

speed over time during driving tasks (Brown, 1965; 1966; Brown, Simmons, and Tickner, 1967; Safford and Rockwell, 1967).

In refinement of previously developed drowsiness detection algorithms, Wierwille et al. (1994) considered three measures related to acceleration and speed. The measures are described here:

- FVELSD: The standard deviation of the forward velocity of a vehicle.
- FACCSD: The standard deviation of the forward acceleration of a vehicle.
- PEDDEV: The standard deviation of accelerator pedal position.

In a simulator study, Wierwille and Muto (1981) investigated vehicle-following behavior. Reaction times increased significantly after subjects drove the simulator for 30, 60, and 150 minutes compared to baseline runs. Unfortunately, problems arise in implementing this measure because a lead car is not always present in a real driving situation. Also, it may be rather difficult to equip a car with a device capable of measuring this behavior.

Algorithm Development

After completion of the Ellsworth, Wreggit, and Wierwille (1993) study, Wreggit, Kirn, and Wierwille (1993) conducted the second phase of the research project. This study considered the physiological measures identified as reliable indicators of drowsiness in the first study, along with the performance measures supported in the literature as potential indicators of drowsiness. The purpose of this second phase was to develop algorithms and procedures applicable to driver drowsiness detection. Twelve residents of the Blacksburg, Virginia area ranging from 18 to 45 years of age volunteered as subjects. As in the first phase of the experiment, subjects were instructed

to awaken by 7:00 A.M. and carry on normal daily activities without taking any naps. An experimenter picked up each subject at 6:00 P.M., took him or her to dinner, and observed him or her in the laboratory to ensure that no napping or ingestion of stimulants took place.

Shortly after 12:00 A.M., each subject began a 150-minute period of driving the Virginia Tech driving simulator. Previous research showed 150 minutes to be the most realistic scenario for fatigued drivers (Wierwille et al., 1992; OCAR, 1991). The subject was instructed to maintain a constant speed of 60 miles per hour and to stay within the boundaries of the right-hand lane of the simulated two-lane roadway. Four of the subjects were also presented with the A/O task. Two labeled buttons on the cross member of the steering wheel ("YES" and "NO") were used by the subject to indicate whether each presented word contained an A or an O.

Two microcomputers with analog-to-digital converter interfaces were employed to collect data and perform on-line computations. Physiological data were collected and used by one of the computers to calculate heart rate, mean brain wave amplitudes, and the proportion of time that the subject's eyes were 80% to 100% closed. Performance measures, including lateral control measures and A/O task responses, were also collected.

Five physiological measures were taken as dependent measures and were analyzed for predictive power against thirty independent, performance-based measures. The following dependent measures were used:

- AVEOBS: Average drowsiness rating of three observers for each one-minute interval. Data for this measure consisted of videotapes of the subjects' faces, viewed and rated after the completion of all experimental sessions.
- EYEMEAS: Mean square of the percentage of the subjects' eye closure.
- PERCLOS: Proportion of time that a subject's eyes were 80% to 100% closed.

- NEWDEF: Operational definition of drowsiness developed by Ellsworth, Wreggit, and Wierwille (1993):

$$Y = 18.45722 (\text{PERCLOS}) - 0.01569 (\text{MNALPHA}) + 0.020173 (\text{MNTHETA}) - 0.00549 (\text{MNBETA}) + 0.000698 (\text{MNSQHRT}).$$

(The constants shown here are B weights instead of the beta weights of the definition described earlier.)

- MASTER: Sum of the standardized values of AVEOBS, EYEMEAS, NEWDEF, and PERCLOS.

Several categories of independent measures were analyzed for all subjects, including steering-related measures, lane-related measures, lateral accelerometer-related measures, and heading-related measures. The A/O task responses were also analyzed for the four subjects who completed that task. Each data point consisted of a one-minute average of data. In preparation for statistical analysis, the first two minutes of data were deleted for all measures; during this time, subjects tended to show inconsistent driving behavior due to “settling into” the driving task. The next ten minutes of data were averaged and subtracted from all subsequent one-minute segments; this served to baseline the independent measures. Subsequent data were averaged across six-minute intervals; previous research had shown these to have high correlation values (Ellsworth, Wreggit, and Wierwille, 1993).

After the data were manipulated, multiple regression and discriminant analysis procedures were used for analysis. Multiple regression analyses were used to minimize the difference between a subject’s predicted level and the actual criterion level reported by the monitoring equipment and the computers. Drowsiness levels were classified as “Awake,” “Questionable,” or “Drowsy”; correct classification was maximized using discriminant analyses.

All five dependent measures were found to be reasonably predictable. The number of large classification errors (“Awake” misclassified as “Drowsy” or vice versa) was relatively small, ranging as low as one or two per 100 cases. Also, the difference between discriminant analysis results and multiple regression results was found to be small and even nonexistent in some cases.

Even with this large degree of algorithm accuracy, Wreggit, Kirn, and Wierwille (1993) note that the models will produce some false alarms, with likely error rates of two to three percent. A two-stage detection system was recommended to overcome this problem. In the first stage, probable drowsiness would be detected based on driver performance measures. Once a preset threshold was passed and this detection took place, the second stage would further discriminate drowsiness based on subsidiary task response measures.

Algorithm Validation

After developing the previously mentioned detection algorithms, Wreggit, Kirn, and Wierwille (1994) tested their validity by applying a subset of them to a new set of data collected from a new group of subjects. Algorithms chosen for validation were chosen from two classes: steering/lateral accelerometer-based algorithms and steering/lateral accelerometer/lane-based algorithms. These algorithms were chosen for two reasons: they contained reliable and unintrusively obtainable measures, and they allowed for a step-up, step-down procedure. The step-up, step-down procedure would be important in an actual driving situation; many roads do not have lane lines at all and many others have lane lines that can not be accurately picked up by a vehicle-mounted sensor. In either of these cases, the detection system can “step-down” to the algorithm employing only steering and lateral accelerometer measures. When lane-related signals become available,

the system can “step-up” to the algorithm that includes lane-related measures as well as steering and lateral accelerometer measures.

The validation experiment was conducted in a manner similar to the detection experiment, except that certain other measures were incorporated into the apparatus and the design so that additional analyses could also be performed on the data (see next section). Similar data manipulation techniques and statistical analyses were used. Results from the validation of the algorithms using driver-vehicle performance measures showed no significant loss in drowsiness prediction accuracy when they were applied to the new data (Wreggit, Kirn, and Wierwille, 1994).

Effects of Cruise Control, Secondary Task, and Velocity-Related Measures

The validation experiment was a multipurpose experiment, designed not only to validate the detection algorithms developed by Wreggit, Kirn, and Wierwille (1993), but also to address additional research issues. The additional issues were researched by Kirn, Wreggit, and Wierwille (1994) using the same data collected in the Wreggit, Kirn, and Wierwille (1994) validation experiment.

Longitudinal control measures were not collected in the Wreggit, Kirn, and Wierwille (1993) algorithm development experiment, but it was observed during that research that drivers tended to vary their speed when they became drowsy. As previously noted, past research linking longitudinal control measures to driver fatigue has been inconclusive. Therefore, three velocity-related measures were implemented in the validation experiment so that their contribution to drowsiness detection and algorithm accuracy could be tested. Validation experiment data were used to develop new algorithms both with and without velocity-related measures, so that a direct assessment of any gain in accuracy afforded by the velocity-related measures could be made.

Further investigation was also carried out to determine if the A/O task had an alerting effect on the driver. The task was originally designed to provide an independent assessment of the level of drowsiness in a two-stage detection system. Questions remained as to whether the A/O task could serve the dual purposes of a detection aid and an alertness aid. In the validation experiment, all subjects performed the A/O task for certain segments of their driving time.

In the validation experiment, all subjects were also subjected to a cruise control condition for certain segments of their driving time. The cruise control condition maintained the speed of the car at 60 miles per hour. Direct comparisons between cruise control and non-cruise control conditions could be made, and the question of whether or not cruise control usage contributed to drowsiness could be answered.

The data analysis showed that velocity-related measures contributed only slightly to the accuracy of detection models, with an average gain in correlation of only 0.010 (Kim, Wreggit, and Wierwille, 1994). It is likely that variances in speed for alert drivers and for drowsy drivers were too similar in nature to provide a reliable indicator of drowsiness. However, velocity-related measures became much better indicators when the subject was not given a secondary task. This would seem to suggest that the secondary A/O task kept subjects more stimulated and thus helped them to monitor and control their speed; perhaps the subjects were under-loaded without the secondary task. However, no significant effect of the A/O task on drowsiness was found. Kim, Wreggit, and Wierwille (1994) concluded that lack of a secondary task may induce inattention rather than drowsiness and that further research concerning a secondary task is necessary. Finally, Kim, Wreggit, and Wierwille (1994) found no significant effect of cruise control on drowsiness.

Further Algorithm Refinement and Optimization Studies

Several attempts at refining and optimizing the drowsiness detection algorithms developed and validated by Wreggit, Kirn, and Wierwille (1993; 1994) were made. All of these studies involved re-analysis of the data collected in the algorithm development and validation experiments (Wreggit, Kirn, and Wierwille, 1993; 1994).

Effects Of Higher-Order Algorithms

Fairbanks and Wierwille (1994) investigated the potential for improvement in algorithm accuracy with the addition of second-order terms to the detection algorithms. On the basis of three groups of independent measures selected from the development experiment, first order, partial second order, and full second order multiple regression analyses were performed. The dependent measure chosen for the analysis was PERCLOS, one of the most reliable definitional measures between subjects (Wreggit, Kirn, and Wierwille, 1993). The algorithms developed were then applied to data from the validation experiment. The results suggested that higher-order algorithms do not result in detection accuracy improvement when applied to a new set of data (Fairbanks and Wierwille, 1994). It was also concluded that more sophisticated detection algorithms using nonlinear optimization techniques, such as pattern recognition or neural networks, would probably not provide any improvement in detection accuracy when applied to new data.

Elimination Of Outliers From A/O Based Algorithms

Wreggit and Wierwille (1994a) investigated the possibility that the significant decrease in R values from the development phase to the validation phase in algorithms using A/O data could be due to the effects of prediction outliers. Certain algorithm

output data from eight previously developed A/O task algorithms and the four definitional measures of drowsiness were re-examined. These data were limited to the minimum and maximum values of the observed data so that no prediction outliers were present in the data set. It was concluded that “clipping” the data in this manner did not increase drowsiness prediction for linear algorithms and that there would be no advantage to limiting the range of the prediction data (Wreggit and Wierwille, 1994a).

False Alarm Rates During Alert-Driver Segments

In a practical warning system, it is extremely important to have a low false alarm rate so that users do not become habituated to frequent false signals, ignoring or disbelieving a legitimate alarm. Considering this, Wreggit and Wierwille (1994b) investigated the false alarm rate produced by several algorithms when applied exclusively to periods of unquestionably alert driving (as determined by the definitional measures PERCLOS and AVEOBS, using more stringent thresholds than those used in the algorithm development and validation phases). Data from the validation experiment were re-examined for this purpose. The false alarm rates were found to be acceptably low during alert segments of driving (Wreggit and Wierwille, 1994b).

Effects Of Time-On-Task Variables

Fairbanks, Fahey, and Wierwille (1995) examined the ability of time-on-driving-task measures to increase predictability when added to driver drowsiness detection algorithms. Six variable groups were selected from data obtained in the development experiment (Wreggit, Kirn, and Wierwille, 1993) and compared to PERCLOS, EYEMEAS, and AVEOBS. A linear measure of time-on-task and a second-order (squared) measure of time-on-task were added to each of the six groups, yielding a total

of 12 groups. The same procedures as those used in previous studies were used to obtain optimal algorithms and apply them to data from the validation experiment. In one-half of the cases, inclusion of time-on-task variables increased the accuracy of the resulting algorithms; however, these contributions were modest at best and not consistent (Fairbanks, Fahey, and Wierwille, 1995).

Comparison of Real-Time Accuracy With Advance Intervals

The drowsiness detection algorithms (Wreggit, Kirn, and Wierwille, 1993) were developed based on six-minute averages of data and an immediate warning system. This led to some degree of “smoothing” of the data and the possibility that warning may not be fast enough as the driver becomes drowsy in on-the-road implementation. Therefore, Fairbanks, Fahey, and Wierwille (1995) examined the predictive ability of the algorithms by developing and validating algorithms to predict driver drowsiness level two, four, six, and twelve minutes before independent measure data collection. Raw data for independent measures from the development experiment (Wreggit, Kirn, and Wierwille, 1993) in the form of one-minute averages was baselined using the same procedure employed in the previous studies. Data files were created for this baselined data for each of the previously mentioned time intervals in advance of the dependent measure data. Multiple regression was used to create the new algorithms, and validation of the algorithms was performed using the same procedure as was used in previous studies. Classification matrices and R values were determined for every data set in each variable group. Minimal degradation in algorithm accuracy was found for prediction intervals of up to six minutes in advance (Fairbanks, Fahey, and Wierwille, 1995). This predictive ability could allow a system more time to alert a driver or to confirm drowsiness using other methods, thereby reducing the false alarm rate.

Time Trend Analysis

Fairbanks, Fahey, and Wierwille (1995) investigated the effect of combining independent measure data collected during the two prior six-minute intervals with data collected during the current six-minute interval on drowsiness detection algorithms. Previous studies had used only concurrently collected independent measure data. Algorithm F4a (Wreggit, Kim, and Wierwille, 1993), one of the most stable of the previously developed algorithms, was chosen for comparison. New algorithms were obtained and tested on the validation data as in previous studies. A small improvement in algorithm accuracy was found using previous time intervals in algorithm development, but the improvement did not remain stable when applied to validation data (Fairbanks, Fahey, and Wierwille, 1995).

Elimination Of Outliers From PERCLOS Data

PERCLOS, selected as a typical dependent measure and used in many of the preceding algorithm optimization attempts, was known to contain some apparent outliers among the development and validation data. As a result, Fairbanks, Fahey, and Wierwille (1995) studied the effect of “clipping” the level of PERCLOS estimated by Algorithm F4a (Wreggit, Kim, and Wierwille, 1993) to an operational range. PERCLOS values greater than 0.25 were “clipped” to a value of 0.25, and multiple regression was performed as in previous studies. The results indicated no improvement in algorithm accuracy due to clipping PERCLOS before regression.

Appropriate Drowsiness Warnings for Motor Vehicles

During the latter stages of the research involving measures of drowsiness and algorithms utilizing those measures to detect drowsiness, concurrent development of the alerting component of the drowsiness detection system was initiated. Questions as to the nature of appropriate warnings served as the starting point for this investigation. In order for any warning to be effective, it must be sensed, received, understood, and heeded (Sanders and McCormick, 1993). For crash avoidance warnings in motor vehicles, there are specific considerations that must be carefully examined. The time available to react and the proper level of warning urgency must be carefully considered in the design. The warning system must be compatible with a variety of drivers and a variety of vehicles. Preliminary human factors guidelines for such warnings have been developed (Lerner, Kotwal, Lyons, and Gardner-Bonneau, 1993). An abbreviated summary of the guidelines is presented here:

- The warning must be intrusive and convey a sense of urgency. It must not be so intrusive and urgent that it excessively startles or annoys the driver, but it must be intrusive and urgent enough to effectively convey its message.
- Multiple warning levels should be used to alleviate the activation of urgent false alarms. The most urgent alarms should be unique in frequency and pitch and should be reserved for imminent crash warning signals, with less urgent and less disturbing signals signifying earlier stages. It should be noted that this concept is compatible with the two-stage drowsiness detection system recommended by Wreggit, Kirn, and Wierwille (1993).
- Alarms signifying immediate danger should be presented through two sensory modalities to improve the likelihood that they will be received, due to differences in drivers' perceptions and differences in driving environments.

- Warning systems should ordinarily be activated automatically every time the vehicle is started, though it may be necessary for the driver to activate the drowsiness detection/warning system for legal reasons. Nevertheless, drivers should be able to deactivate the system since it is possible that unforeseeable situations will occur where the warning system is not necessary and would produce false or nuisance alarms if left engaged.
- Built-in diagnostic tests should be engaged every time the system is activated and the driver should be notified immediately of system failure. However, this notification should not be presented in such a way that it can be easily confused with a warning requiring the driver's immediate attention.
- Multiple settings should be available on warning alarms; these settings should be adjustable for different drivers and different environments. When the vehicle is started the settings should default to a predetermined setting of the current settings should be prominently displayed.

There are a variety of ways in which a driver in a drowsy condition may be alerted. Such warning systems should follow the guidelines presented above, as well as certain additional specifications presented here.

Auditory Displays

Generally, auditory displays are preferred for their effectiveness in alerting (Horowitz and Dingus, 1992). However, special precautions must be taken to avoid excessively startling or annoying the driver. Onset rates of 10 dB/ms and higher should be avoided, and sounds emanating from a single area should be avoided unless they are consistent with the direction of the hazard (Lerner et al., 1993). Fundamental frequencies should be in the range of 500 to 3000 Hz, and frequencies easily masked by the ambient

noise of the environment should not be used (Lerner et al., 1993). Characteristics such as a high repetition rate, high intensity, high fundamental frequency, and large frequency oscillations can increase the perceived urgency of a warning signal (Edworthy, Loxely, and Dennis, 1991). Speech warnings may be used, but they are generally slower to convey a message than other modes of alerting, and they are often not distinguishable from the voice of a passenger or a radio message. Speech warnings are most effective when they are combined with other types of warnings.

Tactile Displays

The lower level of vibration in simulators than in actual vehicles may explain why subjects participating in simulator studies have shown a greater reduction in alertness level than subjects participating in road studies. Thus, it is possible that a tactile display, such as vibration in the driver's seat or steering column, may have an alerting effect on drivers. The recommended vibration frequency for such a device is between 100 and 300 Hz (Lerner et al., 1993). However, frequencies much lower than this have been used and found effective (Fairbanks, Fahey, and Wierwille, 1995).

Visual Displays

Use of visual displays in a drowsy driver alerting system is problematic because a drowsy or inattentive driver may be less likely to perceive a visual warning in time to react appropriately. Also, use of a visual display may overload the sensory channel of vision due to the inherently highly visual nature of the driving task. Any visual signal used as part of a warning should be presented within 15 degrees of the driver's normal line of sight of the roadway (Lerner et al., 1993).

Termination of Warnings

Lerner et al. (1993) suggest that warnings automatically triggered by a specific condition be presented for one second or until the triggering condition no longer exists, whichever is greater. This should give the driver enough time to recognize the purpose of the alarm. The system should be reactivated immediately after the alarm is terminated in order to detect any quickly reoccurring decrease in alertness. If the warning may be terminated by the driver, the mode of termination should not have a high degree of accessibility. Termination of warning signals should not become habit, as has happened all too often in previous attempts at drowsiness alerting devices. Some physical motion should be required by the driver.

Countermeasures to Drowsiness

Following the alerting of a driver as to his or her drowsy condition, some method of counteracting the drowsiness for a time so the driver can reach a safe stopping area is necessary. Without such countermeasures, the possibility exists that the driver could immediately revert to a state of drowsiness, rendering the warning signal useless. There are many types of potential countermeasures to driver drowsiness, which are presented here in an abbreviated summary of a literature review by Fairbanks, Fahey, and Wierwille (1995).

Simple Countermeasures

Simple countermeasures involve simply avoiding the factors that contribute to the onset of fatigue: inadequate sleep or rest, prolonged hours of driving, and the intake of certain foods or drugs (Haworth, Heffernan, and Horne, 1989). These types of countermeasures fall into several categories (Haworth and Vulcan, 1990):

- Driver-oriented countermeasures: Education of drivers to recognize the signs and dangers of drowsiness while driving; often not practical due to situations prevalent in modern society where driving for long periods of time after long periods of work is unavoidable.
- Vehicle-based countermeasures: Consist of listening to the radio or providing ventilation. Cool air has been shown to have a preventive effect (Mackie and O'Halon, 1977) but has not been shown to alleviate existing fatigue.
- Environmental countermeasures: Rest breaks or pavement treatments. Rest breaks may have a refreshing effect before fatigue has developed, but have little beneficial value once fatigue sets in (Harris and Mackie, 1972). Pavement treatments, specifically rumble strips, have been shown to be effective countermeasures (Haworth and Vulcan, 1990; Wood, 1994).
- Other simple countermeasures: These include chewing gum, singing along with the radio, sitting on something hard, or removing the right shoe (Harris, 1967). These methods do not work consistently for all people.

Chemical Countermeasures

Caffeine is favored by many drivers as a way to remain alert over extended periods of driving. Subjects given a dosage of caffeine have been shown to have significantly lower reaction times for a subsidiary task while driving (Lisper, Tornros, and Van Loon, 1981) and to exhibit significantly reduced attention lapses (Baker and Theologus, 1982) when compared to subjects not under the influence of caffeine. However, Childs (1978) concluded that a large dose of caffeine may disrupt visual attention in those who do not normally consume caffeine, while enhancing visual attention in subjects who normally consume large amounts of caffeine.

Stronger stimulants, such as amphetamines, have been shown to improve tracking, concentration, and attention, particularly when a subject is sleep-deprived; however, they also have been found to increase risk taking behaviors, making them inappropriate for a vigilance task such as driving (Haworth and Vulcan, 1990). Nicotine has been shown to have a positive effect on the performance of vigilance tasks (Wesnes and Warburton, 1978); however, nicotine has been shown to be a highly habit-forming substance.

It should be noted that all the aforementioned stimulants can have serious health effects and that reactions to the substances differ both between and within individuals. Therefore, they should probably not be considered viable sources of drowsiness countermeasures. However, certain other chemical substances may be effective countermeasures to drowsiness. Kaneda, Iizuka, Ueno, Hiramatsu, Taguchi, and Tsukino (1994) investigated the refreshing effect of four scents: lavender, lemon, jasmine, and peppermint. The results showed that peppermint has the greatest refreshing effect. The scent appeared to be most effective when released immediately after the first signs of a decrease in alertness and immediately after the sounding of a buzzer. However, no statistical analysis of the Kaneda et al. (1994) data was made, and the experimental method used was not clearly described. Therefore, Fairbanks, Fahey, and Wierwille (1995) concluded that further investigation into the use of scents as a drowsiness countermeasure was necessary.

Mechanical and Electronic Countermeasures

Certain mechanical and electronic drowsiness “countermeasures” were implemented along with some of the previously mentioned attempts at drowsiness detection systems. However, most of these are simply alarms designed to warn a driver

of a detected drowsy condition; none have been shown to alleviate drowsiness in experimental research.

Secondary tasks produced by electronic means have been investigated in previous research. It has been noted that Kirn, Wreggit, and Wierwille (1994) investigated the A/O task as a possible drowsiness countermeasure, concluding that more research was necessary. Another secondary task that could be used to maintain driver alertness is the lane minder. The lane minder device consists of sensors mounted in the vehicle-that are able to sense the boundaries of the lane. If the vehicle exceeds those boundaries, an alarm is sounded.

Vibration, previously mentioned as a possible tactile display for alerting drivers, could also be an effective countermeasure if the vibration were continued for an extended period of time. Many accidents involving drowsy drivers occur when the driver allow the vehicle to drift off the side of the road. In 1986, fifty-seven percent of all accidents on the Pennsylvania Turnpike were classified as drift-off-road (DOR) accidents (Wood, 1994). In 1987, authorities began installing rumble strips along the shoulders of the Pennsylvania Turnpike. When a vehicle's tires pass over the grooved, uneven pavement of a rumble strip, vibration and a loud sound are produced. Data collected after installation of the rumble strips on the Pennsylvania Turnpike showed a 70% reduction in DOR accidents (Wood, 1994). The California Department of Transportation reported favorable results with a similar installation (Chaudoin and Nelson, 1985). Though shoulder rumble strip installations are effective, the cost of installing them on every road would be prohibitive. Also, the inside boundary of a lane is difficult to equip with rumble strips because of intentional lane changes such as passing. However, a simulated rumble strip effect, produced from within a vehicle and combining vibration and sound, could possibly be a very effective countermeasure to driver drowsiness.

Daimler-Benz (1994) conducted a study in which lane position monitoring and simulated rumble strip effects were implemented. An image processing system for lane position recognition was implemented with three types of warnings: acoustic, haptic (tactile), and corrective haptic. The acoustic warning was a simulated rumble strip sound emanating from the side of the vehicle found to be excessively close to crossing a lane boundary. 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. Participants in the study reported that they found the acoustic warning “enlivening”, suggesting that it may have a sustained alerting effect as well as an immediate alerting effect.

Optimization of Advisory and Alarm Stimuli

Fairbanks, Fahey, and Wierwille (1995) investigated various configurations of advising/alarming stimuli to be used as the second and third stages in a three-stage drowsy driver detection, alerting, and alarming system. The first stage in such a system, the detection phase, would use the algorithms developed and validated by Wierwille et al. (1994). A detection by this phase would initiate the second stage, re-alerting the driver. The second stage would begin with an auditory stimulus informing the driver that a drowsy condition has been detected, activating a full alarm unless the driver presses a reset button. The third stage would allow the driver to engage a drowsiness countermeasure to maintain alertness while searching for a safe rest area. These three stages will be discussed in more detail later.

Eight graduate students at Virginia Tech, each of whom had completed at least one semester of course work in human factors engineering, were used as subjects. Subjects were selected from this group because of their familiarity with behavioral methods. As in

prior studies, subjects were instructed to awaken by 7:00 A.M. and carry on normal daily activities without napping. An experimenter picked up each subject at 6:00 P.M., took him or her to dinner, and observed him or her in the laboratory to ensure that no napping or ingestion of stimulants took place.

For this experiment, the ambient noise level in the simulator was adjusted to a level of 75.5 dBA at a distance 5 inches from the right ear of the driver. This level was based on a small road study which found the average ambient noise level of five different vehicles. Initial advisory tones were produced using a function generator. Voice messages were recorded on a custom-designed digital voice recording/playback system, capable of recording and playing 10 seconds of audio information on each of four channels. Full alarm sounds were produced using various function generators. They were recorded on and played back from a personal computer equipped with a 16-bit digital audio card. Advisory and alarm signals were mixed, amplified, and then applied to speaker enclosures located at mid-dash height to the right and left of the straight-ahead position.

Several countermeasures and peripheral stimuli were available for use in the simulator. Vibration was produced in the driver's seat using eccentrics (unbalanced rotational masses) driven by servo motors and located in the seat pan and seat back. Vibration was also produced in the steering wheel by means of a periodic signal applied to the active servo system that provides steering feel. A switch at the investigator station allowed a simulated brake pulse effect. A lane sensing device was available in the simulator which activated a chirping/beeping alarm when the vehicle exceeded the lane boundaries. A peppermint scent was supplied through a reusable aerosol can that could be used to dispense vapors in front of the driver.

At approximately 12:00 A.M., after an initial period of acclimation to the simulator, each subject began an extended period of driving. The subject was asked to optimize and rate eight different tones, consisting of combinations of four different frequencies and two different wave shapes. The eight tones were counterbalanced across subjects. The amplitude and duration of each tone were set to a preferred level by the subject, and each tone was rated by the subject as to its effectiveness in advising a driver of a subsequent alarm. In a similar manner, the subject was then asked to optimize and rate two different voice messages, choosing the appropriate amplitude and then rating for effectiveness. Both messages consisted of the same words (“Possible drowsiness has been detected; press reset now”). One voice message utilized a male voice, while the other utilized a female voice. The subject was then asked to optimize the amplitude of each of eight different alarm sounds. Following this optimization, the method of paired comparisons was used to determine the preferred alarm sound. Each of the 28 possible pairs of alarms was played for each subject. The pairs were presented in a different order for each subject to minimize possible order effects.

Peripheral stimuli, including steering vibration, seat back vibration, seat pan vibration, and the simulated brake pulse were individually presented to each of the subjects and rated for effectiveness. Each peripheral stimulus was also presented in combination with each alarm sound, and each combination was rated for effectiveness. Presentation order differed for each of the eight subjects. Finally, possible countermeasures were presented individually to each subject. These countermeasures included possible combinations of seat pan vibration and seat back vibration (using lower amplitudes and frequencies than those used for the peripheral alarm stimuli), as well as the lane minder, the A/O task, and the peppermint scent. Each countermeasure was rated by each subject for effectiveness.

Data collected during the experiment included effectiveness ratings, data from the paired comparisons, volume data, and duration data. T-tests were used to test the significance of the volume level and the effectiveness ratings for the voice messages, as well as differences within subjects for the duration of the advisory tones. Analysis of variance (ANOVA) was used to test for significant differences with respect to volumes of the tones and alarms as well as for all effectiveness ratings. Frequencies of preference for the paired comparisons of the eight alarm sounds were converted into a proportion matrix and transformed into an interval scale. The data were then analyzed using a one-way analysis of variance, with alarm as a single factor with eight levels. Where significant differences were found in any of the ANOVAs, Newman-Keuls post-hoc tests were conducted to identify the specific sources of those differences. Alpha levels of 0.05 were used for all tests.

Initial Advisory Tones

For the initial advisory tones, no significant differences within subjects were found by Fairbanks, Fahey, and Wierwille (1995) for tone duration. The average final tone duration chosen by the subjects was 0.80 second. The effect of tone amplitude was found to be significant, and post-hoc tests revealed many cases where there were significant differences between specific advisory tones. Similarly, several specific differences were found between ratings of the tones. It was found that rectangular-wave tones at higher frequencies were rated most effective by the subjects and achieved this effectiveness at lower preferred volume levels than sinusoidal-wave tones. On the basis of these results, Fairbanks, Fahey, and Wierwille (1995) made recommendations for an optimal advisory tone. The recommended characteristics are presented here:

- Duration: 0.8 second.

- Wave shape: Rectangular.
- Frequency: 700 to 1000 Hz.
- Amplitude: Approximately 8.1 dBA below the ambient noise level in the vehicle. (The tone is still audible at this level due to concentration at specific frequencies.)

Voice Messages

For the voice messages, no significant difference between the male voice and the female voice was found for either amplitude or effectiveness rating. Fairbanks, Fahey, and Wierwille (1995) concluded that either voice could be used effectively as part of a drowsy driver warning system when presented at a peak amplitude of approximately 10 dBA above the ambient noise level in the vehicle (average amplitude of 5 dBA below ambient noise level).

Auditory Alarm Stimuli

The effect of amplitude for the alarm stimuli was found to be significant, and post-hoc tests revealed specific differences. The scaled values for the paired comparisons were found to possess additivity; thus, the usual variance ratio test was usable for tests of significance. It was found that the subjects had significant differences in preference among the eight alarms. The two most effective alarms were an on-off tone and a dual tone. Both of these alarms are high-frequency square waves with alternating periods of no sound or lower-frequency sound, and both have a repetition frequency of 3 Hz. Fairbanks, Fahey, and Wierwille (1995) recommended that such an alarm be used as part of a drowsiness warning system. The recommended amplitudes for the on-off tone and

the dual tone were 3.5 dBA and 10 dBA above the ambient noise level in the vehicle, respectively.

Peripheral Stimuli

Fairbanks, Fahey, and Wierwille (1995) found no significant differences between the ratings of the peripheral stimuli when they were presented by themselves. However, for the combinations of alarm stimuli and peripheral stimuli, significant results were found for the main effects of alarm type and peripheral stimulus type as well as for the interaction between the two. Specific differences were isolated with Newman-Keuls tests. The brake pulse significantly increased the effectiveness of all auditory alarms. The combination of seat back and seat pan vibration increased the effectiveness of all the auditory alarms, but the increase was not significant for either of the two recommended alarms. It was noted that this lack of significance could be due to initial high ratings given to the auditory alarms by themselves; subjects may have run out of room to rate the combinations higher than the alarms by themselves on the rating scales. Fairbanks, Fahey, and Wierwille (1995) recommended that the auditory alarm be accompanied by a simulated brake pulse for maximum effectiveness. If brake pulses can not be implemented, combined seat vibration should be used as an effective substitute (Fairbanks, Fahey, and Wierwille, 1995).

Countermeasures

No significant differences among the effectiveness ratings of the countermeasures were found. In addition, no investigated countermeasure rated far above the midpoint of the rating scale used in the study. Fairbanks, Fahey, and Wierwille (1995) concluded that

further research was necessary concerning countermeasures in a drowsy driver detection and alerting system.

PHASE ONE
REMOVAL OF BASELINING TECHNIQUES FROM DROWSY DRIVER
DETECTION ALGORITHMS

Overview

Development of a drowsy driver detection, alarm, and countermeasures system (DDDACs) is dependent upon on-line, accurate measurement of the drowsiness level of a driver. In addition, it is desirable that the method of drowsiness measurement be unintrusive to the driver and have the ability to adapt to varying conditions. Recent research has led to the development of drowsy driver detection algorithms based on driver-vehicle performance measures. The algorithms are accurate, unintrusive estimators of definitional measures of drowsiness found in the literature and in previous research (Ellsworth, Wreggit, and Wierwille, 1993).

Wreggit, Kirn, and Wierwille (1993) conducted a simulator study in which five dependent definitional measures were analyzed for estimation accuracy against thirty independent performance measures. Each data gathering session was 150 minutes in length. After the data gathering sessions were completed in that study, a data manipulation procedure was performed. Initially, all measures were computed over one-minute intervals. Then, the first two minutes of data for all measures were removed. This was done because it was thought that in the first two minutes of driving, some subjects demonstrated inconsistencies concerning their driving behavior, reactions, and physiological measures. For each measure, the next ten minutes of data (minutes 3 through 12) were then averaged to create baselines. The baselines were then subtracted from every subsequent one-minute average of data within each independent measure (minutes 3 through 150). After completion of the baselining procedure, the data were

averaged across six-minute intervals to complete data manipulation. The first two intervals were five-minute averages to compensate for the earlier removal of the first two minutes of data. In all, 25 data intervals were created for each subject. Following data manipulation, multiple regression and discriminant analyses were performed on the collected data to determine the best predictors of drowsiness. Through these analyses, a number of detection algorithms were developed.

A subsequent simulator study (Kim, Wreggit, and Wierwille, 1994) was conducted to validate the recently developed detection algorithms. Data were collected over data-gathering sessions of 156 minutes in length. These data were manipulated in a manner similar to that in the algorithm development experiment. Initially, all measures were computed over one-minute intervals. The first two minutes of data for all measures were removed. For each measure, the next ten minutes of data (minutes 3 through 12) were averaged to create baselines; however, these data were removed after baseline creation. The baselines were then subtracted from every subsequent one-minute average within each independent measure (minutes 13 through 156). After completion of the baselining procedure, the data were averaged across six-minute intervals to create a total of 24 data intervals. Following data manipulation, the previously developed algorithms were applied to the new data set. Correlation analyses between the algorithm outputs and the definitional measure (observed) values were performed, allowing a comparison between R values attained from the original data and the new data. Classification analyses were also performed for purposes of comparison.

Baselined regression techniques for drowsy driver detection algorithms were compared against unbaselined techniques by Ellsworth, Wreggit, and Wierwille (1993). Since pooled data make no distinction as to the subject number and do not account for individual differences in initial values for many measures, it was thought that baselining

could improve the prediction of drowsiness. It was thought that baselining techniques could be easily accomplished on-line by a computer during an experimental run. The computer could simply calculate averages for all measures during the initial portion of the study and then subtract the average values from all subsequent values. Baselined techniques and unbaselined techniques were shown to produce nearly equal R-values. However, the baselined methods were shown to require an average of 5.0 independent variables to achieve the regression model, while the unbaselined methods required an average of 7.8 independent variables. It was concluded that baselined regression methods are more efficient than unbaselined methods. When attempts to reduce the number of terms in regression models were made while retaining nearly the same values of R as full models, unbaselined data showed greater vulnerability. It was concluded that baselining is an advantageous procedure because it provides small improvements in the models. For this reason, baselined regression was used in the previously summarized algorithm development and validation experiments.

Baselining techniques can be easily implemented on-line if a single detection algorithm is used throughout an experimental session. However, it is desirable to be able to use “step-up, step-down” procedures. Such a system allows the DDDACS to switch algorithms and continue to function when all parameters are not available. In the DDDACS test and evaluation experiment, lane-related measures were available during some intervals, but not others. Also, a secondary task, the A/O task (Wierwille et al., 1994), was operating during some portions of the experiment, but not others. These two conditions determined which detection algorithm was used for any one-minute interval in each experimental session. In all, a combination of four types of detection algorithms must be used. Of the baselined detection algorithms developed and validated by Wierwille et al. (1994), the following were chosen to fulfill these requirements. All are

estimators of PERCLOS, the proportion of time a subject's eyes are 80 to 100 percent closed. PERCLOS was shown to be one of the most reliable measures of drowsiness in previous studies (Wreggit, Kim, and Wierwille, 1993). The four algorithms are briefly described here:

- Algorithm D4a: Includes only steering-related and lateral-accelerometer-related measures. This algorithm type was used during periods when the lane-related measures were not available and the A/O task was not being performed.
- Algorithm F4a: Includes steering, lateral-accelerometer, and lane-related measures. This algorithm type was used when lane-related measures were available and the A/O task was not being performed.
- Algorithm J4a: Includes steering, lateral-accelerometer, and A/O task-related measures. This algorithm type was used when lane-related measures were not available and the A/O task was being performed.
- Algorithm L3a: Includes steering, lateral-accelerometer, lane-related and A/O task-related measures. This algorithm type was used when lane-related measures were available and the A/O task was being performed.

Summaries of regression and classification analyses for the above algorithms are shown in Figures 1, 2, 3, and 4.

In the test and evaluation experiment, availability of lane-related measures for algorithm computation was determined on a recurring 30-minute cycle. Lane-related measures were available for 25 of the one-minute intervals in each cycle. They were unavailable for the remaining five minutes (minutes 4, 17, 18, 19, and 20). These intervals were chosen to simulate the level of performance that a lane-sensing device on

Baselined Regression Summary for Dependent Variable: PERCLOS

$R = 0.78910$ $R^2 = 0.62268$ Adjusted $R^2 = 0.61626$

$F(5,294) = 97.038$ $p < 0.0000$ Std. Error of estimate: 0.06065

	BETA	St. Err. of BETA	B	St. Err. of B	t(294)	p-level
Intercept			0.014132	0.005	2.974	0.003
INTACDEV	-0.134	0.038	-0.084597	0.024	-3.488	0.001
LGREV	0.395	0.050	0.040055	0.005	7.845	0.000
STEXED	0.146	0.042	74.427007	21.231	3.506	0.001
NMRHOLD	-0.427	0.054	-0.007378	0.001	-7.985	0.000
THRSHLD	0.450	0.047	0.416209	0.044	9.491	0.000

		Predicted			
Original Observed	Group	% Correct	Awake	Questionable	Drowsy
	Awake	88.29	181	22	2
	Questionable	43.18	11	19	14
	Drowsy	52.94	2	22	27
	Total	75.67	194	63	43

PERCLOS (R Value = 0.789)

Apparent Accuracy Rate (large misclassifications): 0.987

Apparent Accuracy Rate (all misclassifications): 0.757

Classification Matrix Generated From Multiple Regression Analysis of **Original PERCLOS** Data Resulting in **Algorithm D4a**. (Independent variables employed included Steering and Accelerometer.)

		Predicted			
New Observed	Group	% Correct	Awake	Questionable	Drowsy
	Awake	79.32	188	40	9
	Questionable	30.00	8	6	6
	Drowsy	90.32	1	2	28
	Total	77.08	197	48	43

PERCLOS (R Value = 0.800)

Apparent Accuracy Rate (large misclassifications): 0.965

Apparent Accuracy Rate (all misclassifications): 0.771

Algorithm D4a Applied to New Data and Compared with New Observed **PERCLOS** Data

Figure 1: Baselined Regression Summary and Classification Matrices Showing Accuracy of Algorithm D4a When Applied to Original Data and New Data (Wierwille et al., 1994).

Baselined Regression Summary for Dependent Variable: PERCLOSR = 0.87159526 R² = 0.75967830 Adjusted R² = 0.75475703

F(6,293) = 154.37 p < 0.0000 Std. Error of estimate: 0.04849

	BETA	St. Err. of BETA	B	St. Err. of B	t(293)	p-level
Intercept			-0.003	0.004053	-0.694	0.488
INTACDEV	-0.109	0.030	-0.069	0.019114	-3.603	0.000
LANDEV	0.873	0.063	0.066	0.004763	13.798	0.000
LNERRSQ	-0.258	0.054	-0.002	0.000410	-4.820	0.000
STEXED	0.090	0.033	45.740	16.818827	2.720	0.007
NMRHOLD	-0.204	0.045	-0.004	0.000785	-4.494	0.000
THRSHLD	0.250	0.041	0.231	0.037904	6.098	0.000

Predicted

Original Observed	Group	% Correct	Awake	Questionable	Drowsy
	Awake	89.76	184	18	3
	Questionable	47.73	7	21	16
	Drowsy	62.75	3	16	32
	Total	79.00	194	55	51

PERCLOS (R Value = 0.872)

Apparent Accuracy Rate (large misclassifications): 0.980

Apparent Accuracy Rate (all misclassifications): 0.790

Classification Matrix Generated From Multiple Regression Analysis of **Original PERCLOS** Data Resulting in **Algorithm F4a**. (Independent variables employed included Steering, Accelerometer, LANDEV/VAR, LNMNSQ, LANEX, & LNERRSQ.)

Predicted

New Observed	Group	% Correct	Awake	Questionable	Drowsy
	Awake	89.03	211	22	4
	Questionable	15.00	12	3	5
	Drowsy	80.65	3	3	25
	Total	82.99	226	28	34

PERCLOS (R Value = 0.862)

Apparent Accuracy Rate (large misclassifications): 0.976

Apparent Accuracy Rate (all misclassifications): 0.830

Algorithm F4a Applied to New Data and Compared with New Observed **PERCLOS** Data

Figure 2: Baselined Regression Summary and Classification Matrices Showing Accuracy of Algorithm F4a When Applied to Original Data and New Data (Wierwille et al., 1994).

Baselined Regression Summary for Dependent Variable: PERCLOS

R = 0.83585799 R² = 0.69865857 Adjusted R² = 0.68262977

F(5,94) = 43.588 p < 0.00000 Std. Error of estimate: 0.05044

	BETA	St. Err. of BETA	B	St. Err. of B	t(94)	p-level
Intercept			0.002620	0.008	0.341	0.734
ACCVAR	-0.182	0.064	-0.007485	0.003	-2.843	0.005
LGREV	0.302	0.093	0.031048	0.010	3.259	0.002
AOTIME	1.234	0.211	0.087985	0.015	5.839	0.000
NMWRONG	-1.028	0.199	-0.251580	0.049	-5.173	0.000
NMNR	0.313	0.092	0.140206	0.041	3.419	0.001

		Predicted			
Original Observed	Group	% Correct	Awake	Questionable	Drowsy
	Awake	94.29	66	3	1
	Questionable	43.75	2	7	7
	Drowsy	42.86	2	6	6
	Total	79.00	70	16	14

PERCLOS (R Value = 0.836)

Apparent Accuracy Rate (large misclassifications): 0.970

Apparent Accuracy Rate (all misclassifications): 0.790

Classification Matrix Generated From Multiple Regression Analysis of **Original PERCLOS** Data Resulting in **Algorithm J4a**. (Independent variables employed included A/O Task, Steering, and Accelerometer.)

		Predicted			
New Observed	Group	% Correct	Awake	Questionable	Drowsy
	Awake	93.33	112	6	2
	Questionable	27.27	6	3	2
	Drowsy	38.46	4	4	5
	Total	83.33	122	13	9

PERCLOS (R Value = 0.599)

Apparent Accuracy Rate (large misclassifications): 0.958

Apparent Accuracy Rate (all misclassifications): 0.833

Algorithm J4a Applied to New Data and Compared with New Observed **PERCLOS** Data

Figure 3: Baselined Regression Summary and Classification Matrices Showing Accuracy of Algorithm J4a When Applied to Original Data and New Data (Wierwille et al., 1994).

Baselined Regression Summary for Dependent Variable: PERCLOSR= 0.87487717 R²= 0.76541006 Adjusted R² = 0.75027522

F(6,93) = 50.573 p< 0.00000 Std. Error of estimate: 0.04474

	BETA	St. Err. of BETA	B	St. Err. of B	t(93)	p-level
Intercept			-0.004308	0.007	-0.648	0.519
LANVAR	0.924	0.246	0.009506	0.003	3.749	0.000
LANEX	0.310	0.105	0.195697	0.066	2.964	0.004
LNERRSQ	-0.641	0.174	-0.007871	0.002	-3.683	0.000
AOTIME	0.548	0.236	0.039067	0.017	2.326	0.022
NMWRONG	-0.591	0.197	-0.144499	0.048	-3.002	0.003
NMNR	0.286	0.095	0.127975	0.042	3.014	0.003

Predicted

Original Observed	Group	% Correct	Awake	Questionable	Drowsy
	Awake	94.29	66	3	1
	Questionable	62.50	2	10	4
	Drowsy	50.00	4	3	7
	Total	83.00	72	16	12

PERCLOS (R value = 0.875)

Apparent Accuracy Rate (large misclassifications): 0.950

Apparent Accuracy Rate (all misclassifications): 0.830

Classification Matrix Generated From Multiple Regression Analysis of **Original PERCLOS** Data Resulting in **Algorithm L3a**. (Independent variables employed included A/O Task, Steering, Accelerometer, LANDEV/VAR, LNMNSQ, LANEX, & LNERRSQ.)

Predicted

New Observed	Group	% Correct	Awake	Questionable	Drowsy
	Awake	94.17	113	5	2
	Questionable	54.55	3	6	2
	Drowsy	46.15	1	6	6
	Total	86.81	117	17	10

PERCLOS (R Value = 0.796)

Apparent Accuracy Rate (large misclassifications): 0.979

Apparent Accuracy Rate (all misclassifications): 0.868

Algorithm L3a Applied to New Data and Compared with New Observed **PERCLOS**

Figure 4: Baselined Regression Summary and Classification Matrices Showing Accuracy of Algorithm L3a When Applied to Original Data and New Data (Wierwille et al., 1994).

an actual vehicle can achieve. Lane boundaries are often either absent or obscured in such a way that they cannot be detected. Difficulty arises in attempting to establish a baseline for lane-related measures when they are not always available. In order to baseline the lane-related measures in a manner compatible with the baselined detection algorithms, lane-related measures must be available for the first twelve minutes of driving. Lane-related measures were not available for minute 4 in the planned test and evaluation experiment; therefore, calculation of baselines for these measures would have been impossible due to incomplete data. In an actual vehicle, it cannot be guaranteed that lane-related measures will be available during the first twelve minutes of driving. Thus, baselining may become difficult or impossible in some situations.

In the test and evaluation experiment, the A/O task was tested as a part of the DDDACS in some of the experimental sessions. It became active as a drowsiness countermeasure after any detection of drowsiness and remained active for twenty-four minutes after the reset button was pressed. If drowsiness was detected again within that twenty-four minute period, the A/O task remained active for twenty-four minutes after the most recent activation of the reset button. Since drowsiness cannot be detected in the first twelve minutes of driving in a baselined system, the A/O task cannot be activated in the first twelve minutes. Therefore, it would be impossible to collect any A/O task data with which a baseline could be calculated.

The on-line computational difficulties presented by baselining suggested that removal of baselining techniques from detection algorithm computation would be greatly desirable. However, only a minimal loss in predictive accuracy of the detection algorithms was considered to be acceptable. Prior research has shown that driver drowsiness detection algorithms gain some efficiency, but only a small degree of accuracy when baselining techniques are used. As a result, new, unbaselined algorithms

were optimized and validated using the original data collected during the algorithm development and validation experiments. These algorithms were then compared directly with the baselined algorithms. In this manner, the loss in predictive accuracy for those algorithms that were used in test and evaluation of the DDDACS was determined.

METHOD

A revised data manipulation procedure was performed on the original data collected from the algorithm development experiment. All data manipulation was performed using Microsoft Excel, version 5.0a. The one-minute averages of data calculated previously were converted directly to six-minute averages. No data were either removed or averaged to create a baseline. It was thought that allowing the first two minutes of data for each subject to remain in the regression analysis would have a minimal effect on the outcome of the regression and classification analyses. Also, it would allow the DDDACS to become active immediately upon the commencement of driving. A total of 25 data intervals, all of which represented six-minute averages, were created for each subject. Six-minute averages for all subjects were pooled, for a total of 300 data points.

Independent measures were selected and placed in the following groupings:

- Steering-related and lateral-accelerometer-related measures only.
- Steering, lateral-accelerometer, and lane-related measures.
- Steering, lateral-accelerometer, and A/O task-related measures.
- Steering, lateral-accelerometer, lane, and A/O task-related measures.

These groupings of measures are equivalent to those originally used to develop the baselined algorithms D4a, F4a, J4a, and L3a. The definitional measure PERCLOS was selected to be the dependent measure for all analyses. This allowed for reliable comparisons between the baselined algorithms and the unbaselined algorithms.

Backwards stepwise multiple regression and resubstitution were performed on each of the four groupings of the unbaselined data to find optimized combinations of variables that would best predict values of PERCLOS. All regressions were performed using STATISTICA for Windows, version 4.5A. Algorithms were produced using the *B*

weights resulting from multiple regression as coefficients for the corresponding independent variables. (In the regression summary tables, the *B* coefficients are raw regression coefficients, while the *Beta* coefficients are standardized regression coefficients. Since the distributions of the independent variables were not standardized to a mean of 0 and a standard deviation of 1, the raw coefficients were used.) Each algorithm was then applied to the six-minute averages of independent measures to predict average values of PERCLOS. In classification analysis, these predicted values were compared against the measured PERCLOS values. The same dual-threshold, three-tiered classification scheme used in the algorithm development study was used. All classification analyses were performed using Microsoft Excel, version 5.0a. Each measured and calculated value of PERCLOS was classified as “Awake”, “Questionable”, or “Drowsy” according to the following criteria:

<u>Classification</u>	<u>PERCLOS Value</u>
Awake	$\text{PERCLOS} \leq 0.075$
Questionable	$0.075 < \text{PERCLOS} \leq 0.15$
Drowsy	$\text{PERCLOS} > 0.15$

Classifications were scored as correct classifications (“hits”), small misclassifications, or large misclassifications. If the predicted classification was one category away from the observed classification, it was considered a small misclassification; if it was two categories away, it was considered a large misclassification. After classification, apparent accuracy rates (APARs) were calculated for large misclassifications and for all misclassifications. Regression R-values, classification matrices, and APAR values for the four new, unbaselined algorithms were compared with values previously obtained for the baselined algorithms.

The algorithms obtained using the development data were then applied to data collected during the validation experiment. The validation data were converted directly from one-minute averages to six-minute averages with no baselining or removal of data. A total of 26 data intervals were created for each subject. The data were pooled to create a total of 312 data points. The new algorithms were applied to the independent measures to estimate PERCLOS values for each six-minute period. Correlation analysis produced r values for comparison. Classification analysis was performed as with the development data to produce classification matrices and APAR values for comparison.

RESULTS AND DISCUSSION

Summaries of regression and classification analyses for unbaselined drowsy driver detection algorithms D4a-N, F4a-N, J4a-N, and L3a-N are presented in Figures 5, 6, 7, and 8. The name of each algorithm denotes the starting set of independent measures used to optimize the algorithm. In this respect, each algorithm corresponds to one of the baselined algorithms summarized in Figures 1, 2, 3, and 4. The B weights (raw regression coefficients) shown in the regression summaries are used as coefficients in calculating algorithm values. Classification matrices are shown for both the development data set and the validation data set to allow for comparison. Direct comparisons for R values, APAR values for large misclassifications, and APAR values for all misclassifications between the baselined algorithms and the unbaselined algorithms are presented in Figure 9. The comparisons in Figure 9 are made for both the development data set and the validation data set.

R values and APAR values in comparison of the results produced by the baselined and unbaselined algorithms suggest that, in general, the unbaselined algorithms are able to predict drowsiness almost as well as the baselined algorithms. The results show both minor losses and minor gains in R values and APAR values for both development data and validation data. In fact, the average of all gains and losses in R values and APAR values is a loss of 0.0004 between baselined algorithms and unbaselined algorithms. The efficiency of the unbaselined algorithms also compares favorably with that of the baselined algorithms, with the unbaselined algorithms requiring an average of 5.25 independent variables and the baselined algorithms requiring an average of 5.5 variables per algorithm.

The results suggest that baselining techniques generally increase the predictive accuracy of the drowsy driver detection algorithms. However, the increase is so small that it does not justify the increased computational complexity that would be required to implement baselining in a “step up, step down” algorithm system. The results also suggest that there is no difference in efficiency between the baselined algorithms and unbaselined algorithms. Based on these conclusions, unbaselined algorithms are clearly the best choice for use in a “step up, step down” detection system as in the planned test and evaluation experiment.

Previous attempts at refining the drowsy driver detection algorithms (Wierwille et al., 1994; Fairbanks, Fahey, and Wierwille, 1995) have suggested that increases in algorithm sophistication do not result in corresponding increases in algorithm accuracy. The results of this phase suggest that the same trend holds for the increase in sophistication from a unbaselined algorithm to a baselined algorithm.

Unbaselined Regression Summary for Dependent Variable: PERCLOS

R = 0.78785881 R² = 0.62072151 Adjusted R² = 0.61427120

F(5,294) = 96.231 p < 0.0000 Std. Error of estimate: 0.06098

	BETA	St. Err. of BETA	B	St. Err. of B	t(294)	p-level
Intercept			0.11788	0.02382	4.94994	0.000001
INTACDEV	-0.137132	0.038813	-0.09201	0.02604	-3.53315	0.000477
LGREV	0.401964	0.048588	0.04060	0.00491	8.27288	0.000000
STEXED	0.136825	0.041765	69.92649	21.34442	3.27610	0.001179
NMRHOLD	-0.292849	0.045575	-0.00684	0.00107	-6.42562	0.000000
THRSHLD	0.458615	0.049091	0.44023	0.04712	9.34208	0.000000

		Predicted			
Original Observed	Group	% Correct	Awake	Questionable	Drowsy
	Awake	86.27	176	25	3
	Questionable	50.00	11	23	12
	Drowsy	62.00	5	14	31
	Total	76.67	192	62	46

PERCLOS (R value = 0.788)

Apparent Accuracy Rate (LARGE misclassifications): 0.973

Apparent Accuracy Rate (ALL misclassifications): 0.767

Classification Matrix Generated From Multiple Regression Analysis of **Original PERCLOS** Data Resulting in **Algorithm D4a-N**. (Independent variables employed included Steering and Accelerometer.)

		Predicted			
New Observed	Group	% Correct	Awake	Questionable	Drowsy
	Awake	72.80	190	58	13
	Questionable	50.00	6	10	4
	Drowsy	83.87	1	4	26
	Total	72.44	197	72	43

PERCLOS (R value = 0.769)

Apparent Accuracy Rate (LARGE misclassifications): 0.955

Apparent Accuracy Rate (ALL misclassifications): 0.724

Algorithm D4a-N Applied to New Data and Compared with New Observed **PERCLOS** Data

Figure 5: Unbaselined Regression Summary and Classification Matrices Showing Accuracy of Algorithm D4a-N When Applied to Original Data and New Data.

Unbaselined Regression Summary for Dependent Variable: PERCLOS

R = 0.85702904 R² = 0.73449877 Adjusted R² = 0.72906189

F(6,293) = 135.10 p < 0.0000 Std. Error of estimate: 0.05111

	BETA	St. Err.of BETA	B	St. Err. of B	t(293)	p-level
Intercept			0.06088	0.02027	3.00296	0.002904
ACCVAR	-0.167453	0.034934	-0.00808	0.00169	-4.79340	0.000003
LANDEV	0.929554	0.076055	0.06550	0.00536	12.22208	0.000000
LNERRSQ	-0.225729	0.057392	-0.00173	0.00044	-3.93310	0.000105
STEXED	0.082472	0.035086	42.14847	17.93123	2.35056	0.019408
NMRHOLD	0.210594	0.038488	-0.00492	0.00090	-5.47166	0.000000
THRSHLD	0.134577	0.050396	0.12918	0.04838	2.67042	0.007999

		Predicted			
Original Observed	Group	% Correct	Awake	Questionable	Drowsy
	Awake	93.63	191	10	3
	Questionable	43.48	12	20	14
	Drowsy	68.00	4	12	34
	Total	81.67	207	42	51

PERCLOS (R value = 0.857)

Apparent Accuracy Rate (LARGE misclassifications): 0.977

Apparent Accuracy Rate (ALL misclassifications): 0.817

Classification Matrix Generated From Multiple Regression Analysis of **Original PERCLOS** Data Resulting in **Algorithm F4a-N**. (Independent variables employed included Steering, Accelerometer, LANDEV/VAR, LNMNSQ, LANEX, & LNERRSQ.)

		Predicted			
New Observed	Group	% Correct	Awake	Questionable	Drowsy
	Awake	90.42	236	22	3
	Questionable	35.00	8	7	5
	Drowsy	70.97	2	7	22
	Total	84.94	246	36	30

PERCLOS (R value = 0.871)

Apparent Accuracy Rate (LARGE misclassifications): 0.984

Apparent Accuracy Rate (ALL misclassifications): 0.849

Algorithm F4a-N Applied to New Data and Compared with New Observed **PERCLOS** Data

Figure 6: Unbaselined Regression Summary and Classification Matrices Showing Accuracy of Algorithm F4a-N When Applied to Original Data and New Data.

Unbaselined Regression Summary for Dependent Variable: PERCLOS

R = 0.84140802 R² = 0.70796745 Adjusted R² = 0.69567134

F(4,95) = 57.577 p < 0.00000 Std. Error of estimate: 0.04992

	BETA	St. Err. of BETA	B	St. Err. of B	t(95)	p-level
Intercept			0.114695	0.026601	4.31160	0.000040
LGREV	0.343015	0.072384	0.035013	0.007388	4.73886	0.000008
NMRHOLD	-0.438793	0.079892	-0.008136	0.001481	-5.49230	0.000000
THRSHLD	0.643609	0.086890	0.440463	0.059465	7.40714	0.000000
NMNR	0.220279	0.068832	0.099736	0.031165	3.20023	0.001867

Predicted					
Original Observed	Group	% Correct	Awake	Questionable	Drowsy
	Awake	91.30	63	5	1
	Questionable	55.56	4	10	4
	Drowsy	61.54	0	5	8
	Total	81.00	67	20	13

PERCLOS (R value = 0.841)

Apparent Accuracy Rate (LARGE misclassifications): 0.990

Apparent Accuracy Rate (ALL misclassifications): 0.810

Classification Matrix Generated From Multiple Regression Analysis of **Original PERCLOS** Data Resulting in **Algorithm J4a-N**. (Independent variables employed included A/O Task, Steering, and Accelerometer.)

Predicted					
New Observed	Group	% Correct	Awake	Questionable	Drowsy
	Awake	78.87	112	24	6
	Questionable	45.45	2	5	4
	Drowsy	76.92	1	2	10
	Total	76.51	115	31	20

PERCLOS (R value = 0.662)

Apparent Accuracy Rate (LARGE misclassifications): 0.958

Apparent Accuracy Rate (ALL misclassifications): 0.765

Algorithm J4a-N Applied to New Data and Compared with New Observed **PERCLOS** Data

Figure 7: Unbaselined Regression Summary and Classification Matrices Showing Accuracy of Algorithm J4a-N When Applied to Original Data and New Data.

Unbaselined Regression Summary for Dependent Variable: PERCLOSR = 0.87153552 R² = 0.75957416 Adjusted R² = 0.74406281

F(6,93) = 48.969 p < 0.00000 Std. Error of estimate: 0.04578

	BETA	St. Err. of BETA	B	St. Err of B	t(93)	p-level
Intercept			0.047027	0.029750	1.58071	0.117340
ACCVAR	-0.179261	0.062437	-0.007425	0.002586	-2.87106	0.005066
LANDEV	0.494766	0.113252	0.032746	0.007496	4.36873	0.000032
STVELV	0.211349	0.073406	0.000764	0.000265	2.87918	0.004947
NMRHOLD	-0.279452	0.079623	-0.005182	0.001476	-3.50970	0.000694
THRSHLD	0.395668	0.107084	0.270781	0.073284	3.69495	0.000371
NMNR	0.139885	0.070148	0.063336	0.031761	1.99413	0.049066

		Predicted			
Original Observed	Group	% Correct	Awake	Questionable	Drowsy
	Awake	95.65	66	1	2
	Questionable	38.89	4	7	7
	Drowsy	61.54	0	5	8
	Total	81.00	70	13	17

PERCLOS (R value = 0.872)

Apparent Accuracy Rate (LARGE misclassifications): 0.980

Apparent Accuracy Rate (ALL misclassifications): 0.810

Classification Matrix Generated From Multiple Regression Analysis of **Original PERCLOS** Data Resulting in **Algorithm L3a-N**. (Independent variables employed included A/O Task, Steering, Accelerometer, LANDEV/VAR, LNMNSQ, LANEX, & LNERRSQ.)

		Predicted			
New Observed	Group	% Correct	Awake	Questionable	Drowsy
	Awake	91.55	130	11	1
	Questionable	45.45	2	5	4
	Drowsy	61.54	1	4	8
	Total	86.14	133	20	13

PERCLOS (R value = 0.785)

Apparent Accuracy Rate (LARGE misclassifications): 0.988

Apparent Accuracy Rate (ALL misclassifications): 0.861

Algorithm L3a-N Applied to New Data and Compared with New Observed **PERCLOS** Data

Figure 8: Unbaselined Regression Summary and Classification Matrices Showing Accuracy of Algorithm L3a-N When Applied to Original Data and New Data.

Algorithm D4a (Steering, Lateral-Accelerometer)

	Development			Validation		
	R	APAR Large	APAR All	R	APAR Large	APAR All
D4a (Baselined)	0.789	0.987	0.757	0.800	0.965	0.771
D4a-N (Unbaselined)	0.788	0.973	0.767	0.769	0.955	0.724
Gain (+) / Loss (-)	-0.001	-0.014	+0.010	-0.031	-0.010	-0.047

Algorithm F4a (Steering, Lateral-Accelerometer, Lane)

	Development			Validation		
	R	APAR Large	APAR All	R	APAR Large	APAR All
F4a (Baselined)	0.871	0.980	0.790	0.862	0.976	0.830
F4a-N (Unbaselined)	0.857	0.977	0.817	0.871	0.984	0.849
Gain (+) / Loss (-)	-0.014	-0.003	+0.027	+0.009	+0.008	+0.019

Algorithm J4a (Steering, Lateral-Accelerometer, A/O Task)

	Development			Validation		
	R	APAR Large	APAR All	R	APAR Large	APAR All
J4a (Baselined)	0.836	0.970	0.790	0.599	0.958	0.833
J4a-N (Unbaselined)	0.841	0.990	0.810	0.662	0.958	0.765
Gain (+) / Loss (-)	+0.005	+0.020	+0.020	+0.063	0	-0.068

Algorithm L3a (Steering, Lateral-Accelerometer, Lane, A/O Task)

	Development			Validation		
	R	APAR Large	APAR All	R	APAR Large	APAR All
L3a (Baselined)	0.875	0.950	0.830	0.796	0.979	0.868
L3a-N (Unbaselined)	0.872	0.980	0.810	0.785	0.988	0.861
Gain (+) / Loss (-)	-0.003	+0.030	-0.020	-0.011	+0.009	-0.007

Figure 9: Comparison of R Values and Apparent Accuracy Rates For Baselined and Unbaselined Algorithms When Applied to Original Data and New Data.

PHASE TWO
SIMULATOR TEST AND EVALUATION OF A COMPLETE DRIVER
STATUS/PERFORMANCE MONITORING ADVISORY-ALARM-
COUNTERMEASURES SYSTEM

Research Objectives

Recent research has resulted in a nearly complete specification of characteristics for a three-stage drowsy driver detection, alarm, and countermeasures system (DDDACS). However, specification of characteristics for each stage has taken place independently of the other stages. Also, specification of detection criteria based on monitoring of driver status and monitoring of driver performance have taken place independently of one another. The major purpose of the test and evaluation phase was to integrate the two types of detection criteria along with advisory, alarm, and countermeasure stages into a complete DDDACS for test and evaluation in a simulator environment.

Through qualitative evaluation of the simulation of the combined elements of the DDDACS, the test and evaluation phase attempted to answer several unresolved questions dealing with the characteristics of the proposed system. The effectiveness of a two-stage “step-up, step-down” detection algorithm system was evaluated. Efforts were made to evaluate how well the detection, alarm, and countermeasure portions of the DDDACS work together. In addition, the effectiveness of the integration of the status and performance components was assessed. It was also possible to evaluate the computational processes in the system, including the use of unbaselined algorithms.

Other unresolved issues pertaining to the complete DDDACS were examined through quantitative analysis. The effectiveness of countermeasures in keeping drivers

alert, investigated in previous studies, were further evaluated. Any possible alerting effect of the alarm itself on the driver was addressed through comparison of data from the present experiment with data from previous experiments. If sufficient bouts of drowsiness had occurred during the experimental sessions, the effects of peripheral stimuli on both alarm effectiveness and sustaining alertness would have been investigated.

System Description

Stage One: Initial Detection of Drowsiness Status or Reduced Performance Level

Unbaselined algorithms using performance measures as independent variables were used in the first stage to monitor both driver status and driver performance. Algorithms based on performance measures are desirable because they are not intrusive to the driver. In the previous phase of the current study, it was shown that unbaselined algorithms produce no appreciable decrement in accuracy when compared with baselined algorithms previously developed by Wreggit, Kirn, and Wierwille (1993). Based on that result, unbaselined algorithms were used because they are more straightforward and easier to use in on-line implementation.

Algorithms designed to monitor both driver status and driver performance were integrated into the system. For monitoring of driver status, four unbaselined algorithms for estimation of the definitional measure PERCLOS were used. These algorithms were developed in the previous phase of this study. PERCLOS is the percentage of time that the eyes of the driver are 80% to 100% closed. Prior research has shown PERCLOS to be one of the most reliable definitional measures of drowsiness between subjects (Wreggit, Kirn, and Wierwille, 1993). “Step-up, step-down” procedures were used, allowing the detection system to switch algorithms and continue to function during periods when all

parameters were not available. Specifically, lane-related measures were specified to be available during some intervals of the experiment, but not others. Also, the secondary A/O task was operating during some portions of selected cells of the experiment, but not others. A brief description of each algorithm and the conditions under which the system used each algorithm are presented here:

- Algorithm D4a-N: Includes only steering-related and accelerometer-related measures. This algorithm was used during periods when the lane-related measures were not available and the A/O task was not being performed.
- Algorithm F4a-N: Includes steering, accelerometer, and lane-related measures. This algorithm was used when lane-related measures were available and the A/O task was not being performed.
- Algorithm J4a-N: Includes steering, accelerometer, and A/O task-related measures. This algorithm was used when lane-related measures were not available and the A/O task was being performed.
- Algorithm L3a-N: Includes steering, accelerometer, lane-related, and A/O task-related measures. This algorithm was used when lane-related measures were available and the A/O task was being performed.

The regression summaries and classification matrices found for each of the preceding algorithms are presented in Figures 5, 6, 7, and 8. The B weights listed in the regression summaries were used as coefficients for the computational formulae for the algorithms.

For monitoring of driver performance, two algorithms developed by Fairbanks, Lewin, and Wierwille (1995) for estimation of the lane measure LANEX were used. LANEX is the proportion of time that any part of the vehicle exceeds either lane boundary. LANEX is a good indicator of driver performance because good driving practices in the United States dictate that a driver should remain within the lane

boundaries (except when changing lanes). When lane-related information is available, LANEX data can be measured directly without the need for an algorithm. Therefore, the “step-up, step-down” procedure allowed the detection system to switch between measured LANEX and the algorithms estimating LANEX when lane-related measures were not available. If the algorithms were in use, switching between the two algorithms took place based on whether the A/O task was operating. A brief description of each algorithm and the conditions under which the system used each algorithm are presented here:

- Algorithm LDV-1: Includes only steering-related and accelerometer-related measures. This algorithm was used during periods when the lane-related measures were not available and the A/O task was not being performed.
- Algorithm LDV-4: Includes steering, accelerometer, and A/O task-related measures. This algorithm was used when lane-related measures were not available and the A/O task was being performed.

The regression summaries and classification matrices found for each of the preceding algorithms (Fairbanks, Lewin, and Wierwille, 1995) are presented in Appendix A. The B weights listed in the regression summaries were used as coefficients for the computational formulae for the algorithms.

Once the system was engaged, the PERCLOS algorithm in use for the current minute computed an estimated PERCLOS (ePERCLOS) value for each one-minute interval. Also, LANEX was measured for each one-minute interval if possible. If it was not possible to measure LANEX, the appropriate algorithm computed an estimated LANEX (eLANEX) value for each one-minute interval. After the first six minutes of the experiment had passed, the set of one-minute averages was used to compute six-minute average values for ePERCLOS and LANEX/eLANEX. For each minute thereafter,

average values were computed for ePERCLOS and LANEX/eLANEX using six-minute moving averages.

The value of each six-minute ePERCLOS average was compared to a threshold value of 0.14. This value was shown by Fairbanks, Lewin, and Wierwille (1995) to correspond to an optimal PERCLOS threshold of 0.125 (based on a trend model of an ideal progression of drowsiness) at a point where errors were minimized. Fairbanks, Lewin, and Wierwille (1995) also reported an optimal LANEX threshold of 0.10 based on a trend model of an ideal progression into performance decrement, with a corresponding eLANEX threshold of 0.12 for minimization of errors. It was also determined that six-minute intervals with less than three minutes of actual lane data produce poor estimates of driver performance. Based on these results, transitional interval thresholds for LANEX/eLANEX were set according to the specific availability of lane information, determined by linear interpolation between the known endpoint threshold values of 0.10 (lane information available all six minutes) and 0.12 (no lane information available). In the present experiment, the value of each six-minute LANEX/eLANEX average was compared to the appropriate transitional interval threshold.

As long as the value of neither the six-minute ePERCLOS moving average nor the six-minute LANEX/eLANEX moving average exceeded its corresponding threshold value, the system remained in stage one. If the current six-minute interval contained less than three minutes of lane data, the system was able to progress to stage two only if the six-minute ePERCLOS average exceeded its threshold value. If the current six-minute interval contained three or more minutes of lane data, the system was able to progress to stage two using “or” logic: if either the six-minute ePERCLOS average exceeded its

threshold or the six-minute LANEX/eLANEX average exceeded its threshold, detection took place.

A logic diagram further detailing the computational procedures required for on-line status and performance monitoring is presented in Appendix B.

Stage Two: Re-Alerting the Driver

The second stage began upon receipt of a signal or “detection flag” from stage one. This signal disengaged cruise control, if it was engaged. It also activated an auditory advisory tone informing the driver that he or she had exhibited a decrease in alertness level or performance level. The tone used was the optimal advisory tone found by Fairbanks, Fahey, and Wierwille (1995). The audible tone was followed by the following voice message: “Possible drowsiness has been detected; press reset now.” A male voice was used, but there was no significant difference in effectiveness found between a male voice and a female voice (Fairbanks, Fahey, and Wierwille, 1995). After the voice message, the driver had the opportunity to press a reset button to avoid unnecessary exposure to full alarms. The reset button was located in a position on the dashboard to the right of the driver.

If the reset button was not pressed by the driver immediately after the voice message, a full alarm was activated. The driver was able to stop the full alarm at any time while it was engaged by pressing the reset button. The auditory component of this alarm was one of the optimal alarms found by Fairbanks, Fahey, and Wierwille (1995): an on-off tone with a repetition frequency of 3 Hz and an amplitude of 3.5 dBA above the ambient sound level in the vehicle. At any time that the auditory alarm was engaged, a simulated brake pulse or a combined seat back/seat pan vibration could also be engaged

to increase the effectiveness of the auditory alarm. These were two of the most effective peripheral tactile stimuli found by Fairbanks, Fahey, and Wierwille (1995).

Once the reset button was pressed, the advisory/alarming stimuli were disengaged for six minutes. It was considered likely that the six-minute moving averages of algorithm output data would still be indicating a noticeable reduction in alertness level and/or performance level after the reset button was pressed. The six-minute delay allowed the moving averages to be purged of single-minute data recorded before the alarm was sounded. The data for any one-minute period during which advisory/alarming stimuli are presented were discarded because erratic and inconsistent data were likely to be recorded during such a minute. Therefore, the first moving average available when the advisory/alarming stimuli were re-engaged was a five-minute average of data collected after the reset button was pressed. If this five-minute average did not indicate driver drowsiness, the subsequent moving averages were standard six-minute averages.

Stage Three: Maintaining Alertness

The third stage of the three-stage system was activated whenever the reset button was pressed in the second stage of the system. In this stage, the driver may be presented with a drowsiness countermeasure to aid in keeping himself or herself awake while searching for a safe rest area. Of the countermeasures investigated by Fairbanks, Fahey, and Wierwille (1995), two were selected for further study: the introduction of the secondary A/O task and the activation of a lane-minder device. In selected cells of the experiment, one of these countermeasures was used.

The A/O task involved presentation of recorded words presented aurally to the driver. The driver was asked to respond by pressing one of two buttons, labeled “YES” or “NO” and located on the cross member of the steering wheel. If the presented word

contained the letter “A” or the letter “O”, the “YES” button was to be depressed; otherwise, the “NO” button was to be depressed. During this task, the algorithms used for status and performance monitoring in the stage one-switched from Algorithms D4a-N, F4a-N, and LDV-1 to Algorithms J4a-N, L3a-N, and LDV-4 (Fairbanks, Lewin, and Wierwille, 1995). This switching of algorithms allowed for the results of the A/O task for each minute of data to be incorporated into the calculated PERCLOS value. Upon a detection of reduced driver alertness or performance level, the A/O task was suspended and the system reverted to the beginning of stage two, sounding the initial advisory tone and voice message.

Once activated, the lane-minder device sounded a warbling tone if the driver allowed the vehicle to exceed the lane boundaries. The alarm sounded through a piezo buzzer to either the driver’s left or right depending on the side on which the vehicle exceeded the lane boundary.

The A/O task and the lane-minder device only operated independently of one another; both countermeasures could not be active at the same time. Once engaged, the countermeasure in use remained active for twenty-five minutes. If a decrease in driver alertness level or performance level was detected again within that twenty-five minute period, the same countermeasure was re-engaged after the reset button was pressed again. The countermeasure then remained active for twenty-five minutes after the most recent pressure of the reset button.

METHOD

Subjects

All participants were volunteers from the Blacksburg, Virginia area ranging in age from 21 to 45 years. This age range corresponds to the population most heavily involved in drowsiness-related accidents (Knipling and Wierwille, 1994).

All subjects were screened according to a questionnaire, which included questions concerning normal sleeping habits, normal working hours, smoking habits, general health, and body size. This questionnaire appears in Appendix C. It was required that all subjects possess a valid driver's license, have 20/40 vision or better (corrected or uncorrected), and have no known hearing problems.

Each subject was compensated for his or her one-time participation in the experiment. At 6:00 P.M., each subject was picked up at his or her home and received \$6.00 for dinner. Each subject was paid \$5.00 per hour from 6:00 P.M. until midnight and \$8.00 per hour from midnight until the end of the experiment.

As originally planned, this experiment was to include eighteen participants (nine males and nine females) from whom valid data could be collected. However, the experiment was stopped after experimental sessions had been conducted with a total of nine subjects.

During the first experimental session, it became evident that certain elements in the experimental protocol needed to be refined. Specifically, the subject was not observing lane boundaries as he would have been in an actual vehicle, causing a large number of detections based on lane excursions. Based on this observation, the experimental instructions were modified to enforce lane-keeping behavior more consistent with what would typically be seen in an actual vehicle. This modification included the addition of

practice with the lane minder device to the training sessions for future subjects. As a result of the change in protocol, the data for Subject 1 were discarded.

Another problem in subject adherence to instructions became evident in the second experimental session. Subject 2 made little attempt to try to fight drowsiness during the session, particularly from the end of the first hour through the end of the experiment. Again, the behavior of the subject was not representative of that which would be exhibited in an actual vehicle. As a result, the data collected from Subject 2 were erratic and were also discarded.

It was originally planned to conduct six experimental sessions with male subjects followed by six experimental sessions with female subjects. However, after the collection of valid data on four male subjects it became evident that the results were going to be much different than expected. It was decided to shorten and counterbalance the experiment. Therefore, data were collected from three female subjects, and valid data from one of the male subjects were discarded to achieve gender balance. This process yielded valid data from three male and three female subjects from a total of nine experimental sessions.

Apparatus

Simulator

The simulator used in this study is a computer-controlled, hydraulically powered moving-base automobile simulator that handles like a mid-sized rear wheel drive automobile. This simulator has been validated by Leonard and Wierwille (1975) and is located at the Vehicle Analysis and Simulation Laboratory at Virginia Tech. Previous research, including the detection algorithm development and validation experiments and

the optimization of advisory and alarm stimuli experiments, has used the same simulator (Wierwille et al., 1994; Fairbanks, Fahey, and Wierwille, 1995).

The simulator has four degrees of freedom of physical motion (roll, yaw, lateral translation, and longitudinal translation). The roadway image was presented using a monochrome CRT viewed through a Fresnel lens. The image was that of a two-lane highway with side markers and a dashed center line. Light horizontal lines were embedded in the horizontal plane to enhance the image of the roadway continuing into the distance. A simulated automobile hood was also included in the image.

Roadway vibration and sounds, such as engine noise, tire squeal on hard turns, and tire screech on hard braking were also presented to the driver. The ambient engine noise level in the simulator at 60 miles per hour was set at 75.5 dBA so that auditory stimuli developed by Fairbanks, Fahey, and Wierwille (1995) could be used in the same auditory context as that in which they were developed.

Eye Closure Monitoring Equipment

A low light level camera (RCA TC1004-UO1) was used to continuously monitor a subject's entire face, including eye movements. The camera can operate at very low light levels and thus be unintrusive. The video signal was passed through a video cassette recorder and viewed by an experimenter using a Sanyo VM 4512A monitor. This experimenter manipulated a specially designed linear potentiometer to track the movement of the subject's eyelids. This tracking produced a value for the measure PERCLOS. Video and audio recordings of each session were made and kept for future reference.

Drowsiness Detection (Stage 1) Equipment

During the experiment, a variety of analog sensors on the simulator were operating. Two serially interfaced microcomputers equipped with special interfaces received the analog data and converted it to digital format, calculated algorithm values and moving averages on-line, stored data for later analysis, and provided a detection flag whenever a reduced level of alertness or performance in the subject was detected.

Timing for the system originated from a Sony MDS-302 MiniDisc recorder, a high-quality digital audio recording device capable of recording two audio channels on a 74-minute optical disc. When played, the disc can be repeated indefinitely without degradation. One audio channel consisted of prerecorded words for the A/O task, and the other consisted of an 18,000 Hz signaling pulse. A word and a pulse occurred simultaneously every 15 seconds. The pulse was either of low amplitude or high amplitude, depending on the correct response ("YES" or "NO") for the word currently being presented. The audio channel containing the words was sent to a power amplifier and then to a speaker on the simulator. When the A/O task is not activated, the signal from the power amplifier was attenuated. Thus, the signaling pulses were used at all times to control interval timing, and the prerecorded words for the A/O task were presented only when the task was activated.

The pulses from the MDS-302 recorder, as well as the actual responses for the A/O task collected from buttons on the simulator steering wheel when the task is activated, were fed into a TRS-80 Model III microcomputer via a custom analog-to-digital (A/D) converter interface. A BASIC program running on the TRS-80 counted the pulses and performed scoring of the A/O task. Upon receipt of every fourth pulse (marking the passage of one minute), the TRS-80 sent a "flag" signal and its A/O task results to a WIN 486-33i microcomputer via a serial RS-232 interface.

The WIN 486-33i microcomputer was equipped with a National Instruments AT-MIO-16 A/D converter interface card. This card allows for rapid digital sampling of analog data on 16 different channels. Measured LANEX and all performance measures necessary to calculate the six algorithms that were used in this phase of the study (except those related to the A/O task) were sampled, converted, and calculated by the WIN 486-33i. The output from the linear potentiometer for the measure PERCLOS was handled in the same manner. The computer was programmed using Microsoft QuickBASIC to collect raw data and calculate the necessary measures on-line from that data. Every minute, the program received the signal and the A/O task data from the TRS-80 via a serial communications port. Upon receipt of the signal, the program averaged the data for that minute and saved it in a data file for subsequent analysis. It then calculated an estimate of PERCLOS for that minute on-line using one of the algorithms. It also either used measured LANEX or calculated an estimate of LANEX for that minute on-line using one of the algorithms. The program automatically selected the proper algorithm using previously supplied data about the experimental conditions. It also constructed six-minute moving averages of the one-minute PERCLOS and LANEX values. If the value of either the six-minute PERCLOS moving average or the six-minute LANEX moving average exceeded its threshold value, the program sent a signal to an IBM 433DX/S computer via another RS-232 serial cable indicating that drowsy status or a performance decrement had been detected (the "detection flag"). The program automatically re-initialized the moving average whenever a detection occurred. It did not send another detection flag until six minutes had passed since the most recent detection. Additionally, the program displayed its status and provided advisory messages to the experimenters via a video display. A more detailed description of the program with a logic diagram appears in Appendix B.

Driver Re-Alerting (Stage 2) Equipment

The IBM 433DX/S microcomputer was equipped with a Sound Blaster 16 digital audio card. This interface card is capable of high-quality audio recording and playback to and from sound files stored on the computer's hard disk drive. The sound file used for playback during the experiment consisted of the optimal advisory tone, followed by a voice message, followed by the optimal auditory alarm (Fairbanks, Fahey, and Wierwille, 1995). A program running on the IBM 433DX/S computer monitored the serial port for a detection flag from the WIN 486-33i computer and played the sound file on the Sound Blaster whenever a detection flag was received.

The line output from the card passed first through a stereo mixer and dual power amplifiers. From there, it passed through a timer-relay system controlled by the reset button on the instrument panel, which was connected in such a way that pressure of the reset button disconnected the signal. Until the reset button was pressed, the signal passed through it and was applied to dual speaker enclosures containing of 4-inch woofers and 1-inch tweeters with a frequency response of 100 to 20,000 Hz. The speaker enclosures were located to the right and left of the driver.

Vibration was produced in the seat back and seat pan with the use of eccentrics (unbalanced rotational masses) driven by high-quality servo motors. This vibration could be engaged at the beginning of the auditory alarm produced by the IBM 433DX/S computer. The stimulus could be disengaged simultaneously with pressure of the reset button.

A simulated brake pulse effect was available and could be activated using a switch connected directly to the simulator. If the switch were activated, the simulator would momentarily lurch backward to produce the feeling of braking, and speed would be

reduced somewhat. This effect could be engaged simultaneously with the beginning of the auditory alarm produced by the IBM 433DX/S computer.

Alertness Maintenance (Stage 3) Equipment

The A/O task was presented via audio output from the MDS-302 MiniDisc recorder and was responded to via steering wheel buttons and the TRS-80 computer, as previously described. An experimenter activated the A/O task after pressure of the reset button by the driver by simply turning up the volume level on the power amplifier. A displayed message from the WIN 486-33i computer prompted the experimenter to turn down the volume on the power amplifier, deactivating the task.

The lane minder was activated via a switch connected directly to the simulator. It sounded a warbling tone when the vehicle exceeded lane boundaries. This sound was presented via dual piezo buzzers. The lane minder was activated manually after the driver presses the reset button and deactivated manually by the experimenter when a message from the WIN 486-33i was displayed. The lane minder was also manually activated during the training session and during the first five minutes of each experimental session.

Experimental Design

As originally designed, the experiment was to employ a 3×3 factorial between-subject design. The design was to therefore have nine cells, with two subjects (one male and one female) assigned to each cell. The two factors of interest were as follows:

- Alarm Stimulus: The type of alarm presented to the subject if the reset button was not pressed after the initial advisory tone and the voice message. The three

levels of this factor were to be Auditory Alarm Alone, Auditory Alarm with Seat Vibration, and Auditory Alarm with Brake Pulse.

- Task After Detection: The type of subsidiary task presented to the subject after pressure of the reset button as a countermeasure to drowsiness. The three levels of this factor were to be No Task, Lane Minder Task, and A/O Task. The tasks were presented for twenty-five minutes after the most recent pressure of the reset button.

In every instance of detection during the experimental sessions, the subject pressed the reset button immediately after the initial advisory tone and voice message. Since the full alarm was never activated, the peripheral stimuli were never activated. As a result, the final experimental design included only one factor of interest, Task After Detection. The design contained three cells with two subjects (one male and one female) assigned to each cell.

The definitional measure PERCLOS, the proportion of time that a driver's eyes are 80% to 100% closed during a one-minute segment, was gathered for analysis in this experiment. Several categories of performance measures were collected in this experiment. The measures collected included LANEX, the proportion of time that any part of the vehicle exceeds a lane boundary. Also, all measures necessary to calculate estimates of PERCLOS and LANEX using any of the previously mentioned algorithms were collected. These measures are described here:

Steering-Related Measures:

- NMRHOLD: The number of times the hold circuit output on the steering wheel exceeds a threshold value (corresponding to holding the steering wheel still for 0.4 second or longer).

- THRSILD: The proportion of total time the hold circuit output on the steering wheel exceeds a threshold value.
- LGREV: The number of times that steering excursion exceeds 15° after steering velocity passes through zero.
- STEXED: The proportion of time that steering velocity exceeds 150° per second.
- STVELV: The variance of steering velocity, where velocity was measured in degrees per second.

Lane-related measures (other than LANEX):

- LANDEV: The standard deviation of lateral position relative to the lane.
- LNNRSQ: The mean square of the difference between the outside edge of the vehicle and the lane edge when the vehicle exceeds the lane. When the vehicle does not exceed the lane, the contribution to the measure is zero.

Accelerometer-related measures:

- ACCDEV: The standard deviation of the smoothed output of the accelerometer, where the output was first converted to feet per second-squared. (Smoothing was accomplished with a low-pass filter having a corner frequency at 7.25 Hz.)
- ACCVAR: The variance of the smoothed output of the accelerometer. (Square of ACCDEV).
- INTACDEV: The standard deviation of the lateral velocity of the vehicle. (This signal will be obtained by passing the smoothed accelerometer signal through an additional low pass filter with a corner frequency of 0.004 Hz.)

A/O Task-related measures (obtainable from two of the subjects):

- AOTIME: Mean response time to a correct response. Incorrect responses and non-responses are specified as 12 seconds.
- NMWRONG: Mean number of incorrect responses. Non-responses are not included in this measure.
- NMNR: Mean number of stimuli for which there is no response.

In an actual driving situation, any lapse in ability of the system to detect lane boundaries during a substantial portion of a one-minute section of driving time would render an algorithm including lane-related measures useless for that entire minute. It would also make LANEX unmeasurable and make it necessary to use an algorithm to estimate LANEX. In this experiment, lane-related measures were assumed to be available for five-sixths of the total driving time and unavailable for the remaining one-sixth. In a recurring thirty-minute cycle, twenty-five one-minute intervals used ePERCLOS algorithms that contain lane-related measures, as well as using measured LANEX. The other five intervals in the cycle used ePERCLOS and eLANEX algorithms that do not contain the lane-related measures. These techniques adequately simulated the small losses in lane boundary detection that are likely to occur in normal driving.

In addition to the collection of objective performance measures, several types of subjective ratings were collected. Immediately following each detection, an experimenter subjectively rated the subject's drowsiness and performance levels. Subjective rating has been demonstrated to be a reliable indicator of drowsiness in previous experimentation (Ellsworth, Wreggit, and Wierwille, 1993). Performance rating was based on apparent lane-keeping behavior. After the subject pressed the reset button, he or she was asked to give a self-rating of his or her drowsiness level. The subject was also asked to give a

rating of the timing of the advisory tone and message, taking his or her drowsiness level into account. This was accomplished by querying the subject, with the experimenter writing down the response. The scales used to collect these subjective ratings appear in Appendix D.

Procedure

Subjects who were selected by the research team after screening were contacted and scheduled for a particular date. On the scheduled day, each subject was asked to awaken by 7:00 A.M. The subject was informed that he or she should carry on normal daily activities, but should not take any naps.

Each participant was picked up by an experimenter at 6:00 P.M. and taken to dinner at a fast food restaurant. At dinner, the subject was reminded not to ingest any caffeinated substances or sugared beverages. The subject was permitted to smoke immediately following dinner, but not thereafter.

The participant was brought to the laboratory after dinner. He or she was allowed to watch television, read, study, watch a movie on a VCR, listen to music on headphones, etc. An experimenter remained with the subject until midnight to ensure no napping. During this time, the experimenter gave the participant an information sheet and an informed consent form describing the events to take place in the experiment. These documents appear in Appendix E. The subject was asked to read and sign the consent form.

At midnight, two rested experimenters arrived and relieved the experimenter who stayed with the subject. Immediately afterward, the subject entered the simulator for a training and practice driving session. The subject was instructed in procedures for terminating the experiment if it became necessary. The subject was also instructed in the

general operation of the DDDACS and informed about the advisory tone, voice message, and alarm stimuli sequence as configured for his or her cell of the experiment. The reset button was shown to the subject. The subject was asked if he or she had any questions. Once all questions were answered, the lights were dimmed and the practice driving session began. While driving during the first few minutes of this session, the subject had the opportunity to change lanes and alter speed on straight and curved roadways.

Once the subject had become accustomed to the simulator, he or she practiced staying in the right-hand lane of the simulated roadway. To help the subject in recognizing where the lane boundaries were, the experimenters activated the lane-minder device. Initially, the experimenters instructed each subject to drive out of the lane on each side to become accustomed to the lane minder. Then, the subject was instructed to stay in the right-hand lane. If the lane minder sounded frequently, the experimenters verbally reminded the subject of the importance of staying in the lane.

If the secondary A/O task was to be used in the subject's cell of the experiment, he or she was given two to three minutes of practice in responding to the aurally presented words. All subjects were presented with the sequence of advisory tone, advisory message, and full alarm in the final phase of the practice session and were able to practice pressing the reset button to stop the alarm stimuli. After the practice session, the subject was excused from the simulator as the experimenters made final preparations for the data-gathering session.

Before the data gathering session, the subject was informed that he or she would be driving for approximately two hours and forty-five minutes. The subject was told that cruise control would be engaged after he or she accelerated to 60 miles per hour (mph). The subject was also told that cruise control would be disengaged if a detection occurred and to maintain a speed of 60 mph whenever cruise control was disengaged. The subject

was informed that he or she would be asked for subjective ratings of drowsiness level and for timing of the alarm following any detection. The rating scales for subjective rating of drowsiness level and for timing of the alarm were explained to the subject.

Once the subject understood all instructions, he or she returned to the simulator, the lights were dimmed, and the data-gathering session began. When the driver reached 60 mph and cruise control was engaged, the data-gathering computational equipment was started. The subject was asked to drive the simulator as he or she would drive an actual midsize car with automatic transmission. The driver was also asked to attempt to stay within the boundaries of the right lane. The computational equipment monitored the performance measures and calculate the driver's alertness and performance levels once every minute. An experimenter also constantly tracked the definitional measure PERCLOS by viewing a video image of the subject's face and tracking eyelid movement with the linear potentiometer. With these data, comparisons could be made between PERCLOS and ePERCLOS.

If either ePERCLOS or LANEX/eLANEX exceeded its threshold value, the advisory tone automatically sounded. As this tone began to sound, an experimenter disengaged cruise control. Immediately, another experimenter subjectively evaluated the drowsiness and performance levels of the subject. The voice message was automatically presented after the advisory tone. If the reset button was not pressed by the driver, the auditory alarm was automatically presented. If the auditory alarm had sounded, an experimenter would have engaged one of the peripheral stimuli (vibration or brake pulse) if it was called for in the subject's cell of the experiment.

When the subject pressed the reset button, the auditory stimuli were automatically disengaged and any peripheral stimulus in use was manually disengaged by an experimenter. At this time, the subject was told by an experimenter to maintain a speed

of 60 miles per hour. An experimenter then asked the subject for a subjective rating of drowsiness at the time that the advisory/alarm was sounded. The experimenter also asked the subject whether the timing of the alarm was too early, correct, or too late based on his or her perceived alertness level. Thereafter, if the subject's experimental cell called for a subsidiary task to be presented, an experimenter informed the subject that either the lane minder task or the A/O task would be activated.

If a subsidiary task was activated, it was manually deactivated by the experimenter upon an advisory message from the computational equipment. If a detection took place while a subsidiary task was already being presented, that task was suspended at the detection and reinstated in the same manner in which it was originally engaged.

At the end of the driving period, cruise control was disengaged and the subject was instructed to slow to a complete stop. The subject exited the simulator and was asked if he or she had any further questions about the experiment. If the subject had no further questions, he or she was paid, thanked, and driven home. The experimenter who drove the subject home was on a different sleep schedule than the subject and therefore was not drowsy.

Data Analysis Overview

The major purpose of this phase of the study was test and evaluation of the DDDACS. Therefore, both quantitative and qualitative data analyses were made. Quantitative data from two sources, numerical data collected from the simulator and subjective evaluations, were evaluated. Numerical on-line data for each minute of the experiment were computed and stored by the WIN 486-33i computer. These data included the measurement of PERCLOS, the estimated value of PERCLOS calculated by the algorithm in use for that minute, the measurement of LANEX, and the performance

measures necessary to calculate all algorithms. These data were imported into both a spreadsheet package (Microsoft Excel, version 5.0a) and a statistical package (STATISTICA for Windows, version 4.5A).

First, algorithm performance was evaluated. All algorithm performance analyses were conducted for each subject individually and for the group of six subjects as a whole. Six-minute averages of PERCLOS, ePERCLOS, LANEX, and eLANEX were calculated from the data set of each subject. These six-minute averages were independent of one another and were calculated in a manner similar to the manner in which six-minute averages were calculated for development of the unbaselined algorithms (Fairbanks, Lewin, and Wierwille, 1995). They should not be confused with the six-minute moving averages used on-line in the detection system. No data for any minute following a detection were considered in the analyses.

Once the averages were calculated, line charts were prepared comparing PERCLOS with Algorithm F4a-N, Algorithm D4a-N, and a composite of all ePERCLOS algorithm outputs selected during the course of each session by the “step up, step down” system. Pearson product-moment correlation coefficients were calculated on the relationship between PERCLOS and ePERCLOS for Algorithm F4a-N, Algorithm D4a-N, and the composite of all ePERCLOS algorithm outputs. Individual analyses for the relationship between PERCLOS and ePERCLOS for Algorithms J4a-N and L3a-N were not performed due to insufficient data. (Algorithms J4a-N and L3a-N could only be calculated while the A/O task was being performed; the A/O task was performed only for limited segments of two of the six experimental sessions.)

For comparisons to be made, both LANEX and eLANEX values were retrieved for every minute of the experimental sessions, regardless of whether the “step up, step down” system selected LANEX or eLANEX for any given minute. eLANEX values were

manually calculated from the component measures, switching between algorithms LDV-1 and LDV-4 based on whether or not the A/O task was operating. Six-minute averages were then calculated for LANEX and a composite of the eLANEX algorithm outputs. Line charts were prepared comparing LANEX with composite eLANEX, and Pearson product-moment correlation coefficients were calculated on the relationship between LANEX and composite eLANEX. Individual analyses of the two algorithms were not performed because the A/O task was performed only for limited segments of two of the six experimental sessions.

Consistency between the data obtained in the current experiment and that obtained in the algorithm development experiment (Wreggit, Kirn, and Wierwille, 1993) was evaluated. PERCLOS, LANEX, and all component measures for the six algorithms used in the current experiment were evaluated. Two-tailed *t*-tests ($\alpha = 0.05$) were used to determine if there was a significant difference between the mean value of each measure in the development (earlier) experiment and the current experiment.

Additional analyses using PERCLOS, LANEX, and LANDEV values were conducted to evaluate the effect of the alarms and the countermeasures on driver status and performance. For the 25 minutes after any detection (advisory tone is sounded and alarm sounds if driver does not press reset button immediately), measured PERCLOS values for subjects presented with no countermeasures, the lane minder task, and the A/O task were compared using analysis of variance (ANOVA) with unequal *n*'s. Similar analyses were performed with LANEX and LANDEV values. These analyses allowed the effectiveness of the two subsidiary tasks in maintaining driver alertness and in maintaining driver performance to be assessed.

Another analysis compared PERCLOS, LANEX, and LANDEV data from the algorithm development experiment (Wreggit, Kirn, and Wierwille, 1993) with data

collected in the present phase of the study. PERCLOS, LANEX, and LANDEV values from both experiments measured during 25-minute periods following a drowsy condition were compiled. The drowsy condition was defined as either a 6-minute PERCLOS average in excess of 0.15 (in data from the development experiment) or a detection by the DDDACS (in data from the present experiment). Comparisons were made between values for each measure in the two experiments for the case in which no countermeasures were presented and for the case in which the A/O task was presented. These comparisons allowed assessment as to whether the advisory tones/alarms were helping to keep drivers awake and/or helping drivers to maintain a satisfactory level of performance.

Qualitative evaluation was made in several areas. The apparent effectiveness of the integration of the status and performance components of the system and the apparent effectiveness of the “step up, step down” detection algorithm system were evaluated. Also, the compatibility of the detection, alarm, and countermeasure components of the DDDACS was qualitatively assessed.

RESULTS

Qualitative Observations

Detection

A total of 16 detections took place during the six experimental sessions. All detections were made because the six-minute moving average of LANEX and eLANEX values exceeded threshold. The six-minute moving average of ePERCLOS values never exceeded threshold. Thus, all detections were based wholly on actual or estimated decreases in driver performance; no detection was based either wholly or partly on eye closure. In addition, several extreme bouts of drowsiness noted during the experimental sessions were not detected by the system.

A history of the 16 detections, including subjective ratings and experimenter ratings, appears in Table 1. Experimenter ratings were measured from the scales in Appendix D and converted to values ranging from 0 to 100. For the drowsiness scale, “Not Drowsy” was assigned a value of zero and “Extremely Drowsy” was assigned a value of 100. For the performance scale, “Excellent” was assigned a value of zero and “Poor” was assigned a value of 100. Experimenter subjective ratings for Subject 4 were not performed during the experimental session and had to be conducted off-line at a later time. Since a videotape image of the driver’s face was the only resource available, no ratings could be made for the performance of Subject 4.

Subjective ratings of drowsiness provided by the subjects immediately following detection ranged through the entire scale from “not drowsy” to “extremely drowsy”. The most common ratings given were “moderately drowsy” and “very drowsy”; one of these two ratings was given for 11 of the 16 detections. Subjective ratings of the timing of the detections ranged through the entire scale from “much too early” to “much too late”. The

Table 1: History of Detections with Subjective Ratings and Experimenter Ratings

Subject	End of Minute	Subjective Ratings		Experimenter Ratings	
		Drowsiness	Timing	Drowsy	Performance
4	100	Moderately	About Right	67.97	not available
4	113	Slightly	About Right	53.91	not available
4	139	Not	Somewhat Early	77.34	not available
5	139	Very	Somewhat Late	78.91	62.50
6	19	Slightly	Much Too Early	3.13	73.83
6	101	Moderately	About Right	56.25	80.47
6	129	Very	Somewhat Late	47.27	43.75
6	144	Moderately	About Right	56.25	60.16
6	150	Moderately	Somewhat Early	44.53	25.00
7	59	Very	Somewhat Late	77.73	65.62
7	113	Very	About Right	60.16	28.91
7	138	Extremely	Much Too Late	88.67	71.87
8	129	Very	Much Too Late	68.75	64.84
9	109	Moderately	About Right	72.27	64.06
9	128	Very	Somewhat Late	84.38	81.64
9	158	Extremely	Somewhat Late	78.13	64.45

most common timing ratings were “about right” and “somewhat late”; one of the two ratings was given for 11 of the 16 detections.

The subjective ratings of drowsiness given by the experimenter followed a pattern similar to that of the ratings given by the subject. Six of the sixteen ratings fell between “very drowsy” and “extremely drowsy”; another six fell between “moderately drowsy” and “very drowsy”. Experimenter subjective ratings of driver performance detections were obtained for five of the six subjects (13 of the 16 detections). The performance ratings ranged from “poor” to “good”; the majority (8 of 13) fell between “fair” and “moderate”.

Advisory, Alarm, and Countermeasures

Based on experimenter observation, the combination of advisory tone and advisory message performed very well. At every detection, the subject pressed the reset button during the advisory message or immediately following it. Neither the full alarm tone nor the peripheral alarm cues following the advisory message were ever activated, because the subject had already pressed the reset button.

Quantitative Analyses

Algorithm Performance

Line charts comparing PERCLOS with Algorithms F4a-N, D4a-N, and a composite of all ePERCLOS algorithms appear in Appendix F. Charts comparing LANEX with a composite of the two eLANEX algorithms appear in Appendix G. For both measures, charts were prepared for each subject and for the group of six subjects as a whole. For all cases in which line charts were prepared, correlation coefficients were also calculated. A summary of the r values obtained in the correlation analyses appears in Table 2. An asterisk (*) in the subject column denotes that the A/O task was used during that subject's experimental session and that composite algorithm outputs for that subject include outputs for the A/O task algorithms (J4a-N, L3a-N, and LDV-4). Correlation coefficients for comparisons involving PERCLOS data for Subject 6 were not applicable because the six-minute average PERCLOS value was zero for the entire experimental session.

Consistency of Data

Mean values for PERCLOS, LANEX, and each component measure were calculated for each subject in data from the current experiment and for each subject in data from the

Table 2: Summary of R Values Obtained in Correlation Analyses.

<i>Subject</i>	<i>Composite ePERCLOS R</i>	<i>Algorithm F4a-N R</i>	<i>Algorithm D4a-N R</i>	<i>Composite eLANEX R</i>
*4	0.2395	0.2770	-0.0011	0.6493
5	0.7843	0.7820	0.6307	0.6648
6	not applicable	not applicable	not applicable	0.5225
*7	0.5549	0.5169	0.2519	0.7225
8	0.6052	0.6267	0.4237	0.4697
9	0.3356	0.2558	0.4599	0.1629
All	0.4839	0.5435	0.2161	0.4408

algorithm development experiment (Wreggit, Kirn, and Wierwille, 1993). Using these means as data points, grand means and standard deviations were calculated for each measure in each experiment. Two-tailed *t*-tests ($\alpha = 0.05$) were conducted on the difference between the two grand means for each measure. A summary of these analyses appears in Table 3. Mean PERCLOS values for each subject were not significantly different between the two experiments. However, mean LANEX values were significantly lower for the new data than for the development data. Some of the mean component measure values for each subject were significantly lower for the new data than for the development data. These measures are denoted by asterisks (*) in Table 3.

Effects of Alarms and Countermeasures on Status and Performance

Summaries for nine single factor, unequal *n*'s analyses of variance (ANOVAs) are presented in Appendix H. Three ANOVAs were calculated on countermeasure type with PERCLOS (Table H1), LANEX (Table H2), and LANDEV (Table H3) as dependent measures. Countermeasure type was shown to be a significant effect on each of the three dependent variables.

Table 3: Summary of Data Consistency Analyses Between Algorithm Development Experiment and Test and Evaluation Experiment.

Development (12 subs)	PERCLOS	LANEX	ACCVAR	ACCDEV
grand mean	0.0660	0.1421	4.1321	1.7156
standard deviation	0.05566	0.09647	0.5696	0.1404
Test & Eval. (6 subs)	PERCLOS	LANEX	ACCVAR	ACCDEV
grand mean	0.01605	0.02429	3.0879	1.3749
standard deviation	0.02022	0.009677	0.2472	0.09666
σ	0.04976	0.08376	0.5160	0.1353
t	2.0075	2.8133	4.0476	5.0376
t _{.025,16}	2.12	2.12	2.12	2.12
* = significant		* (p<0.0125)	* (p<0.0010)	* (p<0.0002)

Development (12 subs)	INTACDEV	LGREV	STEXED	NMRHOLDS
grand mean	0.3198	0.8778	5.991×10^{-5}	23.686
standard deviation	0.01025	0.5672	7.876×10^{-5}	2.8134
Test & Eval. (6 subs)	INTACDEV	LGREV	STEXED	NMRHOLDS
grand mean	0.2657	0.2673	5.643×10^{-6}	24.748
standard deviation	0.009038	0.2504	1.382×10^{-5}	3.4855
σ	0.01046	0.5146	6.873×10^{-5}	3.2391
t	10.342	2.3728	1.5792	-0.6563
t _{.025,16}	2.12	2.12	2.12	2.12
* = significant	* (p<0.0001)	* (p<0.0306)		

Development (12 subs)	THRSHLDS	STVELV	LANDEV	LNERRSQ
grand mean	0.2268	44.137	1.9968	4.3446
standard deviation	0.08318	17.749	0.9560	6.4054
Test & Eval. (6 subs)	THRSHLDS	STVELV	LANDEV	LNERRSQ
grand mean	0.2029	26.763	0.8964	0.02222
standard deviation	0.06887	9.0637	0.06311	0.004717
σ	0.08347	16.343	0.8288	5.5472
t	0.5707	2.1262	2.6551	1.5584
t _{.025,16}	2.12	2.12	2.12	2.12
* = significant		* (p<0.0494)	* (p<0.0173)	

Six ANOVAs compared the current experiment to the algorithm development experiment (Wreggit, Kirn, and Wierwille, 1993) for periods of twenty-five minutes after a drowsy condition, defined as either a detection or a six-minute PERCLOS average in excess of 0.10. For data taken when no countermeasure was operating, ANOVAs were calculated with PERCLOS (Table H4), LANEX (Table H5), and LANDEV (Table H6) as dependent measures. For data taken when the A/O task was operating, ANOVAs were also calculated with PERCLOS (Table H7), LANEX (Table H8), and LANDEV (Table H9) as dependent measures. Each dependent measure was shown to have a significantly lower value in the current experiment than in the algorithm development experiment for both the no-task case and the A/O task case.

DISCUSSION

Qualitative Observations

Detection

The DDDACS was designed to alert the driver based on a decrease in the alert status of the driver (as estimated by a six-minute average of ePERCLOS) or a decrease in the driver performance level (as estimated by a combined six-minute average of LANEX and eLANEX). Since all detections were based wholly on performance (that is, LANEX/eLANEX), the performance criterion is considered to be much stricter than status. In fact, it completely dominated the detection process. The appearance that extreme bouts of drowsiness often went undetected was further substantiated with examination of the line charts comparing PERCLOS and ePERCLOS that appear in Appendix F. The combination of these observations suggests that the ePERCLOS algorithms were not operating correctly. For this reason, subsequent quantitative analyses were undertaken to locate the source of the problem.

Drowsiness levels at detections were classified as “moderately drowsy” or “very drowsy” by a majority of the subjects; the experimenter classified a majority of the same detections between “moderately drowsy” and “extremely drowsy”. This suggests that though the system may have missed some detections, there were very few “false alarms”. The majority of timing ratings given by the subjects were either “about right” or “somewhat late”; this suggests that the system tended to be late in making its detections.

Advisory, Alarm, and Countermeasures

The sequence of advisory tone and advisory message was very effective in alerting the subject; neither the full alarm tone nor the peripheral alarm stimuli were necessary to

awaken the subject. Driver alertness and performance appeared to improve following the advisory sequence and with use of the countermeasures; subsequent quantitative analyses were undertaken to determine whether the effects were significant.

Quantitative Analyses

Algorithm Performance

The charts appearing in Appendices F and G demonstrate that neither the ePERCLOS algorithms nor the eLANEX algorithms are tracking well with the measures they were designed to predict. The correlation coefficients reported in Table 2 are much lower than expected. Algorithm F4a-N (containing lane-related measures) had an R value of 0.857 when applied to development PERCLOS data and 0.871 when applied to validation PERCLOS data; it had an R value of only 0.544 when applied to the current PERCLOS data set. Algorithm D4a-N (no lane-related measures included) had an R value of 0.788 when applied to development PERCLOS data and 0.769 when applied to validation PERCLOS data; it had a value of only 0.216 when applied to the current PERCLOS data set. The eLANEX algorithms LDV-1 and LDV-4 had R values of 0.830 and 0.833, respectively, when applied to combined development and validation LANEX data; the composite of the two algorithms had an R value of only 0.441 when applied to the current LANEX data set. Clearly, the algorithms did not perform as intended in the current experiment. An examination of Table 2 and Appendices F and G also shows little consistency in the performance of the algorithms between subjects, with R values and apparent performance characteristics ranging widely. These results led to a comparative examination of current data and development data in an attempt to determine if significant differences exist between the two data sets.

Consistency of Data

The *t*-test results in Table 3 do not show a significant difference in the grand mean of PERCLOS between the development experiment and the current experiment. This is expected because the subjects were sleep-deprived in the same manner in the two experiments and PERCLOS was tracked in the same manner. LANEX and LANDEV were significantly lower in the current experiment than in the development experiment. This suggests that lane-keeping behavior was different in the development experiment than in the current experiment. In fact, each subject received training in lane-keeping behavior before his or her experimental session in the current experiment, resulting in strict adherence to lane boundaries reflective of that which would be exhibited in an actual vehicle. This training was not used in the algorithm development and validation experiments; in those experiments, subjects exhibited large lane deviations which are now recognized as atypical of actual driving.

The other measures which were significantly lower in the current experiment (ACCVAR, ACCDEV, INTACDEV, LGREV, STVELV) are not measured directly from the position of the vehicle relative to the lane. However, they all are in some way related to lateral movement of the vehicle, as measured by a lateral accelerometer, by variances in steering velocity, or by large steering reversals. All four ePERCLOS and both eLANEX algorithms rely heavily on measures that were significantly lower in the current experiment than in the development experiment.

Effects of Alarms and Countermeasures on Status and Performance

The ANOVA results in Appendix H show that the countermeasures had a significant alerting effect on subjects. For cases in which the no countermeasure was operating and for cases in which the A/O task was operating, the advisory tone-message

sequence had a positive effect on driver alertness. Levels of LANEX and LANDEV were shown to be significantly lower in the current experiment than in the development experiment for periods following a defined drowsy condition. However, this significant difference was also shown in comparison of the whole data sets; therefore, it is not possible to determine whether the advisory tone-message sequence had a positive effect on driver performance.

Conclusions

The performance aspect of the system dominated the detection process in this experiment. Investigation of this unanticipated result showed that algorithms developed previously as estimators of PERCLOS and LANEX did not function as intended. Correlations between the algorithm outputs and the measures they predict (eye closure and lane excursions) were lower than expected. The mean values of many of the measures upon which the algorithms were based, particularly those related directly or indirectly to lateral position, were shown to be significantly lower in the current experiment than in the development experiment. Training procedures in the current experiment enforced lane behavior typical of driving in an actual vehicle. Therefore, the algorithms were developed based on measures calculated from data for which the subjects were more tolerant of lane errors than they would have been in an actual vehicle. It appears that the algorithms do not function well when lane errors are controlled to a level more closely reflecting full scale. They produce inaccurate estimates of LANEX and PERCLOS.

The advisory tone/advisory message combination is sufficient to alert drowsy subjects without the need for any further auditory or peripheral alarms. The advisory tone/advisory message combination and the lane minder and A/O task countermeasures

appear to have beneficial effects on both driver status and performance. However, the power of statistical tests conducted in investigation of these effects is weak because there is not a sufficient number of data points. Also, the inconsistency of data between previous experiments and the current experiment confounds comparisons as to whether the alarms and countermeasures had an appreciable effect on driver performance.

Recommendations

Originally, the present phase of the study was to include 18 participants. However, the data already gathered on six participants are sufficient to demonstrate that algorithms developed from data collected in the original algorithm development experiment are not reliable within the paradigm of the current experiment. The likely cause is that drivers were more tolerant of lane errors in the simulator during that experiment than they would have been in an actual vehicle.

Data gathering in this experiment should be discontinued. In its place, a new experiment, conducted in a manner similar to the algorithm development and validation experiments (Wierwille et al., 1994), should be designed. The new experiment should differ from the previous experiments in that drivers would be required to perform the lane-keeping task in a manner more closely reflecting performance in an actual vehicle. New data should be collected under such a paradigm, and new algorithms should be developed from the new data. It is possible that some data collected in the present experiment can be integrated with the new data in algorithm development.

The new experiment should not include the use of advisory tones, alarms, or countermeasures as they were used in the present experiment. As a result, no further analysis of the effects of the advisory sequence or countermeasures on driver status and

performance would be possible in the new experiment. Instead, emphasis in the new experiment should be on development of algorithms having the highest probability of success when used in full-scale (field) tests.

PHASE THREE

DEVELOPMENT OF REVISED ALGORITHMS FOR DETECTION OF DRIVER DROWSINESS

Research Objectives

In Phase Two of the present study, a complete drowsy driver detection, alarm, and countermeasures system (DDDACS) was tested and evaluated. The system was designed to detect the drowsy status of a driver or a reduced driver performance level. Driver drowsiness levels were determined by the output of one of a number of algorithms designed to estimate PERCLOS (the proportion of time that a driver's eyes are 80% to 100% closed). Driver performance levels were measured by LANEX (the proportion of time any part of the vehicle exceeds the lane boundary) or an algorithm estimate of LANEX. The system was designed to be able to operate with or without the availability of lane-related information by switching between algorithms in a "step-up, step-down" process.

The results of the test and evaluation phase were quite different than what was expected. All detections that took place during the experiment were based wholly on driver performance; no detection was based wholly or partly on drowsiness levels. Neither the driver status (drowsiness) algorithms nor the driver performance (LANEX estimation) algorithms tracked well with the measures they were designed to predict; correlations were much lower than expected. Comparisons between independent measures from the data collected in the current experiment and the data previously collected in the algorithm development experiment (Wreggit, Kirn, and Wierwille, 1993) revealed significant differences in the values of the means for many of the measures. Specifically, measures related directly or indirectly to the position of the vehicle relative

to the lane had significantly lower mean values in the test and evaluation experiment than in the algorithm development experiment.

It was concluded that subjects were more tolerant of lane errors in the previous algorithm development experiment than they would have been in an actual vehicle. It appeared that algorithms developed from that data set did not function well when lane errors were controlled to a level more closely reflecting full scale. Therefore, data gathering in the test and evaluation experiment was discontinued.

The primary goal of this phase of the present study was to develop new algorithms that would have a higher probability of success than the previous algorithms when used in full scale. To this end, an experiment similar in nature to the original algorithm development experiment (Wierwille et al., 1994) was designed, in which no detections, alarms, or countermeasures were used. The primary difference between the current experiment and the original algorithm development experiment is the requirement that drivers perform the lane-keeping task in a manner reflecting performance in an actual vehicle. This was accomplished through the implementation of a lane departure warning system, which will be described in detail later in this document. Also, several new definitional measures of drowsiness were developed and implemented in this experiment, so that a larger variety of algorithms could be developed, examined, and compared.

METHOD

Subjects

Twelve volunteer subjects (eight male and four female) were used in this phase of the study. The use of a two to one ratio of male to female subjects has been determined to be a more accurate representation of the population at high risk for driver drowsiness than a one to one ratio of males to females. Recent research has shown that males substantially outnumber females in drowsiness-related automobile accidents (Knipling and Wierwille, 1994). All participants were volunteers from the Blacksburg, Virginia area ranging in age from 21 to 45 years. This age range corresponds to the population most heavily involved in drowsiness-related accidents (Knipling and Wierwille, 1994).

All subjects were screened according to a questionnaire, which included questions concerning normal sleeping habits, normal working hours, smoking habits, general health, and body size. This questionnaire appears in Appendix C. It was required that all subjects possess a valid driver's license, have 20/40 vision or better (corrected or uncorrected), and have no known hearing problems.

Each subject was compensated for his or her one-time participation in the experiment. At 6:00 P.M., each subject was picked up at his or her home and received \$6.00 for dinner. Each subject was paid \$5.00 per hour from 6:00 P.M. until midnight and \$8.00 per hour from midnight until the end of the experiment. Each subject also had the opportunity to earn a bonus of up to \$12.00 for good driving performance.

During the one of the experimental sessions, it became evident that the signal from the drowsiness control (see Apparatus section) was not reaching the WIN 486-33i microcomputer and that no data from the drowsiness control were entering the data file. Since this resulted in an incomplete data set, the experimental session was terminated and

the subject's data were not used. Another subject experienced symptoms of oncoming uneasiness early in the run. The experimental session was terminated at that subject's request and the subject's data were not used. The two problem subjects were paid for their time and replaced with two additional subjects, resulting in collection of valid data from a total of twelve subjects.

Apparatus

Simulator

The simulator used in this study is a computer-controlled, hydraulically powered moving-base automobile simulator that handles like a mid-sized rear wheel drive automobile. This simulator has been validated by Leonard and Wierwille (1975) and is located at the Vehicle Analysis and Simulation Laboratory at Virginia Tech. Previous studies, including the original detection algorithm development and validation experiments, the optimization of advisory and alarm stimuli experiments, and the test and evaluation experiment, have used the same simulator (Wierwille et al., 1994; Fairbanks, Fahey, and Wierwille, 1995).

The simulator has four degrees of freedom of physical motion (roll, yaw, lateral translation, and longitudinal translation). The roadway image was presented using a monochrome CRT viewed through a Fresnel lens. The image was that of a two-lane highway with side markers and a dashed center line. Light horizontal lines were embedded in the horizontal plane to enhance the image of the roadway continuing into the distance. Both straight and curved roadways were presented. A simulated automobile hood was also included in the image.

Roadway vibration and sounds, such as engine noise, tire squeal on hard turns, and tire screech on hard braking were also presented to the driver. The ambient sound level in the simulator at 60 miles per hour was set at 75.5 dBA.

Eye Closure Monitoring Equipment

A low light level camera (RCA TC1004-UO1) was used to continuously monitor a subject's entire face, including eye movements. The camera could operate at very low light levels and thus be unintrusive. The video signal was passed through a video cassette recorder and viewed by an experimenter using a Sanyo VM 4512A monitor. This experimenter manipulated a specially designed linear potentiometer to track the movement of the subject's eyelids. This tracking produced values for the measures PERCLOS and EYEMEAS. Video and audio recordings of each session were made and kept for future reference.

Subjective Drowsiness Rating Equipment

Another potentiometer was mounted on the center console of the simulator, to the right of the driver's seat. This potentiometer was labeled with the same range of drowsiness levels previously used for subjective ratings in the test and evaluation experiment ("Not Drowsy", "Slightly Drowsy", "Moderately Drowsy", "Very Drowsy", and "Extremely Drowsy"). The subject manipulated this potentiometer during the course of the experiment to rate his or her perceived level of drowsiness. An LED mounted on the dashboard flashed once per minute as a reminder to the subject to consider moving the potentiometer if there had been a change in his or her drowsiness level. The light output of the LED was of a yellow hue and its luminance was adjusted so as not to be

intrusive. Output from the potentiometer produced values for the dependent measure DRVDROW.

Lane Enforcement Equipment

During the test and evaluation experiment, it became apparent that drivers were more tolerant of lane errors in the simulator than they would have been in an actual vehicle. As a result, controls that required subjects to perform the lane-keeping task in a manner more closely reflecting performance in an actual vehicle were used in the present experiment. A lane minder was activated during the training session and during the first five minutes of each experimental session. This device was activated via circuitry connected directly to the simulator and sounded a tone whenever the vehicle approached lane boundaries. The tone would increase in amplitude as the simulated vehicle approached the lane edge. As it passed over the lane edge, the tone would begin to warble. Maximum amplitude would occur when the vehicle exceeded the lane edge by more than two feet. The tone was presented via dual piezo buzzers, one to the left and the other to the right of the driver. The usage of the lane minder served to acclimate each subject to staying within the lane boundaries.

A second lane enforcement system, the “lane departure warning system”, alerted drivers to extreme exceedance of lane boundaries, using seat vibration to simulate the effect of a rumble strip. Installations of rumble strips along the shoulders of roadways have proved to be highly effective in reducing drift-off-road accidents (Wood, 1994, Chaudoin and Nelson, 1985), and seat vibration has been shown to be an effective peripheral alarm stimulus for driver drowsiness (Fairbanks, Fahey, and Wierwille, 1995). The system was used throughout each experimental session and was automatically activated whenever the vehicle crossed boundaries set at equal distances approximately

three feet beyond the left and right boundaries of the right lane. When these boundaries were crossed, vibration was produced in the seat back and seat pan with the use of eccentrics (unbalanced rotational masses) driven by high-quality servo motors. These motors were powered by two power supplies connected to the simulator via a signal processor.

Data Collection Equipment

During the experiment, a variety of analog sensors on the simulator were operating. Two serially interfaced microcomputers equipped with special interface cards received the analog data and converted it to digital format, calculated necessary measures on-line, and stored data for later analysis.

Timing for the system originated from a Sony MDS-302 MiniDisc recorder, a high-quality digital audio recording device capable of recording two audio channels on a 74-minute optical disc. When played, the disc can be repeated indefinitely without degradation. A signaling pulse of 18,000 Hz occurred every 15 seconds to control interval timing.

The pulses from the MDS-302 recorder were fed into a TRS-80 Model III microcomputer via a custom analog-to-digital (A/D) converter interface. A BASIC program running on the TRS-80 counted the pulses. Upon receipt of every fourth pulse (marking the passage of one minute), the TRS-80 sent a "flag" signal to a WIN 486-33i microcomputer via a serial RS-232 interface.

The WIN 486-33i microcomputer was equipped with a National Instruments AT-MIO-16 A/D converter interface card. This card allowed for rapid digital sampling of analog data on 16 different channels. The computer was programmed using Microsoft QuickBASIC to collect and store raw data and to calculate the necessary measures on-line

from that data. Every minute, upon receipt of the “flag” signal from the TRS-80, the program computed the measures for that minute and saved them in a file for subsequent analysis.

All signals from various sensors on the simulator from which independent measures were derived and subsequently used in algorithm redevelopment were sampled, converted, and calculated by the WIN 486-33i. The output signals from the two potentiometers for the measures PERCLOS, EYEMEAS and DRVDROW were handled in the same manner. Also, a signal from the seat vibration system was processed by the WIN 486-33i, recording when and for how long the seat vibration was activated.

Data Analysis Equipment

All manipulation and analysis of data and charting of results were performed on IBM PC-compatible 486 microcomputers running the Microsoft Windows operating system. All multiple regressions were performed using STATISTICA for Windows, version 4.5A. All data manipulation procedures, correlation analyses, distribution analyses, classification analyses, and charting were performed using Microsoft Excel, version 5.0a.

Experimental Design

The experimental design employed a multiple regression approach to data analysis. The experiment was conducted expressly for the purpose of gathering data for the development of new algorithms; therefore, all subjects were exposed to the same treatment conditions. A set of definitional measures of drowsiness was collected, as well as a set of independent driving performance measures. The independent performance

measures were then used in regression analysis in various configurations to predict the definitional drowsiness measures.

Two of the definitional measures of drowsiness collected were also collected in the previous algorithm development and validation experiments (Wierwille et al., 1994).

- EYEMEAS: The mean square of the percentage of the subject's eye closure.
- PERCLOS: The proportion of time that a subject's eyes were 80% to 100% closed.

Four other definitional measures of drowsiness were collected and/or calculated for the first time in the current experiment.

- DRVDROW: The subject's online subjective rating of drowsiness, which he or she inputted using the drowsiness control. Values ranged from zero (not drowsy) to 100 (extremely drowsy).
- SLEEPER1: The sum of weighted values of EYEMEAS and DRVDROW. Weighting was accomplished using values for each measure considered to signify a "very drowsy" level of drowsiness:

$$\text{SLEEPER1} = (1/2500) (\text{EYEMEAS}) + (1/75) (\text{DRVDROW})$$

- SLEEPER2: The sum of weighted values of EYEMEAS and DRVDROW. Weighting was accomplished using the mean values of three-minute averages for each measure:

$$\text{SLEEPER2} = (1/1179.8) (\text{EYEMEAS}) + (1/55.6) (\text{DRVDROW})$$

- SLEEPER3: The sum of weighted values of PERCLOS and DRVDROW. Weighting was accomplished using values for each measure considered to signify a "very drowsy" level of drowsiness:

$$\text{SLEEPER3} = (1/0.014) (\text{PERCLOS}) + (1/75) (\text{DRVDROW})$$

A listing of the independent measures collected follows, grouped by category. All of these measures were collected in the previous algorithm development and validation experiments (Wierwille et al., 1994).

Steering-Related Measures:

- STVELV: The variance of steering velocity, where velocity was measured in degrees per second.
- LGREV: The number of times that steering excursion exceeds 15° after steering velocity passes through zero.
- MDREV: The number of times that steering excursion exceeds 5° , but does not exceed 15° , after steering velocity passes through zero.
- SMREV: The number of times that steering excursion exceeds 1° , but does not exceed 5° , after steering velocity passes through zero.
- STEXED: The proportion of time that steering velocity exceeds 125° per second.
- NMRHOLD: The number of times the hold circuit output on the steering wheel exceeds a threshold value (corresponding to holding the steering wheel still for 0.4 second or longer).
- THRSHLD: The proportion of total time the hold circuit output on the steering wheel exceeds a threshold value.

Lane-related measures:

- LNMNSQ: The mean square of lane position; “zero” position is defined as that position occurring when the vehicle is centered in the lane. (Lane position was measured in feet.)

- LANVAR: The variance of lateral position relative to the lane. (Lane position was measured in feet.)
- LANDEV: The standard deviation of lateral position relative to the lane; the square root of LANVAR.
- LANEX: The proportion of time that any part of the vehicle exceeds the lane boundary.
- LNERRSQ: The mean square of the difference (in feet) between the outside edge of the vehicle and the lane edge when the vehicle exceeds the lane. When the vehicle does not exceed the lane, the contribution to the measure is zero.
- LNRTVAR: The variance of the time derivative of lane position. (Lane position was measured in feet.)
- LNRTDEV: The standard deviation of the time derivative of lane position (square root of LNRTVAR).

Accelerometer-related measures:

- ACCVAR: The variance of the smoothed output of the accelerometer, where the output was first converted to feet per second-squared. (Smoothing was accomplished with a low-pass filter having a corner frequency at 7.25 Hz.)
- ACCDEV: The standard deviation of the smoothed output of the accelerometer. (Square root of ACCVAR).
- INTACVAR: The variance of the lateral velocity of the vehicle. (This signal will be obtained by passing the smoothed accelerometer signal through an additional low pass filter with a corner frequency of 0.004 Hz. The unit of output was volts in which one unit (volt) corresponds to a smoothed lateral velocity of 73.34 feet per second.)

- INTACDEV: The standard deviation of the lateral velocity of the vehicle. (Square root of INTACVAR).
- ACXEED: The proportion of time that the magnitude of lateral acceleration exceeded 0.3 g (9.66 ft/second²).

Heading-related measures:

- HPHDGVAR: The variance of the high-pass heading signal in degrees. (The heading signal was passed through a single-pole high-pass filter with a corner frequency of 0.016 Hz.)
- HPHDGDEV: The standard deviation of the high-pass heading signal (square root of HPHDGVAR).

Lane-departure measures:

- VIBPROP: The proportion of time that the seat vibration system was activated.
- NUMVIB: The number of times that the seat vibration system was activated.

Procedure

Subjects who were selected by the research team after screening were contacted and scheduled for a particular date. On the scheduled day, each subject was asked to awaken by 7:00 A.M. The subject was informed that he or she should carry on normal daily activities, but should not take any naps.

Each participant was picked up by an experimenter at 6:00 P.M. and taken to dinner at a fast food restaurant. At dinner, the subject was reminded not to ingest any caffeinated substances or sugared beverages. The subject was permitted to smoke immediately following dinner, but not thereafter.

The participant was brought to the laboratory after dinner. He or she was allowed to watch television, read, study, watch a movie on a VCR, listen to music on headphones, etc. An experimenter remained with the subject until midnight to ensure no napping. During this time, the experimenter gave the participant an information sheet and an informed consent form describing the events to take place in the experiment. These documents appear in Appendix E. The subject was asked to read and sign the consent form.

At midnight, two rested experimenters arrived and relieved the experimenter who stayed with the subject. Immediately afterward, the subject entered the simulator for a training and practice driving session. The subject was instructed in procedures for terminating the experiment if it became necessary. The subject was also instructed in the operation of the drowsiness control device, with an emphasis on the importance of providing the experimenters with accurate data. The rating scale for subjective rating of drowsiness level was explained to the subject. However, the seat vibration system was not introduced to the subject before or during the practice session; this was done so that the subject would not become overly familiar with the outer lane excursion limits used in

that system. The subject was asked if he or she had any questions. Once all questions were answered, the lights were dimmed and the practice driving session began. While driving during the first few minutes of this session, the subject had the opportunity to change lanes and alter speed on straight and curved roadways.

Once the subject had become accustomed to the simulator, he or she practiced staying in the right-hand lane of the simulated roadway. To help the subject in recognizing where the lane boundaries were, the experimenters activated the lane-minder device. Initially, the experimenters instructed each subject to drive out of the lane on each side to become accustomed to the lane minder. Then, the subject was instructed to stay in the right-hand lane. If the lane minder sounded frequently, the experimenters verbally reminded the subject of the importance of staying in the lane.

After a few minutes of lane-minder practice, the subject was given a few minutes of practice in adjusting the drowsiness control and using the dashboard-mounted LED as a reminder to adjust the control. After four to five more minutes, the practice session ended. The subject was excused from the simulator as the experimenters made final preparations for the data-gathering session.

Before the data gathering session, the subject was informed that cruise control would be engaged after he or she accelerated to 60 miles per hour (mph). The subject was also told that cruise control would be disengaged if a large lane excursion occurred and to maintain a speed of 60 mph whenever cruise control was disengaged. The subject was reminded to drive as if leaving the right lane could result in accident or injury and to assume that oncoming traffic was possible in the left lane.

At this time, the subject was told about the lane departure warning system and informed that excessive lane excursion on either side would result in a “rumble effect”. The subject was told that he or she would have the opportunity to earn a maximum of \$12

in bonus money and that the bonus was offered as an incentive to reinforce and reward good driving performance. The subject was told that \$4 would be subtracted from the bonus each time the “rumble effect” came on. After three such incidents, the subject’s bonus would be gone.

Once the subject understood all instructions, he or she returned to the simulator, the lights were dimmed, and the data-gathering session began. When the driver reached 60 mph and cruise control was engaged, the data-gathering computational equipment was started. The subject was asked to drive the simulator as he or she would drive an actual midsize car with automatic transmission, staying within the boundaries of the right-hand lane. The computational equipment monitored the performance measures and averaged and recorded each measure once each minute. An experimenter constantly tracked eye closure by viewing a video image of the subject's face and tracking eyelid movement with the linear potentiometer. The subject inputted his or her perceived drowsiness level throughout the run with the drowsiness control.

The lane minder system operated during the first five minutes of the experimental run. Afterwards, it was turned off, and the subject was told that the lane minder had been deactivated. If the outer lane excursion limits were exceeded at any time during the experimental session, seat back and seat pan vibration were automatically activated and remained in operation until the driver brought the vehicle back within the outer limits. Occurrences of vibration and the proportion of time within each minute that vibration was active were recorded in the data file. If vibration occurred, an experimenter immediately subjectively evaluated the drowsiness level of the subject.

At the end of the driving period, cruise control was disengaged and the subject was instructed to slow to a complete stop. The subject exited the simulator and was asked if he or she had any further questions about the experiment. If the subject had no further

questions, he or she was paid, thanked, and driven home. The experimenter who drove the subject home was on a different sleep schedule than the subject and therefore was not drowsy.

Data Analysis Overview

The one-minute average data computed and stored by the WIN 486-33i computer for each minute of the experiment served as the starting point for data analysis. A data truncation procedure was then undertaken to ensure that only realistic data were used in subsequent analyses. An examination of the data sets for each subject revealed that if a subject exceeded the outer lane excursion limits and caused the seat vibration effect to be activated more than four times, subsequent vibration activations became frequent and sustained extreme drowsiness levels were noted. As a result, it was concluded that drivers of actual vehicles likely would have either pulled off the road to rest or become involved in an accident by the fourth activation of the seat vibration system. Using this as a guideline, data taken after the fourth vibration for each subject who activated more than four vibrations were deleted and not considered in subsequent analyses. The point at which data were truncated is shown by the double line in Table 4. Following truncation, the data for each measure were plotted and examined. One subject had a minute of data (Subject 12, minute 60) that contained outlying data points for several measures; these data points were clipped, limiting them to more realistic maximum values.

Additional data sets were then created in which the data were averaged over two-minute, three-minute, and six-minute intervals. The two-minute interval data were calculated by taking an average of the data for two one-minute intervals for each variable. For example, one-minute intervals one and two were averaged to give two-minute

Table 4: History of Vibration Activations (data truncated for each subject after fourth vibration)

Subject	Minute of Vibration Occurrence							
	1st	2nd	3rd	4th	5th	6th	7th	8th and higher
1	100	140	148	154				
2	57	78	124	136	137	147		
3	6 [†]	20 [†]	55 [†]	70	74 [‡]	75 [‡]	76 [‡]	many more
4	32	72	84	85	89	93	93	3 more
5	100							
6	61 [†]	88	149					
8*	45	47	56	57	60	63	65	many more
9								
10	28	49	86	112	114	115	124	3 more
11	100	104	113	126	131	133	133	2 more
12	45	48	60	71	71	73	75	many more
13	95							

* Data for Subject 7 were not used; Subject 7 was replaced with Subject 13.

[†] Not considered as a “drowsy” activation of lane departure warning system in distribution analysis.

[‡] These activations of the lane departure warning system were used in the distribution analysis in place of earlier activations not considered to be “drowsy” activations.

interval one, one-minute intervals three and four were averaged to give two-minute interval two, and so on. Three-minute and six-minute interval data were calculated likewise with their respective numbers of intervals used for averaging.

Following averaging, correlation analyses were performed on the four resulting data sets. Values for three driver status measures (EYEMEAS, PERCLOS, and DRVDROW) and two driver performance measures (LANDEV and LANEX) were correlated with all other values in the data set. The objectives were to determine which performance measures were the best predictors of drowsiness, to determine how well the definitional measures of drowsiness correlated with one another, and to determine the best time

interval over which to average the data (1, 2, 3, or 6 minutes). Later, a similar analysis was performed in which the composite status measures SLEEPER1, SLEEPER2, and SLEEPER3 were correlated with all other measures in the data set.

For the driver status measures PERCLOS, EYEMEAS, and DRVDROW and the driver performance measures LANEX, LANDEV, and LNMNSQ, a set of distribution analyses were performed. The one-minute interval and three-minute interval data sets were used for these analyses. First, any activations of the seat vibration system for which neither the subject nor the experimenter reported that the subject had a high level of drowsiness were identified. There were four such activations, which are noted in Table 4. The activations were replaced with subsequent activations of the seat vibration system if they were available (also noted in Table 4). Driver drowsiness and performance data points recorded immediately before the remaining activations of the seat vibration system were isolated into a separate data set. In the remaining one-minute and three-minute data sets, data points occurring near activations of the seat vibration system were eliminated. In the one-minute data set, all data points occurring within 10 minutes before a vibration occurrence or within 5 minutes after a vibration occurrence were eliminated; in the three-minute data, all data points occurring within 9 minutes (3 intervals) before a vibration occurrence or within 6 minutes (2 intervals) after a vibration occurrence were eliminated. Once the data sets were separated and manipulated in this manner, histograms were charted so that the distributions of values for each of the six measures during drowsy periods and non-drowsy periods, as determined by activation of the seat vibration system, could be compared with one another.

The multiple regression analyses from which new driver status and performance algorithms were developed were performed using the three-minute interval and six-minute interval data sets (with truncation of unrealistic data, but without the data

separation that was used for the histograms). As in previous experimentation, driver status algorithms were developed with the purpose of finding optimized combinations of independent measures obtainable while driving that would best predict “drowsiness” during driving sessions. Multiple regression has many advantages for use in a drowsy driver detection system that have been documented previously (Wierwille et al., 1994).

Backward stepwise multiple regression analyses were performed on all twelve subjects. The beta weights of the various measures were first examined, allowing for the removal of measures that were strongly linearly related. Any measures that contained large, equal and opposite coefficients (and thus contained approximately the same predictive information) were eliminated one at a time. Then, the elimination of measures with $p > 0.05$ significance levels began, starting with the measure having the smallest F-ratio. This continued until only significant measures remained in the analysis. Afterward, each measure which had been removed was substituted back into the set to determine if it substantially improved the regression.

Several sets of algorithms intended to detect the drowsy status of a driver in an on-board detection system implementation were developed. Algorithms in each set were developed to estimate PERCLOS, EYEMEAS, DRVDROW, SLEEPER1, SLEEPER2, and SLEEPER3. Each set of algorithms used a slightly different set of independent measures so that the loss of certain measures for which data collection problems might arise in an actual vehicle on actual roadways will not cause the failure of the entire detection system. Specifically, it is anticipated that lane-related measures would be collectable for a majority of the time that a detection system is in operation, but would not be collectable for certain intervals. Previous studies have shown that the best driver status algorithms usually contain lane-related measures, but alternative algorithms that do not contain lane-related measures are desirable for use during intervals when the lane

Table 5: Sets of Measures Used In Multiple Regression Analyses for Each Dependent (Definitional) Variable.

Independent Measures		Dependent Measures					
		PERCLOS	EYEMEAS	DRVDROW	SLEEPER1	SLEEPER2	SLEEPER3
D	Steering and Accelerometer						
E	Steering, Accelerometer, & HPHDGDEV/VAR						
F	Steering, Accelerometer, LANDEV/VAR, LNMNSQ, LANEX, & LNERRSQ						
G	Steering, Accelerometer, & all lane measures (includes LNRTDEV/VAR)						

boundaries are not detectable. This concept has been referred to as the “step up, step down” approach. Table 5 shows the different sets of measures used in the regression analyses. (In Table 5, “accelerometer” refers to lateral accelerometer.)

Following development of the algorithms, classification analyses were performed. For all algorithms with correlation coefficients higher than 0.4, scatterplots were generated with algorithm output values on one axis and the measured values the algorithms were designed to predict on the other. Thresholds were set based on examination of the scatterplots, and classification matrices were subsequently constructed for each algorithm. The classification matrices examined the predictability of awake and drowsy categories of wakefulness at the threshold value.

Following development and classification of the driver status algorithms, the effect of combining driver status and driver performance on the classification matrices was examined. The arrangement was similar to the “or” criterion used previously in the test and evaluation experiment. In this arrangement, the driver was classified as impaired if the drowsiness level (status algorithm output) exceeded its threshold, if the performance level (measured by LANEX or LNMNSQ) exceeded a threshold, or if both measures

exceeded their threshold values. For those independent measure groups where lane-related measures were not included, new algorithms were developed using the independent measure groups to predict LANEX and LNMNSQ, and the status algorithm outputs were “or”ed with the resulting eLANEX (estimated LANEX) and eLNMNSQ (estimated LNMNSQ) outputs.

RESULTS AND DISCUSSION

Correlation Analyses

Tables 6a, 6b, 6c, and 6d show the Pearson product-moment correlation results for EYEMEAS, PERCLOS, DRVDROW, LANDEV, and LANEX for the 1, 2, 3, and 6-minute interval data sets, respectively. As in previous studies, it was found that a trend existed toward increasing correlations with longer averages. However, there were two noteworthy differences between the correlations for the current data set and previous correlation results: the correlation values were generally much lower, and the incremental correlation values between time intervals (that is, between the 1-minute interval data set and the 2-minute interval data set, between the 2-minute interval data set and the 3-minute data set, and between the 3-minute interval data set and the 6-minute interval data set) were lower.

It was hypothesized that the differences in properties of the correlation coefficients between the current study and previous studies were a result of the lane enforcement procedures. Large lane excursions that took place in the previous studies did not occur in this study due to the controls placed upon lane keeping. These large, uncontrolled lane excursions corresponded with drowsy periods of driving, and resulted in high correlations between lane-related measures and definitional drowsiness measures. Control of large lane excursions resulted in substantially lower correlations. Also, it is likely that lower incremental differences between correlations for the 1, 2, 3, and 6-minute interval data sets are a result of the lower magnitudes of the correlations.

It was also noted that incremental differences between correlations tended to become smaller as the time interval over which data were averaged increased. In some

Table 6a: Correlation Coefficients Between Measures for 1-Minute Interval Data Set.

r Values*	EYEMEAS	PERCLOS	DRVDROW	LANDEV	LANEX
EYEMEAS		0.445	0.503	0.234	0.237
PERCLOS	0.445		0.177	0.230	0.224
DRVDROW	0.503	0.177		0.236	0.127
STVELV	-0.039	0.024	0.373	0.316	0.050
LGREV	-0.016	0.015	0.214	0.370	0.115
MDREV	-0.077	-0.010	0.395	0.410	0.113
SMREV	-0.108	-0.065	0.149	-0.102	-0.045
STEXED	0.009	-0.012	0.024	0.100	0.013
NMRHOLD	0.085	0.019	-0.321	-0.217	-0.077
THRSHLD	0.100	0.080	-0.212	-0.096	-0.046
LANDEV	0.234	0.230	0.236		0.513
LANVAR	0.230	0.253	0.217	0.961	0.544
LNMN SQ	0.294	0.249	0.188	0.663	0.873
LANEX	0.237	0.224	0.127	0.513	
LNERRSQ	0.143	0.186	0.089	0.372	0.451
LNRTDEV	0.109	0.145	0.298	0.760	0.334
LNRTVAR	0.092	0.151	0.279	0.711	0.322
ACCDEV	-0.004	0.000	0.045	0.507	0.333
ACCVAR	-0.006	-0.012	0.017	0.453	0.330
INTACDEV	-0.016	-0.020	-0.020	0.373	0.332
INTACVAR	-0.010	-0.020	-0.013	0.342	0.355
ACEXEED	-0.034	-0.027	0.019	0.242	0.261
HPHDGDEV	0.098	0.133	0.302	0.745	0.323
HPHDGVAR	0.081	0.138	0.284	0.699	0.310
VIBPROP	0.103	0.195	0.072	0.316	0.252
NUMVIB	0.094	0.169	0.075	0.318	0.267

* Note: All correlation coefficients with $|r| \geq 0.052$ are significant ($\alpha = 0.05$).

Table 6b: Correlation Coefficients Between Measures for 2-Minute Interval Data Set.

r Values*	EYEMEAS	PERCLOS	DRVDROW	LANDEV	LANEX
EYEMEAS		0.495	0.520	0.271	0.302
PERCLOS	0.495		0.228	0.308	0.332
DRVDROW	0.520	0.228		0.274	0.156
STVELV	-0.040	0.022	0.381	0.327	0.038
LGREV	-0.020	0.019	0.239	0.387	0.098
MDREV	-0.080	-0.005	0.416	0.435	0.113
SMREV	-0.120	-0.086	0.159	-0.126	-0.042
STEXED	0.011	-0.018	0.022	0.116	0.011
NMRHOLD	0.100	0.034	-0.343	-0.213	-0.085
THRSHLD	0.103	0.090	-0.230	-0.083	-0.037
LANDEV	0.271	0.308	0.274		0.543
LANVAR	0.281	0.359	0.261	0.968	0.571
LNMNSQ	0.373	0.371	0.231	0.704	0.888
LANEX	0.302	0.332	0.156	0.543	
LNERRSQ	0.223	0.343	0.125	0.401	0.488
LNRTDEV	0.126	0.198	0.328	0.798	0.349
LNRTVAR	0.111	0.210	0.314	0.749	0.332
ACCDEV	-0.009	0.013	0.052	0.524	0.344
ACCVAR	-0.011	-0.003	0.022	0.467	0.336
INTACDEV	-0.019	-0.013	-0.018	0.396	0.327
INTACVAR	-0.013	-0.014	-0.013	0.370	0.332
ACEXEED	-0.026	-0.002	0.026	0.282	0.262
HPHDGDEV	0.112	0.180	0.331	0.783	0.338
HPHDGVAR	0.098	0.192	0.318	0.736	0.318
VIBPROP	0.164	0.322	0.104	0.348	0.287
NUMVIB	0.135	0.257	0.099	0.351	0.286

* Note: All correlation coefficients with $|r| \geq 0.074$ are significant ($\alpha = 0.05$).

Table 6c: Correlation Coefficients Between Measures for 3-Minute Interval Data Set.

r Values*	EYEMEAS	PERCLOS	DRVDROW	LANDEV	LANEX
EYEMEAS		0.530	0.530	0.287	0.346
PERCLOS	0.530		0.248	0.314	0.360
DRVDROW	0.530	0.248		0.291	0.176
STVELV	-0.034	0.026	0.392	0.340	0.034
LGREV	-0.005	0.020	0.255	0.410	0.111
MDREV	-0.078	-0.001	0.429	0.447	0.111
SMREV	-0.122	-0.082	0.175	-0.171	-0.057
STEXED	0.035	-0.023	0.035	0.128	0.000
NMRHOLD	0.098	0.023	-0.367	-0.197	-0.066
THRSHLD	0.101	0.068	-0.253	-0.054	-0.028
LANDEV	0.287	0.314	0.291		0.562
LANVAR	0.298	0.359	0.277	0.972	0.601
LNMSNSQ	0.405	0.391	0.248	0.718	0.903
LANEX	0.346	0.360	0.176	0.562	
LNERRSQ	0.244	0.334	0.140	0.413	0.564
LNRTDEV	0.128	0.198	0.341	0.811	0.368
LNRTVAR	0.113	0.208	0.328	0.763	0.351
ACCDEV	-0.002	0.012	0.057	0.505	0.347
ACCVAR	-0.007	0.000	0.020	0.452	0.334
INTACDEV	-0.020	-0.029	-0.024	0.364	0.326
INTACVAR	-0.018	-0.028	-0.021	0.339	0.322
ACEXEED	-0.040	-0.017	0.019	0.293	0.249
HPHDGDEV	0.115	0.182	0.344	0.796	0.353
HPHDGVAR	0.099	0.190	0.332	0.750	0.337
VIBPROP	0.177	0.330	0.127	0.359	0.365
NUMVIB	0.164	0.287	0.130	0.369	0.312

* Note: All correlation coefficients with $|r| \geq 0.090$ are significant ($\alpha = 0.05$).

Table 6d: Correlation Coefficients Between Measures for 6-Minute Interval Data Set.

r Values*	EYEMEAS	PERCLOS	DRVDROW	LANDEV	LANEX
EYEMEAS		0.561	0.545	0.308	0.423
PERCLOS	0.561		0.279	0.363	0.464
DRVDROW	0.545	0.279		0.326	0.212
STVELV	-0.028	0.022	0.408	0.350	0.013
LGREV	0.002	0.032	0.293	0.418	0.072
MDREV	-0.078	-0.006	0.453	0.450	0.085
SMREV	-0.139	-0.112	0.188	-0.259	-0.079
STEXED	0.060	-0.028	0.074	0.207	0.035
NMRHOLD	0.113	0.046	-0.405	-0.145	-0.027
THRSHLD	0.118	0.091	-0.282	0.012	0.001
LANDEV	0.308	0.363	0.326		0.561
LANVAR	0.323	0.410	0.316	0.978	0.596
LNMNSQ	0.467	0.472	0.290	0.716	0.927
LANEX	0.423	0.464	0.212	0.561	
LNERRSQ	0.312	0.401	0.184	0.447	0.619
LNRTDEV	0.134	0.206	0.370	0.830	0.360
LNRTVAR	0.116	0.205	0.359	0.784	0.336
ACCDEV	-0.021	0.000	0.092	0.456	0.313
ACCVAR	-0.028	-0.018	0.046	0.390	0.302
INTACDEV	-0.035	-0.054	-0.016	0.273	0.270
INTACVAR	-0.027	-0.049	-0.011	0.256	0.255
ACEXEED	-0.040	-0.029	0.041	0.277	0.202
HPHDGDEV	0.120	0.189	0.373	0.814	0.343
HPHDGVAR	0.102	0.188	0.363	0.769	0.320
VIBPROP	0.225	0.380	0.190	0.400	0.405
NUMVIB	0.202	0.347	0.193	0.438	0.349

* Note: All correlation coefficients with $|r| \geq 0.127$ are significant ($\alpha = 0.05$).

Table 7a: Correlation Coefficients Between SLEEPER1, SLEEPER2, and SLEEPER3, and All Other Measures for 3-Minute Interval Data Set.

r Values*	SLEEPER1	SLEEPER2	SLEEPER3
EYEMEAS	0.863	0.918	0.646
PERCLOS	0.438	0.471	0.927
DRVDROW	0.886	0.822	0.594
STVELV	0.216	0.160	0.174
LGREV	0.149	0.116	0.115
MDREV	0.213	0.148	0.165
SMREV	0.038	0.000	-0.001
STEXED	0.040	0.040	-0.006
NMRHOLD	-0.165	-0.105	-0.123
THRSHLD	-0.096	-0.050	-0.041
LANDEV	0.331	0.329	0.374
LANVAR	0.328	0.329	0.405
LNMSNSQ	0.370	0.388	0.421
LANEX	0.294	0.314	0.367
LNERRSQ	0.217	0.229	0.332
LNRTDEV	0.274	0.246	0.297
LNRTVAR	0.257	0.229	0.299
ACCDEV	0.033	0.025	0.032
ACCVAR	0.008	0.005	0.008
INTACDEV	-0.025	-0.024	-0.033
INTACVAR	-0.022	-0.022	-0.031
ACEXEED	-0.010	-0.018	-0.006
HPHDGDEV	0.268	0.238	0.284
HPHDGVAR	0.252	0.222	0.287
VIBPROP	0.173	0.178	0.323
NUMVIB	0.167	0.171	0.288

* Note: All correlation coefficients with $|r| \geq 0.090$ are significant ($\alpha = 0.05$).

Table 7b: Correlation Coefficients Between SLEEPER1, SLEEPER2, and SLEEPER3, and All Other Measures for 6-Minute Interval Data Set.

r Values*	SLEEPER1	SLEEPER2	SLEEPER3
EYEMEAS	0.865	0.920	0.677
PERCLOS	0.470	0.504	0.914
DRVDROW	0.892	0.830	0.645
STVELV	0.228	0.172	0.190
LGREV	0.176	0.139	0.149
MDREV	0.229	0.160	0.187
SMREV	0.038	-0.004	-0.009
STEXED	0.077	0.075	0.009
NMRHOLD	-0.181	-0.115	-0.135
THRSHLD	-0.105	-0.054	-0.047
LANDEV	0.361	0.358	0.427
LANVAR	0.363	0.362	0.460
LNMNSQ	0.425	0.446	0.498
LANEX	0.355	0.381	0.459
LNERRSQ	0.279	0.294	0.397
LNRTDEV	0.293	0.262	0.321
LNRTVAR	0.277	0.245	0.315
ACCDEV	0.044	0.030	0.039
ACCVAR	0.012	0.003	0.006
INTACDEV	-0.028	-0.031	-0.049
INTACVAR	-0.021	-0.023	-0.043
ACEXEED	0.003	-0.007	-0.005
HPHDGDEV	0.288	0.255	0.308
HPHDGVAR	0.272	0.238	0.303
VIBPROP	0.235	0.238	0.383
NUMVIB	0.225	0.225	0.358

* Note: All correlation coefficients with $|r| \geq 0.127$ are significant ($\alpha = 0.05$).

cases, the correlation coefficients for the 3-minute interval data set were actually higher than those for the 6-minute interval data set. However, both the three-minute and six-minute interval data sets consistently yielded higher correlation coefficients than either the one-minute interval data set or the two-minute interval data set. For this reason, it was decided to remove the one-minute and two-minute interval data sets from further consideration in multiple regression and classification analyses, since the three-minute and six-minute interval data sets were considered likely to yield algorithms with more predictive power.

A final observation made from the correlation results shown in Tables 6a, 6b, 6c, and 6d was that DRVDROW, for which data were collected for the first time in this phase of the study, showed promise as a reliable definitional measure of drowsiness on a level comparable to that of PERCLOS and EYEMEAS. DRVDROW correlated relatively well with EYEMEAS in the three-minute and six-minute interval data sets ($r = 0.530$ and $r = 0.545$, respectively). Also, DRVDROW correlations with some of the independent measures, particularly steering-reversal measures and heading-related measures, were much higher than correlations between either PERCLOS or EYEMEAS and the same measures. This pointed to the possibility that DRVDROW might be able to predict some aspects of driver drowsiness better than PERCLOS or EYEMEAS. Furthermore, it suggested that it might be advantageous to try developing new variables that combine DRVDROW and either PERCLOS and EYEMEAS so that the aspects of drowsiness that are represented well by eye closure and the aspects of drowsiness that are represented well by driver subjective opinion are combined into a single measure.

Therefore, the measures SLEEPER1, SLEEPER2, and SLEEPER3 (described previously) were developed. SLEEPER1 and SLEEPER2 combine DRVDROW and EYEMEAS, while SLEEPER3 combines DRVDROW and PERCLOS. Tables 7a and 7b

show the Pearson product-moment correlation results between the new SLEEPER1, SLEEPER2, and SLEEPER3 variables and all other variables in the data set. All three SLEEPER variables had correlation results comparable to those of PERCLOS, EYEMEAS, and DRVDROW, signifying that they should be retained for further analysis.

Distribution Analyses

Because the correlation analyses produced lower r values than those observed in previous studies, it was decided to investigate an alternate approach to algorithm development concurrently with algorithm development through the previously used multiple regression analyses. It was desired to determine if characteristics of data taken for intervals immediately prior to an activation of the seat vibration system were substantially different than characteristics of data taken during “non-drowsy” periods. This led to the distribution analyses, which have been described previously.

The resulting histograms for the one-minute interval data set are shown in Appendix I, and those for the three-minute interval data set are shown in Appendix J. The number under each column in each histogram represents the upper limit of the bin range for that column, while the number under the column immediately to the left of each column represents the lower limit of the bin range for that column.

It was hoped that the results of the distribution analyses would show a clear separation between the distributions for “drowsy” data taken immediately before activations of the seat vibration system and “non-drowsy” data taken at other times during the experimental sessions. Unfortunately, an examination of Appendices I and J reveals that this is not the case. For all six measures and both time intervals, the two distributions essentially completely overlap one another.

The results demonstrate that it is not possible to accurately predict when a driver will experience extreme lane exceedances using eye-closure-related measures, conventional lane-related measures, or the subjective opinion of the driver as to his or her level of drowsiness. Therefore, it is not possible to develop any type of driver drowsiness detection algorithm based on differing characteristics between data taken near occurrences of extreme lane deviation and data taken at other times. As a result, further efforts toward algorithm development were concentrated on using the multiple regression approach with whole data sets that has been used in previous studies.

Multiple Regression

Driver Status (Drowsiness) Algorithms

Multiple regression analyses were performed on three-minute and six-minute data sets. No baselining techniques were applied to the data sets. Summaries of descriptive statistics for each dependent measure and each independent measure in each of the two data sets used in the analyses appear in Appendix K. The tables include minimum, maximum, median, and mean values for each measure, as well as the standard deviation for each measure. These summaries help to characterize the ranges over which each independent and dependent variable varied in the new algorithm development data, and thereby define the ranges for which the resulting regression models are applicable.

Tables 8a and 8b are summaries of results that were obtained from the multiple regression analyses performed on three-minute and six-minute interval data sets, respectively. More complete results of the regression analyses for three-minute data can be found in Appendix L, and more complete results for six-minute data can be found in Appendix M. In the regression summaries, both standardized regression coefficients (*Beta* weights) and raw regression coefficients (*B* weights) are shown. Since the

independent variables are not standardized to a mean of 0 and a standard deviation of 1, the raw (B) coefficients were used as the coefficients for the independent variables in the subsequent analyses.

An examination of the R values in Tables 8a and 8b shows a substantial difference between the algorithms developed from data collected in the original algorithm development experiment (Wierwille et al., 1994) and algorithms developed from data collected in the current experiment. The R values for the six dependent variables collected in the current 3-minute data set ranged from 0.140 to 0.539, and the R values

Table 8a: Summary Table of Multiple Regression Analyses Showing R Values for 3-Minute Interval Data Set and Dependent Driver Status Measures.

Independent Measures		Dependent Measures					
		PERCLOS	EYEMEAS	DRVDROW	SLEEPER1	SLEEPER2	SLEEPER3
D	Steering and Accelerometer	0.140	0.152	0.467	0.216	0.160	0.186
E	Steering, Accelerometer, & HPHDGDEV/VAR	0.386	0.315	0.512	0.351	0.303	0.428
F	Steering, Accelerometer, LANDEV/VAR, LNMNSQ, LANEX, & LNERRSQ	0.482	0.493	0.554	0.481	0.475	0.534
G	Steering, Accelerometer, & all lane measures (includes LNRTDEV/VAR)	0.491	0.505	▲	▲	▲	0.539

NOTE: ▲ denotes that introduction of the variable did not improve the R value. See entry directly above for model with same R and fewer terms.

Table 8b: Summary Table of Multiple Regression Analyses Showing R Values for 6-Minute Interval Data Set and Dependent Driver Status Measures.

Independent Measures		Dependent Measures					
		PERCLOS	EYEMEAS	DRVDROW	SLEEPER1	SLEEPER2	SLEEPER3
D	Steering and Accelerometer	0.214	0.178	0.499	0.223	0.133	0.214
E	Steering, Accelerometer, & HPHDGDEV/VAR	0.404	0.340	0.538	0.326	0.325	0.422
F	Steering, Accelerometer, LANDEV/VAR, LNMNSQ, LANEX, & LNERRSQ	0.575	0.569	0.583	0.547	0.532	0.614
G	Steering, Accelerometer, & all lane measures (includes LNRTDEV/VAR)	▲	▲	▲	▲	0.544	▲

NOTE: ▲ denotes that introduction of the variable did not improve the R value. See entry directly above for model with same R and fewer terms.

for the same variables collected in the current six-minute data set ranged from 0.133 to 0.614. In contrast, the R values for the five dependent variables collected in the original algorithm development experiment for independent variable groups D through G ranged from 0.677 to 0.886. Clearly, the previously developed algorithms were able to predict drowsiness levels more accurately than the new algorithms developed from lane-enforced data. Also, there appeared to be no appreciable difference in the ability to predict

drowsiness levels between the three-minute algorithms and the six-minute algorithms developed from the lane-enforced data.

As in previous studies, it is apparent that algorithms that include lane-related measures (using regressor groups F and G) predict drowsiness better than algorithms that do not include lane-related measures (regressor groups D and E). However, it was also found that the inclusion of the heading-related measures HPHDGDEV and HPHDGVAR (regressor group E) yielded algorithms with higher R values than those developed with only steering and accelerometer-related measures (regressor group D).

Though the relatively low R values of the new set of algorithms are disappointing results, they are not surprising. The set of drowsiness algorithms developed from the data collected in the original algorithm development experiment performed poorly in the test and evaluation experiment. It was hypothesized that the reason for this poor performance was the enforcement of the lane positioning of the vehicle in that experiment, and it was recommended that new algorithms be developed under lane-enforced conditions. The current results show that algorithms developed under lane-enforced conditions do not provide R values as high in estimating drowsiness as those developed under conditions where large lane excursions were not tightly controlled. Therefore, it appears that the R values of the algorithms will be lower when lane deviations are controlled in such a manner as to reflect actual driving.

Driver Performance Algorithms

Tables 9a and 9b are summaries of results obtained from multiple regression analyses on three-minute and six-minute data sets, respectively, using the performance measures LANEX and LNMNSQ as dependent variables. These variables can be measured directly when lane-related information is available, but it is desirable to have

Table 9a: Summary Table of Multiple Regression Analyses Showing R Values for 3-Minute Interval Data Set and Dependent Driver Performance Measures.

Independent Measures		Dependent Measures	
		LANEX	LNMSQ
D	Steering and Accelerometer	0.367	0.493
E	Steering, Accelerometer, & HPHDGDEV/VAR	0.497	0.664

Table 9b: Summary Table of Multiple Regression Analyses Showing R Values for 6-Minute Interval Data Set and Dependent Driver Performance Measures.

Independent Measures		Dependent Measures	
		LANEX	LNMSQ
D	Steering and Accelerometer	0.348	0.485
E	Steering, Accelerometer, & HPHDGDEV/VAR	0.514	0.680

algorithms to predict them for situations where lane information is not available. Such algorithms yield performance measures which can be used in “OR” classification analyses. Complete regression summaries for the eight algorithms appear in Appendix N. As can be seen in Tables 9a and 9b, the R values for the algorithms developed range from 0.348 to 0.680. Again, these R values are somewhat low, but they are in the same general range as the R-values for the new driver status algorithms, and most have R values high enough to be used in further “OR” classification matrix analyses.

Classification Matrices

Driver Status (Drowsiness)

Unconditional classification matrix analyses were performed for all driver status algorithms that had correlation coefficients of 0.4 or above. (The word “unconditional” is used here to describe classifications in which there is no “or”ing with performance.) An examination of algorithms with R values below 0.4 showed that the algorithms did not merit further analysis; often, the backwards stepwise regression analyses only left a single independent variable in the regression equation.

Threshold values for each of the six measures were based on examination of scatterplots constructed with the algorithm output on one axis and the measured variable on the other. The thresholds were set as follows:

PERCLOS	0.012
EYEMEAS	2000
DRVDFROW	70
SLEEPER1	1.7
SLEEPER2	3
SLEEPER3	1.4

Each algorithm output data point and each measured data point was classified as “drowsy” if it was greater than or equal to the threshold value and as “awake” if it was less than the threshold value. Points were then paired on a casewise basis and grouped into categories of correct detections, correct rejections, false alarms, and missed detections. Apparent accuracy rates (APARs) were determined by dividing the number of correct classifications (correct detections and rejections) by the total number of paired data points.

The unconditional classification matrices for the drowsiness measures appear directly beneath the regression summaries for their respective algorithms in Appendices L and M. Summaries of the apparent accuracy rates obtained for the classification analyses appear in Tables 10a and 10b for three-minute average data and six-minute average data, respectively.

The APAR values obtained here are surprisingly high. Even though the R values for the algorithms developed in the current experiment are much lower than those developed previously, the APAR values are in a similar range to those obtained from previously developed algorithms. PERCLOS algorithms yielded some of the highest APAR values when classified. As was the case with the R values, there appears to be no appreciable difference between the performance of three-minute algorithms and six-minute algorithms in the ability to classify driver status correctly based on unconditional APARs.

Table 10a: Summary Table of Unconditional Classification Matrix Analyses Showing Apparent Accuracy Rates for Driver Status Algorithms developed from 3-Minute Interval Data Set.

Independent Measures		Dependent Measures					
		PERCLOS	EYEMEAS	DRVDROW	SLEEPER1	SLEEPER2	SLEEPER3
D	Steering and Accelerometer			0.692			
E	Steering, Accelerometer, & HPHDGDEV/VAR			0.703			0.854
F	Steering, Accelerometer, LANDEV/VAR, LNMNSQ, LANEX, & LNERRSQ	0.928	0.857	0.719	0.785	0.852	0.850
G	Steering, Accelerometer, & all lane measures (includes LNRTDEV/VAR)	0.928	0.863	▲	▲	▲	0.854

KEY:
▲ denotes that introduction of the variable did not improve the R value. See entry directly above for model with same APAR and fewer terms.
Blank cells indicate that no classification matrices were generated for the algorithm because its R value was less than 0.4.

Table 10b: Summary Table of Unconditional Classification Matrix Analyses Showing Apparent Accuracy Rates for Driver Status Algorithms developed from 6-Minute Interval Data Set.

Independent Measures		Dependent Measures					
		PERCLOS	EYEMEAS	DRVDROW	SLEEPER1	SLEEPER2	SLEEPER3
D	Steering and Accelerometer			0.702			
E	Steering, Accelerometer, & HPHDGDEV/VAR	0.916		0.714			0.853
F	Steering, Accelerometer, LANDEV/VAR, LNMNSQ, LANEX, & LNERRSQ	0.937	0.861	0.735	0.782	0.840	0.840
G	Steering, Accelerometer, & all lane measures (includes LNRTDEV/VAR)	▲	▲	▲	▲	0.832	▲

KEY:

▲ denotes that introduction of the variable did not improve the R value. See entry directly above for model with same APAR and fewer terms.
Blank cells indicate that no classification matrices were generated for the algorithm because its R value was less than 0.4.

An examination of the unconditional classification matrices in Appendices K and L reveals that false alarm rates (where the algorithm classified the driver as drowsy but the measured variable classified the driver as awake) were much lower than missed detection rates (where the algorithm classified the driver as awake but the measured variable classified the driver as drowsy). Though it is not desirable to have a system that misses detections, it is less desirable to have a system that produces a large number of false alarms. In an actual system, a high false alarm rate will likely translate to a low acceptance rate of the system by the driving population.

Driver Status or Performance

Because there were missed detections in the unconditional classification matrices, a process in which two criteria could be used in detection instead of one was implemented in new sets of classification matrices, called “conditional” classification matrices. In

these matrices, each algorithm output data point and each measured data point were classified as before. However, data for LANEX and LNMNSQ were also classified as “awake” or “drowsy”. For independent measure sets F and G, LANEX and LNMNSQ were directly measurable. Therefore, all LANEX and LNMNSQ data points were classified as either correct detections or correct detections; since no algorithm needed to be used, there were no missed detections or false alarms. Thresholds were set at 0.066667 for LANEX and at 3 for LNMNSQ. The effect of combining a drowsiness algorithm and measured LANEX or LNMNSQ was to move data points from the drowsiness algorithm classification matrices from classification as missed detections, false alarms or correct rejections to classification as correct detections if the LANEX or LNMNSQ data point for that minute was above threshold. The result of this was more correct classifications than the unconditional classification matrices, resulting in higher APARs. These results can be seen in Tables 11a, 11b, 11c, and 11d.

In independent measure sets D and E, LANEX and LNMNSQ were not available; therefore, in the conditional classification matrices the eLANEX and eLNMNSQ algorithms that appear in Appendix N were used. The same threshold values as before were used. The disadvantage in using these algorithms was that they contained incorrect classifications; they introduced additional missed detections and false alarms when the matrices were “or”ed. The result of this, which can be seen by comparing Tables 11a, 11b, 11c, and 11d with Tables 10a and 10b, was that the apparent accuracy rates were lower for the composite “or”ed matrices than for the unconditional classification matrices in every case where eLANEX or eLNMNSQ algorithms were used. In contrast, introducing measured LANEX or measured LNMNSQ resulted in a higher apparent accuracy rate in each case.

Table 11a: Summary Table of Conditional Classification Matrix Analyses Showing Apparent Accuracy Rates for Driver Status Algorithms developed from 3-Minute Interval Data Set -OR- LANEX/eLANEX.

Independent Measures		Dependent Measures					
		PERCLOS	EYEMEAS	DRVDROW	SLEEPER1	SLEEPER2	SLEEPER3
D	Steering and Accelerometer			✦			
E	Steering, Accelerometer, & HPHDGDEV/VAR			0.650*			0.802*
F	Steering, Accelerometer, LANDEV/VAR, LNMNSQ, LANEX, & LNERRSQ	0.966	0.890	0.757	0.823	0.882	0.899
G	Steering, Accelerometer, & all lane measures (includes LNRTDEV/VAR)	0.966	0.897	▲	▲	▲	0.903

KEY:

▲ denotes that introduction of the variable did not improve the R value. See entry directly above for model with same APAR and fewer terms.

✦ denotes that the classification matrix was not generated because the R value of the eLANEX algorithm was less than 0.4.

* indicates that an eLANEX algorithm was used in the OR classification matrix. For all other values, measured LANEX was used.

Blank cells indicate that no classification matrices were generated for the algorithm because its R value was less than 0.4

Table 11b: Summary Table of Conditional Classification Matrix Analyses Showing Apparent Accuracy Rates for Driver Status Algorithms developed from 3-Minute Interval Data Set -OR- LNMNSQ/eLNMNSQ.

Independent Measures		Dependent Measures					
		PERCLOS	EYEMEAS	DRVDROW	SLEEPER1	SLEEPER2	SLEEPER3
D	Steering and Accelerometer			0.654*			
E	Steering, Accelerometer, & HPHDGDEV/VAR			0.667*			0.808*
F	Steering, Accelerometer, LANDEV/VAR, LNMNSQ, LANEX, & LNERRSQ	0.966	0.890	0.764	0.829	0.884	0.899
G	Steering, Accelerometer, & all lane measures (includes LNRTDEV/VAR)	0.964	0.897	▲	▲	▲	0.903

KEY:

▲ denotes that introduction of the variable did not improve the R value. See entry directly above for model with same APAR and fewer terms.

* indicates that an eLNMNSQ algorithm was used in the OR classification matrix. For all other values, measured LNMNSQ was used.

Blank cells indicate that no classification matrices were generated for the algorithm because its R value was less than 0.4

Table 11c: Summary Table of Conditional Classification Matrix Analyses Showing Apparent Accuracy Rates for Driver Status Algorithms developed from 6-Minute Interval Data Set -OR- LANEX/eLANEX.

Independent Measures		Dependent Measures					
		PERCLOS	EYEMEAS	DRVDROW	SLEEPER1	SLEEPER2	SLEEPER3
D	Steering and Accelerometer			✦			
E	Steering, Accelerometer, & HPHDGDEV/VAR	0.853*		0.681*			0.815*
F	Steering, Accelerometer, LANDEV/VAR, LNMNSQ, LANEX, & LNERRSQ	0.962	0.891	0.769	0.828	0.878	0.887
G	Steering, Accelerometer, & all lane measures (includes LNRTDEV/VAR)	▲	▲	▲	▲	0.874	▲

KEY:

- ▲ denotes that introduction of the variable did not improve the R value. See entry directly above for model with same APAR and fewer terms.
 - ✦ denotes that the classification matrix was not generated because the R value of the eLANEX algorithm was less than 0.4.
 - * indicates that an eLANEX algorithm was used in the OR classification matrix. For all other values, measured LANEX was used.
- Blank cells indicate that no classification matrices were generated for the algorithm because its R value was less than 0.4.

Table 11d: Summary Table of Conditional Classification Matrix Analyses Showing Apparent Accuracy Rates for Driver Status Algorithms developed from 6-Minute Interval Data Set -OR- LNMNSQ/eLNMNSQ.

Independent Measures		Dependent Measures					
		PERCLOS	EYEMEAS	DRVDROW	SLEEPER1	SLEEPER2	SLEEPER3
D	Steering and Accelerometer			0.672*			
E	Steering, Accelerometer, & HPHDGDEV/VAR	0.857*		0.689*			0.824*
F	Steering, Accelerometer, LANDEV/VAR, LNMNSQ, LANEX, & LNERRSQ	0.966	0.891	0.769	0.824	0.874	0.891
G	Steering, Accelerometer, & all lane measures (includes LNRTDEV/VAR)	▲	▲	▲	▲	0.870	▲

KEY:

- ▲ denotes that introduction of the variable did not improve the R value. See entry directly above for model with same APAR and fewer terms.
 - * indicates that an eLNMNSQ algorithm was used in the OR classification matrix. For all other values, measured LNMNSQ was used.
- Blank cells indicate that no classification matrices were generated for the algorithm because its R value was less than 0.4.

A comparison of Tables 11a, 11b, 11c, and 11d reveals that the performance of LANEX is roughly equivalent to the performance of LNMNSQ in raising the APAR values from the unconditional APAR values. Also, the performance of three-minute interval data is roughly equivalent to the performance of six-minute interval data.

Conclusions

The primary focus of this experiment was to collect a new set of data from sleep-deprived subjects, using lane enforcement procedures, and develop new algorithms from that data for the detection of driver drowsiness. The restrictions placed on lane-keeping behavior and the truncation of the resulting data produced a data set with characteristics believed to be much closer to those of actual, on-the-road driving than data sets that were collected in previous experiments. However, the lack of large lane deviations resulted in much lower correlations between individual measures, specifically between drowsiness measures (PERCLOS, EYEMEAS, and DRVDROW) and performance-related measures, than had been observed previously. Based on the outcome of the test and evaluation experiment, this result was expected. Also, correlations did not substantially improve with lengthening of the interval over which data were averaged, and the incremental correlation values between averaging intervals became smaller as the interval became longer. All differences in the properties of the correlations are apparently a result of the lane-enforcement procedures and are not surprising results.

The distribution analyses demonstrated that “lane busts”, or exceedances of the lane boundaries by three feet or more on either side, could not be predicted reliably by measures such as PERCLOS, EYEMEAS, DRVDROW, or LANDEV. Some missed detections and/or false alarms will be unavoidable, even if PERCLOS, EYEMEAS, or DRVDROW can be detected with perfect accuracy. Therefore, to avoid run-off-road

accidents in a drowsy driver detection system, an auxiliary system that can measure lane position directly and nearly instantaneously must be used.

The new driver status (drowsiness) algorithms developed under lane-enforced conditions in this experiment all had lower multiple-R values than the algorithms developed in previous experiments. Again, this was not surprising given the results of the test and evaluation experiment. Multiple-R values for most algorithms developed without lane-related variables in the set of regressors were so low that the algorithms have fundamentally no predictive ability. Therefore, a detection system using the set of new algorithms cannot operate reliably when lane information is not available. The new driver performance algorithms developed in this experiment also had lower multiple-R values than the algorithms developed previously.

Although the multiple-R values for the new set of algorithms was relatively low, the unconditional classification accuracy of the new algorithms was much higher than expected and in a comparable range to that of the algorithms developed previously. False alarm rates for the new algorithms were particularly low; this bodes well for public acceptance of an actual detection system. When conditional classification matrices combining driver status and performance were constructed, classification accuracy improved in those cases where lane-related information (measured LANEX or LNMNSQ) was available, but degraded in those cases where lane-related information was not available and algorithm estimates of LANEX or LNMNSQ were used. This further demonstrates the inability of algorithms developed under conditions when lane information was unavailable to correctly classify drivers as “awake” or “drowsy”.

Throughout the multiple regression and classification analyses, it was shown that three-minute average algorithms perform nearly as well as six-minute average algorithms. Apparently, there is no fundamental difference between the accuracy levels for the two

time intervals. This result is encouraging because three-minute averages would allow for a faster-acting system than the system tested and evaluated in Phase Two of this study without an appreciable loss in accuracy from that of six-minute averages.

The addition of the subjective measure DRVDROW to the set of definitional measures examined produced encouraging results through the correlation and multiple regression analyses. In multiple regression, algorithms developed with DRVDROW as the dependent variable produced the highest multiple-R values when lane information was not available. When lane information was available, DRVDROW produced R values comparable to those produced by PERCLOS and EYEMEAS. The composite definitional measures SLEEPER1, SLEEPER2, and SLEEPER3 also performed well through correlation and multiple regression analyses; however, SLEEPER1 and SLEEPER2 in particular tended not to surpass the performance of their component measures (EYEMEAS and DRVDROW). In classification, however, PERCLOS yielded the highest APAR values, followed by SLEEPER3. Since SLEEPER3 is a composite of PERCLOS and DRVDROW, there may be some cases in which SLEEPER3 can classify driver status more accurately than PERCLOS, and both should be retained for further study.

Recommendations

The combined results of this experiment and previous experiments are sufficient to specify the design of a complete drowsy driver detection/alarm/countermeasures system (DDACS) for use in full-scale testing in an actual vehicle. The system should be a three-stage system, similar to those proposed previously. It should have a detection phase, an alarm phase, and a countermeasure phase.

The detection phase should be based on a composite criterion of driver status (drowsiness) or driver performance. Driver status should be monitored using a detection algorithm that includes steering, lateral accelerometer, and lane-related measures in its set of independent measures. Driver performance should be monitored using a directly-measured lane-related measure. When lane information is not available, reasonable detection accuracy cannot be achieved and the system should not be active. A “step up, step down” system similar to the system studied in the test and evaluation experiment is not feasible due to the lack of detection accuracy in algorithms when lane-related measures are not included in the independent variable set.

Because the differences in the results between three-minute averages and six-minute averages are so slight, three-minute averages should be used in the system because they allow for a faster-acting system. “Pipelining” of the averages should be accomplished in the same manner as in the test and evaluation experiment. Each minute, new moving averages should be calculated from the three most recent one-minute averages for each measure.

Based on the regression and classification results, the driver status algorithm used in the detection phase should be either Algorithm F4e-3 (see Appendix L, Figure L13) or Algorithm SLEEPER3-F-3 (see Appendix L, Figure L18). These algorithms have been designed to predict the definitional drowsiness measures PERCLOS and SLEEPER3, respectively. However, the algorithms are designed to function properly only when the independent variables from which they were developed remain within similar ranges to their ranges in the revised algorithm development data set (see Appendix K, Figure K1). If the values of any independent variables vary significantly outside these ranges, the algorithms will not be valid.

Threshold values should be set at 0.012 for PERCLOS and at 1.4 for SLEEPER3. Both algorithms include steering, lateral accelerometer, and lane-related independent measures. Both algorithms have high apparent classification accuracy and relatively high multiple-R values. The classification accuracy of both algorithms is improved by including “or”ing with a driver performance measure. Either measured LANEX or measured LNMNSQ may be used as the driver performance measure. Thresholds should be set at 0.06667 for LANEX or at 3.0 for LNMNSQ. Detection should take place if the three-minute moving average value of estimated drowsiness level exceeds threshold, if the three-minute moving average value of the performance measure exceeds threshold, or both values exceed their respective thresholds. The simultaneous monitoring of status and performance in this manner should produce reasonable detection accuracy when lane information is available.

“Lane busts” are prima facie evidence of driving performance that can result in accidents. However, it has been shown here that they cannot be reliably predicted using any of the dependent drowsiness variables. Therefore, an auxiliary lane departure warning system should be included in the detection phase. This system should use a seat vibration system similar to the one used in this experiment, which was presented using eccentrics (rotational masses) attached to electric servo motors installed in the seat back and seat pan of the driver’s seat. The vibration should be activated instantaneously when the vehicle exceeds either lane boundary by more than a specified amount and deactivated when the vehicle is brought back within the boundaries.

The alarm phase should be similar in nature to the alarm phase implemented in the test and evaluation experiment. It should begin with an advisory tone, immediately followed by a voice message (“Possible drowsiness has been detected; press reset now”). This message should be followed by a full alarm, an on-off tone which will only be

activated in the advisory tone and message fail to awaken the driver and will only remain active until the driver presses a reset button. Following this, a choice of countermeasures should be made available to the driver, and the countermeasure of the driver's choice should be active for a period of at least 25 minutes.

It is hoped that the moving average-based detection/alarm/countermeasures system in conjunction with the faster-acting lane departure warning system will provide the best possible system for driver status and performance monitoring. The fast-acting lane departure warning component should serve to correct extreme lapses in driver performance rapidly, while the slower-acting moving average-based system should serve to correct more gradual losses in performance and increases in levels of drowsiness.

REFERENCES

- Baker, W. J. and Theologus, G. C. (1972). Effects of caffeine on visual monitoring. Journal of Applied Psychology, 56, 422-427.
- Brown, I. D. (1965). A comparison of two subsidiary tasks used to measure fatigue in car drivers. Ergonomics, 8, 467-473.
- Brown, I. D. (1966). Effects of prolonged driving upon driving skill and performance of a subsidiary task. Industrial Medicine and Surgery, 35, 760-765.
- Brown, I. D., Simmonds, D. C., and Tickner, A. H. (1967). Measurement of control skills, vigilance, and performance on a subsidiary task during 12 hours of car driving. Ergonomics, 13, 239-242.
- Carroll, J. S., Blisewise, D. L., and Dement, W. C. (1989). A method for checking interobserver reliability in observational sleep studies. Sleep, 12(4), 363-367.
- Chaudoin, J. H. and Nelson, G. (1985). Interstate routes 15 and 40 shoulder rumble strips. (Report Caltrans-08-85-1). California Department of Transportation.
- Childs, J. M. (1978). Caffeine consumption and target scanning performance. Human Factors, 20, 91-96.
- Daimler-Benz. (1994, May). Driver drowsiness warning system. Slide presentation materials available upon request from Daimler Benz, Forschung und Technik, Stuttgart, Germany.
- Dingus, T. A., Hardee, L. H., and Wierwille, W. W. (1985). Detection of drowsy and intoxicated drivers based on highway driving performance measures. (Technical Report #8402). Blacksburg, VA: Virginia Polytechnic Institute and State University, IEOR Department.

- Edworthy, J., Loxely, S., and Dennis, I. (1991). Improving warning design: Relationship between warning sound parameters and perceived urgency. Human Factors, 33(2), 205-231.
- Ellsworth, L. A., Wreggit, S. S., and Wierwille W. W. (1993). Research on vehicle-based driver status/performance monitoring (Report # 93-02). Blacksburg, VA: Virginia Polytechnic Institute and State University, ISE Department.
- 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.
- Fairbanks, R. J., Fahey, S. E., and Wierwille, W. W. (1995). Research on vehicle-based driver status/performance monitoring (Report # 95-06). Blacksburg, VA: Virginia Polytechnic Institute and State University, ISE Department.
- Fairbanks, R. J., Lewin, M.G., and Wierwille, W. W. (1995). Research on vehicle-based driver status/performance monitoring (Report # 95-23). Blacksburg, VA: Virginia Polytechnic Institute and State University, ISE Department.
- Fairbanks, R. J. and Wierwille, W. W. (1994, October). Effects of using higher-order algorithms on drowsy driver detection accuracy (supplemental research report). Blacksburg, VA: Virginia Polytechnic Institute and State University, ISE Department.
- Fell, D. (1994, July). Safety Update: Problem Definition and Countermeasure Summary: Fatigue. New South Wales Road Safety Bureau, RUS No. 5. Cited in Knippling, R. R. and Wang, J. (1995, October). Revised estimates of the U.S. drowsy driver crash problem size based on General Estimates System case

reviews. Paper presented at the 39th Annual Proceedings of the Association for the Advancement of Automotive Medicine, Chicago, IL.

- Hardee, L. H., Dingus, T. A., and Wierwille, W. W. (1985). A comparison of three subsidiary tasks used as driver drowsiness countermeasures. (Technical Report # 8505). Blacksburg, VA: Virginia Polytechnic Institute and State University, IEOR Department.
- 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-71-2). Washington, D. C.: Bureau of Motor Carrier Safety. (NTIS No. PB213-963).
- 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. and Vulcan, P. (1990). Testing of commercially available fatigue monitors. (Draft Report). Melbourne, Australia: Monash University, Accident Research Centre.
- Haworth, N. L., Heffernan, C. J., and Horne, E. J. (1989). Fatigue in truck accidents. Melbourne, 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.
- Huntley, M. S. and Centybear, T. M. (1974). Alcohol, sleep deprivation, and driving speed effects upon control use during driving. Human Factors, 16, 19-28.
- 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 International Technical Conference on the Enhanced Safety of Vehicles, Munich, Germany.
- Kearney, P. W. (1966). Highway homicide. New York: Cromwell.
- Khaldi, S. and Vallet, M. (1994, May). Drowsiness of the driver: E.E.G. and vehicle parameters interaction. Presented at the 14th International Technical Conference on the Enhanced Safety of Vehicles, Munich, Germany.
- Kim, C. L., Wreggit, S. S., and Wierwille, W. W. (1994). Research on vehicle-based driver status/performance monitoring (Report # 94-02). Blacksburg, VA: Virginia Polytechnic Institute and State University, ISE Department.
- Knipling, R. R. and Wang, J. (1995, October). Revised estimates of the U.S. drowsy driver crash problem size based on General Estimates System case reviews. Paper presented at the 39th Annual Proceedings of the Association for the Advancement of Automotive Medicine, Chicago, IL.

- Knipling, R. R. and Wierwille, W. W. (1994, April). Vehicle-based drowsy driver detection: Current status and future prospects. Paper presented at the IVHS America Fourth Annual Meeting, Atlanta, GA.
- 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.
- Lerner, 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. DTNH 22-91-C-07004). Washington, D.C.: NHTSA, OCAR.
- 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., Tornros, 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). Melbourne, 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, Melbourne, Australia.
- Mast, T. M., Jones, H. V., and Heimstra, N. W. (1966). Effects of fatigue on performance in a driving device. Highway Research Record, 122, 93

- (Abridgment). Cited in Haworth, N. L., Vulcan P., Triggs, T. J., and Fildes, B. N. (1989). Driver fatigue research: Development of methodology. Melbourne, Australia: Monash University, Accident Research Centre.
- Office of Crash Avoidance Research (1991, 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.
- Planque, S., Chaput, D., Petit, C., Tarterre, C., and Chabanon, C. (1991, November). Analysis of EOG and EEG signals to detect lapses of alertness in car simulation driving. Presented at the 13th ESV Conference, Paris, France.
- Riemersma, J. B. J., Sanders, A. F., Wildervanck, C., and Gaillard, A. W. (1977). Performance decrement during prolonged night driving. 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). Melbourne, Australia: Monash University, Accident Research Centre.
- Ryder, J. M., Malin, S. A., and Kinsley, C. H. (1981). The effects of fatigue and alcohol on highway safety. National Highway Traffic Safety Administration Report No. DOT-HS-805-854. Cited in Dingus, T. A., Hardee, L. H., and Wierwille, W. W. (1985). Detection of drowsy and intoxicated drivers based on highway driving performance measures. (Technical Report # 8402). Blacksburg, VA: Virginia Polytechnic Institute and State University, IEOR Department.
- Sanders, M. S. and McCormick, E. J. (1993). Human Factors in Engineering and Design. New York, NY: McGraw-Hill, Inc.
- Safford, R. and Rockwell, T. H. (1967). Performance decrements in twenty-four hour driving. Highway Research Record, 163, 68-79.

- Seko, Y. (1984). Present technological status of detecting drowsy driving patterns. Jidosha Gijutsu, 30(5), 547-554.
- Skipper, J. W., Wierwille, W. W., and Hardee, L., (1984) An investigation of low-level stimulus-induced measures of driver drowsiness (Technical Report # 8402). Blacksburg, VA: Virginia Polytechnic Institute and State University, IEOR Department.
- Sugarman, R. D and Cozad, C. P. (1972). Road test of alertness variables (NTIS No. PB-215 450/8). Washington, DC: National Highway Traffic Safety Administration, U.S. Department of Transportation.
- Sussman, R. D., Sugarman, R. C., and Knight, J. R. (1971). Use of simulation in a study investigating alertness during long-distance, low-event driving. Highway Research Record, 364, 27-32.
- 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.
- Tilley, 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.
- 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). Melbourne, 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., and Mitchell, M. W. (1992). Research on vehicle-based driver status/performance monitoring (Report #92-01). Blacksburg, VA: Virginia Polytechnic Institute and State University, ISE Department.

Wierwille, W. W., Wreggit, S. S., Kirn, C. L., Ellsworth, L. A., and Fairbanks, R. J. (1994). Research on vehicle-based driver status/performance monitoring (Report #94-04). Blacksburg, VA: Virginia Polytechnic Institute and State University, ISE Department.

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. (1994). Research on vehicle-based driver status/performance monitoring (Report # 94-01). Blacksburg, VA: Virginia Polytechnic Institute and State University, ISE Department.

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.

Wreggit, S. S. and Wierwille, W. W. (1994a, August). A comparison of R values obtained from the application of algorithms to original A/O data, new A/O data, and new clipped A/O data (supplemental research report). Blacksburg, VA: Virginia Polytechnic Institute and State University, ISE Department.

Wreggit, S. S. and Wierwille, W. W. (1994b, August). An investigation of false alarm rates when applying detection algorithms to alert-driver segments (supplemental research report). Blacksburg, VA: Virginia Polytechnic Institute and State University, ISE Department.

Yabuta, K., Iizuka, H., Yanaqishima, T., Kataoka, Y., and Seno, T. (1985). The development of drowsiness warning devices. The Tenth International Technical Conference on Experimental Safety Vehicles.

APPENDIX A

Regression Summaries and Classification Matrices for Unbaselined LANEX Algorithms
Used In Test and Evaluation Phase
(Fairbanks, Lewin, and Wierwille, 1995)

Regression Summary for Dependent Variable: LANEX

R = 0.83026466 R² = 0.68933941 Adjusted R² = 0.68677620

F(5,606) = 268.94 p < 0.0000 Std. Error of estimate: 0.07246

	BETA	St. Err. of BETA	B	St. Err. of B	t(606)	p-level
Intercept			-0.006938	0.020656	-0.33587	0.737084
ACCDEV	0.282832	0.028478	0.059770	0.006018	9.93171	0.000000
STVELV	-0.177676	0.049757	-0.000422	0.000118	-3.57086	0.000384
LGREV	0.672433	0.052391	0.062359	0.004859	12.83487	0.000000
NMRHOLD	-0.170117	0.029684	-0.004275	0.000746	-5.73091	0.000000
THRSHLD	0.381693	0.027722	0.416193	0.030228	13.76859	0.000000

Predicted

Observed	Group	% Correct	Satisfactory	Borderline	Unsatisfactory
	Satisfactory	65.41	174	91	1
	Borderline	77.78	28	161	18
	Unsatisfactory	69.06	1	42	96
	Total	70.42	203	294	115

LANEX (R value = 0.8302647)

Apparent Accuracy Rate (LARGE misclassifications): **0.997**

Apparent Accuracy Rate (ALL misclassifications): **0.704**

Thresholds Used:

Satisfactory/Borderline: 0.075

Borderline/Unsatisfactory: 0.200

Classification Matrix Generated From Multiple Regression Analysis of **Original and New LANEX** Data (24 Subjects) Resulting in **Algorithm LDV-1**. (Independent variables employed included Steering and Accelerometer.)

Figure A1: Regression Summary and Classification Matrix Showing Accuracy of Algorithm LDV-1 When Applied to Combined Data Set (24 Subjects).

Regression Summary for Dependent Variable: LANEX

R = 0.83251125 R² = 0.69307499 Adjusted R² = 0.68717258

F(5,260) = 117.42 p<0.0000 Std. Error of estimate: 0.07485

	BETA	St. Err. of BETA	B	St. Err. of B	t(260)	p-level
Intercept			-0.210593	0.023034	-9.14274	0.000000
ACCDEV	0.263242	0.041927	0.057733	0.009195	6.27851	0.000000
LGREV	0.319591	0.051864	0.036813	0.005974	6.16206	0.000000
THRSHLD	0.316681	0.035752	0.308162	0.034790	8.85784	0.000000
AOTIME	0.682708	0.086778	0.097192	0.012354	7.86727	0.000000
NMWRONG	-0.362757	0.075683	-0.204829	0.042734	-4.79308	0.000003

Predicted

Observed	Group	% Correct	Satisfactory	Borderline	Unsatisfactory
	Satisfactory	78.23	97	27	0
	Borderline	74.70	14	62	7
	Unsatisfactory	74.58	1	14	44
	Total	76.32	112	103	51

LANEX (R value = 0.83251125)

Apparent Accuracy Rate (LARGE misclassifications): **0.996**

Apparent Accuracy Rate (ALL misclassifications): **0.763**

Thresholds:

Satisfactory/Borderline: 0.075

Borderline/Unsatisfactory: 0.200

Classification Matrix Generated From Multiple Regression Analysis of **Original and New LANEX** Data (24 Subjects) Resulting in **Algorithm LDV-4**. (Independent variables employed included A/O Task, Steering, and Accelerometer.)

Figure A2: Regression Summary and Classification Matrix Showing Accuracy of Algorithm LDV-4 When Applied to Combined Data Set (24 Subjects).

APPENDIX B

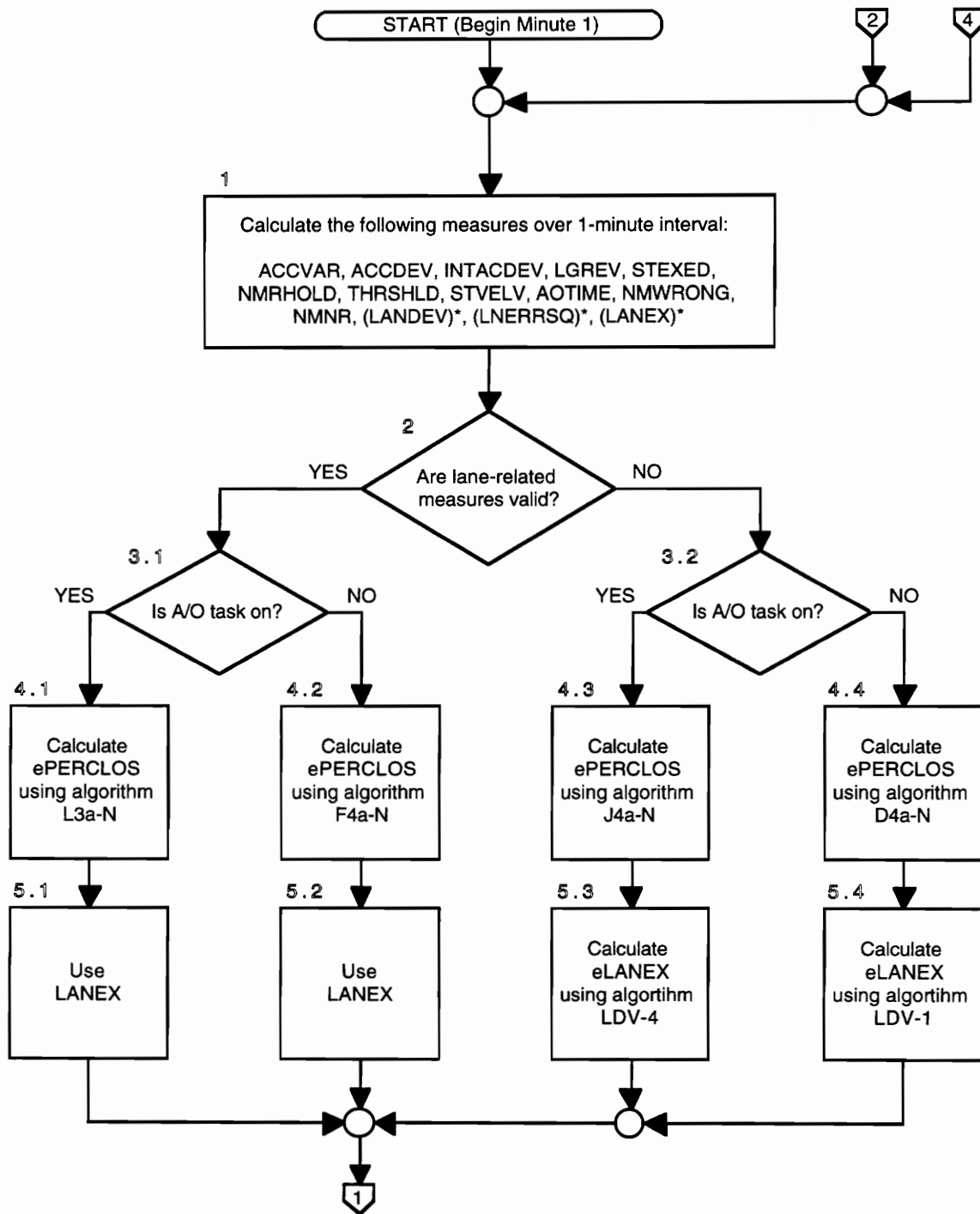
Logic Diagram Representation of Test and Evaluation System Specifications

LOGIC DIAGRAM REPRESENTATION OF TEST AND EVALUATION SYSTEM SPECIFICATIONS

The system recommendations specify what measures should be used for detection, what algorithms should be used to calculate estimates for those measures, when estimates should be used, what thresholds should be used, and what type of logic should be used for monitoring. A computational procedure meeting these specifications is required to accomplish on-line vehicle-based driver status/performance monitoring in the test and evaluation system. The structure of the procedure to be used is presented here in a logic diagram.

The basic structure of the computational procedure will be that of an endless loop. The only way to terminate the loop will be by turning off the vehicle. A timing mechanism will begin the measurement of one-minute intervals. Over each one-minute interval, steering-related measures (LGREV, STEXED, NMRHOLD, THRSOLD, STVELV) and lateral-accelerometer-related measures (ACCVAR, ACCDEV, INTACDEV) will be calculated (Figure B1, Cell 1). If the A/O task is operating, A/O task-related measures (AOTIME, NMWRONG, NMNR) will be calculated. Finally, if lane-related measures (LANDEV, LNERRSQ, LANEX) are valid, they will be calculated over the one-minute interval. Lane-related measures will be considered valid if lane boundaries can be detected for a substantial portion of the one-minute interval.

Following the calculation of one-minute measures, output values for ePERCLOS and either LANEX or eLANEX will be calculated and selected. Selection will be based on two criteria: the validity of lane-related measures (Figure B1, Cell 2), and the status of the A/O task (Figure B1, Cells 3.1, 3.2). Selection will take place as follows:



* Measures in parenthesis can only be calculated when lane position information is available for a substantial portion of the one-minute interval.

Figure B1. Logic Diagram for Detection System Computational Procedure

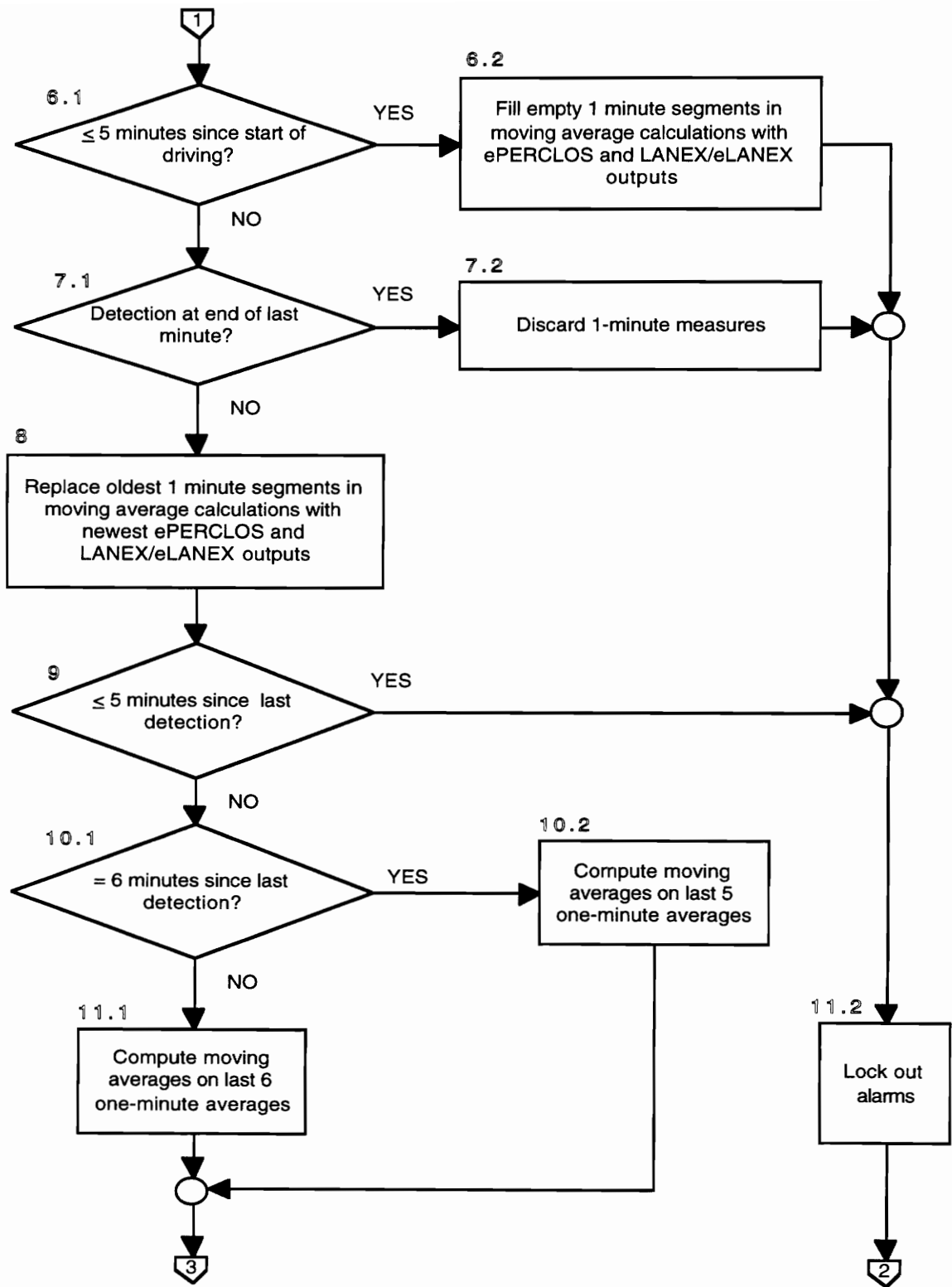


Figure B2. Logic Diagram for Detection System Computational Procedure

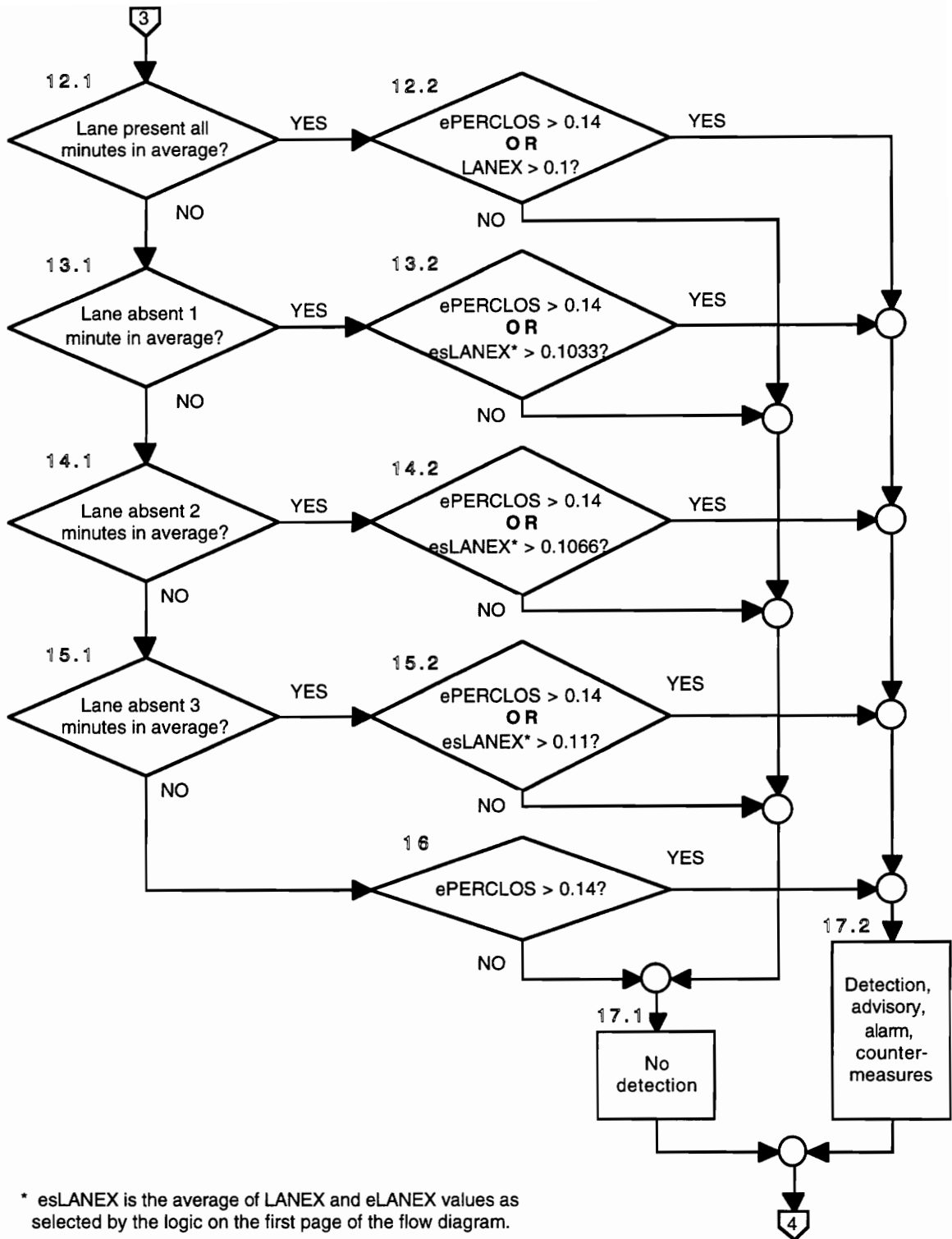
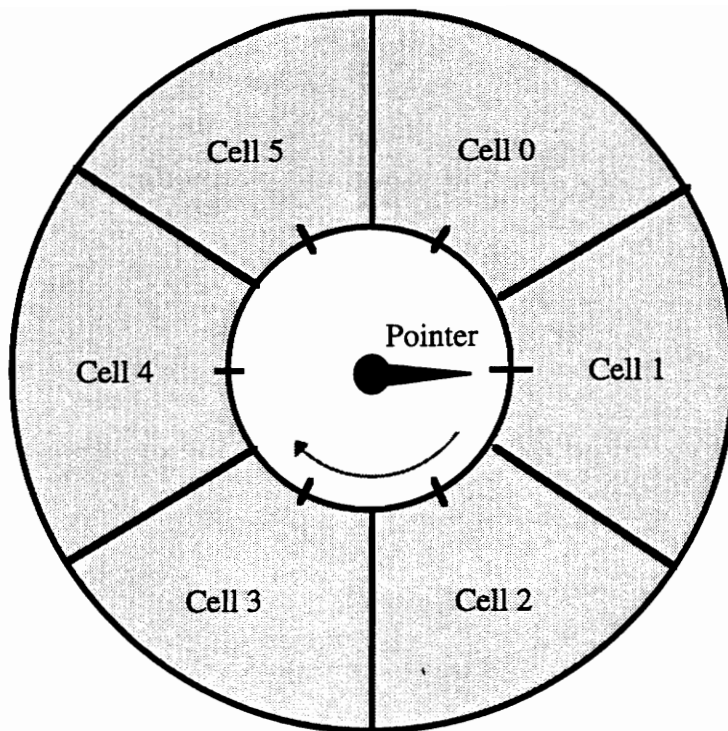


Figure B3. Logic Diagram for Detection System Computational Procedure

- If lane-related measures are *valid* and the A/O task is *on*, ePERCLOS will be calculated using algorithm L3a-N (Figure B1, Cell 4.1) and LANEX will be used (Figure B1, Cell 5.1).
- If lane-related measures are *valid* and the A/O task is *off*, ePERCLOS will be calculated using algorithm F4a-N (Figure B1, Cell 4.2) and LANEX will be used (Figure B1, Cell 5.2).
- If lane-related measures are *not valid* and the A/O task is *on*, ePERCLOS will be calculated using algorithm J4a-N (Figure B1, Cell 4.3) and eLANEX will be calculated using algorithm LDV-4 (Figure B1, Cell 5.3).
- If lane-related measures are *not valid* and the A/O task is *off*, ePERCLOS will be calculated using algorithm D4a-N (Figure B1, Cell 4.4) and eLANEX will be calculated using algorithm LDV-4 (Figure B1, Cell 5.4).

Once all output values are selected, the selected values will generally be entered into moving average calculations. The only exception will be output values for any minute immediately following a detection; these data will be discarded (Figure B2, Cell 7.2). A rotating array of six cells will contain the ePERCLOS and LANEX/eLANEX output values for the six most recent minutes of operation. The location in the array where output values will be placed for any given minute will be determined through modular arithmetic (Figure B4).

If five minutes or less have passed since detection system startup, some of the cells in the array will be empty and will need to be filled for the first time with ePERCLOS and LANEX/eLANEX outputs (Figure B2, Cells 6.1, 6.2). If there was a detection at the end of the last minute, the data obtained for that minute will be discarded (Figure B2, Cells 7.1, 7.2). In this case, values of zero will be substituted for the output values in the array. Under either condition, no moving averages will be calculated. In addition, detections and



The pointer moves from one cell to the next each minute that the detection system operates.

The position of the pointer is determined through modular arithmetic:

The minute number is divided by 6.

The remainder from the division calculation will be a value of 0, 1, 2, 3, 4, or 5.

The pointer will point to the cell address having the calculated remainder value.

Each cell will contain one-minute output values for ePERCLOS and LANEX/eLANEX.

Each minute, the ePERCLOS and LANEX/eLANEX output values will be placed into the cell toward which the pointer is pointing. The only exception to this rule will be for minutes in which the data are discarded (the one-minute intervals immediately following detection). For these minutes, values of zero will be placed into the cell to which the pointer is pointing.

Under normal circumstances, moving averages will be calculated by dividing the summation of the six cells for ePERCLOS by 6 and the summation of the six cells for LANEX/eLANEX by 6. If the first moving averages since a detection are being calculated, they will be five-minute averages. In this event, moving averages will be calculated by dividing the summation of the six cells for ePERCLOS by 5 and the summation of the six cells for LANEX/eLANEX by 5.

Figure B4. Procedure for Calculation of Moving Averages

alarms will be locked out for that minute, since there will be insufficient data to make a detection (Figure B2, Cell 11.2). Also, since no detection can be made, the program will loop back to the beginning to begin calculation of measures over the next one-minute interval.

If neither of the above two conditions are satisfied, the oldest one-minute segments in the moving average will be filled with the newest ePERCLOS and LANEX/eLANEX outputs (Figure B2, Cell 8). After this is done, if five minutes or less have passed since the most recent detection, some of the cells in the array will contain data that were collected before the detection (Figure B2, Cell 9). These pre-detection output values must be “purged” from the moving average calculations before another detection can be made. In this situation, no moving averages will be calculated. In addition, detections and alarms will be locked out for that minute, since there will be insufficient data to make a detection (Figure B2, Cell 11.2). Also, since no detection can be made, the program will loop back to the beginning to begin calculation of measures over the next one-minute interval.

If exactly 6 minutes have passed since the most recent detection, moving averages for ePERCLOS and esLANEX will be calculated for the first time since that detection (Figure B2, Cell 10.1). (esLANEX is the average of LANEX and eLANEX values as selected by the logic in Figure B1.) However, since the outputs for the first minute after detection were discarded (Figure B2, Cell 7.2), the moving averages will be five-minute averages (Figure B2, Cell 10.2). If more than six minutes have passed since the most recent detection, or there has not yet been a detection and six minutes or more have passed since system startup, six-minute moving averages for ePERCLOS and esLANEX will be calculated (Figure B2, Cell 11.1).

If moving averages for ePERCLOS and esLANEX are calculated, selection and application of the proper threshold values will follow. If lane is present for all minutes in

the averages (Figure B3, Cell 12.1), then the esLANEX moving average is actually a LANEX average. There will be a detection if either the ePERCLOS average exceeds 0.14 or the LANEX average exceeds 0.10 (Figure B3, Cells 12.2, 17.2). If neither average exceeds threshold, there will not be a detection (Figure B3, Cell 17.1).

If lane is absent for exactly one minute in the moving averages (Figure B3, Cell 13.1), there will be a detection if either the ePERCLOS average exceeds 0.14 or the esLANEX average exceeds 0.1033 (Figure B3, Cells 13.2, 17.2). If lane is absent for exactly two minutes in the moving averages (Figure B3, Cell 14.1), there will be a detection if either the ePERCLOS average exceeds 0.14 or the esLANEX average exceeds 0.1066 (Figure B3, Cells 14.2, 17.2). If lane is absent for exactly three minutes in the moving averages (Figure B3, Cell 15.1), there will be a detection if either the ePERCLOS average exceeds 0.14 or the esLANEX average exceeds 0.11 (Figure B3, Cells 15.2, 17.2). In each of the three cases, there will not be a detection if neither average exceeds threshold (Figure B3, Cell 17.1).

If lane is absent for four or more minutes in the moving averages, the esLANEX average will no longer be a reliable measure of driver performance. As a result, detection will only occur if the ePERCLOS average exceeds 0.14 (Figure B3, Cells 16, 17.2).

If a detection occurs, the sequence of advisory tone, alarm, and countermeasures (if applicable) will be initiated (Figure B3, Cell 17.2). This sequence will take place independently of the main loop of the program so that the loop can continue on a regular one-minute cycle. Either a detection or a no-detection will result in the program reverting back to the beginning to begin calculation of measures over the next one-minute interval.

APPENDIX C

Potential Subject Screening Questionnaire

Subject Screening Questionnaire

- 1) Name _____
- 2) Telephone Number _____
- 3) Do you have a valid driver's license? YES NO
- 4) Are you a student? YES NO Major? _____
- 5) Age _____
- 6) Gender: M F
- 7) Do you ordinarily wear glasses or contact lenses?
- | | | |
|----------|-----|----|
| Glasses | YES | NO |
| Contacts | YES | NO |
- 8) Do you have any problems with your hearing?
- Explain: _____
- 9) What are your usual sleeping hours?
- Retire ____ : ____ AM PM Awake ____ : ____ AM PM
- 10) Have you ever had any trouble staying awake while driving? YES NO
- If YES, how often.....
- | | | | | |
|-------|-----------------|--------------|---------------------|-------|
| never | almost
never | occasionally | moderately
often | often |
|-------|-----------------|--------------|---------------------|-------|

11) Have you ever had an automobile accident or "near miss" due to drowsiness behind the wheel?

YES

NO

12) On the average, how many cups of coffee do you drink per day? _____

13) On average, how many caffeinated soft drinks do you drink per day? _____

14) How often do you take naps during the day?

	almost		moderately	
never	never	occasionally	often	often

15) Do you smoke cigarettes? YES NO

If YES, how many per day... _____ cigarettes OR _____ packs

16) Do you use other types of smoking materials such as a pipe or cigar?

YES

NO

If YES..... Type: _____ How often: _____

17) When do you ordinarily eat supper? _____ : _____ PM

18) If you snack at night, please describe what you eat and when.

Snack: _____ Time: _____ : _____ PM

19) What is your height and weight? HT: _____ ft _____ in WT: _____ lbs

APPENDIX D

Subject and Experimenter Rating Scales Used In Test and Evaluation Experiment

Subjective Ratings - Test and Evaluation Experiment

<i>How drowsy were</i>	Not Drowsy
<i>you when the tone</i>	Slightly Drowsy
<i>and advisory</i>	Moderately Drowsy
<i>message began?</i>	Very Drowsy
	Extremely Drowsy

Considering your level of drowsiness, do you think the message timing was

Too Early	→	Much Too Early
		Somewhat Early
About Right		Somewhat Late
Too Late	→	Much Too Late

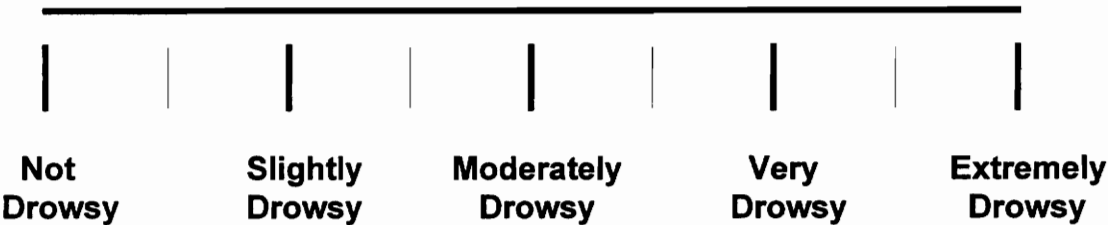
Figure D1: Subjective Rating Scale used by Subject at each Detection during Test and Evaluation Experiment.

Experimenter Ratings - Test and Evaluation Experiment

Subject # _____ End of Minute _____

Detection by: PERCLOS LANEX

How drowsy was the subject just before the tone and advisory message began?



How good was the subject's driving performance just before the tone and advisory message began?



Figure D2: Subjective Rating Scale used by Experimenter at each Detection during Test and Evaluation Experiment.

APPENDIX E

Information Sheet Regarding Procedures for Experiment and Participant Informed Consent Form

Introduction to the Study

The purpose of this research is to test and evaluate a drowsy driver detection system for possible future use in an automobile and to determine an appropriate configuration for drowsiness alarms and countermeasures. 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 Mark Lewin and Terry Fairbanks. The two researchers are graduate students in the Department of Industrial and Systems Engineering. Dr. Walter W. Wierwille is the principal 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. The screen in front of you will show a roadway on which you must drive.

If you decide to participate in this study, you must awake at 7:00 AM or before and go through your normal daytime activities without resting or napping. Then, at about 6:00 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 caffeinated or sugared beverages, such as coffee or cola. 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, drink caffeinated coffee, or drink caffeinated soft drinks, since these may effect the outcome of the experiment. However, you will be permitted to drink water or non-caffeinated, diet soft drinks. A member of the research team will remain with you during all of this time and will prevent you from napping.

Shortly after midnight the experimental session will begin. You will have a period of time (10 to 15 minutes) to get used to the simulator. After that, you will have a short break while the experimenters make final preparations. Then, the data gathering session will begin.

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.

You will be asked to drive the simulator in the same way as you would drive an actual automobile. If the drowsiness detection system detects that you are becoming too drowsy to drive safely, you will be presented with various alerting and alarming stimuli. You will be asked to react to the stimuli by pressing a reset button on the dashboard. Your opinion will be sought

as to the timeliness and effectiveness of the alerting and alarming stimuli. If drowsiness is detected, you may be presented with drowsiness countermeasures. The purpose of these countermeasures will be to help you maintain alertness.*

If possible, we would like you to complete the entire data gathering experiment, which will take a little less than 3 hours. You may, however, withdraw from the experiment at any time if you do not wish to continue for any reason.

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 be on a different schedule and will not be drowsy at this time.

Payment for the experiment will be \$5 per hour between 6:00 PM and midnight, and \$8 per hour from midnight until approximately 3:00 A.M. If you complete the experiment you will receive approximately \$54. 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 be paid for the time actually spent.

Initially, you will be asked to take a simple hearing test and a simple vision test. You will also be asked to 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.

* This paragraph is shown above in the form used for Phase Two (the test and evaluation experiment). For Phase Three (the algorithm revision experiment), the paragraph was replaced with the following:

“You will be asked to drive the simulator in the same way as you would drive an actual automobile. Once the data run has begun you must remain in the right lane. Your cruise control will be set at 60 mph.”

With the preceding exception, the same “Introduction to the Study” document was used for both Phase Two and Phase Three.

VIRGINIA POLYTECHNIC INSTITUTE AND STATE UNIVERSITY

Informed Consent for Participants of Investigative Projects

Title of Project: Simulator Test and Evaluation of a Drowsy Driver
Detection/Alarm/Countermeasures System (DDDACS)

Investigators: Mark G. Lewin, Rollin J. Fairbanks, Dr. Walter W. Wierwille

I. The Purpose of this Research/Project

The nature of this study and the purpose for conducting the research are contained in the document Introduction to the Study, which you have already read.

II. Procedures

The research procedures with which you will be involved are detailed in the document Introduction to the Study, which you have already read. By now you should have a clear understanding of what will be expected of you.

III. Risks

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. The simulator will be stopped, and you will then be guided out of the simulator.

The discomforts are:

- Possible discomfort associated with trying to drive while tired or drowsy.
- Possible discomfort associated with sitting in one seat for a long period of time.
- Possible minor motion sickness due to the movement of the simulator.

In order to minimize these risks to both yourself and the research team, 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 your being sleep deprived and staying up until approximately 3:00 AM.

IV. Benefits of this Project

There are no direct benefits to you from this research (other than payment). No promise or guarantee of any benefits to you (other than payment) have been made to encourage you to participate in this experiment. However, you may find the experiment interesting, and it may be beneficial to society. Your participation and that of other volunteers should aid in the implementation of an effective drowsiness detection and warning system in future automobiles.

V. Extent of Anonymity and Confidentiality

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

VI. Compensation

You will be paid at a rate of \$5.00 per hour between 6 P.M. and midnight and \$8.00 per hour after midnight. If you complete your participation you will be paid \$54.00. Cash payment will be made shortly after you have finished your participation.

VII. Freedom to Withdraw

You should know that at any time you are free to withdraw from participation in this research program, for any reason, without penalty. If you choose to withdraw, you will be compensated for the portion of the time of the study completed. You are free not to answer any questions or respond to experimental situations that you choose without penalty.

There may be circumstances under which the investigators may determine that you should not continue as a subject. If this occurs, you will be compensated for the portion of the project completed.

VIII. Approval of Research

This research project has been approved, as required, by the Institutional Review Board for Research Involving Human Subjects at Virginia Polytechnic Institute and State University.

IX. Subject's Responsibilities

I voluntarily agree to participate in this study. I have the following responsibilities:

- I agree to awake at or before 7:00 AM on the day of the experiment.
- I agree not to take any naps after 7:00 AM on the day of the experiment.
- I agree not to drink caffeinated coffee, drink caffeinated soft drinks, or ingest any other type of stimulant between 6:00 PM on the day of the experiment and the conclusion of the experiment.
- Once seated in the simulator, I agree not to attempt to leave the simulator until I have allowed the investigators to stop the simulator and guide me in exiting.
- I agree to allow one of the investigators to drive me home following the experiment.

X. Subject's Permission

I have read and understand the Informed Consent and conditions of this project. I have had all my questions answered. I hereby acknowledge the above and give my voluntary consent for participation in this project.

If I participate, I may withdraw at any time without penalty. I agree to abide by the rules of this project.

Signature

Date

Should I have any questions about this research or its conduct, I may contact:

Mark G. Lewin
Investigator

(540) 231-9084
Phone

Rollin J. Fairbanks
Investigator

(540) 231-9084
Phone

Walter W. Wierwille
Principal Investigator
Faculty Advisor

(540) 231-7952
Phone

E. R. Stout
Chair, Institutional Review Board
Research Division

(540) 231-9359
Phone

(A copy of this signed Informed Consent form is to be given to the research participant.)

APPENDIX F

Line Charts Comparing Measured PERCLOS with ePERCLOS Algorithm Outputs

$R = 0.48391$

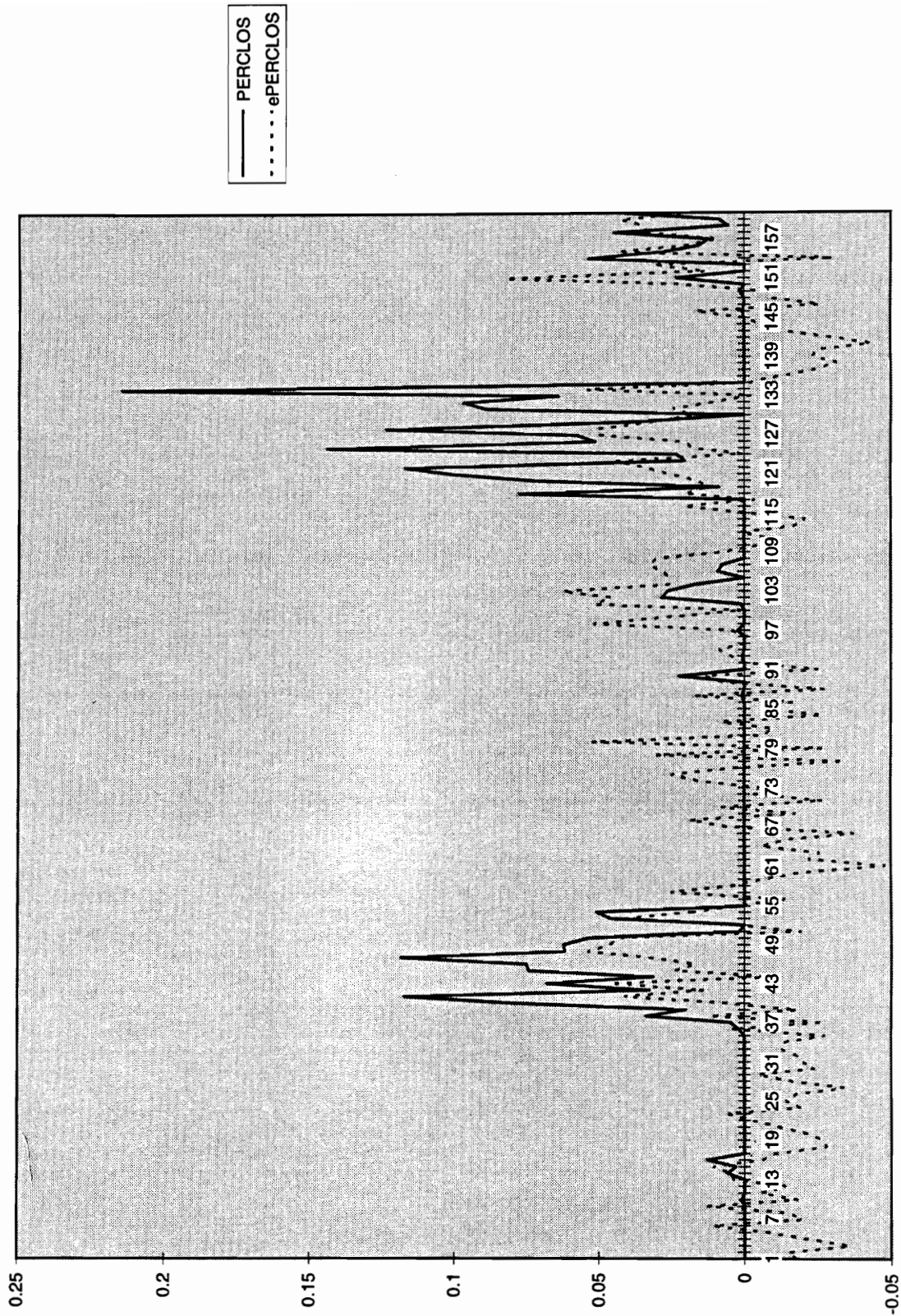


Figure F1: All Subjects: PERCLOS vs. ePERCLOS (straight 6-minute averages, minutes following detection removed)

$R = 0.54350$

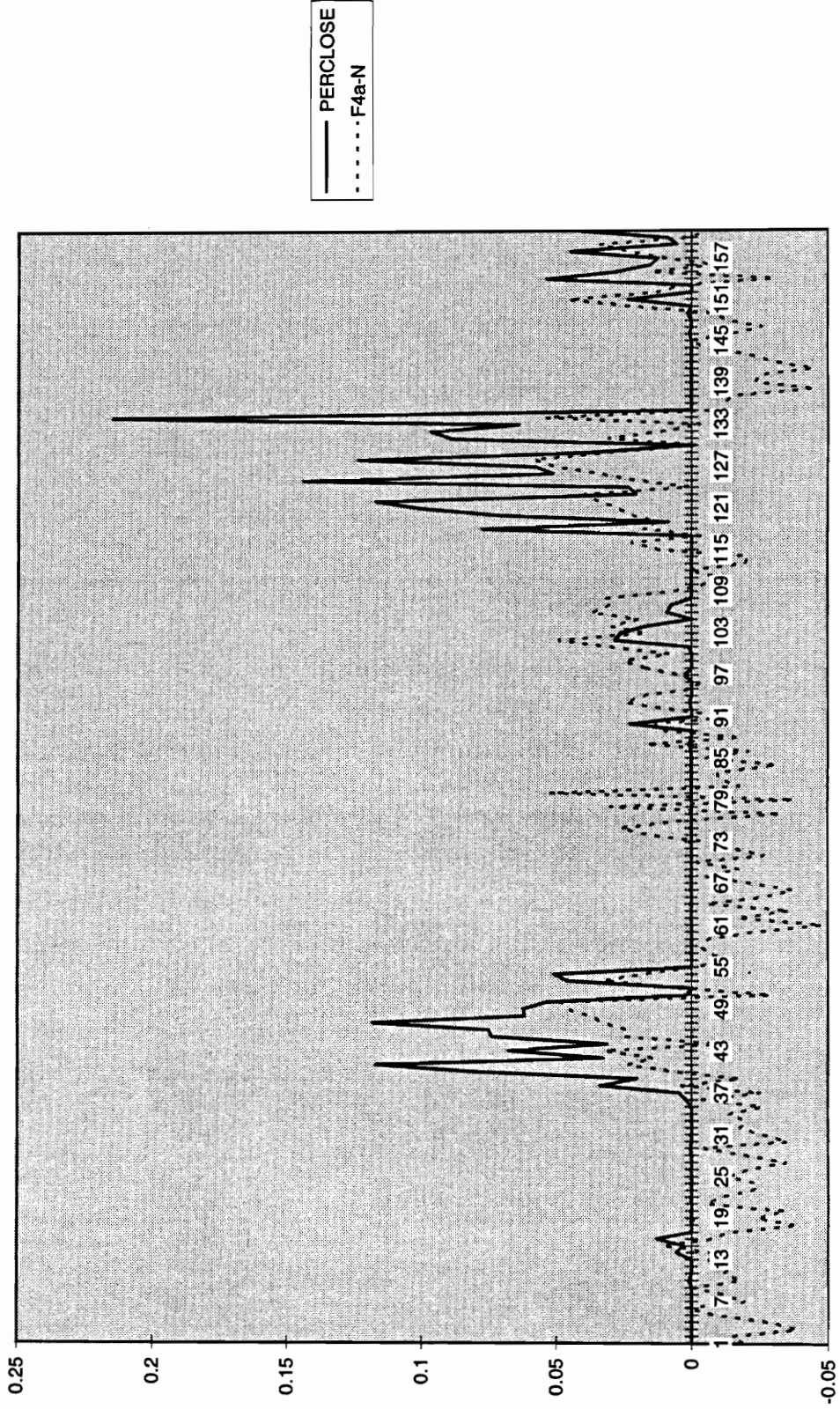


Figure F2: All Subjects: Measured PERCLOSE vs. Algorithm F4a-N (straight 6-minute averages, minutes following detection removed)

$R = 0.21614$

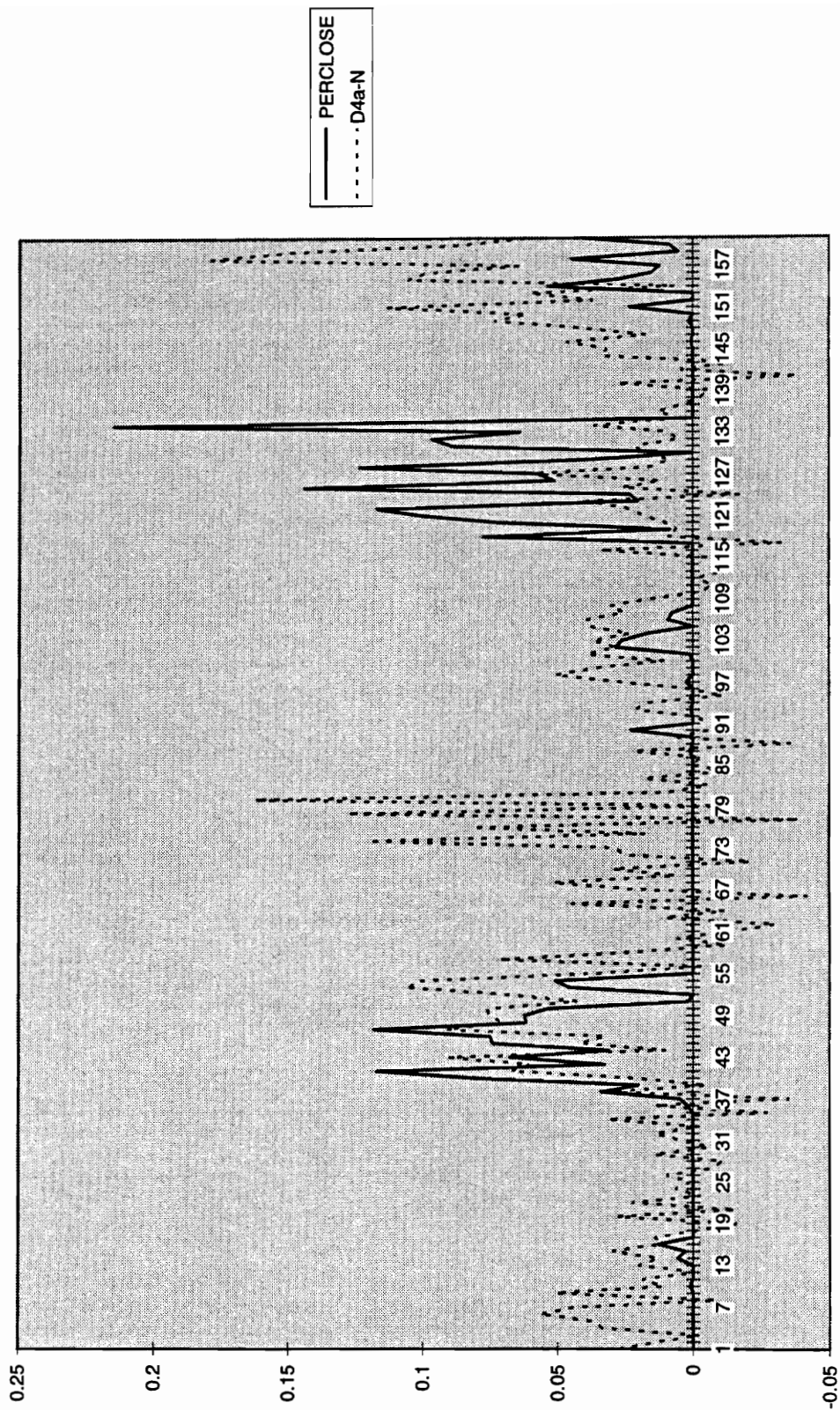


Figure F3: All Subjects: Measured PERCLOS vs. Algorithm D4a-N (straight 6-minute averages, minutes following detection removed)

$R = 0.2395$

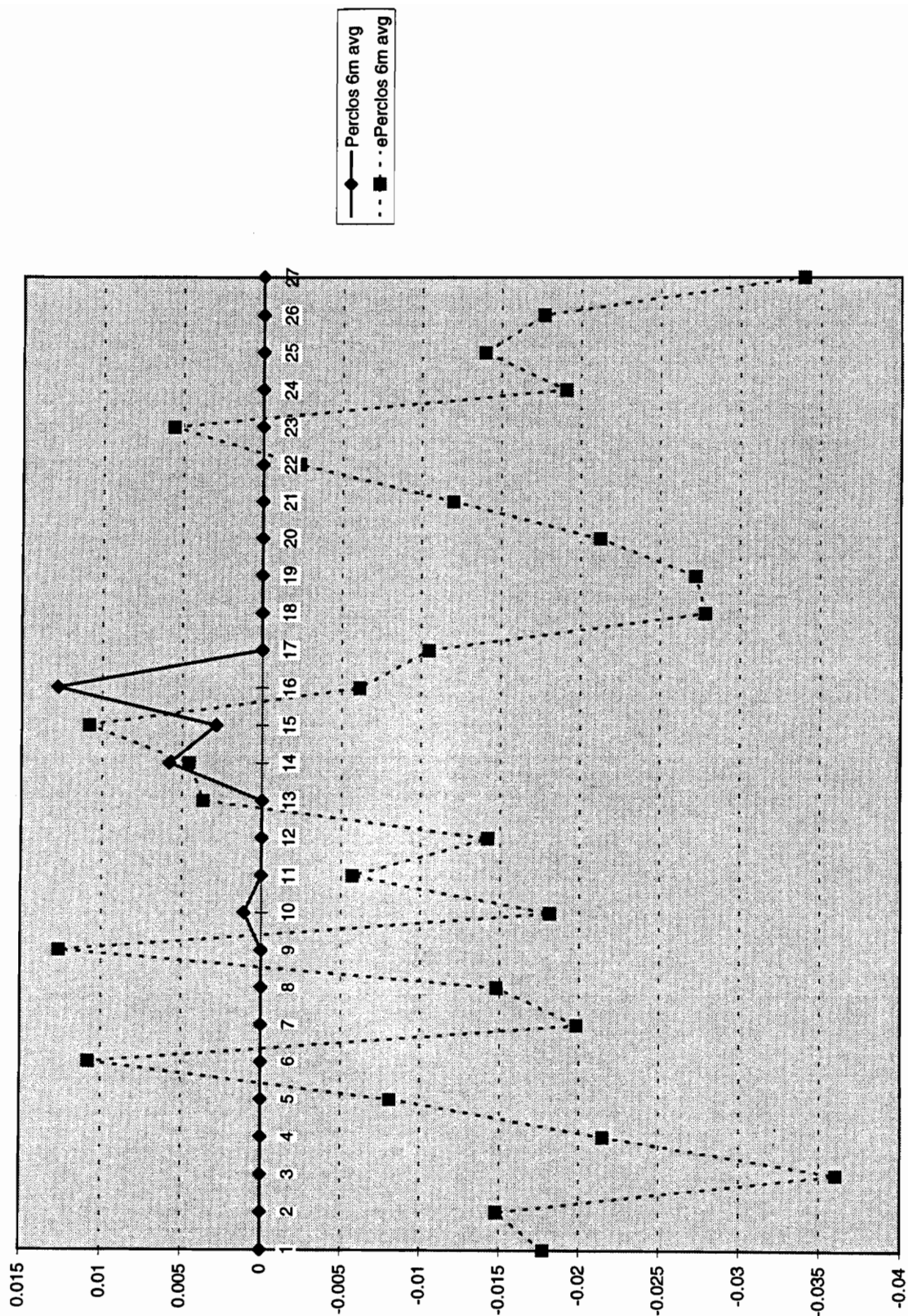


Figure F4: Subject #4: PERCLOS vs. ePERCLOS (straight 6-minute averages, minutes following detection removed)

F4a-N: $R = 0.2770$, D4a-N: $R = -0.0011$

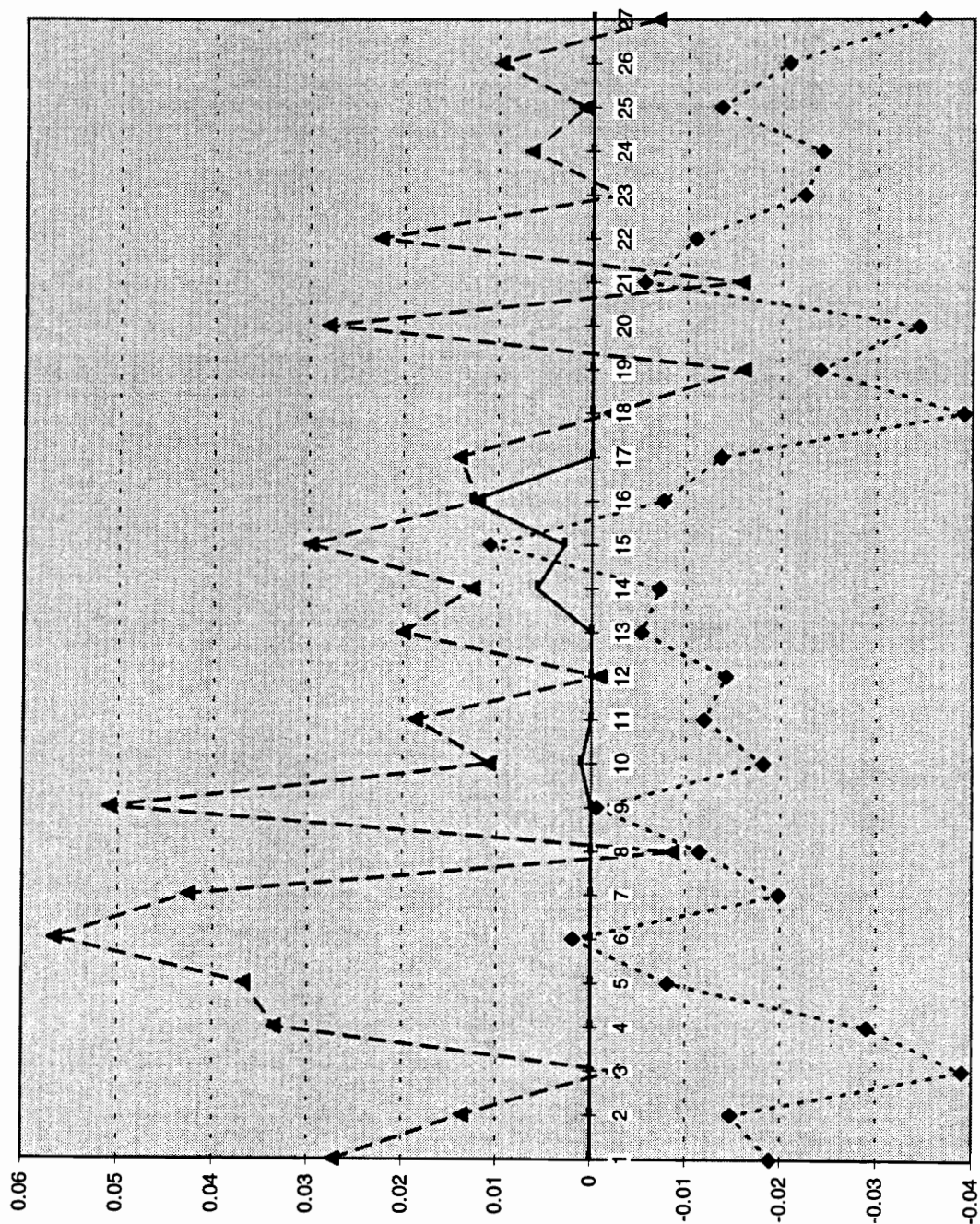


Figure F5: Subject #4: Measured PERCLOS vs. Algorithms F4a-N and D4a-N (straight 6-minute averages, minutes following detection removed)

$R = 0.7843$

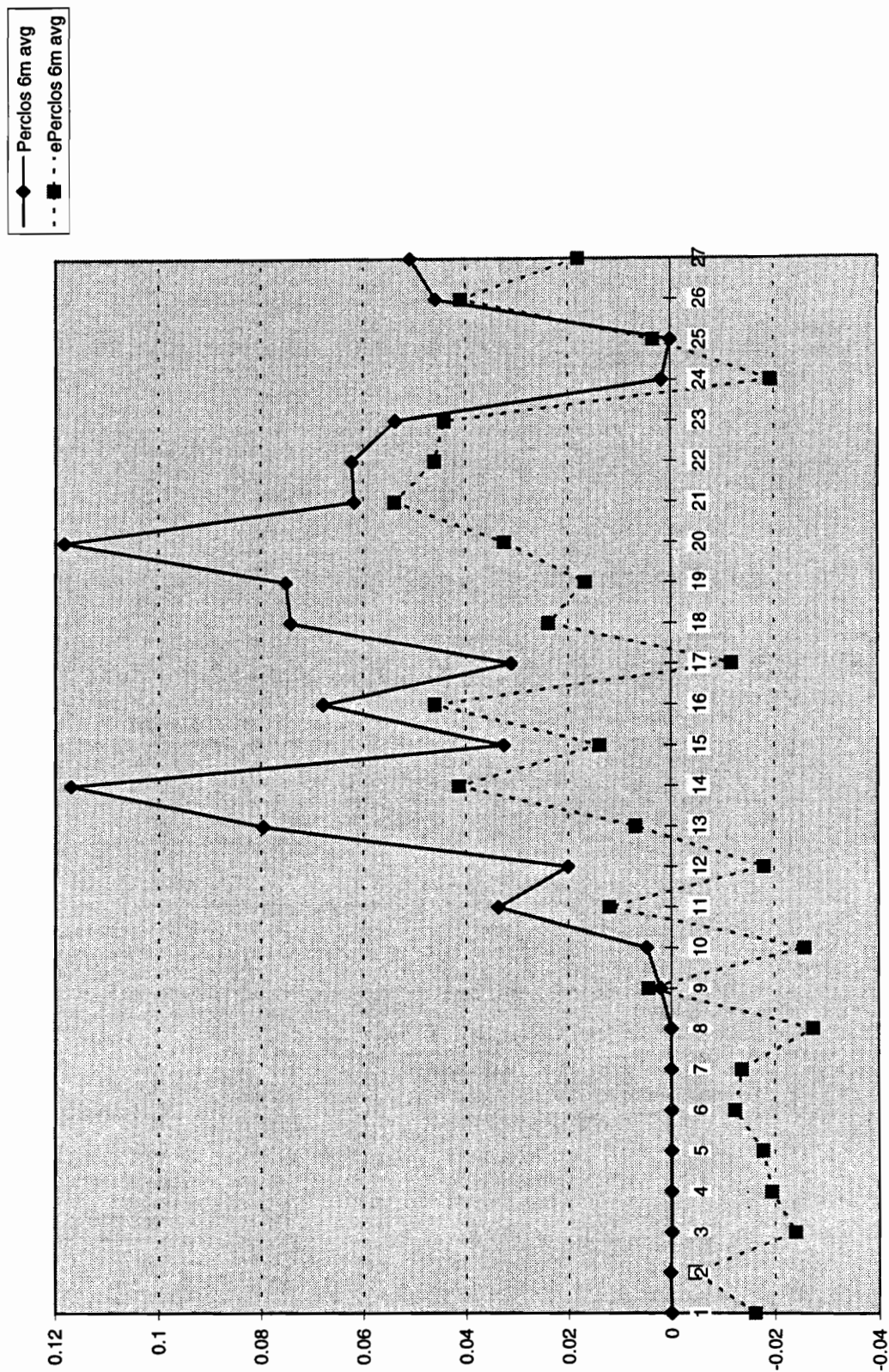


Figure F6: Subject #5: PERCLOS vs. ePERCLOS (straight 6-minute averages, minutes following detection removed)

$F4a-N R = 0.7820, D4a-N R = 0.6307$

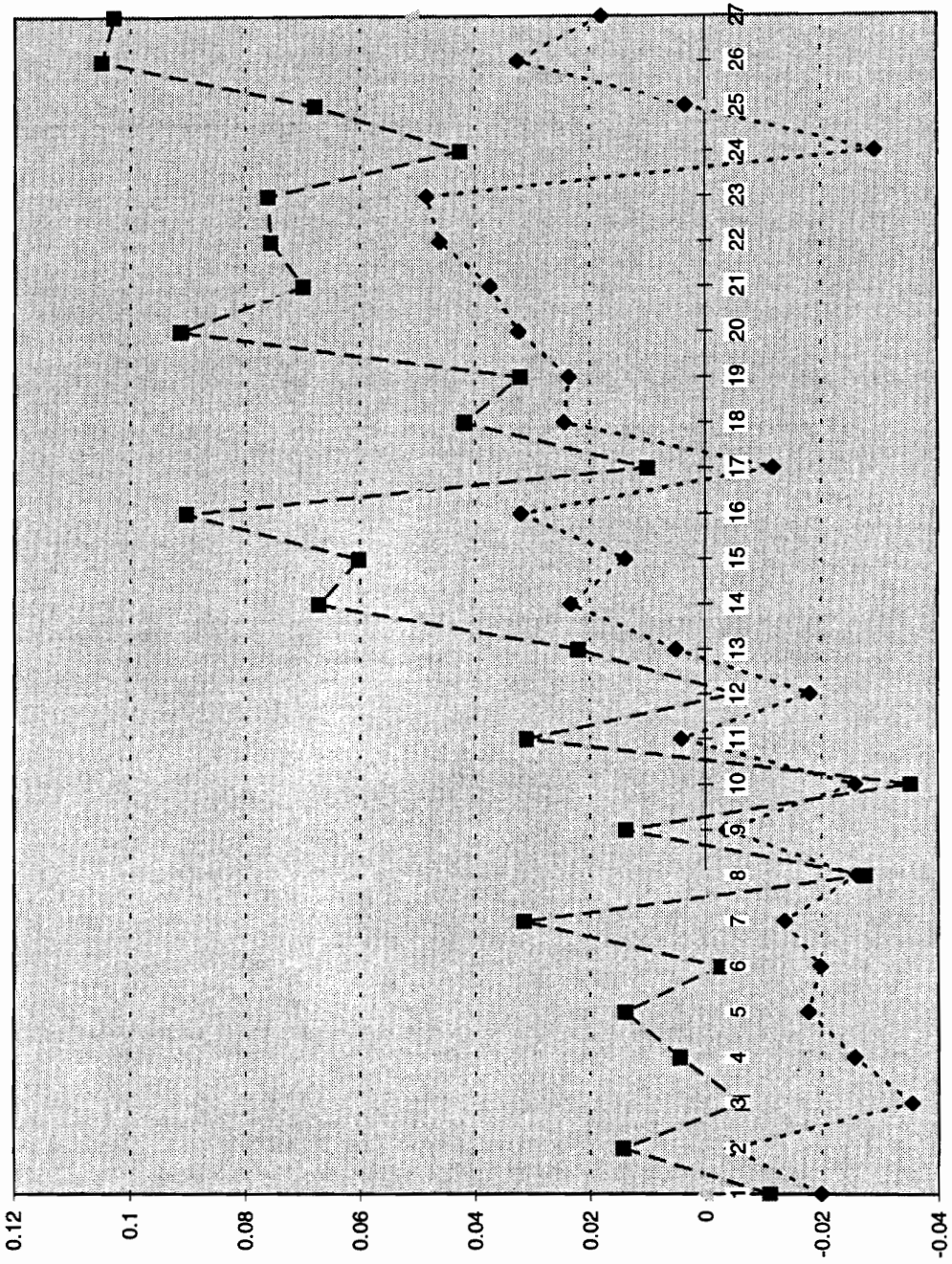


Figure F7: Subject #5: Measured PERCLOS vs. Algorithms F4a-N and D4a-N (straight 6-minute averages, minutes following detection removed)

R value not applicable

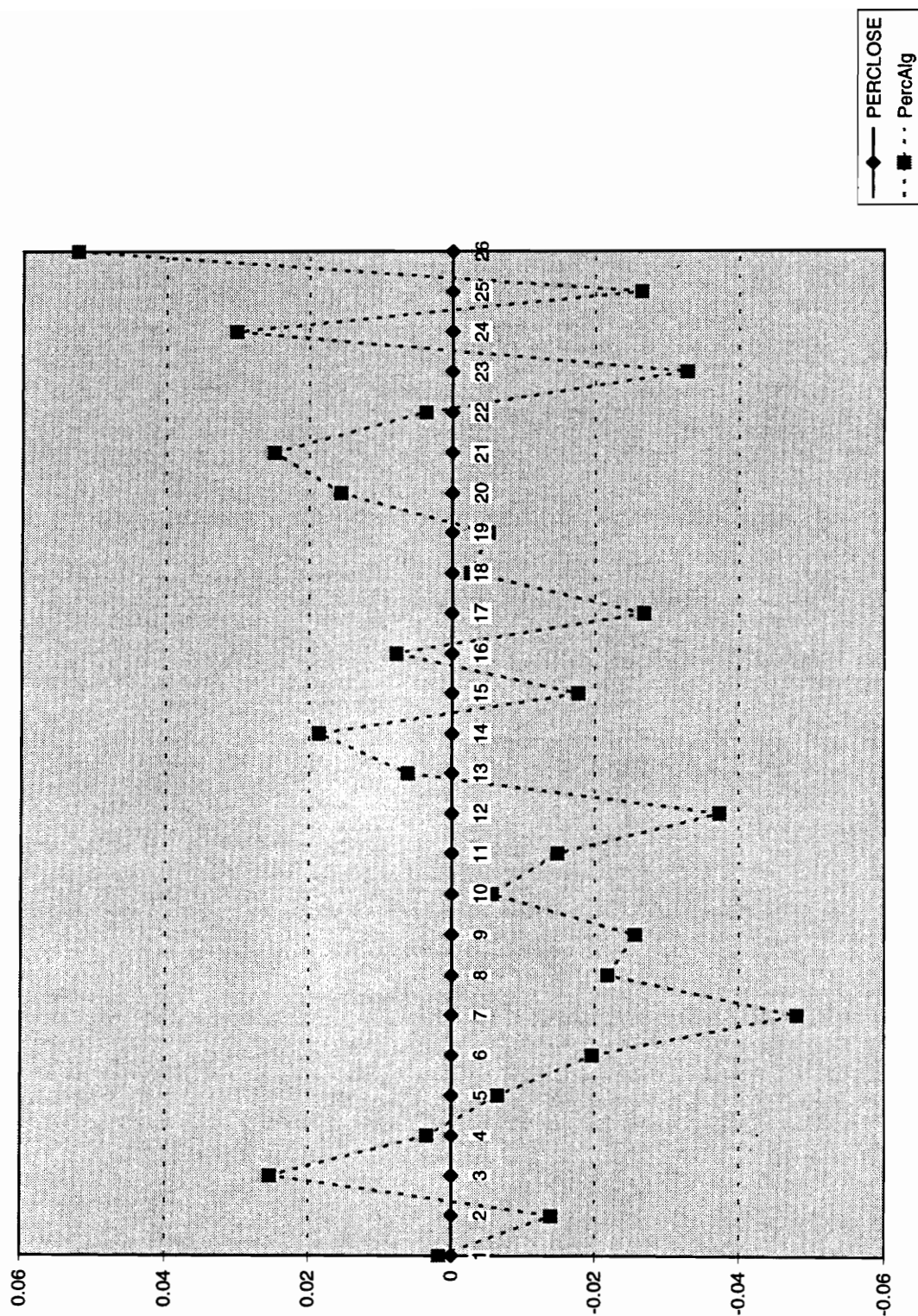


Figure F8: Subject #6: PERCLOS vs. ePERCLOS (straight 6-minute averages, minutes following detection removed)

R values not applicable

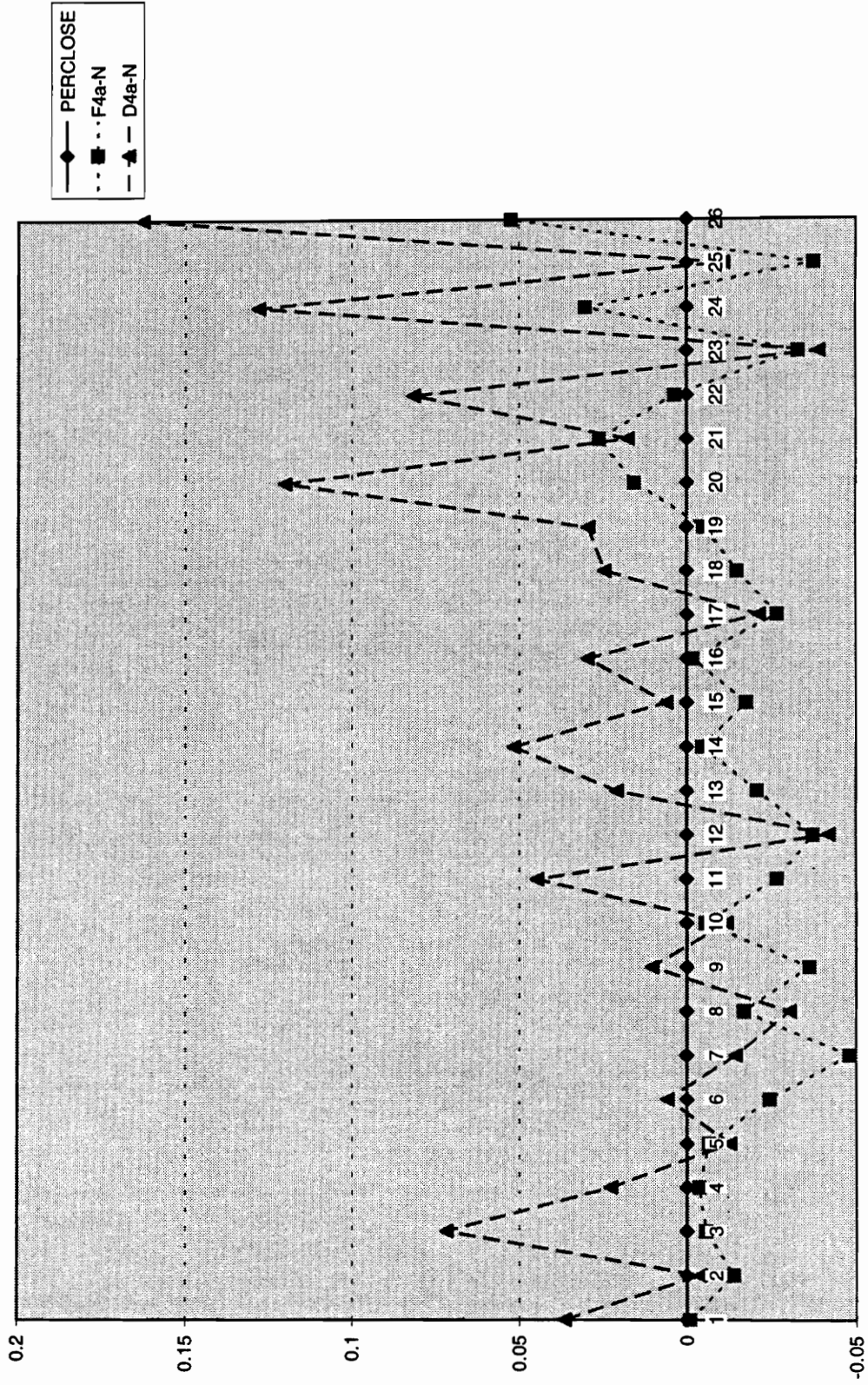


Figure F9: Subject #6: Measured PERCLOSE vs. Algorithms F4a-N and D4a-N (straight 6-minute averages, minutes following detection removed)

$R = 0.5549$

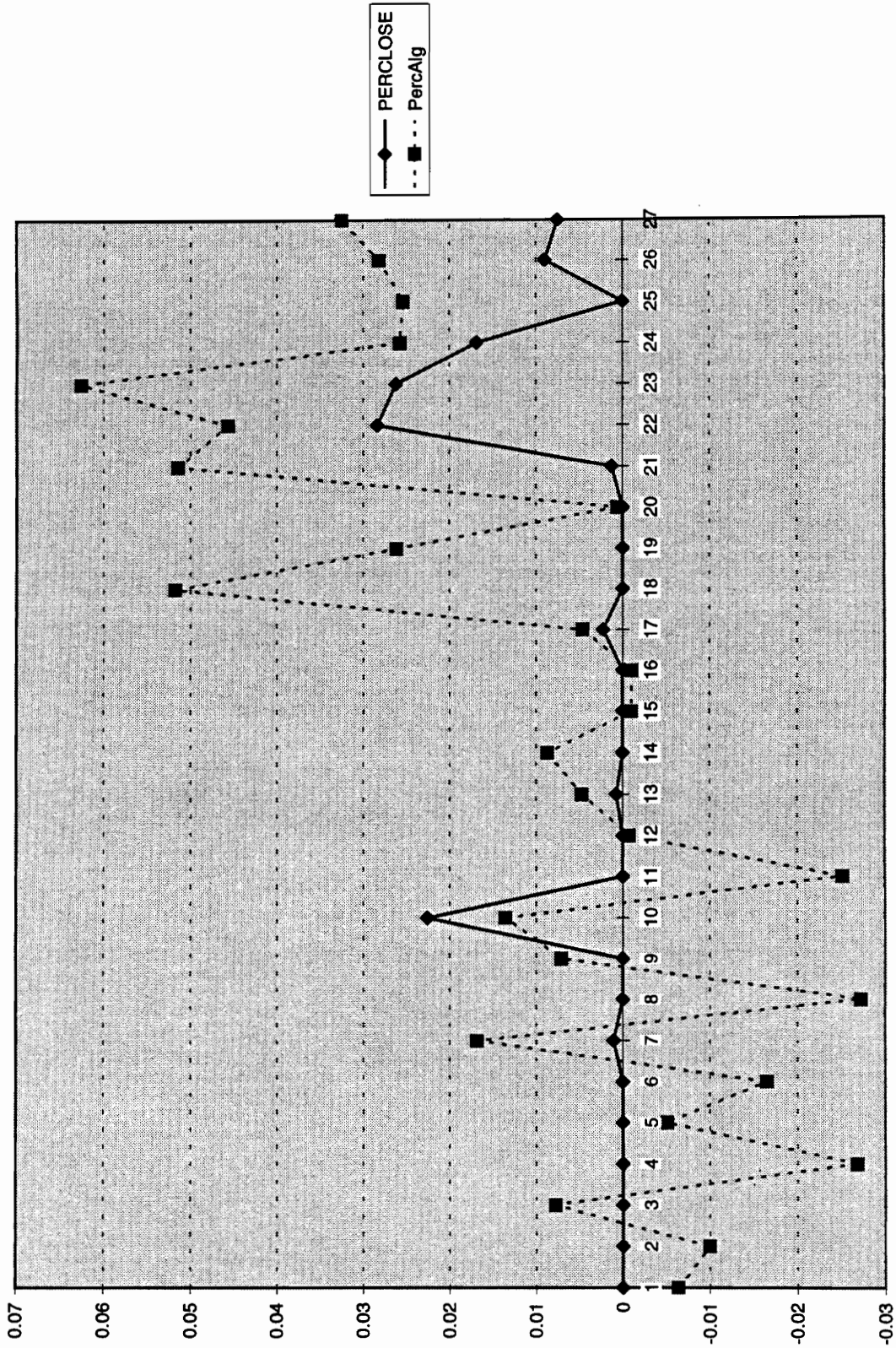


Figure F10: Subject #7: PERCLOS vs. ePERCLOS (straight 6-minute averages, minutes following detection removed)

F4a-N: $R = 0.5169$, D4a-N: $R = 0.2519$

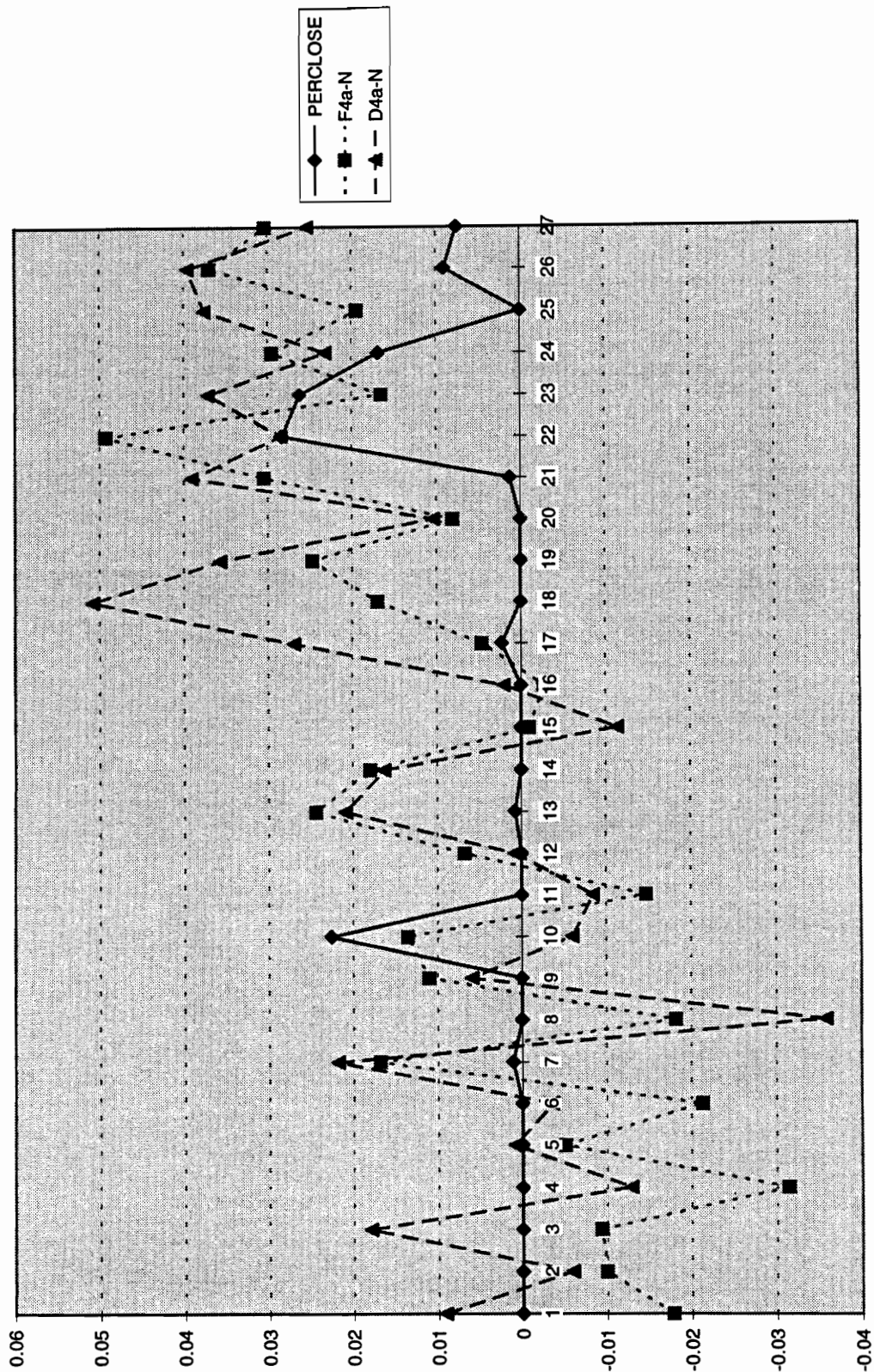


Figure F11: Subject #7: Measured PERCLOSE vs. Algorithms F4a-N and D4a-N (straight 6-minute averages, minutes following detection removed)

$R = 0.6052$

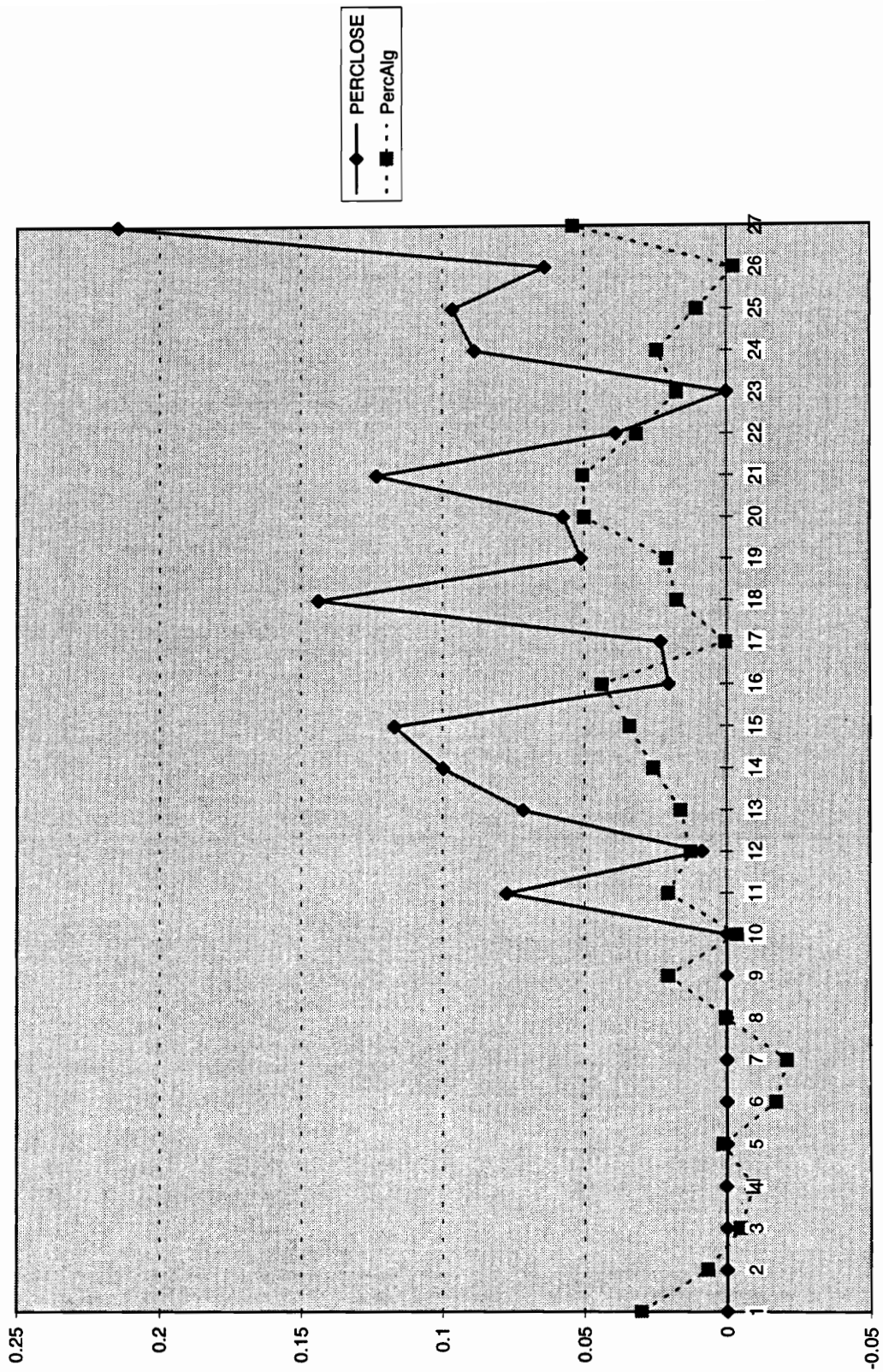


Figure F12: Subject #8: PERCLOS vs. ePERCLOS (straight 6-minute averages, minutes following detection removed)

F4a-N: $R = 0.6267$, D4a-N: $R = 0.4237$

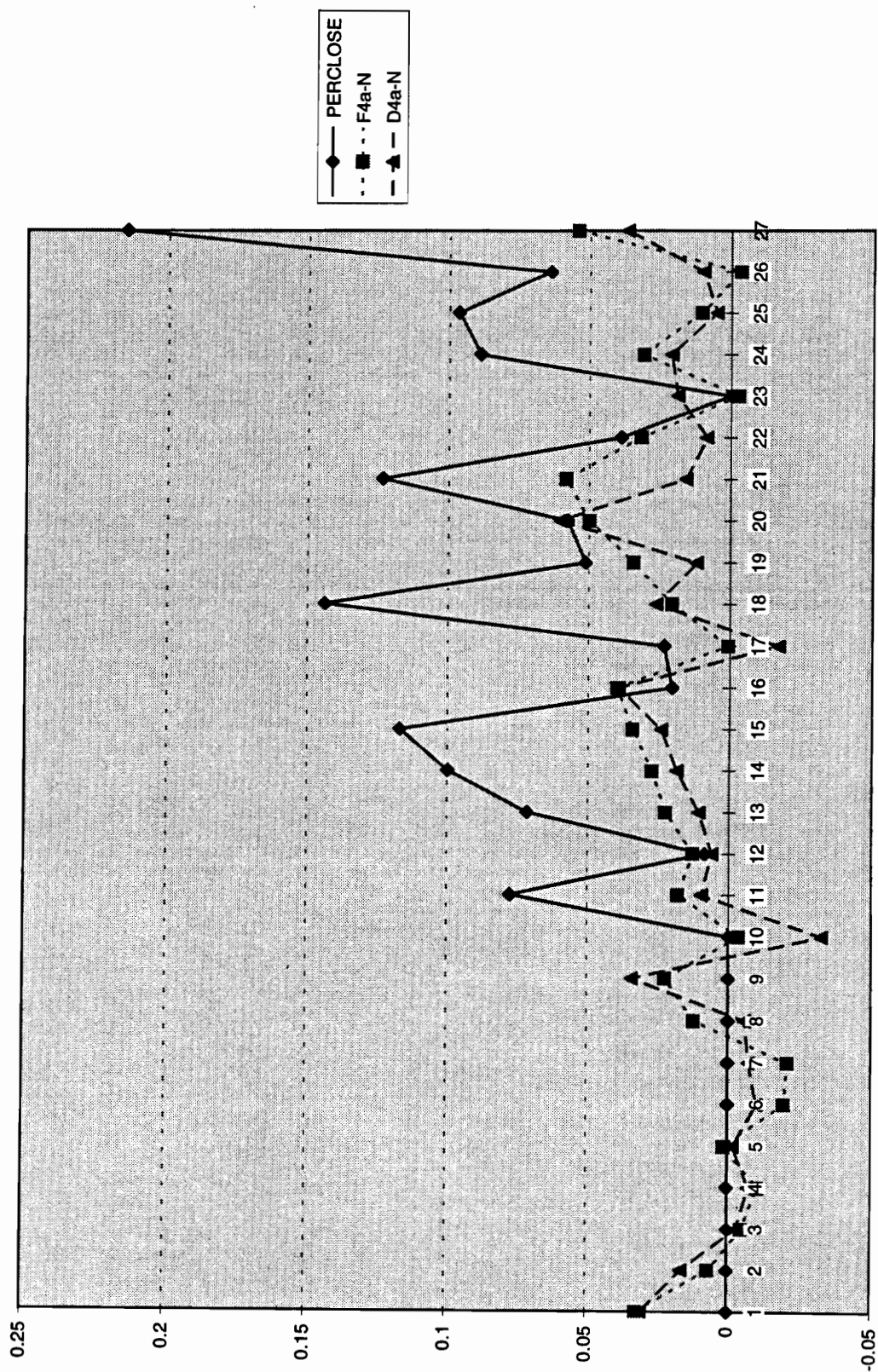


Figure F13: Subject #8: Measured PERCLOSE vs. Algorithms F4a-N and D4a-N (straight 6-minute averages, minutes following detection removed)

$R = 0.3356$

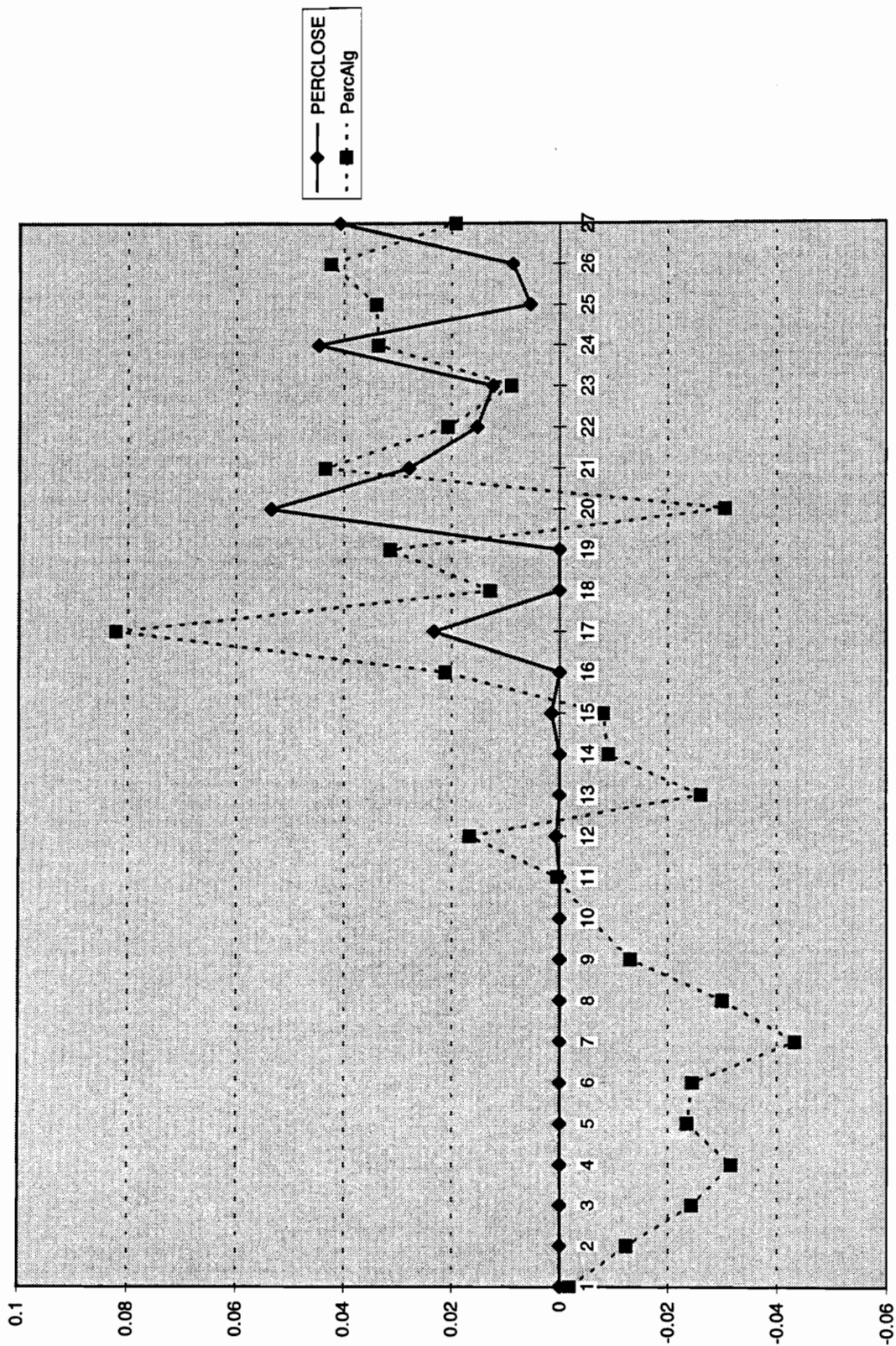


Figure F14: Subject #9: PERCLOS vs. ePERCLOS (straight 6-minute averages, minutes following detection removed)

F4a-N: $R = 0.2558$, D4a-N: $R = 0.4599$

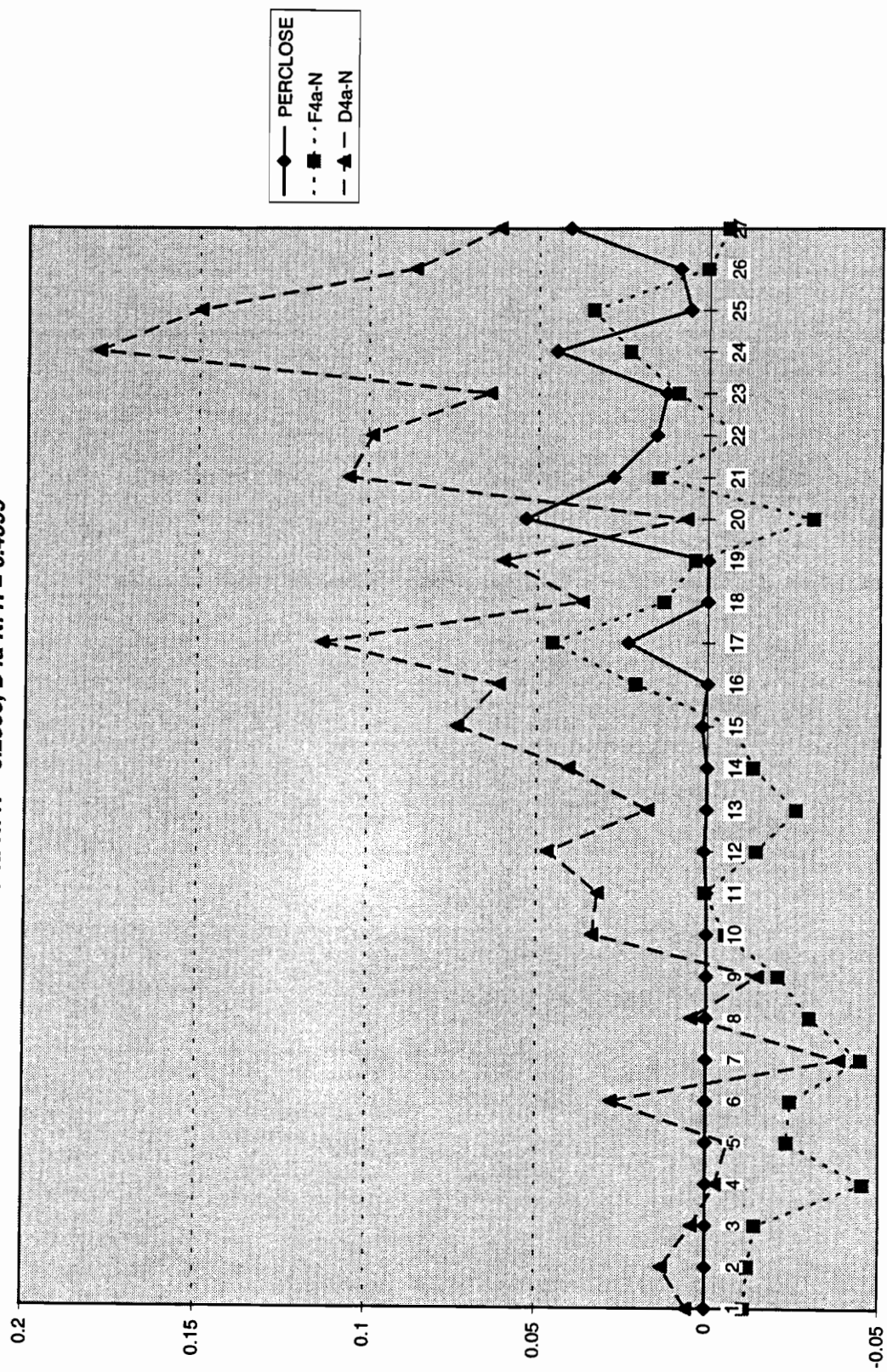


Figure F15: Subject #9: Measured PERCLOSE vs. Algorithms F4a-N and D4a-N (straight 6-minute averages, minutes following detection removed)

APPENDIX G

Line Charts Comparing Measured LANEX with eLANEX Algorithm Outputs

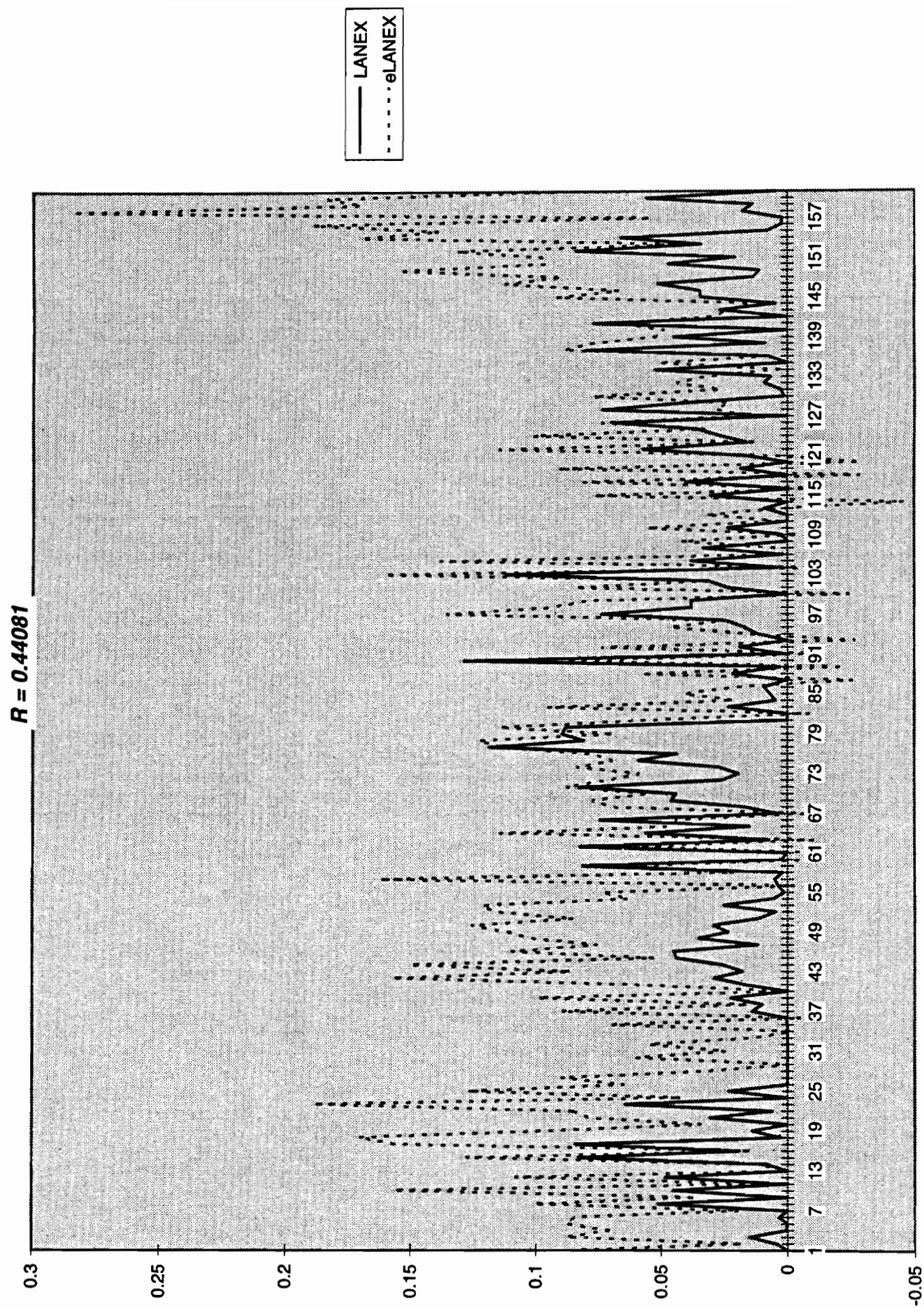


Figure G1: All Subjects: LANEX vs. eLANEX (straight 6-minute averages, minutes following detection removed)

$R = 0.6493$

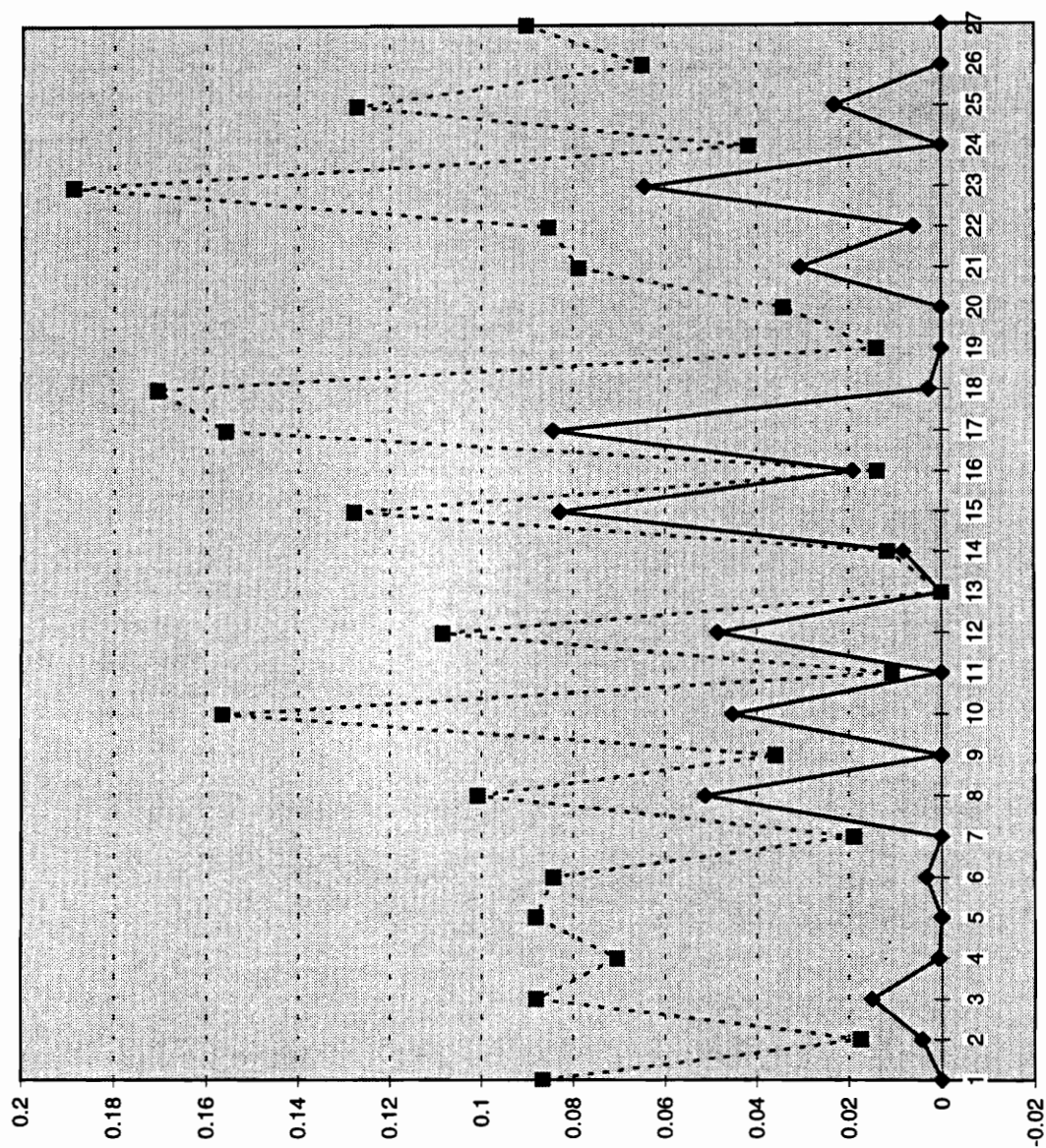


Figure G2: Subject #4: LANEX vs. eLANEX (straight 6-minute averages, minutes following detection removed)

$R = 0.6648$

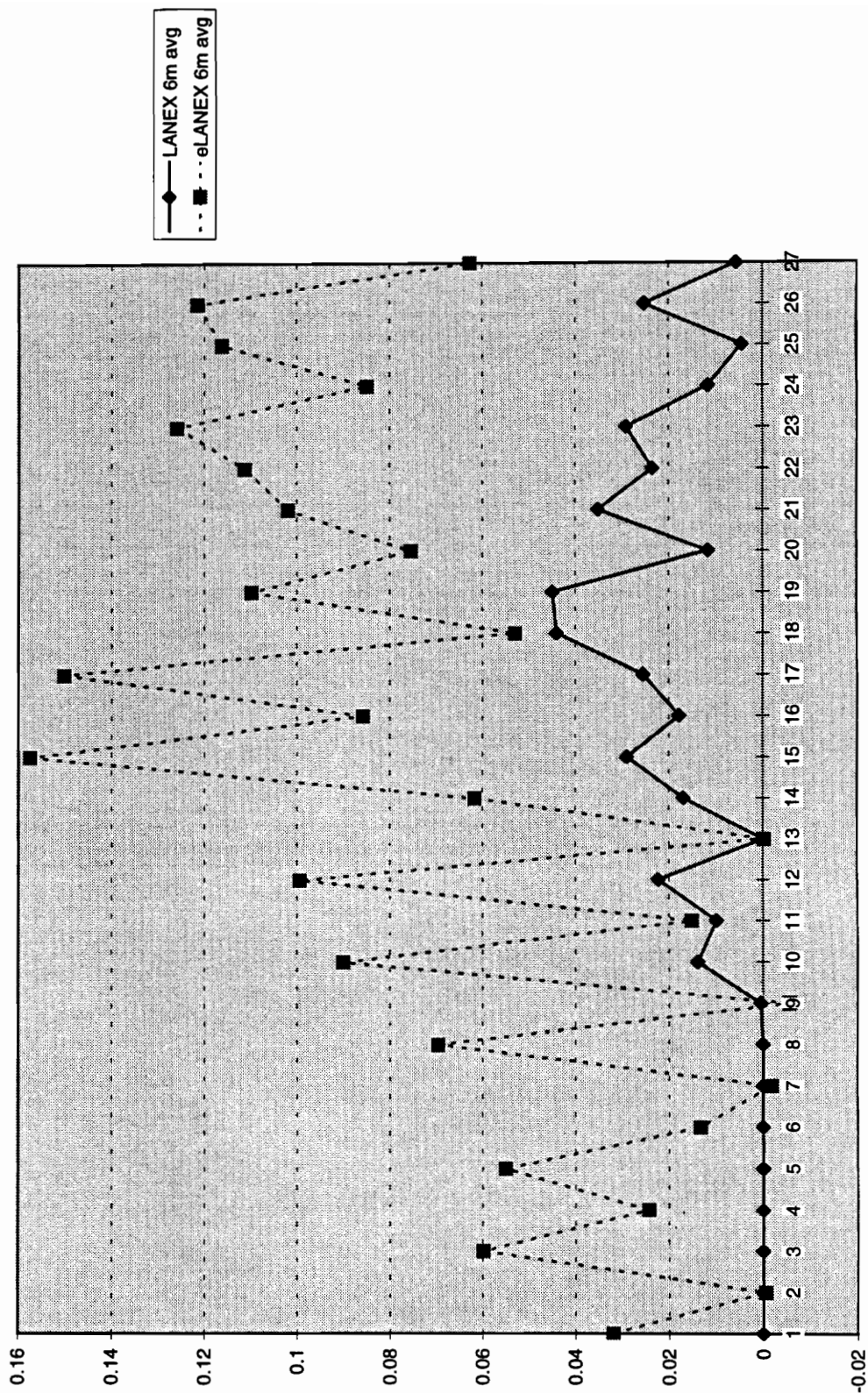


Figure G3: Subject #5: LANEX vs. eLANEX (straight 6-minute averages, minutes following detection removed)

$R = 0.5225$

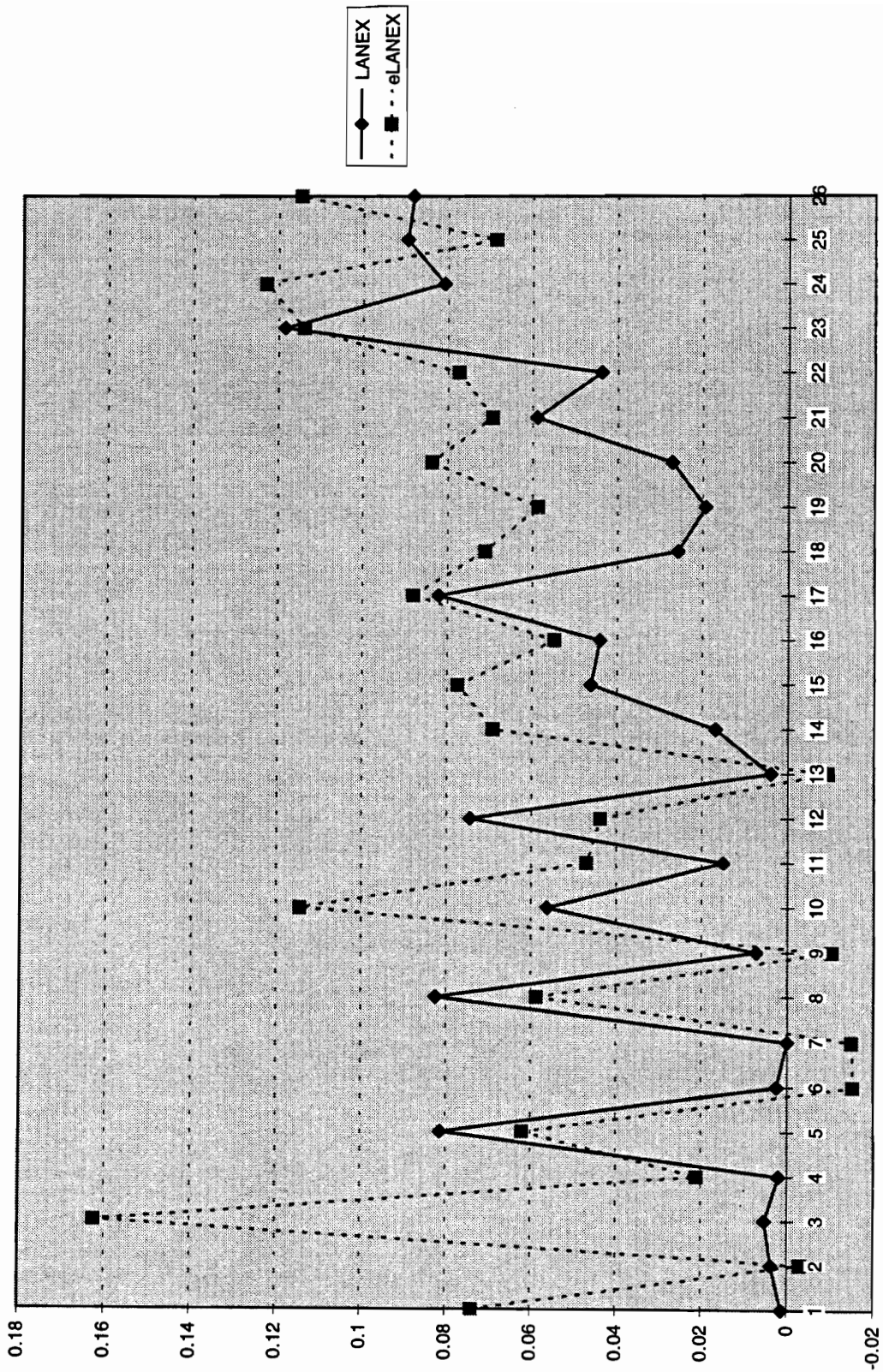


Figure G4: Subject #6: LANEX vs. eLANEX (straight 6-minute averages, minutes following detection removed)

$R = 0.7225$

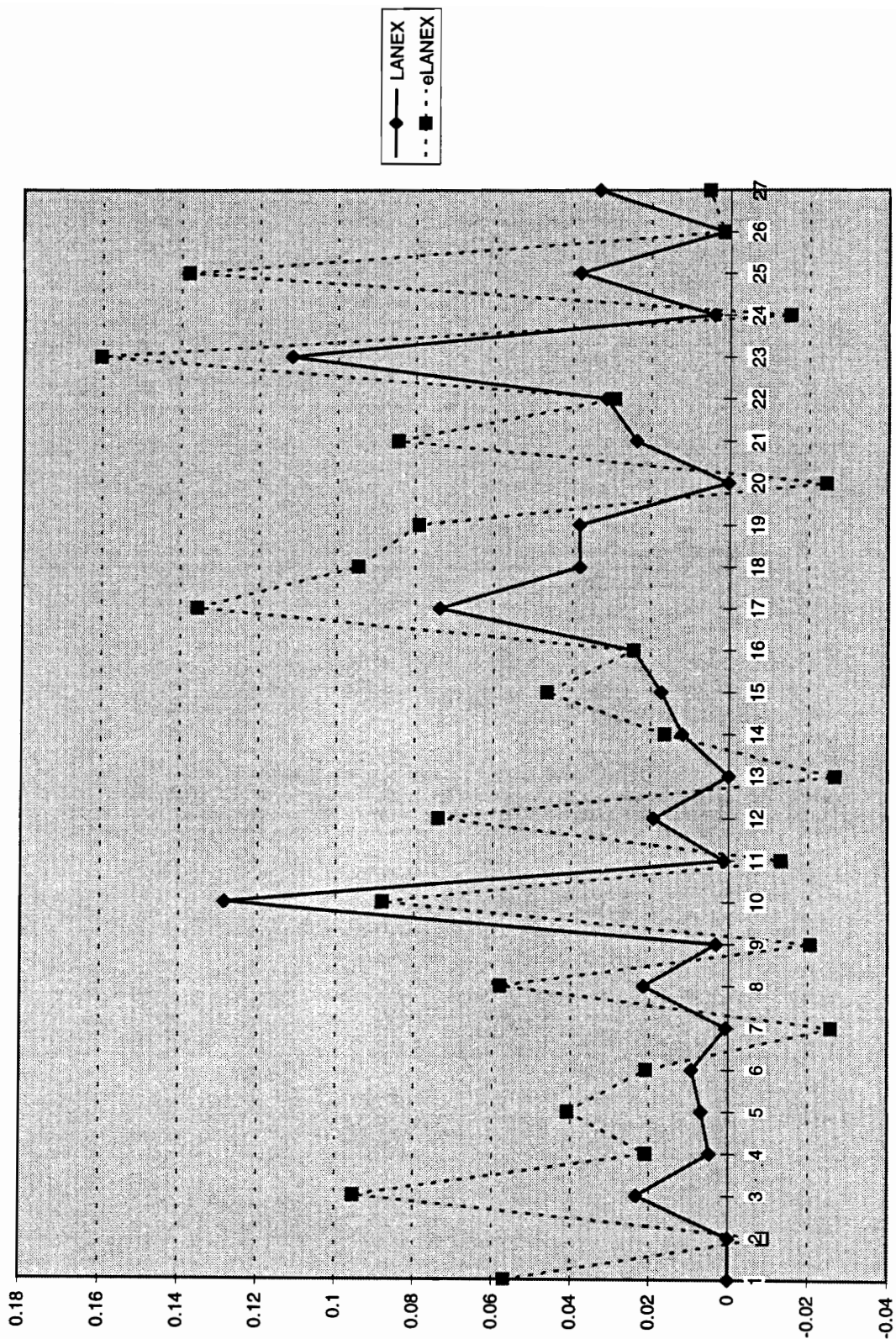


Figure G5: Subject #7: LANEX vs. eLANEX (straight 6-minute averages, minutes following detection removed)

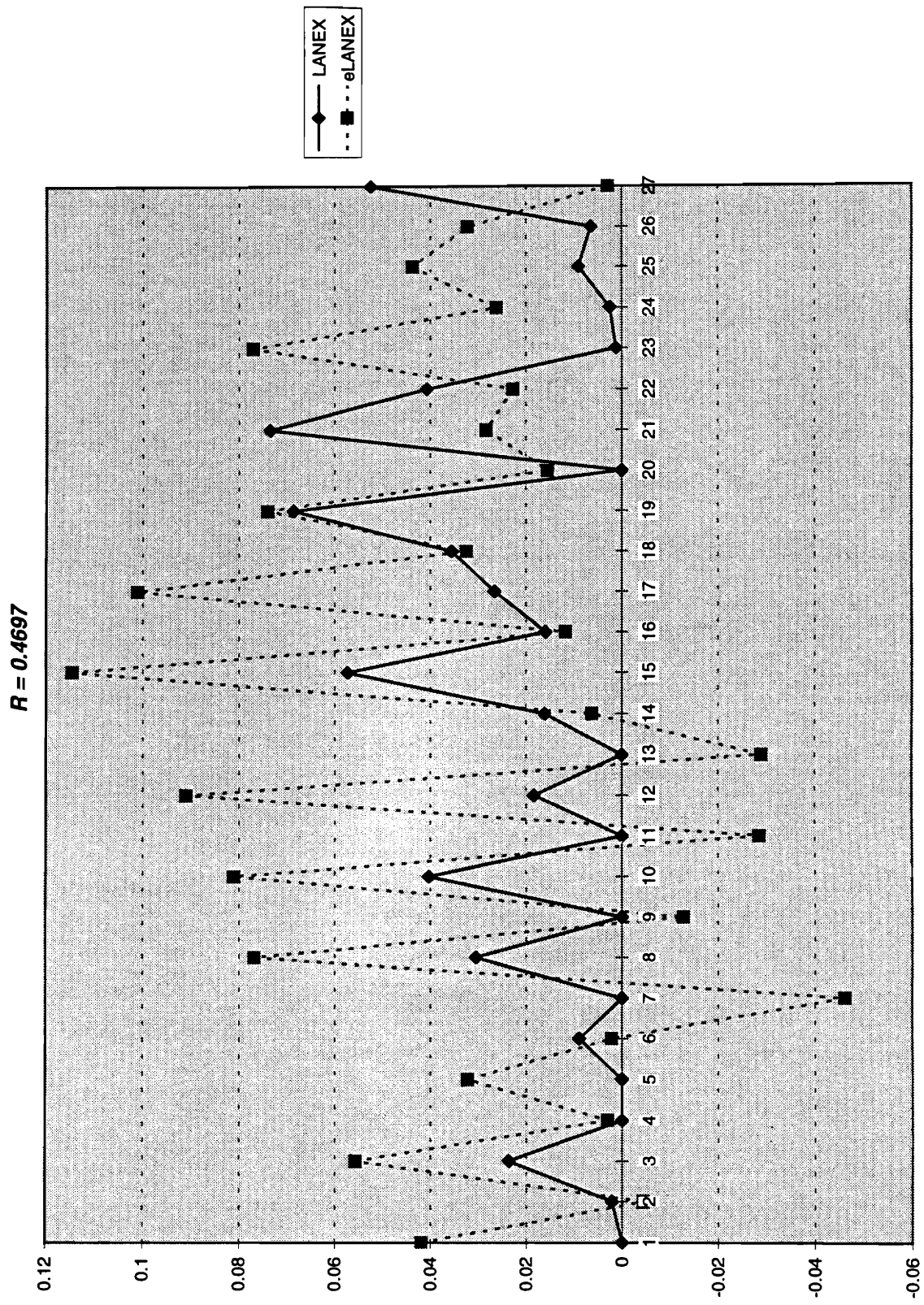


Figure G6: Subject #8: LANEX vs. eLANEX (straight 6-minute averages, minutes following detection removed)

$R = 0.1629$

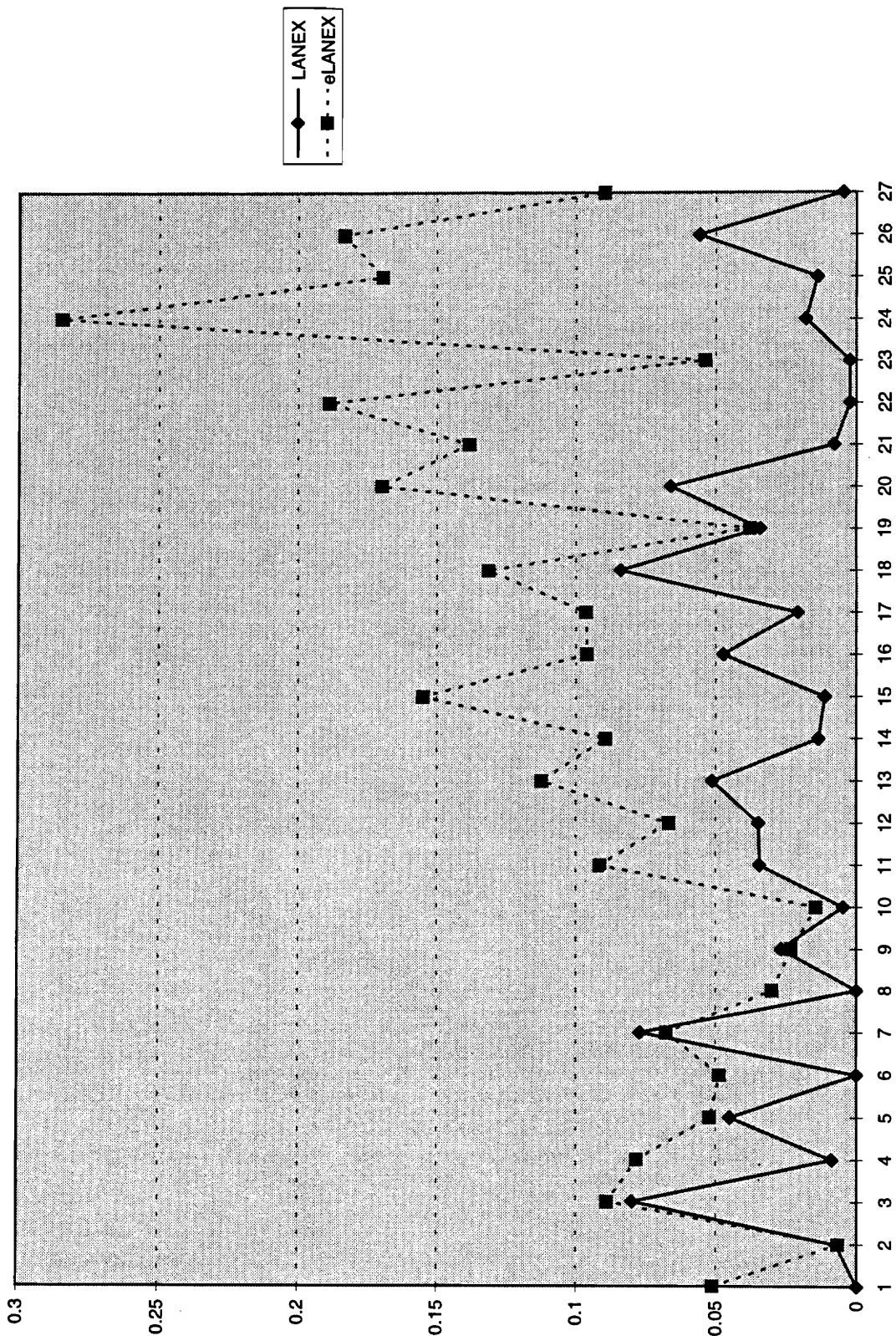


Figure G7: Subject #9: LANEX vs. eLANEX (straight 6-minute averages, minutes following detection removed)

APPENDIX H

Analyses of Variance for Evaluation of Effects of Alarms and Countermeasures on Driver Status and Performance

Table H1: ANOVA Summary Table for Countermeasure Type After Detection
(Dependent Variable: PERCLOS)

SUMMARY

<i>Groups</i>	<i>Count</i>	<i>PERCLOS Sum</i>	<i>PERCLOS Average</i>	<i>PERCLOS Variance</i>
none	110	0.627391384	0.005703558	0.000401958
a/o	138	0.539078935	0.003906369	0.000197857
lane	76	2.715141013	0.03572554	0.005264562

ANOVA

<i>Source of Variation</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>	<i>F crit</i>
Between Groups	0.0561813	2	0.0280906	19.359882	1.153×10^{-8}	3.0238638
Within Groups	0.4657619	321	0.0014510			
Total	0.5219432	323				

Table H2: ANOVA Summary Table for Countermeasure Type After Detection
(Dependent Variable: LANEX)

SUMMARY

<i>Groups</i>	<i>Count</i>	<i>LANEX Sum</i>	<i>LANEX Average</i>	<i>LANEX Variance</i>
none	110	4.329441916	0.039358563	0.004093243
a/o	138	2.501431561	0.018126316	0.001945325
lane	76	0.887098706	0.011672351	0.00074692

ANOVA

<i>Source of Variation</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>	<i>F crit</i>
Between Groups	0.0422478	2	0.0211239	8.8211863	0.0001865	3.0238638
Within Groups	0.7686920	321	0.0023947			
Total	0.8109398	323				

Table H3: ANOVA Summary Table for Countermeasure Type After Detection
(Dependent Variable: LANDEV)

SUMMARY

<i>Groups</i>	<i>Count</i>	<i>LANDEV Sum</i>	<i>LANDEV Average</i>	<i>LANDEV Variance</i>
none	110	92.2138669	0.838307881	0.125016736
a/o	138	122.9491528	0.89093589	0.119288757
lane	76	85.2669987	1.121934193	0.579010266

ANOVA

<i>Source of Variation</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>	<i>F crit</i>
Between Groups	3.9327067	2	1.9663534	8.600015	0.0002300	3.0238638
Within Groups	73.395154	321	0.2286453			
Total	77.327861	323				

Table H4: ANOVA Summary Table for Test and Evaluation Experiment vs.
Development Experiment, Drowsy Condition, No Countermeasure Operating
(Dependent Variable: PERCLOS)

SUMMARY

<i>Groups</i>	<i>Count</i>	<i>Sum</i>	<i>Average</i>	<i>Variance</i>
Test & Eval. PERCLOS	19	0.104565231	0.005503433	0.000217676
Development PERCLOS	62	9.339155777	0.150631545	0.017202707

ANOVA

<i>Source of Variation</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>	<i>F crit</i>
Between Groups	0.3063115	1	0.3063115	22.974457	7.575×10^{-6}	3.9619010
Within Groups	1.0532833	79	0.0133327			
Total	1.3595948	80				

Table H5: ANOVA Summary Table for Test and Evaluation Experiment vs.
Development Experiment, Drowsy Condition, No Countermeasure Operating
(Dependent Variable: LANEX)

SUMMARY				
<i>Groups</i>	<i>Count</i>	<i>Sum</i>	<i>Average</i>	<i>Variance</i>
Test & Eval. LANEX	19	0.779056309	0.041002964	0.001549067
Development LANEX	62	10.21316994	0.164728547	0.011476526

ANOVA						
<i>Source of Variation</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>	<i>F crit</i>
Between Groups	0.2226277	1	0.2226277	24.160397	4.714×10 ⁻⁶	3.9619010
Within Groups	0.7279513	79	0.0092146			
Total	0.9505790	80				

Table H6: ANOVA Summary Table for Test and Evaluation Experiment vs.
Development Experiment, Drowsy Condition, No Countermeasure Operating
(Dependent Variable: LANDEV)

SUMMARY				
<i>Groups</i>	<i>Count</i>	<i>Sum</i>	<i>Average</i>	<i>Variance</i>
Test & Eval. LANDEV	19	16.06613985	0.845586308	0.043000252
Development LANDEV	62	123.9100933	1.998549893	3.265676166

ANOVA						
<i>Source of Variation</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>	<i>F crit</i>
Between Groups	19.332653	1	19.332653	7.6371520	0.0071117	3.9619010
Within Groups	199.98025	79	2.5313956			
Total	219.31290	80				

Table H7: ANOVA Summary Table for Test and Evaluation Experiment vs. Development Experiment, Drowsy Condition, A/O Task Operating (Dependent Variable: PERCLOS)

SUMMARY				
<i>Groups</i>	<i>Count</i>	<i>Sum</i>	<i>Average</i>	<i>Variance</i>
Test & Eval. PERCLOS	23	0.089846489	0.003906369	7.09043×10^{-5}
Development PERCLOS	20	3.618572834	0.180928642	0.00921886

ANOVA						
<i>Source of Variation</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>	<i>F crit</i>
Between Groups	0.3352318	1	0.3352318	77.776371	5.126×10^{-11}	4.0785437
Within Groups	0.1767182	41	0.0043102			
Total	0.5119500	42				

Table H8: ANOVA Summary Table for Test and Evaluation Experiment vs. Development Experiment, Drowsy Condition, A/O Task Operating (Dependent Variable: LANEX)

SUMMARY				
<i>Groups</i>	<i>Count</i>	<i>Sum</i>	<i>Average</i>	<i>Variance</i>
Test & Eval. LANEX	23	0.41690526	0.018126316	0.000602849
Development LANEX	20	4.891557224	0.244577861	0.026083578

ANOVA						
<i>Source of Variation</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>	<i>F crit</i>
Between Groups	0.5485800	1	0.5485800	44.201139	5.181×10^{-8}	4.0785437
Within Groups	0.5088507	41	0.0124110			
Total	1.0574306	42				

Table H9: ANOVA Summary Table for Test and Evaluation Experiment vs. Development Experiment, Drowsy Condition, A/O Task Operating (Dependent Variable: LANDEV)

SUMMARY

<i>Groups</i>	<i>Count</i>	<i>Sum</i>	<i>Average</i>	<i>Variance</i>
Test & Eval. LANDEV	23	20.49152547	0.89093589	0.057672623
Development LANDEV	20	47.80940916	2.390470458	1.015238466

ANOVA

<i>Source of Variation</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>	<i>F crit</i>
Between Groups	24.054833	1	24.054833	47.973167	2.094×10^{-8}	4.0785437
Within Groups	20.558329	41	0.5014226			
Total	44.613161	42				

APPENDIX I

Distribution Analyses for EYEMEAS, PERCLOS, DRVDROW, LANDEV, LANEX,
and LNMNSQ in One-Minute Interval Data Set

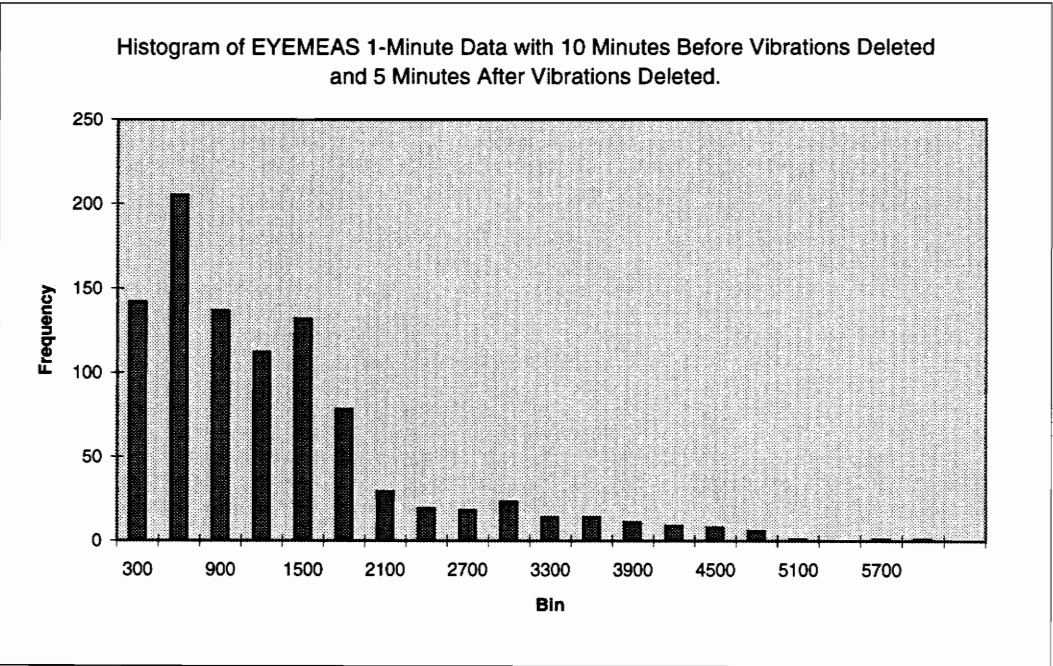
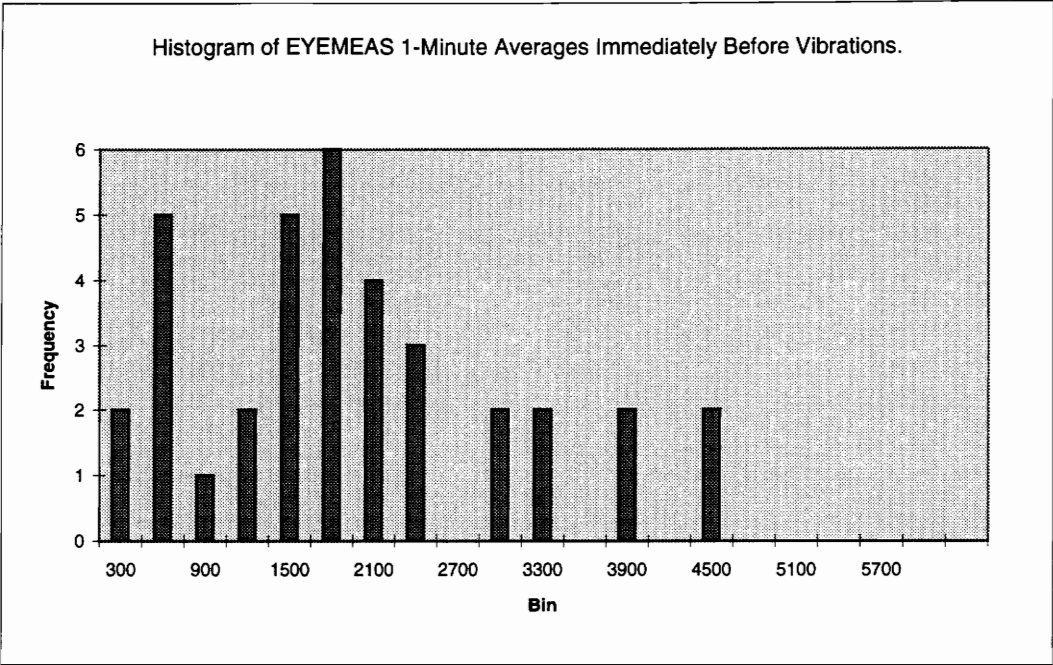


Figure I1: EYEMEAS 1-minute histograms.

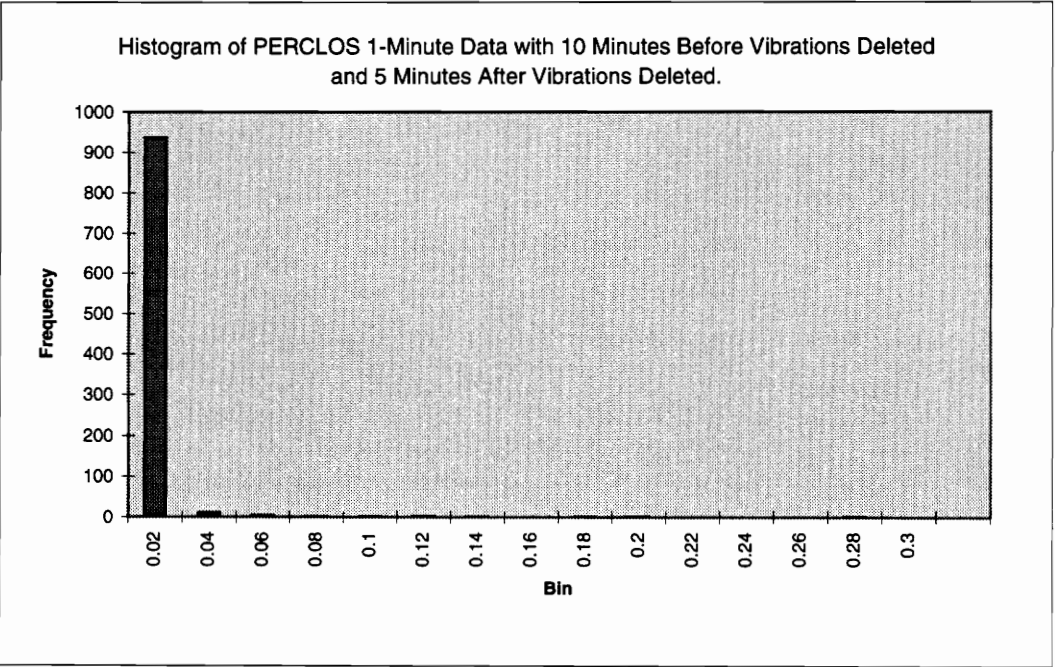
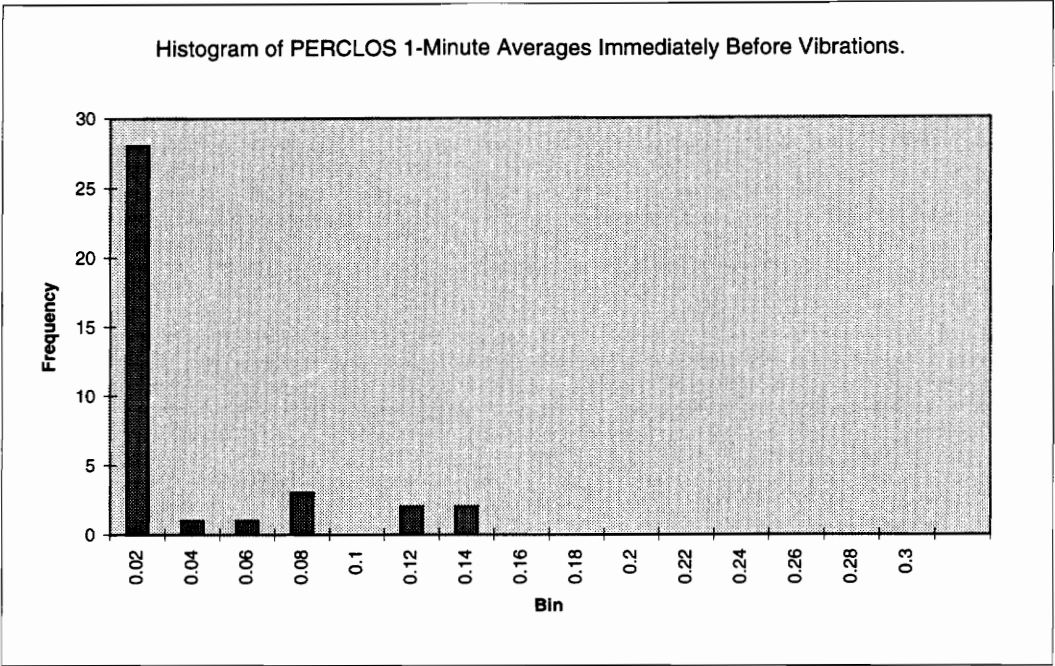


Figure I2: PERCLOS 1-minute histograms.

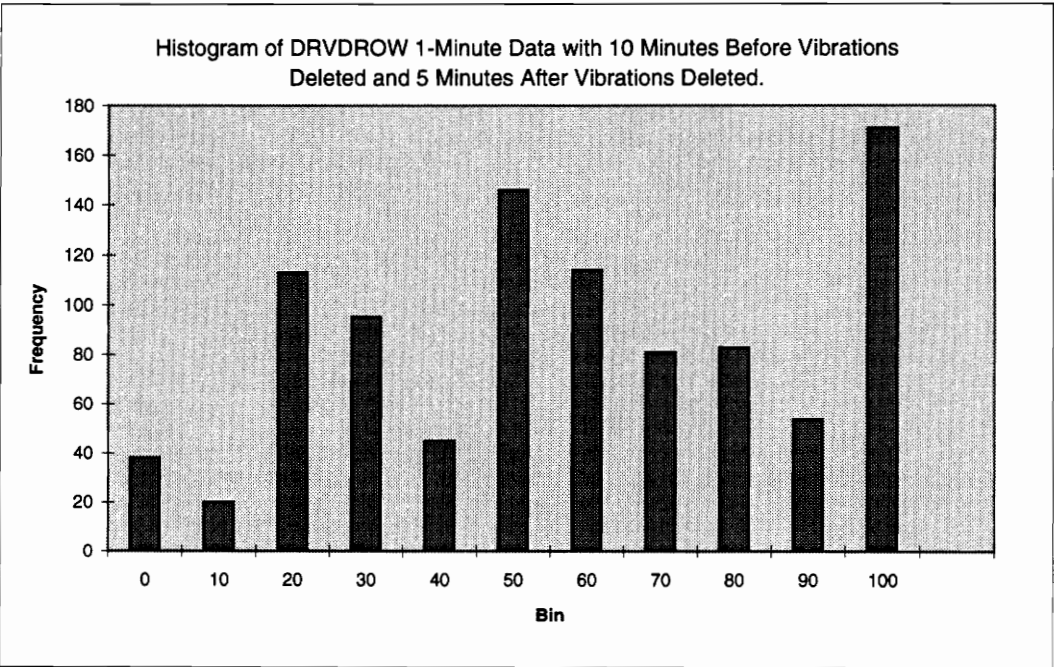
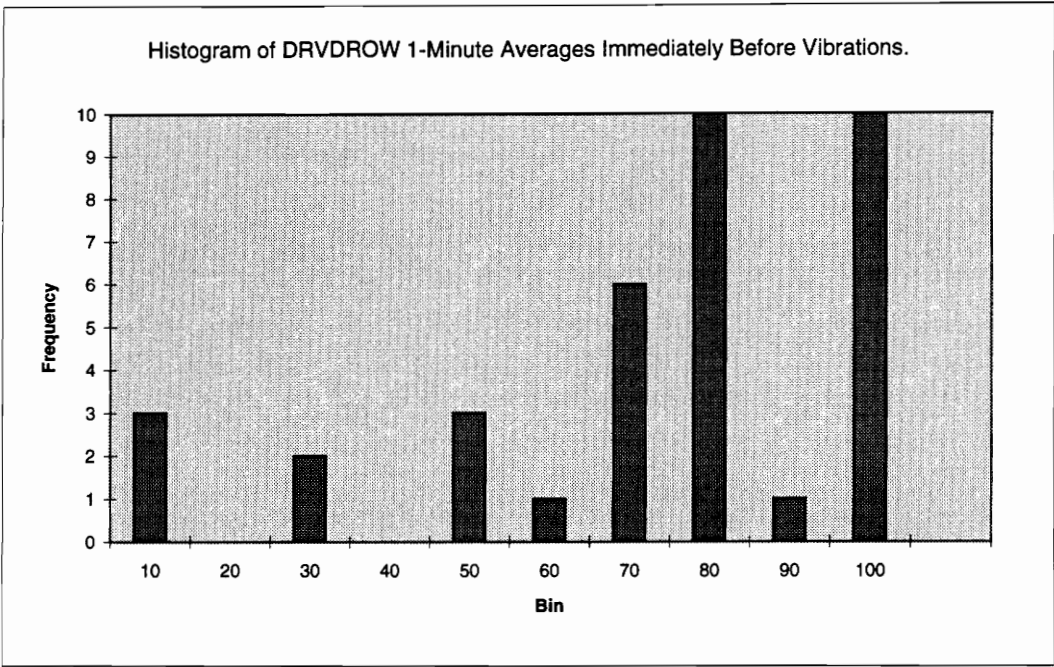


Figure I3: DRVDROW 1-minute histograms.

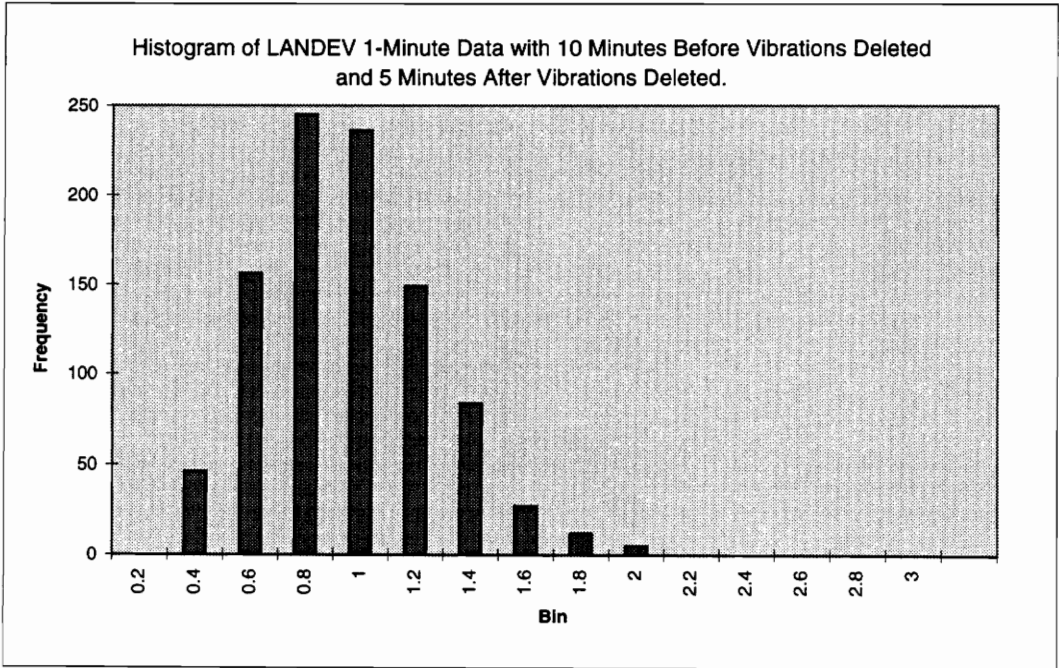
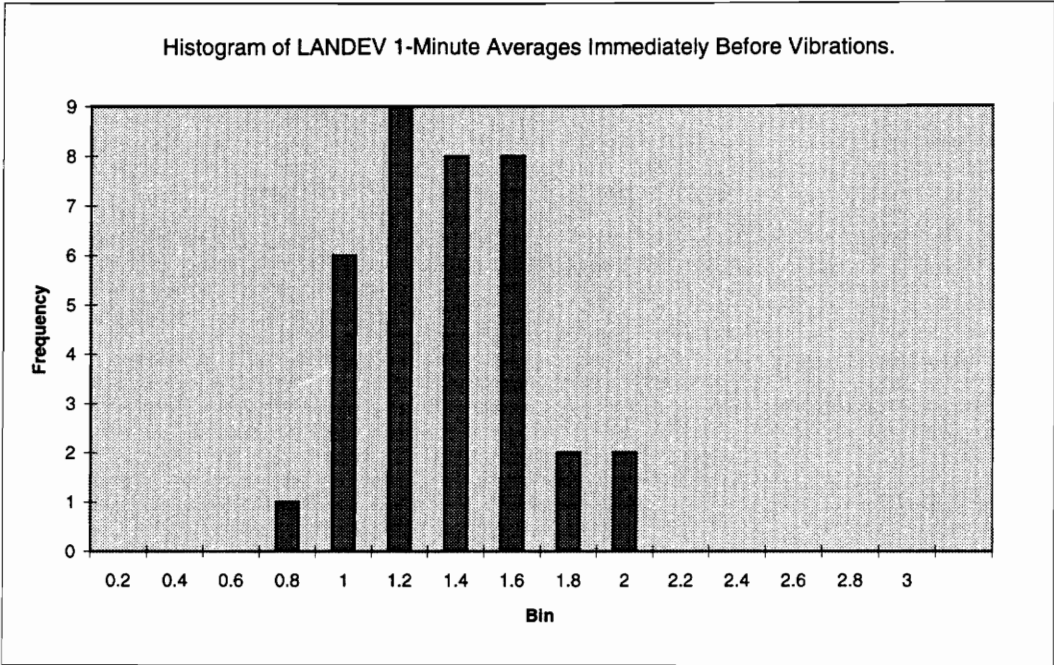


Figure I4: LANDEV 1-minute histograms.

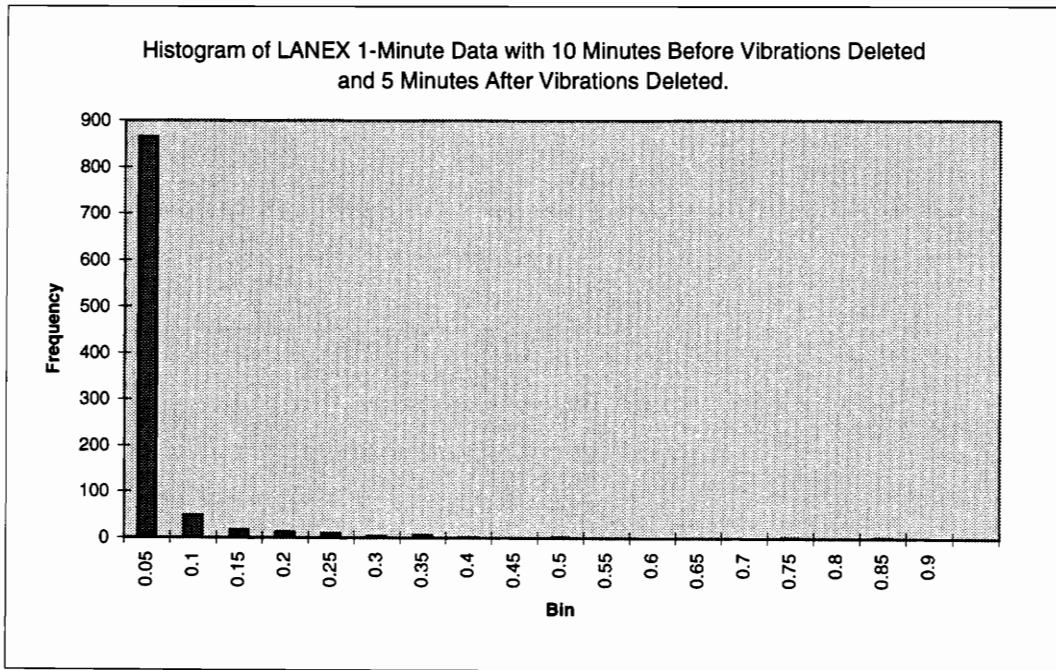
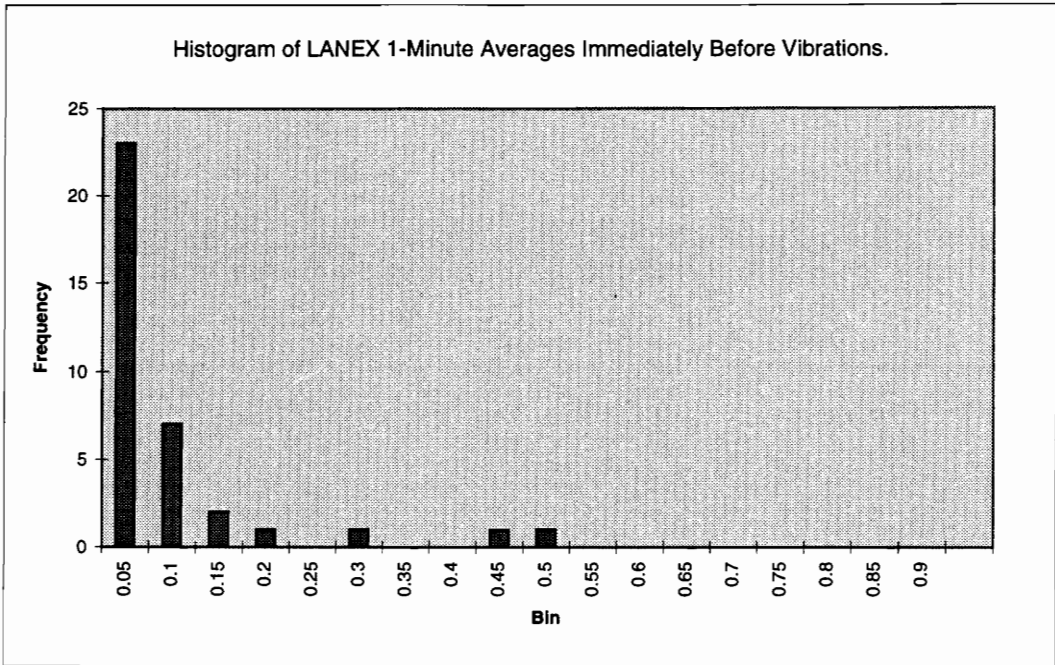


Figure I5: LANEX 1-minute histograms.

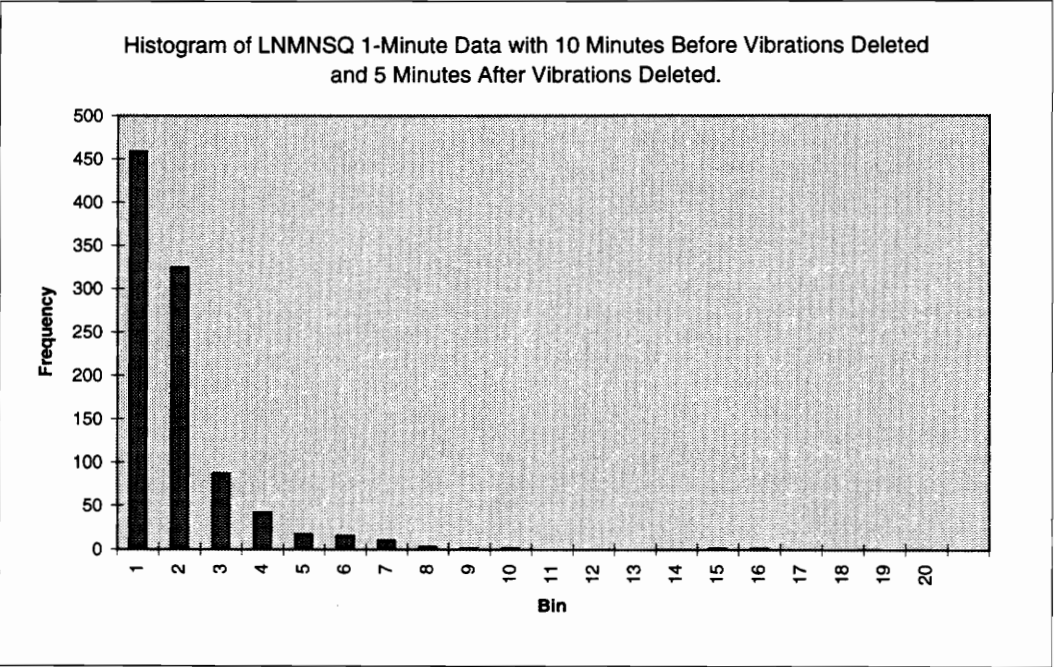
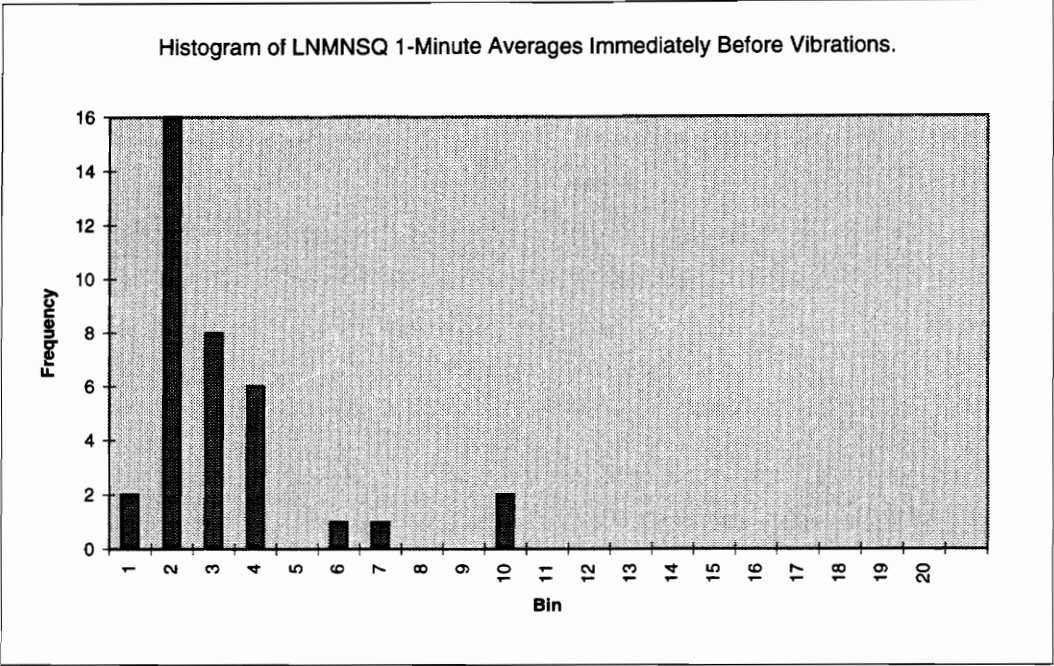


Figure I6: LNMNSQ 1-minute histograms.

APPENDIX J

Distribution Analyses for EYEMEAS, PERCLOS, DRVDROW, LANDEV, LANEX,
and LNMNSQ in Three-Minute Interval Data Set

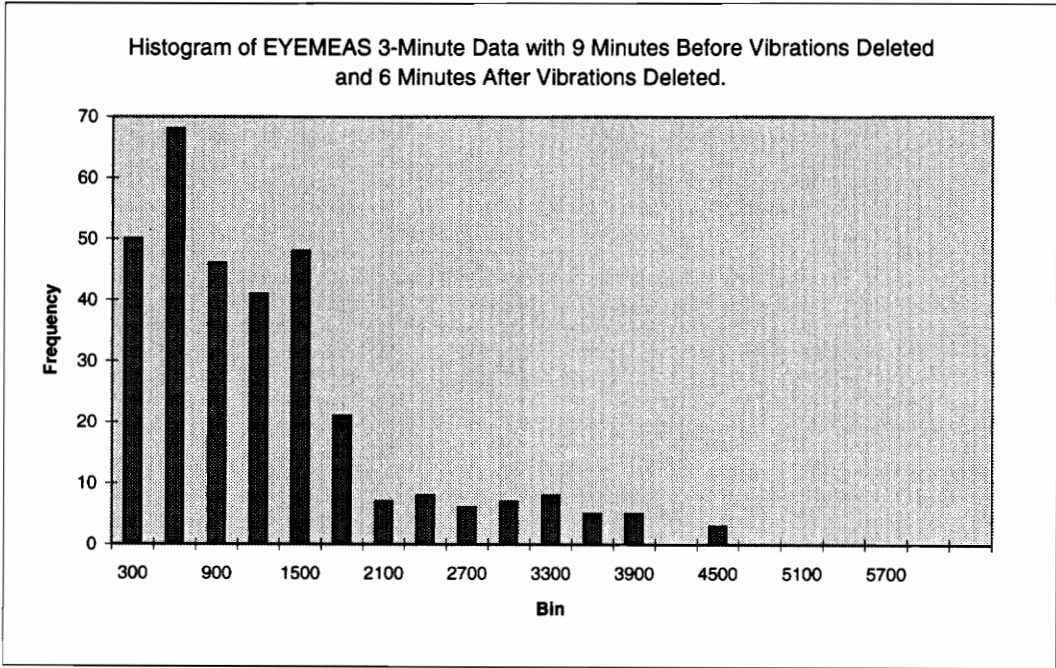
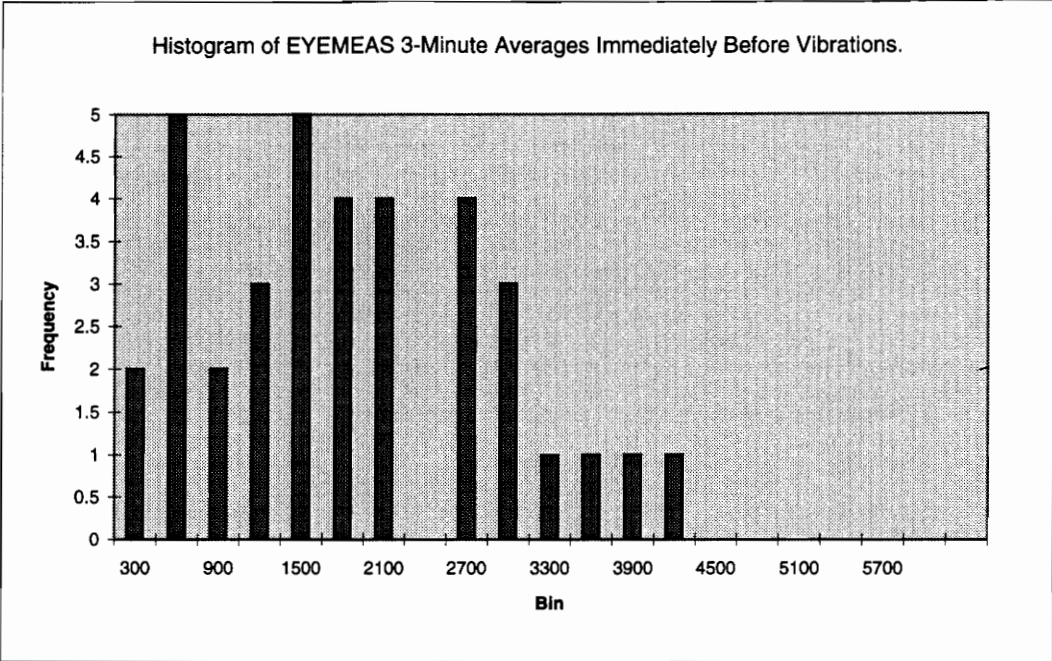


Figure J1: EYEMEAS 3-minute histograms.

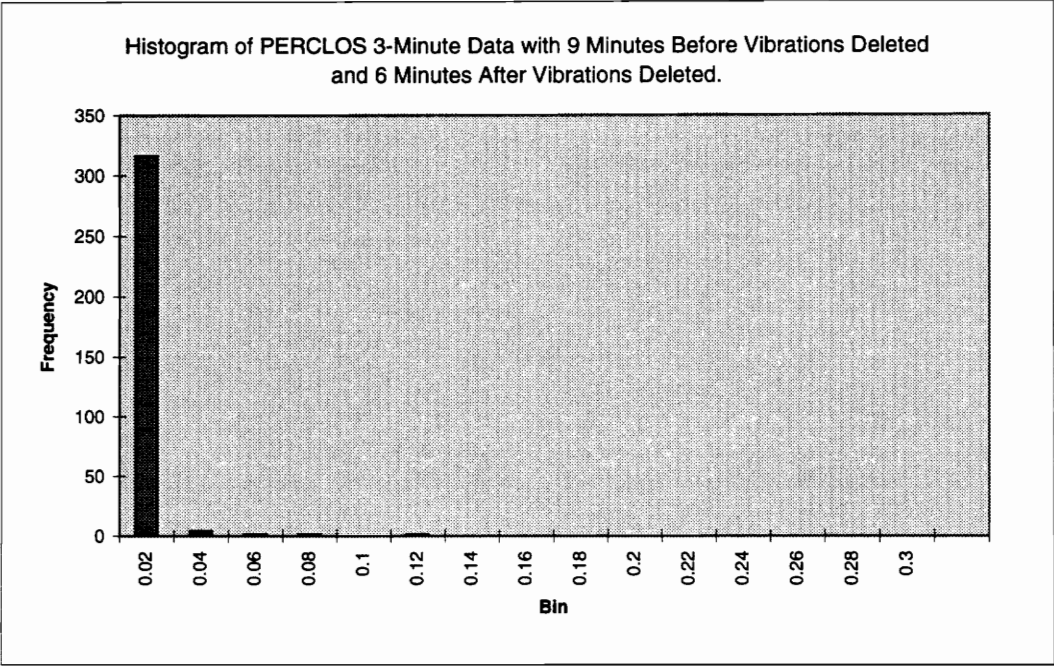
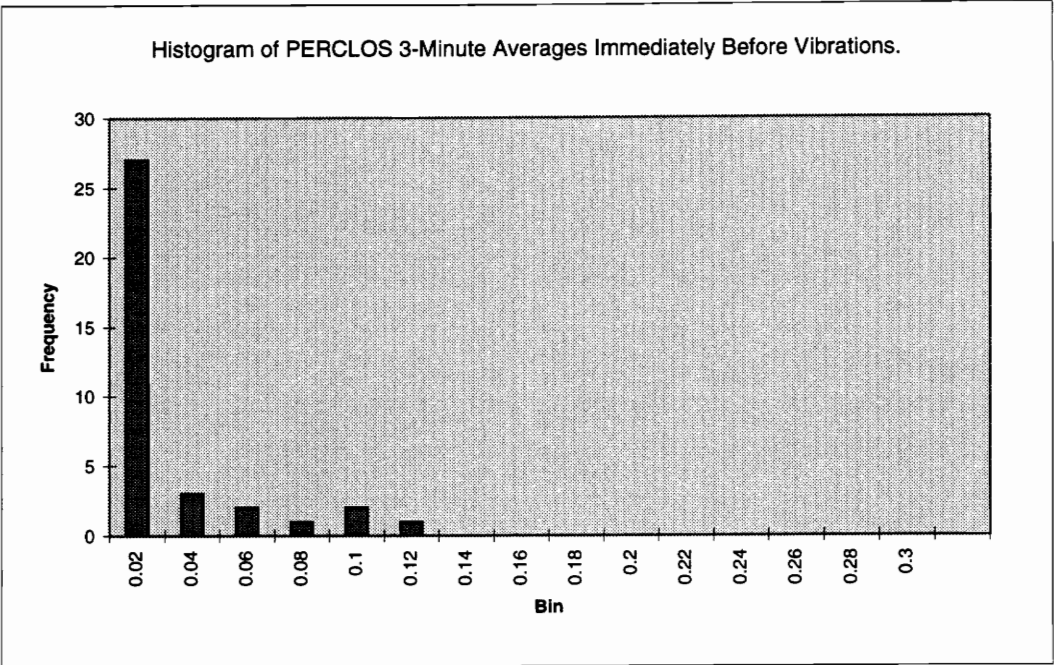


Figure J2: PERCLOS 3-minute histograms.

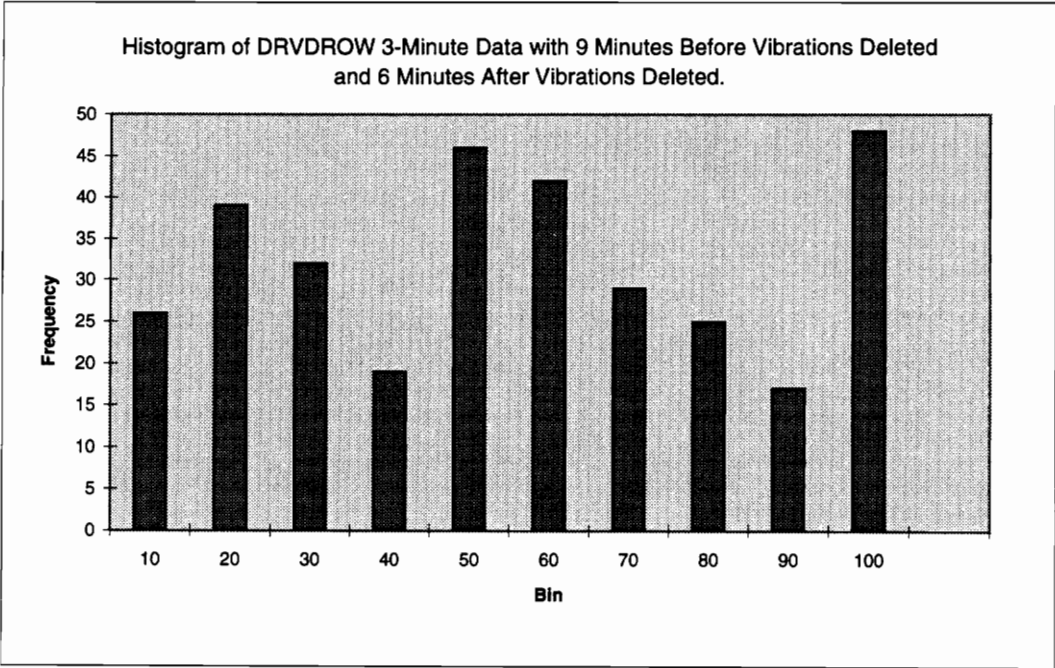
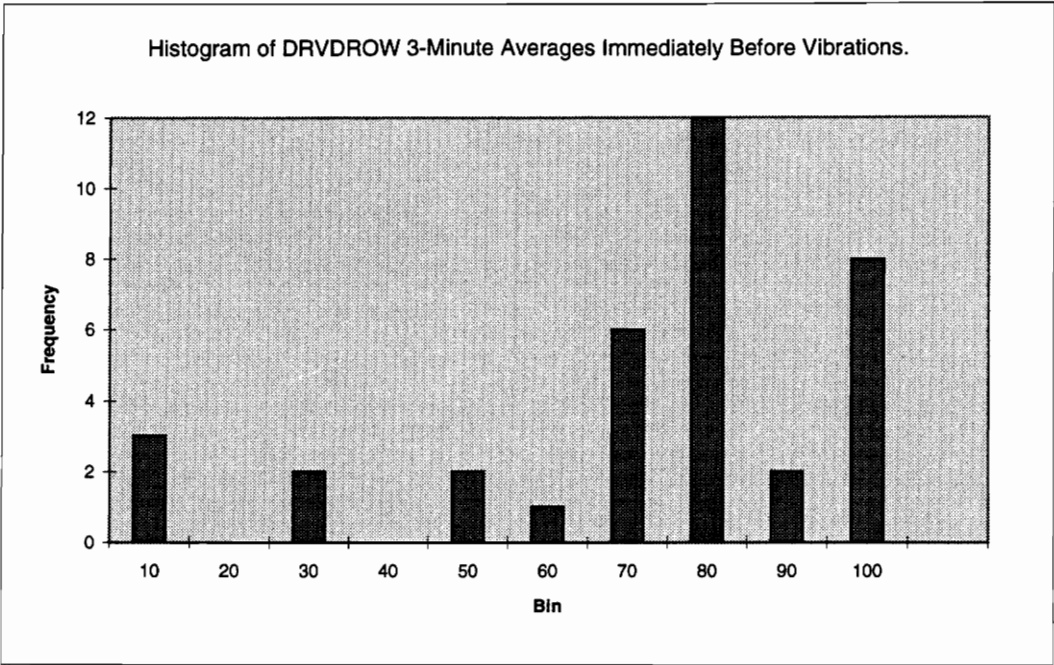


Figure J3: DRVDROW 3-minute histograms.

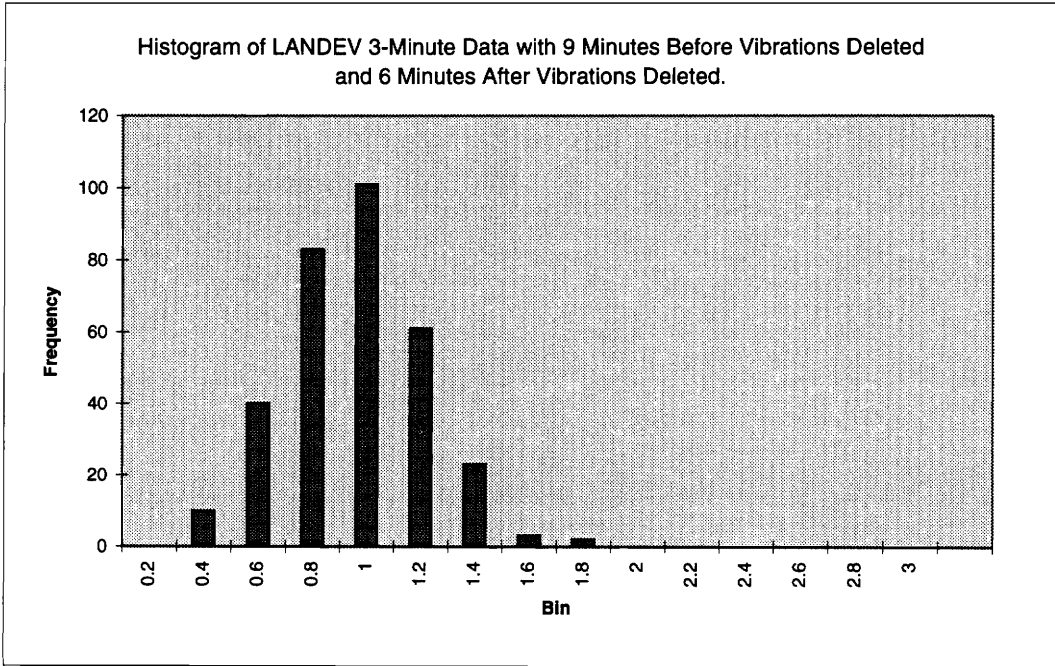
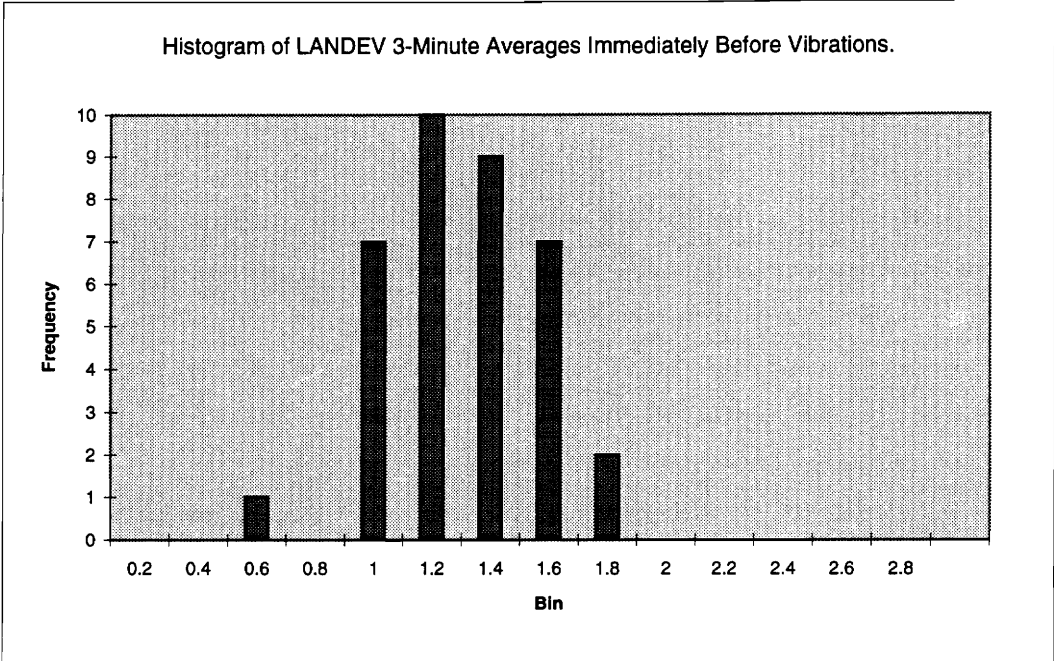


Figure J4: LANDEV 3-minute histograms.

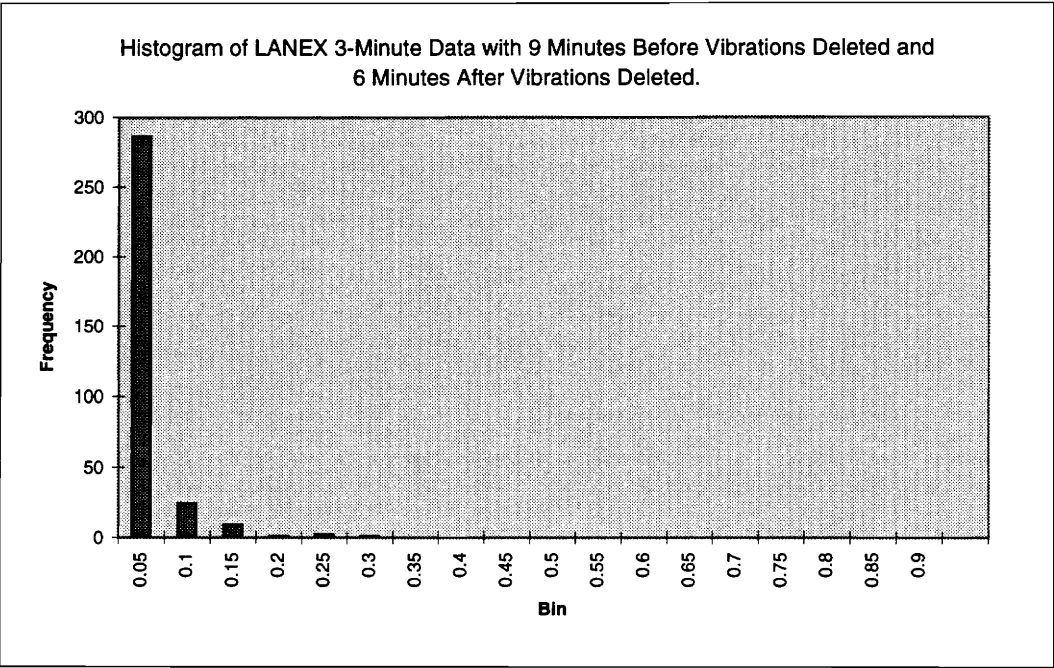
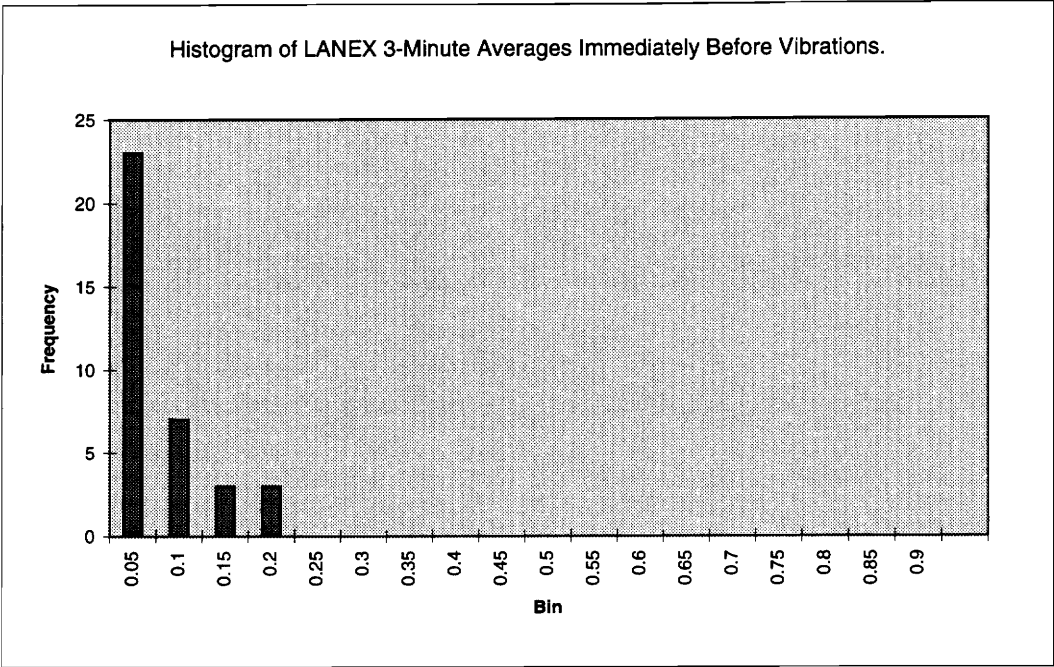


Figure J5: LANEX 3-minute histograms.

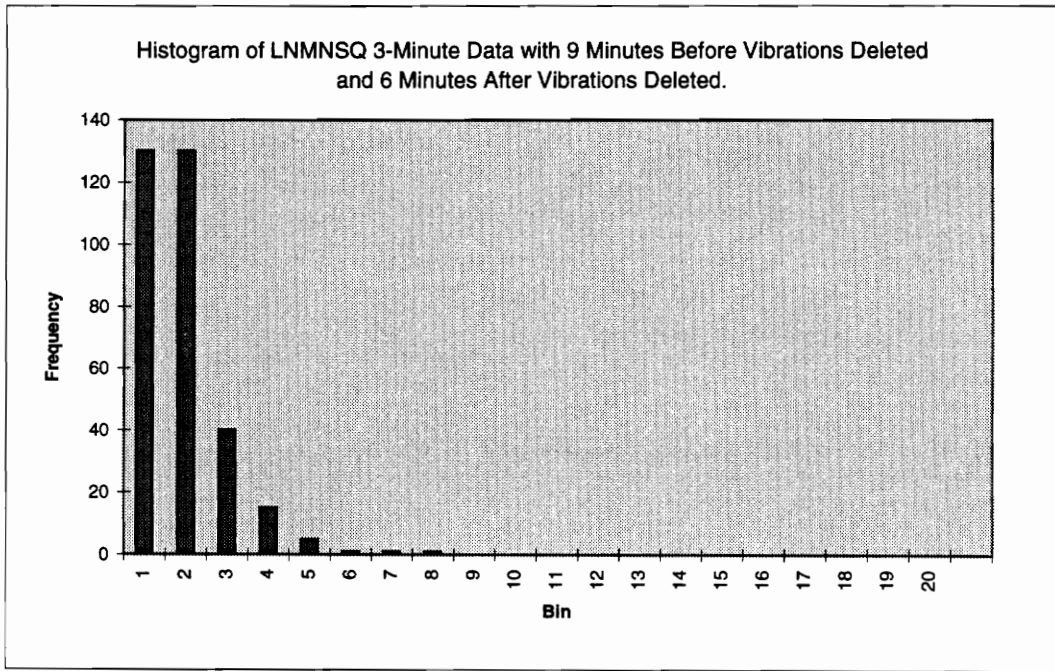
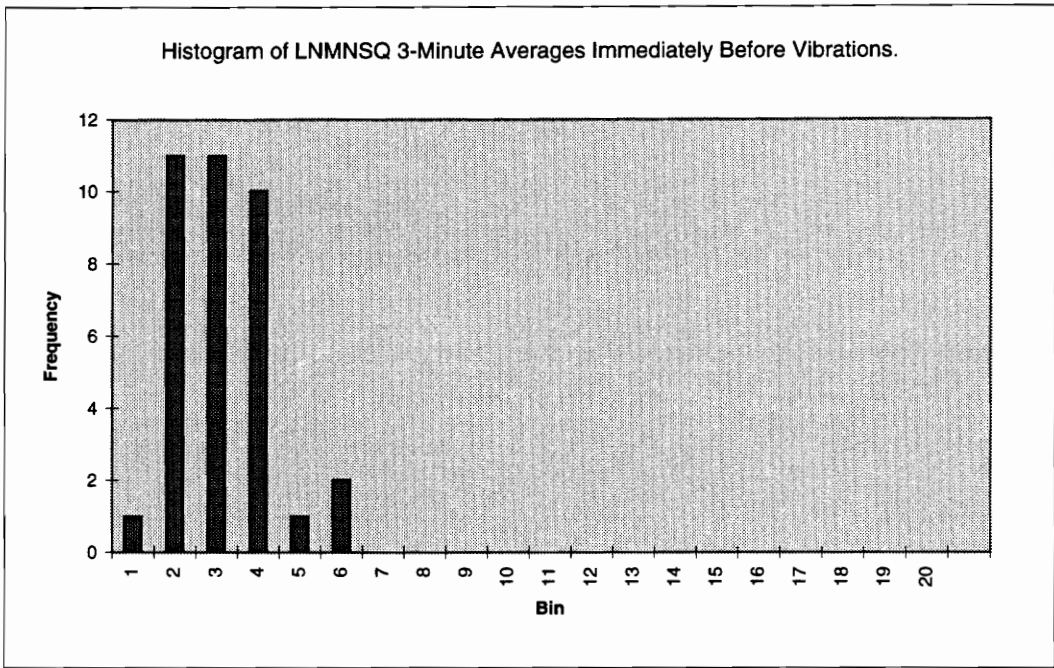


Figure J6: LNMNSQ 3-minute histograms.

APPENDIX K

**Descriptive Statistics for Individual Measures in Three-Minute and Six-Minute Interval
Data Sets**

Table K1: Descriptive Statistics for Individual Measures for 3-Minute Interval Data Set
(474 Observations)

	Minimum Value	Maximum Value	Median Value	Mean Value	Standard Deviation
EYEMEAS	59.99304	4820.026	951.1268	1179.85423	930.102
PERCLOS	0	0.10532	0	0.00305671	0.0121812
DRVDROW	0	99.33762	58.122925	55.6722257	30.463289
SLEEPER1	0.02399721	3.13153333	1.15926058	1.21423803	0.6808258
SLEEPER2	0.05085018	5.70891734	1.85236316	2.00134498	1.1744824
SLEEPER3	0	8.83413806	0.78384527	0.96063268	1.0475755
STVELV	2.925356	508.5606	28.290725	56.9805034	85.938325
LGREV	0	10.33333	0	0.82208155	1.607327
MDREV	0	50.33333	8	11.1652601	10.635023
SMREV	15.33333	90.66667	52.33333	53.42827	13.625779
STEXED	0	0.002659	0	0.000014369	0.000153
NMRHOLD	0.666667	38.66667	20	19.0548523	8.1704472
THRSHLD	0.000997	0.389037	0.115626	0.12304262	0.0804003
LANDEV	0.252426	1.973964	0.910056	0.9287548	0.278821
LANVAR	0.063957	4.779862	0.8659035	0.97945834	0.5900335
LNMNSQ	0.080778	10.2979	1.3937165	1.67258873	1.2216521
LANEX	0	0.422149	0.0053415	0.02332108	0.0436477
LNERRSQ	0	2.20476	0.000092	0.02847129	0.1316377
LNRTDEV	0.145478	1.318519	0.511132	0.5356831	0.1920822
LNRTVAR	0.021366	1.642882	0.2683885	0.33350706	0.2431076
ACCDEV	0.18578	3.613586	1.4021675	1.49202132	0.8876287
ACCVAR	0.035837	13.18893	2.490739	3.41837186	3.2440716
INTACDEV	0.0127	0.908678	0.2213915	0.27988475	0.2399936
INTACVAR	0.000162	0.878027	0.121636	0.16337053	0.1840483
ACEXEED	0	0.089727	0	0.004352	0.0122171
HPHDGDEV	0.10614	0.851191	0.3593135	0.37912521	0.1341202
HPHDGVAR	0.01139	0.808236	0.132416	0.16661768	0.1200566
VIBPROP	0	0.034311	0	0.0006391	0.0029216
NUMVIB	0	0.666667	0	0.02742614	0.0990809

Table K2: Descriptive Statistics for Individual Measures for 6-Minute Interval Data Set
(238 Observations)

	Minimum Value	Maximum Value	Median Value	Mean Value	Standard Deviation
EYEMEAS	60.23657	4365.508	954.4698	1181.21	908.1712
PERCLOS	0	0.090667	0	0.003034	0.010584
DRVDROW	0	99.33388	59.26078	55.76671	30.15598
SLEEPER1	0.024095	2.949728	1.175902	1.21604	0.672886
SLEEPER2	0.051057	5.32367	1.85736	2.004194	1.158275
SLEEPER3	0	7.679734	0.793256	0.960301	0.950268
STVELV	5.940183	476.0987	29.7438	57.46586	85.40905
LGREV	0	8.166667	0.25	0.838235	1.502256
MDREV	0	47.83333	7.833333	11.17157	10.2412
SMREV	26.33333	89.16667	51.75	53.36905	12.53614
STEXED	0	0.001579	0	0.0000154	0.000123
NMRHOLD	1.333333	35.33333	20.16667	19.05322	7.575497
THRSHLD	0.002993	0.319097	0.122577	0.123399	0.073028
LANDEV	0.369721	1.790847	0.907333	0.929564	0.252451
LANVAR	0.13782	3.686875	0.869744	0.981723	0.526161
LNMN SQ	0.305324	7.347464	1.450243	1.674575	1.064554
LANEX	0	0.238969	0.009525	0.023295	0.036269
LNERRSQ	0	1.196218	0.001858	0.02838	0.098979
LNRTDEV	0.206563	1.140118	0.520368	0.536523	0.180653
LNRTVAR	0.043449	1.285429	0.285285	0.334691	0.226819
ACCDEV	0.261574	3.134769	1.442398	1.488887	0.660705
ACCVAR	0.069881	10.50476	3.109475	3.407546	2.372011
INTACDEV	0.013047	0.646428	0.259199	0.278012	0.180247
INTACVAR	0.000171	0.494016	0.142362	0.162314	0.135557
ACEXEED	0	0.053117	0	0.004332	0.008963
HPHDGDEV	0.148163	0.786953	0.366945	0.37972	0.126289
HPHDGVAR	0.022377	0.625138	0.142856	0.167213	0.112325
VIBPROP	0	0.017156	0	0.000643	0.002078
NUMVIB	0	0.333333	0	0.027311	0.070669

APPENDIX L

Regression Summaries, Unconditional Classification Matrices, and Conditional “OR”
Classification Matrices for Driver Status Algorithms Developed Using Three-Minute
Interval Data Set

Regression Summary for Dependent Variable: PERCLOS

R = 0.13988993 R² = 0.01956919 Adjusted R² = 0.00909450

F(5,468) = 1.8682 p < 0.09845 Std. Error of estimate: 0.01213

	BETA	St. Err. of BETA	B	St. Err. of B	t(468)	p-level
Intercept			0.019718	0.009782	2.01572	0.0444
MDREV	-0.16564	0.103934	-0.00019	0.000119	-1.59374	0.11167
SMREV	-0.2	0.095133	-0.00018	0.000085	-2.10231	0.03606
NMRHOLDS	-0.21795	0.144721	-0.00033	0.000216	-1.506	0.132742
ACCDEV	0.191304	0.118654	0.002625	0.001628	1.61228	0.107575
INTACDEV	-0.19115	0.109916	-0.0097	0.005579	-1.73907	0.08268

Note: Classification matrices not developed for this algorithm.

Figure L1: Regression Summary for Algorithm D4e-3.

Regression Summary for Dependent Variable: EYEMEAS

R = 0.15229754 R² = 0.02319454 Adjusted R² = 0.01695961

F(3,470) = 3.7201 p < 0.01150 Std. Error of estimate: 922.18

	BETA	St. Err. of BETA	B	St. Err. of B	t(470)	p-level
Intercept			2708.663	708.4829	3.82319	0.000149
MDREV	-0.19993	0.102887	-17.486	8.9981	-1.94324	0.052584
SMREV	-0.23685	0.092914	-16.167	6.3424	-2.54911	0.011116
NMRHOLDS	-0.21657	0.140019	-24.654	15.9394	-1.54674	0.122599

Note: Classification matrices not developed for this algorithm.

Figure L2: Regression Summary for Algorithm D2e-3.

Regression Summary for Dependent Variable: DRVDROW

R = 0.46730576 R² = 0.21837467 Adjusted R² = 0.21338557

F(3,470) = 43.770 p < 0.00000 Std. Error of estimate: 27.018

	BETA	St. Err. of BETA	B	St. Err. of B	t(470)	p-level
Intercept			31.8411	5.065182	6.28626	0
MDREV	0.448109	0.042569	1.2836	0.121936	10.52666	0
SMREV	0.128083	0.041936	0.2864	0.093756	3.05428	0.002384
INTACDEV	-0.16325	0.042715	-20.7222	5.421954	-3.82191	0.00015

Predicted Algorithm Output	Observed			
		Low	High	Total
	High	15	47	62
	Low	281	131	412
	Total	296	178	474
	% Correct	94.93	26.40	69.20

DRVDROW R Value = 0.46731

Apparent Accuracy Rate: 0.692

Threshold: 70

Classification Matrix Generated from Multiple Regression Analysis of Lane-Enforced DRVDROW Data Resulting in Algorithm D6e-3. (Independent variables employed included Steering and Accelerometer.)

Predicted Algorithm Output	Observed			
		Low	High	Total
	High	19	55	74
	Low	255	145	400
	Total	274	200	474
% Correct	93.07	27.50	65.40	

Algorithm D6e (Threshold = 70) - OR - eLNMNSQ-D (Threshold = 3.00000)

Apparent Accuracy Rate: 0.654

Classification Matrix Generated from Multiple Regression Analysis of Lane-Enforced DRVDROW Data Resulting in Algorithm D6e-3 -OR- Algorithm eLNMNSQ-D-3 Data.

Figure L3: Regression Summary and Classification Matrices for Algorithm D6e-3.

Regression Summary for Dependent Variable: SLEEPER1
R = 0.21560006 R² = 0.04648338 Adjusted R² = 0.04446322
F(1,472) = 23.010 p < 0.00000 Std. Error of estimate: 0.66552

	BETA	St. Err. of BETA	B	St. Err. of B	t(472)	p-level
Intercept			1.116913	0.036689	30.4428	0
STVELV	0.2156	0.044946	0.001708	0.000356	4.79685	0.000002

Note: Classification matrices not developed for this algorithm.

Figure L4: Regression Summary for Algorithm SLEEPER1-D-3.

Regression Summary for Dependent Variable: SLEEPER2

R = 0.16040012 R² = 0.02572820 Adjusted R² = 0.02366406

F(1,472) = 12.464 p < 0.00046 Std. Error of estimate: 1.1605

	BETA	St. Err. of BETA	B	St. Err. of B	t(472)	p-level
Intercept			1.876437	0.063977	29.33004	0
STVELV	0.1604	0.045433	0.002192	0.000621	3.5305	0.000455

Note: Classification matrices not developed for this algorithm.

Figure L5: Regression Summary for Algorithm SLEEPER2-D-3.

Regression Summary for Dependent Variable: SLEEPER3

R = 0.18592858 R² = 0.03456944 Adjusted R² = 0.03046995

F(2,471) = 8.4326 p < 0.00025 Std. Error of estimate: 1.0315

	BETA	St. Err. of BETA	B	St. Err. of B	t(471)	p-level
Intercept			1.41847	0.130668	10.85556	0
NMRHOLDS	-0.37762	0.094318	-0.04842	0.012093	-4.00369	0.000072
THRSHLDS	0.289884	0.094318	3.777037	1.228919	3.07346	0.002239

Note: Classification matrices not developed for this algorithm.

Figure L6: Regression Summary for Algorithm SLEEPER3-D-3.

Regression Summary for Dependent Variable: PERCLOS
 $R = 0.38622797$ $R^2 = 0.14917204$ Adjusted $R^2 = 0.14008200$
 $F(5,468) = 16.410$ $p < 0.00000$ Std. Error of estimate: 0.01130

	BETA	St. Err. of BETA	B	St. Err. of B	t(468)	p-level
Intercept			-0.0024	0.001042	-2.26722	0.023832
STVELV	0.572219	0.123366	0.0001	0.000017	4.6384	0.000005
LGREV	-0.56672	0.099348	-0.0043	0.000753	-5.70439	0
MDREV	-0.63109	0.106843	-0.0007	0.000122	-5.90672	0
STEXED	-0.13252	0.047687	-10.549	3.796015	-2.77898	0.005672
HPHDGVAR	0.742369	0.083434	0.0753	0.008465	8.89767	0

Note: Classification matrices not developed for this algorithm.

Figure L7: Regression Summary for Algorithm E4e-3.

Regression Summary for Dependent Variable: EYEMEAS

R = 0.31524564 R² = 0.09937982 Adjusted R² = 0.09169862

F(4,469) = 12.938 p < 0.00000 Std. Error of estimate: 886.43

	BETA	St. Err. of BETA	B	St. Err. of B	t(469)	p-level
Intercept			251.823	165.5025	1.52157	0.128791
STVELV	0.452565	0.126812	4.898	1.3725	3.56877	0.000396
LGREV	-0.34083	0.095935	-197.227	55.5141	-3.55273	0.00042
MDREV	-0.64568	0.112268	-56.468	9.8186	-5.75118	0
HPHDGDEV	0.548294	0.081092	3802.328	562.3579	6.7614	0

Note: Classification matrices not developed for this algorithm.

Figure L8: Regression Summary for Algorithm E2e-3.

Regression Summary for Dependent Variable: DRVDROW
R = 0.51213868 R² = 0.26228602 Adjusted R² = 0.25440446
F(5,468) = 33.278 p < 0.00000 Std. Error of estimate: 26.304

	BETA	St. Err. of BETA	B	St. Err. of B	t(468)	p-level
Intercept			-25.3836	15.387	-1.64968	0.099679
MDREV	0.411958	0.070696	1.18	0.2025	5.8272	0
SMREV	0.433871	0.081249	0.97	0.18165	5.34001	0
THRSHLDS	0.310563	0.093526	117.671	35.43647	3.32062	0.000968
ACCDEV	-0.21968	0.048847	-7.5392	1.67644	-4.49717	0.000009
HPHDGVAR	0.303346	0.068175	76.9715	17.29883	4.44952	0.000011

Observed			
	Low	High	Total
Predicted High	18	55	73
Predicted Low	278	123	401
Predicted Total	296	178	474
% Correct	93.92	30.90	70.25

DRVDROW R Value = 0.51214
Apparent Accuracy Rate: 0.703
Threshold: 70

Classification Matrix Generated from Multiple Regression Analysis of Lane-Enforced DRVDROW Data Resulting in Algorithm E6e-3. (Independent variables employed included Steering, Accelerometer, and HPHDGDEV/VAR.)

Observed			
	Low	High	Total
Predicted High	24	61	85
Predicted Low	247	142	389
Predicted Total	271	203	474
% Correct	91.14	30.05	64.98

Algorithm E6e (Threshold = 70) - OR - eLANEX-E (Threshold = 0.06667)
Apparent Accuracy Rate: 0.650
Classification Matrix Generated from Multiple Regression Analysis of Lane-Enforced DRVDROW Data Resulting in Algorithm E6e-3 -OR- Algorithm eLANEX-E-3 Data.

Observed			
	Low	High	Total
Predicted High	25	67	92
Predicted Low	249	133	382
Predicted Total	274	200	474
% Correct	90.88	33.50	66.67

Algorithm E6e (Threshold = 70) - OR - eLNMNSQ-E (Threshold = 3.00000)
Apparent Accuracy Rate: 0.667
Classification Matrix Generated from Multiple Regression Analysis of Lane-Enforced DRVDROW Data Resulting in Algorithm E6e-3 -OR- Algorithm eLNMNSQ-E-3 Data.

Figure L9: Regression Summary and Classification Matrices for Algorithm E6e-3.

Regression Summary for Dependent Variable: SLEEPER1
 $R = 0.35060881$ $R^2 = 0.12292654$ Adjusted $R^2 = 0.11165793$
 $F(6,467) = 10.909$ $p < 0.00000$ Std. Error of estimate: 0.64169

	BETA	St. Err. of BETA	B	St. Err. of B	t(467)	p-level
Intercept			0.177286	0.197792	0.89632	0.370541
STVELV	0.412909	0.130455	0.003271	0.001033	3.16515	0.001651
LGREV	-0.301	0.098328	-0.1275	0.041649	-3.0612	0.002332
MDREV	-0.28887	0.121945	-0.01849	0.007807	-2.36881	0.018251
SMREV	0.117505	0.052283	0.005871	0.002612	2.24746	0.025077
ACCDEV	-0.14109	0.055113	-0.10821	0.042273	-2.5599	0.010784
HPHDGDEV	0.524602	0.086037	2.663005	0.436742	6.09744	0

Note: Classification matrices not developed for this algorithm.

Figure L10: Regression Summary for Algorithm SLEEPER1-E-3.

Regression Summary for Dependent Variable: SLEEPER2

R = 0.30313550 R² = 0.09189113 Adjusted R² = 0.08414607

F(4,469) = 11.864 p < 0.00000 Std. Error of estimate: 1.1240

	BETA	St. Err. of BETA	B	St. Err. of B	t(469)	p-level
Intercept			0.732611	0.209855	3.49104	0.000527
STVELV	0.479586	0.127339	0.006554	0.00174	3.76623	0.000187
LGREV	-0.37941	0.096333	-0.27723	0.070391	-3.93848	0.000094
MDREV	-0.36181	0.112734	-0.03996	0.01245	-3.2094	0.001421
HPHDGDEV	0.472684	0.081428	4.139266	0.713061	5.80492	0

Note: Classification matrices not developed for this algorithm.

Figure L11: Regression Summary for Algorithm SLEEPER2-E-3.

Regression Summary for Dependent Variable: SLEEPER3
 $R = 0.42763183$ $R^2 = 0.18286899$ Adjusted $R^2 = 0.17059449$
 $F(7,466) = 14.898$ $p < 0.00000$ Std. Error of estimate: 0.95404

	BETA	St. Err. of BETA	B	St. Err. of B	t(466)	p-level
Intercept			-0.05873	0.2196	-0.26745	0.78924
STVELV	0.547861	0.128913	0.006678	0.0016	4.24985	0.000026
LGREV	-0.5313	0.101234	-0.34628	0.066	-5.24823	0
MDREV	-0.48592	0.115719	-0.04787	0.0114	-4.19919	0.000032
SMREV	0.115847	0.050402	0.008907	0.0039	2.29846	0.021978
STEXED	-0.13603	0.047085	-931.284	322.3619	-2.88894	0.004046
INTACDEV	-0.1159	0.049048	-0.50592	0.2141	-2.36304	0.018535
HPHDGVAR	0.782086	0.08595	6.824233	0.75	9.09927	0

Observed			
	Low	High	Total
Predicted High	39	18	57
Predicted Low	387	30	417
Output Total	426	48	474
% Correct	90.85	37.50	85.44

SLEEPER3 R Value = 0.42763
Apparent Accuracy Rate: 0.854
Threshold: 1.4

Classification Matrix Generated from Multiple Regression Analysis of Lane-Enforced SLEEPER3 Data Resulting in Algorithm SLEEPER3-E-3. (Independent variables employed included Steering, Accelerometer, and HPHDGDEV/VAR.)

Observed			
	Low	High	Total
Predicted High	34	23	57
Predicted Low	357	60	417
Output Total	391	83	474
% Correct	91.30	27.71	80.17

Algorithm SLEEPER3-E (Threshold = 1.4) - OR - eLANEX-E (Threshold = 0.06667)
Apparent Accuracy Rate: 0.802

Classification Matrix Generated from Multiple Regression Analysis of Lane-Enforced SLEEPER3 Data Resulting in Algorithm SLEEPER3-E-3 -OR- Algorithm eLANEX-E-3 Data.

Observed			
	Low	High	Total
Predicted High	36	28	64
Predicted Low	355	55	410
Output Total	391	83	474
% Correct	90.79	33.73	80.80

Algorithm SLEEPER3-E (Threshold = 1.4) - OR - eLNMNSQ-E (Threshold = 3.00000)
Apparent Accuracy Rate: 0.808

Classification Matrix Generated from Multiple Regression Analysis of Lane-Enforced SLEEPER3 Data Resulting in Algorithm SLEEPER3-E-3 -OR- Algorithm eLNMNSQ-E-3 Data.

Figure L12: Regression Summary and Classification Matrices for Algorithm SLEEPER3-E-3.

Regression Summary for Dependent Variable: PERCLOS

R = 0.48232223 R² = 0.23263473 Adjusted R² = 0.22277565

F(6,467) = 23.596 p < 0.00000 Std. Error of estimate: 0.01074

	BETA	St. Err. of BETA	B	St. Err. of B	t(467)	p-level
Intercept			-0.00304	0.001063	-2.85703	0.004467
STVELV	0.387123	0.118343	0.000055	0.000017	3.27121	0.00115
LGREV	-0.20208	0.077889	-0.00153	0.00059	-2.5945	0.009771
MDREV	-0.33165	0.0933	-0.00038	0.000107	-3.55464	0.000417
LNMSNSQ	0.333604	0.064803	0.003326	0.000646	5.14795	0
LANVAR	0.253799	0.06782	0.00524	0.0014	3.74227	0.000205
INTACDEV	-0.15691	0.047731	-0.00796	0.002423	-3.28727	0.001088

Predicted Algorithm Output	Observed		
		Low	High
	High	17	16
	Low	424	17
	Total	441	33
	% Correct	96.15	48.48
	Total	474	474

PERCLOS R Value = 0.48232

Apparent Accuracy Rate: 0.928

Threshold: 0.012

Classification Matrix Generated from Multiple Regression Analysis of Lane-Enforced PERCLOS Data Resulting in Algorithm F4e-3. (Independent variables employed included Steering, Accelerometer, LANDEV/VAR, LNMNSQ, LANEX, & LNERRSQ.)

Predicted Algorithm Output	Observed		
		Low	High
	High	1	56
	Low	402	15
	Total	403	71
	% Correct	99.75	78.87
	Total	474	474

Algorithm F4e (Threshold = 0.012) - OR - LANEX (Threshold = 0.06667)

Apparent Accuracy Rate: 0.966

Classification Matrix Generated from Multiple Regression Analysis of Lane-Enforced PERCLOS Data Resulting in Algorithm F4e-3 -OR- Measured LANEX Data.

Predicted Algorithm Output	Observed		
		Low	High
	High	2	57
	Low	401	14
	Total	403	71
	% Correct	99.50	80.28
	Total	474	474

Algorithm F4e (Threshold = 0.012) - OR - LNMNSQ (Threshold = 3.00000)

Apparent Accuracy Rate: 0.966

Classification Matrix Generated from Multiple Regression Analysis of Lane-Enforced PERCLOS Data Resulting in Algorithm F4e-3 -OR- Measured LNMNSQ Data.

Figure L13: - Regression Summary and Classification Matrices for Algorithm F4e-3.

Regression Summary for Dependent Variable: EYEMEAS

R = 0.49305382 R² = 0.24310207 Adjusted R² = 0.23501556

F(5,468) = 30.063 p < 0.00000 Std. Error of estimate: 813.50

	BETA	St. Err. of BETA	B	St. Err. of B	t(468)	p-level
Intercept			658.4794	91.138	7.22505	0
STVELV	0.309549	0.083595	3.350214	0.905	3.70295	0.000239
MDREV	-0.43026	0.086316	-37.6294	7.549	-4.98478	0.000001
LNMSQ	0.719422	0.098876	547.73	75.279	7.27602	0
LANEX	-0.22521	0.095389	-4799.08	2032.675	-2.36097	0.018636
ACEXED	-0.16174	0.043547	-12313.3	3315.304	-3.71406	0.000229

Predicted Algorithm Output	Observed		
		Low	High
	High	7	16
	Low	390	61
	Total	397	77
	% Correct	98.24	20.78
			85.65

EYEMEAS R Value = 0.49303

Apparent Accuracy Rate: 0.857

Threshold: 2000

Classification Matrix Generated from Multiple Regression Analysis of Lane-Enforced EYEMEAS Data Resulting in Algorithm F2e-3. (Independent variables employed included Steering, Accelerometer, LANDEV/VAR, LNMSQ, LANEX, & LNERRSQ.)

Predicted Algorithm Output	Observed		
		Low	High
	High	2	53
	Low	369	50
	Total	371	103
	% Correct	99.46	51.46
			89.03

Algorithm F2e (Threshold = 2000) - OR - LANEX (Threshold = 0.06667)

Apparent Accuracy Rate: 0.890

Classification Matrix Generated from Multiple Regression Analysis of Lane-Enforced EYEMEAS Data Resulting in Algorithm F2e-3 -OR- Measured LANEX Data.

Predicted Algorithm Output	Observed		
		Low	High
	High	2	54
	Low	368	50
	Total	370	104
	% Correct	99.46	51.92
			89.03

Algorithm F2e (Threshold = 2000) - OR - LNMSQ (Threshold = 3.00000)

Apparent Accuracy Rate: 0.890

Classification Matrix Generated from Multiple Regression Analysis of Lane-Enforced EYEMEAS Data Resulting in Algorithm F2e-3 -OR- Measured LNMSQ Data.

Figure L14: Regression Summary and Classification Matrices for Algorithm F2e-3.

Regression Summary for Dependent Variable: DRVDROW

R = 0.55351317 R² = 0.30637683 Adjusted R² = 0.29746518

F(6,467) = 34.379 p < 0.00000 Std. Error of estimate: 25.534

	BETA	St. Err. of BETA	B	St. Err. of B	t(467)	p-level
Intercept			-19.0802	14.84457	-1.28533	0.199314
STVELV	0.248908	0.086348	0.0882	0.03061	2.88261	0.004126
MDREV	0.351189	0.091147	1.006	0.26108	3.853	0.000133
SMREV	0.385826	0.078104	0.8626	0.17462	4.93992	0.000001
THRSHLDS	0.275862	0.093295	104.5228	35.3491	2.95687	0.003265
LNMSNSQ	0.302007	0.044025	7.5309	1.0978	6.85996	0
ACCDEV	-0.25487	0.049259	-8.747	1.69055	-5.17404	0

Predicted Algorithm Output	Observed		
		Low	High
	High	20	65
	Low	276	113
	Total	296	178
	% Correct	93.24	36.52

DRVDROW R Value = 0.55351

Apparent Accuracy Rate: 0.719

Threshold: 70

Classification Matrix Generated from Multiple Regression Analysis of Lane-Enforced DRVDROW Data Resulting in Algorithm F6e-3. (Independent variables employed included Steering, Accelerometer, LANDEV/VAR, LNMSNSQ, LANEX, & LNERRSQ.)

Predicted Algorithm Output	Observed		
		Low	High
	High	15	103
	Low	256	100
	Total	271	203
	% Correct	94.46	50.74

Algorithm F6e (Threshold = 70) - OR - LANEX (Threshold = 0.06667)

Apparent Accuracy Rate: 0.757

Classification Matrix Generated from Multiple Regression Analysis of Lane-Enforced DRVDROW Data Resulting in Algorithm F6e-3 -OR- Measured LANEX Data.

Predicted Algorithm Output	Observed		
		Low	High
	High	15	103
	Low	259	97
	Total	274	200
	% Correct	94.53	51.50

Algorithm F6e (Threshold = 70) - OR - LNMSNSQ (Threshold = 3.00000)

Apparent Accuracy Rate: 0.764

Classification Matrix Generated from Multiple Regression Analysis of Lane-Enforced DRVDROW Data Resulting in Algorithm F6e-3 -OR- Measured LNMSNSQ Data.

Figure L15: Regression Summary and Classification Matrices for Algorithm F6e-3.

Regression Summary for Dependent Variable: SLEEPER1

R = 0.48144126 R² = 0.23178568 Adjusted R² = 0.22357827

F(5,468) = 28.241 p < 0.00000 Std. Error of estimate: 0.59991

	BETA	St. Err. of BETA	B	St. Err. of B	t(468)	p-level
Intercept			0.357781	0.172959	2.06859	0.039133
STVELV	0.206212	0.045059	0.001634	0.000357	4.57644	0.000006
SMREV	0.138629	0.043307	0.006927	0.002164	3.2011	0.001462
LNMSNSQ	0.361975	0.06053	0.201728	0.033734	5.98004	0
LANDEV	0.175862	0.067209	0.42942	0.164111	2.61665	0.009167
ACCDEV	-0.29967	0.050265	-0.22986	0.038554	-5.96191	0

Predicted Algorithm Output	Observed		
		Low	High
	High	15	24
	Low	348	87
	Total	363	111
	% Correct	95.87	21.62
		Total	474

SLEEPER1 R Value = 0.48144

Apparent Accuracy Rate: 0.785

Threshold: 1.7

Classification Matrix Generated from Multiple Regression Analysis of Lane-Enforced SLEEPER1 Data Resulting in Algorithm SLEEPER1-F-3. (Independent variables employed included Steering, Accelerometer, LANDEV/VAR, LNMSNSQ, LANEX, & LNERRSQ.)

Predicted Algorithm Output	Observed		
		Low	High
	High	8	60
	Low	330	76
	Total	338	136
	% Correct	97.63	44.12
		Total	474

Algorithm SLEEPER1-F (Threshold = 1.7) - OR - LANEX (Threshold = 0.06667)

Apparent Accuracy Rate: 0.823

Classification Matrix Generated from Multiple Regression Analysis of Lane-Enforced SLEEPER1 Data Resulting in Algorithm SLEEPER1-F-3 -OR- Measured LANEX Data.

Predicted Algorithm Output	Observed		
		Low	High
	High	7	62
	Low	331	74
	Total	338	136
	% Correct	97.93	45.59
		Total	474

Algorithm SLEEPER1-F (Threshold = 1.7) - OR - LNMSNSQ (Threshold = 3.00000)

Apparent Accuracy Rate: 0.829

Classification Matrix Generated from Multiple Regression Analysis of Lane-Enforced SLEEPER1 Data Resulting in Algorithm SLEEPER1-F-3 -OR- Measured LNMSNSQ Data.

Figure L16: Regression Summary and Classification Matrices for Algorithm SLEEPER1-F-3.

Regression Summary for Dependent Variable: SLEEPER2

R = 0.47484712 R² = 0.22547978 Adjusted R² = 0.21720499

F(5,468) = 27.249 p < 0.00000 Std. Error of estimate: 1.0391

	BETA	St. Err. of BETA	B	St. Err. of B	t(468)	p-level
Intercept			1.325161	0.127258	10.41318	0
STVELV	0.182294	0.043048	0.002491	0.000588	4.23462	0.000028
LNMSNSQ	0.681952	0.099898	0.655621	0.096041	6.82651	0
LANEX	-0.2091	0.095768	-5.62658	2.57695	-2.18343	0.0295
ACCDEV	-0.19145	0.051094	-0.25332	0.067606	-3.74695	0.000201
ACEXEED	-0.12712	0.047267	-12.2209	4.544002	-2.68945	0.007413

Predicted Algorithm Output	Observed		
		Low	High
	High	7	21
	Low	383	63
	Total	390	84
% Correct		98.21	25.00
		474	85.23

SLEEPER2 R Value = 0.47485

Apparent Accuracy Rate: 0.852

Threshold: 3

Classification Matrix Generated from Multiple Regression Analysis of Lane-Enforced SLEEPER2 Data Resulting in Algorithm SLEEPER2-F-3. (Independent variables employed included Steering, Accelerometer, LANDEV/VAR, LNMSNSQ, LANEX, & LNERRSQ.)

Predicted Algorithm Output	Observed		
		Low	High
	High	4	58
	Low	360	52
	Total	364	110
% Correct		98.90	52.73
		474	88.19

Algorithm SLEEPER2-F (Threshold = 3) - OR - LANEX (Threshold = 0.06667)

Apparent Accuracy Rate: 0.882

Classification Matrix Generated from Multiple Regression Analysis of Lane-Enforced SLEEPER2 Data Resulting in Algorithm SLEEPER2-F-3 -OR- Measured LANEX Data.

Predicted Algorithm Output	Observed		
		Low	High
	High	3	58
	Low	361	52
	Total	364	110
% Correct		99.18	52.73
		474	88.40

Algorithm SLEEPER2-F (Threshold = 3) - OR - LNMSNSQ (Threshold = 3.00000)

Apparent Accuracy Rate: 0.884

Classification Matrix Generated from Multiple Regression Analysis of Lane-Enforced SLEEPER2 Data Resulting in Algorithm SLEEPER2-F-3 -OR- Measured LNMSNSQ Data.

Figure L17: Regression Summary and Classification Matrices for Algorithm SLEEPER2-F-3.

Regression Summary for Dependent Variable: SLEEPER3
 $R = 0.53394126$ $R^2 = 0.28509327$ Adjusted $R^2 = 0.27435433$
 $F(7,466) = 26.548$ $p < 0.00000$ Std. Error of estimate: 0.89238

	BETA	St. Err. of BETA	B	St. Err. of B	t(466)	p-level
Intercept			0.868176	0.232043	3.74145	0.000206
STVELV	0.393598	0.114568	0.004798	0.001397	3.43549	0.000644
LGREV	-0.20171	0.0763	-0.13147	0.049729	-2.64363	0.008479
MDREV	-0.2913	0.105169	-0.02869	0.010359	-2.7698	0.005833
NMRHOLDS	-0.1756	0.065891	-0.02252	0.008448	-2.66502	0.007965
LNMN SQ	0.357999	0.062946	0.306987	0.053977	5.68737	0
LANVAR	0.265219	0.067738	0.470883	0.120265	3.91538	0.000104
INTACDEV	-0.24404	0.04804	-1.06524	0.209694	-5.07998	0.000001

		Observed		
Predicted Algorithm Output		Low	High	Total
	High	51	28	79
	Low	375	20	395
	Total	426	48	474
	% Correct	88.03	58.33	85.02

SLEEPER3 R Value = 0.53394
Apparent Accuracy Rate: 0.850
Threshold: 1.4

Classification Matrix Generated from Multiple Regression Analysis of Lane-Enforced SLEEPER3 Data Resulting in Algorithm SLEEPER3-F-3. (Independent variables employed included Steering, Accelerometer, LANDEV/VAR, LNMNSQ, LANEX, & LNERRSQ.)

		Observed		
Predicted Algorithm Output		Low	High	Total
	High	31	66	97
	Low	360	17	377
	Total	391	83	474
	% Correct	92.07	79.52	89.87

Algorithm SLEEPER3-F (Threshold = 1.4) - OR - LANEX (Threshold = 0.06667)
Apparent Accuracy Rate: 0.899

Classification Matrix Generated from Multiple Regression Analysis of Lane-Enforced SLEEPER3 Data Resulting in Algorithm SLEEPER3-F-3 -OR- Measured LANEX Data.

		Observed		
Predicted Algorithm Output		Low	High	Total
	High	31	66	97
	Low	360	17	377
	Total	391	83	474
	% Correct	92.07	79.52	89.87

Algorithm SLEEPER3-F (Threshold = 1.4) - OR - LNMNSQ (Threshold = 3.00000)
Apparent Accuracy Rate: 0.899

Classification Matrix Generated from Multiple Regression Analysis of Lane-Enforced SLEEPER3 Data Resulting in Algorithm SLEEPER3-F-3 -OR- Measured LNMNSQ Data.

Figure L18: Regression Summary and Classification Matrices for Algorithm SLEEPER3-F-3.

Regression Summary for Dependent Variable: PERCLOS
 $R = 0.49073253$ $R^2 = 0.24081841$ Adjusted $R^2 = 0.22941440$
 $F(7,466) = 21.117$ $p < 0.00000$ Std. Error of estimate: 0.01069

	BETA	St. Err. of BETA	B	St. Err. of B	t(466)	p-level
Intercept			-0.00299	0.001052	-2.83806	0.004737
STVELV	0.519969	0.124176	0.00007	0.000018	4.18737	0.000034
LGREV	-0.36289	0.09742	-0.00275	0.000738	-3.72499	0.000219
MDREV	-0.49255	0.104825	-0.00056	0.00012	-4.6988	0.000003
STEXED	-0.10572	0.045236	-8.41549	3.60093	-2.33703	0.01986
LNMSQ	0.386739	0.054188	0.00386	0.00054	7.13703	0
LNRTVAR	0.365293	0.089166	0.0183	0.004468	4.09675	0.000049
INTACDEV	-0.14288	0.047827	-0.00725	0.002428	-2.98742	0.002962

Observed			
	Low	High	Total
Predicted			
High	18	17	35
Low	423	16	439
Algorithm			
Output	441	33	474
	95.92	51.52	92.83
% Correct			

PERCLOS R Value = 0.49073
Apparent Accuracy Rate: 0.928
Threshold: 0.012

Classification Matrix Generated from Multiple Regression Analysis of Lane-Enforced PERCLOS Data Resulting in Algorithm G4e-3. (Independent variables employed included Steering, Accelerometer, and all lane measures including LNRTDEV/VAR.)

Observed			
	Low	High	Total
Predicted			
High	2	57	59
Low	401	14	415
Algorithm			
Output	403	71	474
	99.50	80.28	96.62
% Correct			

Algorithm G4e (Threshold = 0.012) - OR - LANEX (Threshold = 0.06667)

Apparent Accuracy Rate: 0.966

Classification Matrix Generated from Multiple Regression Analysis of Lane-Enforced PERCLOS Data Resulting in Algorithm G4e-3 -OR- Measured LANEX Data.

Observed			
	Low	High	Total
Predicted			
High	3	57	60
Low	400	14	414
Algorithm			
Output	403	71	474
	99.26	80.28	96.41
% Correct			

Algorithm G4e (Threshold = 0.012) - OR - LNMNSQ (Threshold = 3.00000)

Apparent Accuracy Rate: 0.964

Classification Matrix Generated from Multiple Regression Analysis of Lane-Enforced PERCLOS Data Resulting in Algorithm G4e-3 -OR- Measured LNMNSQ Data.

Figure L19: Regression Summary and Classification Matrices for Algorithm G4e-3.

Regression Summary for Dependent Variable: EYEMEAS

R = 0.50518315 R² = 0.25521001 Adjusted R² = 0.24402218

F(7,466) = 22.811 p < 0.00000 Std. Error of estimate: 808.70

	BETA	St. Err. of BETA	B	St. Err. of B	t(466)	p-level
Intercept			402.0784	164.862	2.43888	0.015105
STVELV	0.355826	0.089313	3.85107	0.967	3.98405	0.000079
MDREV	-0.38587	0.090308	-33.7468	7.898	-4.27282	0.000023
LNMSNSQ	0.672622	0.119174	512.0988	90.733	5.64403	0
LANDEV	0.236721	0.078085	789.6631	260.478	3.03159	0.002568
LANEX	-0.22762	0.099335	-4850.4	2116.771	-2.29141	0.022385
LNRTVAR	-0.22138	0.083261	-846.976	318.549	-2.65886	0.00811
ACCDEV	-0.1658	0.049411	-173.733	51.776	-3.35549	0.000857

Observed

Predicted Algorithm Output	Low	High	Total
High	7	19	26
Low	390	58	448
Total	397	77	474
% Correct	98.24	24.68	86.29

EYEMEAS R Value = 0.50518

Apparent Accuracy Rate: 0.863

Threshold: 2000

Classification Matrix Generated from Multiple Regression Analysis of Lane-Enforced EYEMEAS Data Resulting in Algorithm G2e-3. (Independent variables employed included Steering, Accelerometer, and all lane measures including LNRTDEV/VAR.)

Observed

Predicted Algorithm Output	Low	High	Total
High	1	55	56
Low	370	48	418
Total	371	103	474
% Correct	99.73	53.40	89.66

Algorithm G2e (Threshold = 2000) - OR - LANEX (Threshold = 0.06667)

Apparent Accuracy Rate: 0.897

Classification Matrix Generated from Multiple Regression Analysis of Lane-Enforced EYEMEAS Data Resulting in Algorithm G2e-3 -OR- Measured LANEX Data.

Observed

Predicted Algorithm Output	Low	High	Total
High	1	56	57
Low	369	48	417
Total	370	104	474
% Correct	99.73	53.85	89.66

Algorithm G2e (Threshold = 2000) - OR - LNMNSQ (Threshold = 3.00000)

Apparent Accuracy Rate: 0.897

Classification Matrix Generated from Multiple Regression Analysis of Lane-Enforced EYEMEAS Data Resulting in Algorithm G2e-3 -OR- Measured LNMNSQ Data.

Figure L20: Regression Summary and Classification Matrices for Algorithm G2e-3.

Regression Summary for Dependent Variable: SLEEPER3
 $R = 0.53863418$ $R^2 = 0.29012678$ Adjusted $R^2 = 0.27791391$
 $F(8,465) = 23.756$ $p < 0.00000$ Std. Error of estimate: 0.89018

	BETA	St. Err. of BETA	B	St. Err. of B	t(465)	p-level
Intercept			-0.16549	0.2039	-0.81147	0.417512
STVELV	0.534174	0.120231	0.006512	0.0015	4.44291	0.000011
LGREV	-0.32983	0.095654	-0.21497	0.0623	-3.44812	0.000616
MDREV	-0.37012	0.106709	-0.03646	0.0105	-3.46854	0.000572
SMREV	0.119326	0.047014	0.009174	0.0036	2.53808	0.011471
STEXED	-0.10653	0.044009	-729.345	301.3027	-2.42064	0.015875
LNMN SQ	0.422557	0.052455	0.362346	0.045	8.05565	0
LNRTVAR	0.368618	0.089118	1.588412	0.384	4.1363	0.000042
INTACDEV	-0.22641	0.047786	-0.98827	0.2086	-4.73796	0.000003

		Observed		
Predicted Algorithm Output		Low	High	Total
	High	48	27	75
	Low	378	21	399
	Total	426	48	474
	% Correct	88.73	56.25	85.44

SLEEPER3 R Value = 0.53863
Apparent Accuracy Rate: 0.854
Threshold: 1.4

Classification Matrix Generated from Multiple Regression Analysis of Lane-Enforced SLEEPER3 Data Resulting in Algorithm SLEEPER3-G-3. (Independent variables employed included Steering, Accelerometer, and all lane measures including LNRTDEV/VAR.)

		Observed		
Predicted Algorithm Output		Low	High	Total
	High	27	64	91
	Low	364	19	383
	Total	391	83	474
	% Correct	93.09	77.11	90.30

Algorithm SLEEPER3-G (Threshold = 1.4) - OR - LANEX (Threshold = 0.06667)
Apparent Accuracy Rate: 0.903

Classification Matrix Generated from Multiple Regression Analysis of Lane-Enforced SLEEPER3 Data Resulting in Algorithm SLEEPER3-G-3 -OR- Measured LANEX Data.

		Observed		
Predicted Algorithm Output		Low	High	Total
	High	27	64	91
	Low	364	19	383
	Total	391	83	474
	% Correct	93.09	77.11	90.30

Algorithm SLEEPER3-G (Threshold = 1.4) - OR - LNMNSQ (Threshold = 3.00000)
Apparent Accuracy Rate: 0.903

Classification Matrix Generated from Multiple Regression Analysis of Lane-Enforced SLEEPER3 Data Resulting in Algorithm SLEEPER3-G-3 -OR- Measured LNMNSQ Data.

Figure L21: Regression Summary and Classification Matrices for Algorithm SLEEPER3-G-3.

APPENDIX M

Regression Summaries, Unconditional Classification Matrices, and Conditional “OR”
Classification Matrices for Driver Status Algorithms Developed Using Six-Minute
Interval Data Set

Regression Summary for Dependent Variable: PERCLOS
R = 0.21424527 R² = 0.04590104 Adjusted R² = 0.01686324
F(7,230) = 1.5807 p < 0.14198 Std. Error of estimate: 0.01049

	BETA	St. Err. of BETA	B	St. Err. of B	t(230)	p-level
Intercept			0.01996	0.016102	1.2394	0.21646
MDREV	-0.27441	0.176541	-0.00028	0.000182	-1.55439	0.121467
SMREV	-0.2107	0.17219	-0.00018	0.000145	-1.22362	0.222347
STEXED	-0.09338	0.069111	-8.01949	5.935241	-1.35116	0.17797
NMRHOLDS	-0.38216	0.257809	-0.00053	0.00036	-1.48233	0.139622
THRSHLDS	0.15323	0.173096	0.02221	0.025087	0.88523	0.376958
ACCDEV	0.416782	0.204128	0.00668	0.00327	2.04177	0.042316
INTACDEV	-0.4071	0.185492	-0.02391	0.010892	-2.19471	0.029185

Note: Classification matrices not developed for this algorithm.

Figure M1: Regression Summary for Algorithm D4e-6.

Regression Summary for Dependent Variable: EYEMEAS

R = 0.17804053 R² = 0.03169843 Adjusted R² = 0.01082986

F(5,232) = 1.5190 p < 0.18467 Std. Error of estimate: 903.24

	BETA	St. Err. of BETA	B	St. Err. of B	t(232)	p-level
Intercept			2893.097	1326.684	2.1807	0.030211
MDREV	-0.27291	0.175504	-24.202	15.563	-1.55503	0.121302
SMREV	-0.25922	0.155916	-18.779	11.295	-1.66253	0.097757
NMRHOLDS	-0.24626	0.241387	-29.523	28.938	-1.0202	0.308696
ACCDEV	0.176989	0.197067	243.28	270.877	0.89812	0.370053
INTACDEV	-0.17065	0.179363	-859.791	903.717	-0.95139	0.342395

Note: Classification matrices not developed for this algorithm.

Figure M2: Regression Summary for Algorithm D2e-6.

Regression Summary for Dependent Variable: DRVDROW
R = 0.49927375 R² = 0.24927428 Adjusted R² = 0.23638628
F(4,233) = 19.342 p < 0.00000 Std. Error of estimate: 26.352

	BETA	St. Err. of BETA	B	St. Err. of B	t(233)	p-level
Intercept			-23.9242	23.69037	-1.00987	0.313605
MDREV	0.604658	0.08666	1.7805	0.25518	6.9774	0
SMREV	0.363384	0.113685	0.8741	0.27347	3.19641	0.001584
THRSHLDS	0.334322	0.140411	138.0536	57.98077	2.38102	0.018069
INTACDEV	-0.08356	0.059355	-13.9804	9.93038	-1.40784	0.16051

Predicted Algorithm Output	Observed			
		Low	High	Total
	High	7	25	32
	Low	142	64	206
	Total	149	89	238
	% Correct	95.30	28.09	70.17

DRVDROW R Value = 0.49927

Apparent Accuracy Rate: 0.702

Threshold: 70

Classification Matrix Generated from Multiple Regression Analysis of Lane-Enforced DRVDROW Data Resulting in Algorithm D6e-6. (Independent variables employed included Steering and Accelerometer.)

Predicted Algorithm Output	Observed			
		Low	High	Total
	High	6	26	32
	Low	134	72	206
	Total	140	98	238
	% Correct	95.71	26.53	67.23

Algorithm D6e (Threshold = 70) - OR - eLNMNSQ-D (Threshold = 3.00000)

Apparent Accuracy Rate: 0.672

Classification Matrix Generated from Multiple Regression Analysis of Lane-Enforced DRVDROW Data Resulting in Algorithm D6e-6 -OR- Algorithm eLNMNSQ-D-6 Data.

Figure M3: Regression Summary and Classification Matrices for Algorithm D6e-6.

Regression Summary for Dependent Variable: SLEEPER1

R = 0.22283711 R² = 0.04965638 Adjusted R² = 0.04156835

F(2,235) = 6.1395 p < 0.00252 Std. Error of estimate: 0.65875

	BETA	St. Err. of BETA	B	St. Err. of B	t(235)	p-level
Intercept			1.633673	0.127913	12.7718	0
NMRHOLDS	-0.44269	0.143256	-0.03932	0.012725	-3.09018	0.002241
THRSHLDS	0.29161	0.143256	2.686911	1.319966	2.03559	0.042913

Note: Classification matrices not developed for this algorithm.

Figure M4: Regression Summary for Algorithm SLEEPER1-D-6.

Regression Summary for Dependent Variable: SLEEPER2

R = 0.13308084 R² = 0.01771051 Adjusted R² = 0.00935060

F(2,235) = 2.1185 p < 0.12250 Std. Error of estimate: 1.1528

	BETA	St. Err. of BETA	B	St. Err. of B	t(235)	p-level
Intercept			1.76262	0.201936	8.72857	0
ACCDEV	0.294471	0.146985	0.51623	0.257677	2.00342	0.046281
INTACDEV	-0.29501	0.146985	-1.89574	0.944529	-2.00707	0.045888

Note: Classification matrices not developed for this algorithm.

Figure M5: Regression Summary for Algorithm SLEEPER2-D-6.

Regression Summary for Dependent Variable: SLEEPER3
R = 0.21449643 R² = 0.04600872 Adjusted R² = 0.03788964
F(2,235) = 5.6667 p < 0.00395 Std. Error of estimate: 0.93209

	BETA	St. Err. of BETA	B	St. Err. of B	t(235)	p-level
Intercept			1.483813	0.180988	8.19841	0
NMRHOLDS	-0.47164	0.14353	-0.05916	0.018004	-3.28598	0.001172
THRSHLDS	0.375983	0.14353	4.892417	1.867665	2.61954	0.009378

Note: Classification matrices not developed for this algorithm.

Figure M6: Regression Summary for Algorithm SLEEPER3-D-6.

Regression Summary for Dependent Variable: PERCLOS

R = 0.40449755 R² = 0.16361826 Adjusted R² = 0.14559280

F(5,232) = 9.0771 p < 0.00000 Std. Error of estimate: 0.00978

	BETA	St. Err. of BETA	B	St. Err. of B	t(232)	p-level
Intercept			-0.0024	0.001377	-1.74202	0.08283
STVELV	0.72866	0.201384	0.0000903	0.000025	3.61825	0.000364
LGREV	-0.63537	0.162345	-0.00448	0.001144	-3.91371	0.000119
MDREV	-0.77627	0.166116	-0.0008	0.000172	-4.67304	0.000005
STEXED	-0.19915	0.069586	-17.1034	5.976072	-2.86197	0.004595
HPHDGVAR	0.839146	0.127311	0.079071	0.011996	6.59131	0

Predicted Algorithm Output	Observed		
		Low	High
	High	4	4
	Low	214	16
	Total	218	20
	% Correct	98.17	20.00
		Total	238

PERCLOS R Value = 0.40450

Apparent Accuracy Rate: 0.916

Threshold: 0.012

Classification Matrix Generated from Multiple Regression Analysis of Lane-Enforced PERCLOS Data Resulting in Algorithm E4e-6. (Independent variables employed included Steering, Accelerometer, and HPHDGDEV/VAR.)

Predicted Algorithm Output	Observed		
		Low	High
	High	4	4
	Low	199	31
	Total	203	35
	% Correct	98.03	11.43
		Total	238

Algorithm E4e (Threshold = 0.012) - OR - eLANEX-E (Threshold = 0.06667)

Apparent Accuracy Rate: 0.853

Classification Matrix Generated from Multiple Regression Analysis of Lane-Enforced PERCLOS Data Resulting in Algorithm E4e-6. -OR- Algorithm eLANEX-E-6 Data.

Predicted Algorithm Output	Observed		
		Low	High
	High	4	7
	Low	197	30
	Total	201	37
	% Correct	98.01	18.92
		Total	238

Algorithm E4e (Threshold = 0.012) - OR - eLNMNSQ-E (Threshold = 3.00000)

Apparent Accuracy Rate: 0.857

Classification Matrix Generated from Multiple Regression Analysis of Lane-Enforced PERCLOS Data Resulting in Algorithm E4e-6. -OR- Algorithm eLNMNSQ-E-6 Data.

Figure M7: Regression Summary and Classification Matrices for Algorithm E4e-6.

Regression Summary for Dependent Variable: EYEMEAS

R = 0.33980182 R² = 0.11546528 Adjusted R² = 0.10028013

F(4,233) = 7.6038 p < 0.00001 Std. Error of estimate: 861.43

	BETA	St. Err. of BETA	B	St. Err. of B	t(233)	p-level
Intercept			116.517	248.6082	0.46868	0.639738
STVELV	0.642393	0.208251	6.831	2.2144	3.08471	0.002284
LGREV	-0.43904	0.156958	-265.414	94.887	-2.79716	0.005586
MDREV	-0.78589	0.1728	-69.691	15.3236	-4.54798	0.000009
HPHDGDEV	0.612748	0.119221	4406.414	857.3442	5.13961	0.000001

Figure M8: Regression Summary for Algorithm E2e-6.

Regression Summary for Dependent Variable: DRVDROWR = 0.53780402 R² = 0.28923316 Adjusted R² = 0.27391491

F(5,232) = 18.882 p < 0.00000 Std. Error of estimate: 25.696

	BETA	St. Err. of BETA	B	St. Err. of B	t(232)	p-level
Intercept			-47.3948	25.14356	-1.88497	0.060684
MDREV	0.403577	0.10556	1.1884	0.31083	3.82321	0.000169
SMREV	0.518145	0.121827	1.2464	0.29306	4.25312	0.000031
THRSHLDS	0.378751	0.139269	156.4	57.50906	2.71957	0.007031
ACCDEV	-0.18118	0.067888	-8.2695	3.09856	-2.66882	0.008149
HPHDGVAR	0.364848	0.10185	97.9508	27.3438	3.58219	0.000415

Predicted Algorithm Output	Observed		
		Low	High
	High	10	31
	Low	139	58
	Total	149	89
	% Correct	93.29	34.83
		Total	238

DRVDROW R Value = 0.53780

Apparent Accuracy Rate: 0.714

Threshold: 70

Classification Matrix Generated from Multiple Regression Analysis of Lane-Enforced DRVDROW Data Resulting in Algorithm E6e-6. (Independent variables employed included Steering, Accelerometer, and HPHDGDEV/VAR.)

Predicted Algorithm Output	Observed		
		Low	High
	High	11	32
	Low	130	65
	Total	141	97
	% Correct	92.20	32.99
		Total	238

Algorithm E6e (Threshold = 70) - OR - eLANEX-E (Threshold = 0.06667)

Apparent Accuracy Rate: 0.681

Classification Matrix Generated from Multiple Regression Analysis of Lane-Enforced DRVDROW Data Resulting in Algorithm E6e-6. -OR- Algorithm eLANEX-E-6 Data.

Predicted Algorithm Output	Observed		
		Low	High
	High	11	35
	Low	129	63
	Total	140	98
	% Correct	92.14	35.71
		Total	238

Algorithm E6e (Threshold = 70) - OR - eLNMNSQ-E (Threshold = 3.00000)

Apparent Accuracy Rate: 0.689

Classification Matrix Generated from Multiple Regression Analysis of Lane-Enforced DRVDROW Data Resulting in Algorithm E6e-6. -OR- Algorithm eLNMNSQ-E-6 Data.

Figure M9: Regression Summary and Classification Matrices for Algorithm E6e-6.

Regression Summary for Dependent Variable: SLEEPER1

R = 0.32646937 $R^2 = 0.10658225$ Adjusted $R^2 = 0.09512817$

F(3,234) = 9.3052 $p < 0.00001$ Std. Error of estimate: 0.64008

	BETA	St. Err. of BETA	B	St. Err. of B	t(234)	p-level
Intercept			0.50387	0.17067	2.95231	0.003475
STVELV	0.27526	0.120789	0.002169	0.000952	2.27884	0.023578
LGREV	-0.32437	0.137987	-0.14529	0.061807	-2.35074	0.019567
HPHDGDEV	0.350601	0.094117	1.868056	0.501469	3.72517	0.000245

Figure M10: Regression Summary for Algorithm SLEEPER1-E-6.

Regression Summary for Dependent Variable: SLEEPER2

R = 0.32456540 $R^2 = 0.10534270$ Adjusted $R^2 = 0.08998378$

F(4,233) = 6.8587 $p < 0.00003$ Std. Error of estimate: 1.1049

	BETA	St. Err. of BETA	B	St. Err. of B	t(233)	p-level
Intercept			0.537309	0.318882	1.68498	0.093332
STVELV	0.620482	0.209439	0.008415	0.00284	2.96259	0.003367
LGREV	-0.4712	0.157853	-0.36331	0.121709	-2.98504	0.003138
MDREV	-0.44372	0.173786	-0.05018	0.019655	-2.55324	0.011311
HPHDGDEV	0.530773	0.119901	4.868061	1.09969	4.42676	0.000015

Figure M11: Regression Summary for Algorithm SLEEPER2-E-6.

Regression Summary for Dependent Variable: SLEEPER3R = 0.42229843 R² = 0.17833596 Adjusted R² = 0.16062769

F(5,232) = 10.071 p < 0.00000 Std. Error of estimate: 0.87061

	BETA	St. Err. of BETA	B	St. Err. of B	t(232)	p-level
Intercept			0.298972	0.1225	2.43984	0.015444
STVELV	0.748381	0.199605	0.008327	0.0022	3.74932	0.000224
LGREV	-0.68534	0.16091	-0.43352	0.1018	-4.25915	0.00003
MDREV	-0.53463	0.164648	-0.04961	0.0153	-3.24712	0.001338
STEXED	-0.16396	0.068971	-1264.18	531.8022	-2.37717	0.018258
HPHDGVAR	0.791618	0.126186	6.697056	1.0675	6.27343	0

Predicted Algorithm Output	Observed		
		Low	High
	High	17	11
	Low	192	18
	Total	209	29
	% Correct	91.87	37.93
		Total	238

SLEEPER3 R Value = 0.42230

Apparent Accuracy Rate: 0.853

Threshold: 1.4

Classification Matrix Generated from Multiple Regression Analysis of Lane-Enforced SLEEPER3 Data Resulting in Algorithm SLEEPER3-E-6. (Independent variables employed included Steering, Accelerometer, and HPHDGDEV/VAR.)

Predicted Algorithm Output	Observed		
		Low	High
	High	15	13
	Low	181	29
	Total	196	42
	% Correct	92.35	30.95
		Total	238

Algorithm SLEEPER3-E (Threshold = 1.4) - OR - eLANEX-E (Threshold = 0.06667)

Apparent Accuracy Rate: 0.815

Classification Matrix Generated from Multiple Regression Analysis of Lane-Enforced SLEEPER3 Data Resulting in Algorithm SLEEPER3-E-6. -OR- Algorithm eLANEX-E-6 Data.

Predicted Algorithm Output	Observed		
		Low	High
	High	14	16
	Low	180	28
	Total	194	44
	% Correct	92.78	36.36
		Total	238

Algorithm SLEEPER3-E (Threshold = 1.4) - OR - eLNMNSQ-E (Threshold = 3.00000)

Apparent Accuracy Rate: 0.824

Classification Matrix Generated from Multiple Regression Analysis of Lane-Enforced SLEEPER3 Data Resulting in Algorithm SLEEPER3-E-6. -OR- Algorithm eLNMNSQ-E-6 Data.

Figure M12: Regression Summary and Classification Matrices for Algorithm SLEEPER3-E-6.

Regression Summary for Dependent Variable: PERCLOS
 $R = 0.57483855$ $R^2 = 0.33043936$ Adjusted $R^2 = 0.31304818$
 $F(6,231) = 19.000$ $p < 0.00000$ Std. Error of estimate: 0.00877

	BETA	St. Err. of BETA	B	St. Err. of B	t(231)	p-level
Intercept			-0.0026	0.001378	-1.92078	0.05599
STVELV	0.357112	0.128601	0	0.000016	2.77689	0.005939
MDREV	-0.44237	0.129947	-0.0005	0.000134	-3.40422	0.000782
STEXED	-0.15776	0.059157	-13.5482	5.080455	-2.66674	0.008201
LNMSQ	0.430096	0.085709	0.0043	0.000852	5.01808	0.000001
LANVAR	0.246026	0.089905	0.0049	0.001809	2.73651	0.006692
INTACDEV	-0.21847	0.058745	-0.0128	0.00345	-3.71897	0.000251

Predicted Algorithm Output	Observed		
		Low	High
	High	8	13
	Low	210	7
	Total	218	20
	% Correct	96.33	65.00
			93.70

PERCLOS R Value = 0.57484

Apparent Accuracy Rate: 0.937

Threshold: 0.012

Classification Matrix Generated from Multiple Regression Analysis of Lane-Enforced PERCLOS Data Resulting in Algorithm F4e-6. (Independent variables employed included Steering, Accelerometer, LANDEV/VAR, LNMNSQ, LANEX, & LNERRSQ.)

Predicted Algorithm Output	Observed		
		Low	High
	High	2	28
	Low	201	7
	Total	203	35
	% Correct	99.01	80.00
			96.22

Algorithm F4e (Threshold = 0.012) - OR - LANEX (Threshold = 0.06667)

Apparent Accuracy Rate: 0.962

Classification Matrix Generated from Multiple Regression Analysis of Lane-Enforced PERCLOS Data Resulting in Algorithm F4e-6. -OR- Measured LANEX Data.

Predicted Algorithm Output	Observed		
		Low	High
	High	1	30
	Low	200	7
	Total	201	37
	% Correct	99.50	81.08
			96.64

Algorithm F4e (Threshold = 0.012) - OR - LNMNSQ (Threshold = 3.00000)

Apparent Accuracy Rate: 0.966

Classification Matrix Generated from Multiple Regression Analysis of Lane-Enforced PERCLOS Data Resulting in Algorithm F4e-6. -OR- Measured LNMNSQ Data.

Figure M13: Regression Summary and Classification Matrices for Algorithm F4e-6.

Regression Summary for Dependent Variable: EYEMEAS

R = 0.56911454 R² = 0.32389136 Adjusted R² = 0.30932005

F(5,232) = 22.228 p < 0.00000 Std. Error of estimate: 754.76

	BETA	St. Err. of BETA	B	St. Err. of B	t(232)	p-level
Intercept			758.79	154.212	4.9204	0.000002
STVELV	0.397864	0.120756	4.23	1.284	3.29477	0.001139
MDREV	-0.4763	0.127599	-42.24	11.315	-3.73278	0.000238
LNMSNSQ	0.918816	0.158481	783.84	135.201	5.79763	0
LANEX	-0.32104	0.150682	-8038.92	3773.096	-2.13059	0.034175
ACCDEV	-0.23169	0.065894	-318.47	90.575	-3.51614	0.000527

Predicted Algorithm Output	Observed		
		Low	High
	High	5	11
	Low	194	28
	Total	199	39
	% Correct	97.49	28.21
EYEMEAS R Value = 0.56911			
Apparent Accuracy Rate: 0.861			
Threshold: 2000			
Classification Matrix Generated from Multiple Regression Analysis of Lane-Enforced EYEMEAS Data Resulting in Algorithm F2e-6. (Independent variables employed included Steering, Accelerometer, LANDEV/VAR, LNMNSQ, LANEX, & LNERRSQ.)			

Predicted Algorithm Output	Observed		
		Low	High
	High	2	25
	Low	187	24
	Total	189	49
	% Correct	98.94	51.02
Algorithm F2e (Threshold = 2000) - OR - LANEX (Threshold = 0.06667)			
Apparent Accuracy Rate: 0.891			
Classification Matrix Generated from Multiple Regression Analysis of Lane-Enforced EYEMEAS Data Resulting in Algorithm F2e-6. -OR- Measured LANEX Data			

Predicted Algorithm Output	Observed		
		Low	High
	High	1	27
	Low	185	25
	Total	186	52
	% Correct	99.46	51.92
Algorithm F2e (Threshold = 2000) - OR - LNMNSQ (Threshold = 3.00000)			
Apparent Accuracy Rate: 0.891			
Classification Matrix Generated from Multiple Regression Analysis of Lane-Enforced EYEMEAS Data Resulting in Algorithm F2e-6. -OR- Measured LNMNSQ Data.			

Figure M14: Regression Summary and Classification Matrices for Algorithm F2e-6.

Regression Summary for Dependent Variable: DRVDROW

R = 0.58317659 R² = 0.34009494 Adjusted R² = 0.32587285

F(5,232) = 23.913 p < 0.00000 Std. Error of estimate: 24.760

	BETA	St. Err. of BETA	B	St. Err. of B	t(232)	p-level
Intercept			3.14	8.7486	0.35896	0.719952
STVELV	0.436211	0.056924	0.154	0.0201	7.66299	0
SMREV	0.264105	0.054703	0.635	0.1316	4.828	0.000003
LNMSNSQ	0.768212	0.155529	21.761	4.4057	4.93935	0.000001
LANEX	-0.39776	0.147715	-330.724	122.819	-2.69277	0.007603
ACCDEV	-0.27764	0.064418	-12.672	2.9402	-4.30996	0.000024

Predicted Algorithm Output	Observed		
		Low	High
	High	10	36
	Low	139	53
	Total	149	89
	% Correct	93.29	40.45
		Total	238

DRVDROW R Value = 0.58318

Apparent Accuracy Rate: 0.735

Threshold: 70

Classification Matrix Generated from Multiple Regression Analysis of Lane-Enforced DRVDROW Data Resulting in Algorithm F6e-6. (Independent variables employed included Steering, Accelerometer, LANDEV/VAR, LNMSNSQ, LANEX, & LNERRSQ.)

Predicted Algorithm Output	Observed		
		Low	High
	High	7	49
	Low	134	48
	Total	141	97
	% Correct	95.04	50.52
		Total	238

Algorithm F6e (Threshold = 70) - OR - LANEX (Threshold = 0.06667)

Apparent Accuracy Rate: 0.769

Classification Matrix Generated from Multiple Regression Analysis of Lane-Enforced DRVDROW Data Resulting in Algorithm F6e-6. -OR- Measured LANEX Data.

Predicted Algorithm Output	Observed		
		Low	High
	High	7	50
	Low	133	48
	Total	140	98
	% Correct	95.00	51.02
		Total	238

Algorithm F6e (Threshold = 70) - OR - LNMSNSQ (Threshold = 3.00000)

Apparent Accuracy Rate: 0.769

Classification Matrix Generated from Multiple Regression Analysis of Lane-Enforced DRVDROW Data Resulting in Algorithm F6e-6. -OR- Measured LNMSNSQ Data.

Figure M15: Regression Summary and Classification Matrices for Algorithm F6e-6.

Regression Summary for Dependent Variable: SLEEPER1R = 0.54728563 R² = 0.29952156 Adjusted R² = 0.28442505

F(5,232) = 19.840 p < 0.00000 Std. Error of estimate: 0.56921

	BETA	St. Err. of BETA	B	St. Err. of B	t(232)	p-level
Intercept			0.410343	0.201123	2.04025	0.042458
STVELV	0.261044	0.058648	0.002057	0.000462	4.45101	0.000013
SMREV	0.135384	0.056359	0.007267	0.003025	2.40216	0.017086
LNMSQ	0.882456	0.160239	0.557785	0.101284	5.50713	0
LANEX	-0.35631	0.152188	-6.61056	2.823517	-2.34125	0.020067
ACCDEV	-0.3168	0.066369	-0.32264	0.067592	-4.77329	0.000003

Predicted Algorithm Output	Observed		
		Low	High
	High	11	14
	Low	172	41
	Total	183	55
% Correct		93.99	25.45
			78.15

SLEEPER1 R Value = 0.54729

Apparent Accuracy Rate: 0.782

Threshold: 1.7

Classification Matrix Generated from Multiple Regression Analysis of Lane-Enforced SLEEPER1 Data Resulting in Algorithm SLEEPER1-F-6. (Independent variables employed included Steering, Accelerometer, LANDEV/VAR, LNMSQ, LANEX, & LNERRSQ.)

Predicted Algorithm Output	Observed		
		Low	High
	High	7	29
	Low	168	34
	Total	175	63
% Correct		96.00	46.03
			82.77

Algorithm SLEEPER1-F (Threshold = 1.7) - OR - LANEX (Threshold = 0.06667)

Apparent Accuracy Rate: 0.828

Classification Matrix Generated from Multiple Regression Analysis of Lane-Enforced SLEEPER1 Data Resulting in Algorithm SLEEPER1-F-6 -OR- Measured LANEX Data.

Predicted Algorithm Output	Observed		
		Low	High
	High	7	31
	Low	165	35
	Total	172	66
% Correct		95.93	46.97
			82.35

Algorithm SLEEPER1-F (Threshold = 1.7) - OR - LNMSQ (Threshold = 3.00000)

Apparent Accuracy Rate: 0.824

Classification Matrix Generated from Multiple Regression Analysis of Lane-Enforced SLEEPER1 Data Resulting in Algorithm SLEEPER1-F-6 -OR- Measured LNMSQ Data.

Figure M16: Regression Summary and Classification Matrices for Algorithm SLEEPER1-F-6.

Regression Summary for Dependent Variable: SLEEPER2
 $R = 0.53228675$ $R^2 = 0.28332919$ $\text{Adjusted } R^2 = 0.27102583$
 $F(4,233) = 23.029$ $p < 0.00000$ $\text{Std. Error of estimate: } 0.98894$

	BETA	St. Err. of BETA	B	St. Err. of B	t(233)	p-level
Intercept			1.318681	0.20206	6.52619	0
STVELV	0.202467	0.059179	0.002746	0.000803	3.42123	0.000736
LNMSNSQ	0.837479	0.15967	0.911209	0.173727	5.24505	0
LANEX	-0.30595	0.152812	-9.7709	4.880212	-2.00215	0.046428
ACCDEV	-0.29521	0.065873	-0.51753	0.115481	-4.48156	0.000012

Predicted Algorithm Output	Observed		
		Low	High
	High	5	11
	Low	189	33
	Total	194	44
	% Correct	97.42	25.00
			84.03

SLEEPER2 R Value = 0.53229
Apparent Accuracy Rate: 0.840
Threshold: 3

Classification Matrix Generated from Multiple Regression Analysis of Lane-Enforced SLEEPER2 Data Resulting in Algorithm SLEEPER2-F-6. (Independent variables employed included Steering, Accelerometer, LANDEV/VAR, LNMSNSQ, LANEX, & LNERRSQ.)

Predicted Algorithm Output	Observed		
		Low	High
	High	2	27
	Low	182	27
	Total	184	54
	% Correct	98.91	50.00
			87.82

Algorithm SLEEPER2-F (Threshold = 3) - OR - LANEX (Threshold = 0.06667)
Apparent Accuracy Rate: 0.878

Classification Matrix Generated from Multiple Regression Analysis of Lane-Enforced SLEEPER2 Data Resulting in Algorithm SLEEPER2-F-6. -OR- Measured LANEX Data.

Predicted Algorithm Output	Observed		
		Low	High
	High	2	29
	Low	179	28
	Total	181	57
	% Correct	98.90	50.88
			87.39

Algorithm SLEEPER2-F (Threshold = 3) - OR - LNMSNSQ (Threshold = 3.00000)
Apparent Accuracy Rate: 0.874

Classification Matrix Generated from Multiple Regression Analysis of Lane-Enforced SLEEPER2 Data Resulting in Algorithm SLEEPER2-F-6. -OR- Measured LNMSNSQ Data.

Figure M17: Regression Summary and Classification Matrices for Algorithm SLEEPER2-F-6.

Regression Summary for Dependent Variable: SLEEPER3

R = 0.61440949 R² = 0.37749902 Adjusted R² = 0.35855334

F(7,230) = 19.925 p < 0.00000 Std. Error of estimate: 0.76107

	BETA	St. Err. of BETA	B	St. Err. of B	t(230)	p-level
Intercept			-0.44069	0.2793	-1.57794	0.115953
STVELV	0.430433	0.130391	0.004789	0.0015	3.30108	0.001116
MDREV	-0.35583	0.137299	-0.03302	0.0127	-2.59166	0.010163
SMREV	0.183382	0.061201	0.013901	0.0046	2.99641	0.003031
STEXED	-0.14003	0.057201	-1079.68	441.0498	-2.44798	0.015115
LNMSQ	0.440137	0.083353	0.392885	0.0744	5.2804	0
LANVAR	0.288321	0.093517	0.520719	0.1689	3.08309	0.002299
INTACDEV	-0.27275	0.057169	-1.43796	0.3014	-4.77096	0.000003

Observed

Predicted Algorithm Output		Low	High	Total
	High	25	16	41
	Low	184	13	197
	Total	209	29	238
	% Correct	88.04	55.17	84.03

SLEEPER3 R Value = 0.61441

Apparent Accuracy Rate: 0.840

Threshold: 1.4

Classification Matrix Generated from Multiple Regression Analysis of Lane-Enforced SLEEPER3 Data Resulting in Algorithm SLEEPER3-F-6. (Independent variables employed included Steering, Accelerometer, LANDEV/VAR, LNMNSQ, LANEX, & LNERRSQ.)

Observed

Predicted Algorithm Output		Low	High	Total
	High	14	29	43
	Low	182	13	195
	Total	196	42	238
	% Correct	92.86	69.05	88.66

Algorithm SLEEPER3-F (Threshold = 1.4) - OR - LANEX (Threshold = 0.06667)

Apparent Accuracy Rate: 0.887

Classification Matrix Generated from Multiple Regression Analysis of Lane-Enforced SLEEPER3 Data Resulting in Algorithm SLEEPER3-F-6. -OR- Measured LANEX Data.

Observed

Predicted Algorithm Output		Low	High	Total
	High	13	31	44
	Low	181	13	194
	Total	194	44	238
	% Correct	93.30	70.45	89.08

Algorithm SLEEPER3-F (Threshold = 1.4) - OR - LNMNSQ (Threshold = 3.00000)

Apparent Accuracy Rate: 0.891

Classification Matrix Generated from Multiple Regression Analysis of Lane-Enforced SLEEPER3 Data Resulting in Algorithm SLEEPER3-F-6. -OR- Measured LNMNSQ Data.

Figure M18: Regression Summary and Classification Matrices for Algorithm SLEEPER3-F-6.

Regression Summary for Dependent Variable: SLEEPER2

R = 0.54417179 R² = 0.29612294 Adjusted R² = 0.28095318

F(5,232) = 19.521 p < 0.00000 Std. Error of estimate: 0.98218

	BETA	St. Err. of BETA	B	St. Err. of B	t(232)	p-level
Intercept			1.251024	0.203366	6.15159	0
STVELV	0.320097	0.082072	0.004341	0.001113	3.90019	0.000126
LNMSNSQ	1.046861	0.188531	1.139025	0.205129	5.55273	0
LANEX	-0.43715	0.164669	-13.961	5.258858	-2.65475	0.008486
LNRTVAR	-0.21054	0.102525	-1.07512	0.523553	-2.0535	0.041146
ACCDEV	-0.27532	0.066136	-0.48266	0.115943	-4.16289	0.000044

Predicted Algorithm Output	Observed		
		Low	High
	High	7	11
	Low	187	33
	Total	194	44
% Correct		96.39	25.00
			83.19

SLEEPER2 R Value = 0.54417

Apparent Accuracy Rate: 0.832

Threshold: 3

Classification Matrix Generated from Multiple Regression Analysis of Lane-Enforced SLEEPER2 Data Resulting in Algorithm SLEEPER2-G-6. (Independent variables employed included Steering, Accelerometer, and all lane measures including LNRTDEV/VAR.)

Predicted Algorithm Output	Observed		
		Low	High
	High	3	27
	Low	181	27
	Total	184	54
% Correct		98.37	50.00
			87.39

Algorithm SLEEPER2-G (Threshold = 3) - OR - LANEX (Threshold = 0.06667)

Apparent Accuracy Rate: 0.874

Classification Matrix Generated from Multiple Regression Analysis of Lane-Enforced SLEEPER2 Data Resulting in Algorithm SLEEPER2-G-6. -OR- Measured LANEX Data.

Predicted Algorithm Output	Observed		
		Low	High
	High	3	29
	Low	178	28
	Total	181	57
% Correct		98.34	50.88
			86.97

Algorithm SLEEPER2-G (Threshold = 3) - OR - LNMSNSQ (Threshold = 3.00000)

Apparent Accuracy Rate: 0.870

Classification Matrix Generated from Multiple Regression Analysis of Lane-Enforced SLEEPER2 Data Resulting in Algorithm SLEEPER2-G-6. -OR- Measured LNMSNSQ Data.

Figure M19: Regression Summary and Classification Matrices for Algorithm SLEEPER2-G-6.

APPENDIX N

**Regression Summaries and Classification Matrices for Driver Performance Algorithms
Developed Using Three-Minute and Six-Minute Interval Data Sets.**

Regression Summary for Dependent Variable: LANEX

R = 0.36680870 $R^2 = 0.13454862$ Adjusted $R^2 = 0.13087367$
F(2,471) = 36.612 $p < 0.00000$ Std. Error of estimate: 0.04069

	BETA	St. Err. of BETA	B	St. Err. of B	t(471)	p-level
Intercept			0.017	0.00781	2.17756	0.02993
SMREV	-0.1208	0.04352	-0.0004	0.00014	-2.7748	0.00574
ACCDEV	0.36788	0.04352	0.01809	0.00214	8.45273	0

Figure N1: Regression Summary for Algorithm eLANEX-D-3.

Regression Summary for Dependent Variable: LNMNSQ

R = 0.49339638 $R^2 = 0.24343999$ Adjusted $R^2 = 0.23535709$
F(5,468) = 30.118 $p < 0.00000$ Std. Error of estimate: 1.0683

	BETA	St. Err. of BETA	B	St. Err. of B	t(468)	p-level
Intercept			0.18557	0.1819	1.02018	0.30817
STVELV	-0.2436	0.08376	-0.0035	0.00119	-2.9085	0.00381
MDREV	0.34451	0.09374	0.03957	0.01077	3.67517	0.00027
THRSHLDS	0.21393	0.0508	3.25059	0.77181	4.21162	0.00003
ACCDEV	0.38236	0.05129	0.52625	0.07059	7.45543	0
ACEXEED	0.1318	0.04628	13.1793	4.62794	2.84776	0.0046

Figure N2: Regression Summary for Algorithm eLNMNSQ-D-3.

Regression Summary for Dependent Variable: LANEX

$R = 0.49743012$ $R^2 = 0.24743672$ Adjusted $R^2 = 0.24101827$

$F(4,469) = 38.551$ $p < 0.00000$ Std. Error of estimate: 0.03803

	BETA	St. Err. of BETA	B	St. Err. of B	t(469)	p-level
Intercept			-0.0174	0.00401	-4.346	0.000017
LGREV	-0.3661	0.06863	-0.0099	0.00186	-5.3348	0
MDREV	-0.2228	0.05969	-0.0009	0.00025	-3.7324	0.00021
ACCDEV	0.26421	0.04661	0.01299	0.00229	5.66828	0
HPHDGVAR	0.6558	0.07429	0.23842	0.02701	8.82751	0

Figure N3: Regression Summary for Algorithm eLANEX-E-3.

Regression Summary for Dependent Variable: LNMNSQ

$R = 0.66422380$ $R^2 = 0.44119326$ Adjusted $R^2 = 0.43642731$

$F(4,469) = 92.572$ $p < 0.0000$ Std. Error of estimate: 0.91711

	BETA	St. Err. of BETA	B	St. Err. of B	t(469)	p-level
Intercept			0.3326	0.08695	3.82541	0.00015
LGREV	-0.481	0.05912	-0.3656	0.04493	-8.1367	0
MDREV	-0.2007	0.05097	-0.0231	0.00586	-3.9366	0.000095
INTACVAR	0.26525	0.03616	1.76064	0.24001	7.33582	0
HPHDGVAR	0.94975	0.06273	9.66436	0.6383	15.1408	0

Figure N4: Regression Summary for Algorithm eLNMNSQ-E-3.

Regression Summary for Dependent Variable: LANEX

$R = 0.34844513$ $R^2 = 0.12141401$ Adjusted $R^2 = 0.11015009$

$F(3,234) = 10.779$ $p < 0.00000$ Std. Error of estimate: 0.03421

	BETA	St. Err. of BETA	B	St. Err. of B	t(234)	p-level
Intercept			0.01677	0.01066	1.57303	0.11706
LGREV	-0.1228	0.07092	-0.003	0.00171	-1.7323	0.08454
SMREV	-0.143	0.06425	-0.0004	0.00019	-2.2258	0.02698
ACCDEV	0.38046	0.06966	0.02089	0.00382	5.46158	0

Figure N5: Regression Summary for Algorithm eLANEX-D-6.

Regression Summary for Dependent Variable: LNMNSQ

$R = 0.48548148$ $R^2 = 0.23569226$ Adjusted $R^2 = 0.22257110$

$F(4,233) = 17.963$ $p < 0.00000$ Std. Error of estimate: 0.93864

	BETA	St. Err. of BETA	B	St. Err. of B	t(233)	p-level
Intercept			1.4963	0.29476	5.07626	0.000001
STVELV	-0.4181	0.13195	-0.0052	0.00165	-3.1689	0.00174
MDREV	0.42108	0.13966	0.04377	0.01452	3.0151	0.00285
SMREV	-0.2131	0.06067	-0.0181	0.00515	-3.5124	0.00053
ACCDEV	0.3979	0.0643	0.64111	0.10361	6.18801	0

Figure N6: Regression Summary for Algorithm eLNMNSQ-D-6.

Regression Summary for Dependent Variable: LANEX

$R = 0.51371514$ $R^2 = 0.26390325$ Adjusted $R^2 = 0.25126639$
 $F(4,233) = 20.884$ $p < 0.00000$ Std. Error of estimate: 0.03138

	BETA	St. Err. of BETA	B	St. Err. of B	t(233)	p-level
Intercept			-0.0485	0.00855	-5.672	0
LGREV	-0.3617	0.09286	-0.0087	0.00224	-3.8956	0.00013
MDREV	-0.2607	0.09117	-0.0009	0.00032	-2.8595	0.00463
ACCVAR	0.22062	0.06269	0.00337	0.00096	3.51901	0.00052
HPHDGDEV	0.71473	0.0974	0.20526	0.02797	7.33813	0

Figure N7: Regression Summary for Algorithm eLANEX-E-6.

Regression Summary for Dependent Variable: LNMNSQ

$R = 0.68032280$ $R^2 = 0.46283912$ Adjusted $R^2 = 0.45361747$
 $F(4,233) = 50.191$ $p < 0.00000$ Std. Error of estimate: 0.78689

	BETA	St. Err. of BETA	B	St. Err. of B	t(233)	p-level
Intercept			-1.1584	0.21443	-5.402	0
LGREV	-0.4376	0.07932	-0.3101	0.05621	-5.5164	0
MDREV	-0.2882	0.07788	-0.03	0.0081	-3.7007	0.00027
ACCVAR	0.25283	0.05356	0.11347	0.02404	4.72084	0.000004
HPHDGDEV	0.95003	0.0832	8.00829	0.70137	11.4181	0

Figure N8: Regression Summary for Algorithm eLNMNSQ-E-6.

VITA

Mark Gustav Lewin

EDUCATION

- Master of Science: Industrial and Systems Engineering, 1996
(Human Factors Engineering Option)
Virginia Polytechnic Institute and State University
Blacksburg, Virginia
- Bachelor of Science: Industrial and Systems Engineering, 1994 (Cum Laude)
Virginia Polytechnic Institute and State University
Blacksburg, Virginia

TEACHING AND RESEARCH EXPERIENCE

Graduate Research Assistant, May 1995 - September 1996
Vehicle Analysis and Simulation Laboratory
Virginia Polytechnic Institute and State University
Blacksburg, Virginia

Graduate Teaching Assistant, August 1994 - May 1995
Department of Industrial and Systems Engineering
Virginia Polytechnic Institute and State University
Blacksburg, Virginia

ADDITIONAL EXPERIENCE

Industrial Engineering Intern, Summer 1993
Burlington Industries, Menswear Division
Halifax, Virginia

PROFESSIONAL ASSOCIATIONS and HONORS

Human Factors and Ergonomics Society
Institute of Industrial Engineers
Tau Beta Pi, Engineering Honor Society
Alpha Pi Mu, Industrial Engineering Honor Society
Golden Key National Honor Society
Scholarships: National Merit, Robert C. Byrd, John Grado (departmental), others

PUBLICATIONS

Lewin, M.G. and Wierwille, W.W. (1996) Research on Vehicle-Based Driver Status/Performance Monitoring: Final Report, Part III: Development of Revised Algorithms for Driver Status/Performance Monitoring (Report No. ISE 96-08). Blacksburg, VA: Virginia Polytechnic Institute and State University, Department of Industrial and Systems Engineering.

Lewin, M.G., Fairbanks, R.J., and Wierwille, W.W. (1996) Research on Vehicle-Based Driver Status/Performance Monitoring: Final Report, Part II: Simulator Test and Evaluation of a Complete Driver Status/Performance Monitoring System (Report No. ISE 96-07). Blacksburg, VA: Virginia Polytechnic Institute and State University, Department of Industrial and Systems Engineering.

Fairbanks, R.J., Lewin, M.G., and Wierwille, W.W. (1995) Research on Vehicle-Based Driver Status/Performance Monitoring: Eighth Semi-Annual Research Report (Report No. ISE 95-23). Blacksburg, VA: Virginia Polytechnic Institute and State University, Department of Industrial and Systems Engineering.

RELEVANT COURSEWORK

Human Factors Systems Design I and II
Human Factors Research Design I and II
Human Physical Capabilities
Human Information Processing
Human Factors Ethics and Professionalism
Occupational Safety and Hazard Control
Work Measurement and Methods Engineering
Plant Layout and Material Handling
Production Planning and Inventory Control

System Safety Analysis
Training Systems Design
Work Physiology
Industrial Systems Design
Statistics for Engineers
Technical Writing
Statistical Quality Control
Operations Research
Manufacturing Processes

