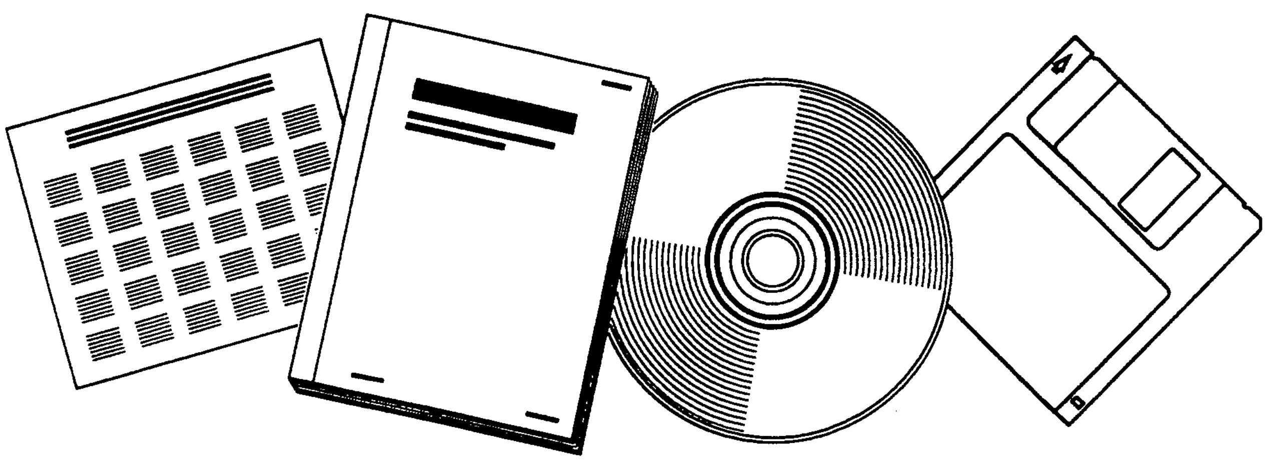
PB98-112329

RESEARCH ON VEHICLE-BASED DRIVER STATUS PERFORMANCE MONITORING, PART II

SEP 96



U.S. Department
of Transportation
National Highway
Traffic Safety
Administration

NHTSA
People Saving People
http://www.nhtsa.dot.gov

DOT HS 808 639 September 1996

Final Report

Research on Vehicle-Based Driver Status/Performance Monitoring, PART II

Technical Report Documentation Page

1. Report No. DOT HS 808 639	2. Government Accession No.	3. Recipients's Catalog No.
4. Title and Subtitle Final Reports: Research on Vehicle-Based Driver Status/Performance Monitoring, Part II		5. Report Date September 1996
		6. Performing Organization Code US DOT/NHTSA/R & D/OCAR/NRD53
7. Author(s) Walter W. Wierwille, Mark G. Lewin, Rollin J. Fairbanks, III		8. Performing Organization Report No.
9. Performing Organization Name and Address Vehicle Analysis & Simulation Laboratory Department of Industrial & Systems Engineering Virginia Polytechnic Institute & State University Blacksburg, VA 24061-0118		10. Work Unit No. (TRAIS)n code
		11. Contract of Grant No. DTNH-22-91-Y-07266
12. Sponsoring Agency Name and Address Office of Crash Avoidance Research, NRD53 National Highway Traffic Safety Administration 400 Seventh Street, S.W. Washington, DC 20590		13. Type of Report and Period Covered NHTSA Contractor Report
		14. Sponsoring Agency Code
15. Supplementary Notes		

16. Abstract

A driver drowsiness detection/alarm/countermeasures system was specified, tested and evaluated, resulting in the development of revised algorithms for the detection of driver drowsiness. Previous algorithms were examined in a test and evaluation study, and were found to be ineffective in detecting drowsiness. These previous algorithms had been developed and validated under simulator conditions that did not emphasize the demand for maintaining the vehicle in the lane as would be expected in normal driving. Revised algorithms were them developed under conditions that encouraged more natural lane-keeping behavior by drivers in the simulator. In these revised algorithms, correlations between dependent drowsiness measures and independent performance-related measures were lower than expected. However, classification accuracy improved when a criterion of "drowsiness *or* performance" was used, with performance assessed directly from a lane-related measure.

17. Key Words Drowsy Driver, Fatigue, Driver Monitoring, Driving Simulation, Vigilance, Driver Impairment, Drowsiness Detection, Fatigue Countermeasures		18. Distribution Statement Document is available to the public through the National Technical Information Service, Springfield, VA 22161	
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No of Pages TBD	22. Price

Form DOT F1700.7 (8-72) Reproduction of completed page authorized

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

EXECUTIVE SUMMARY

This document reports on the test and evaluation of a complete driver status/performance monitoring advisory-alarm-countermeasures system. The system tested used unbaselined drowsy driver detection algorithms and unbaselined driver performance algorithms developed by Fairbanks, Lewin, and Wierwille (1995). These algorithms were developed from previously collected development and validation data and based on previously developed baselined algorithms (Wierwille, Wreggit, Kirn, Ellsworth, and Fairbanks, 1994). The system was equipped to switch between algorithms and continue to function during periods when all parameters were not available. Best estimates and/or measurements of eye closure and lane excursions were collected directly or calculated from independent measures during each minute of each experimental session. The one-minute averages were used to calculate six-minute moving averages, which were compared on-line to optimal threshold values (Fairbanks, Lewin, and Wierwille, 1995). At the end of each minute, detection took place if either the eye closure threshold or the lane excursion threshold (or both) were exceeded.

The system was designed to re-alert drivers using the optimal advisory tone, advisory message, and alarm stimuli determined by Fairbanks, Fahey, and Wierwille (1995). A reset button was used by the driver to respond to the stimuli, signifying that he or she had been successfully re-alerted. For periods of twenty-five minutes after detection, some subjects received one of two drowsiness countermeasures designed to maintain alertness.

Nine sleep-deprived subjects drove an automobile simulator from approximately 12:15 A.M. to 3:00 A.M. Following refinements in the experimental protocol, usable data were obtained from six of these subjects. Each subject received a training session including practice in carefully observing the lane boundaries. A lane-minder device, which produced a warbling tone when the lane boundary was exceeded on either side or the vehicle, was used in the training.

Objective performance measures and subjective ratings were collected during the experimental sessions. Performance measures included all dependent and independent measures included in all driver status and performance algorithms. Drowsiness level and timing of the advisory tone and message were subjectively rated by the subjects following each detection, while drowsiness level and performance level were independently rated by an experimenter.

The results of this experiment were quite different from what was expected. All detections that took place during the experiment were based wholly on decreases in driver performance; no detection was based wholly or partly on eye closure (drowsiness). The performance aspect of the system dominated the detection process. The combination of advisory tone and advisory message were sufficient to alert the driver at every detection; no full auditory or peripheral alarm stimuli were activated during the experiment.

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 current experiment than in the algorithm development experiment. All six status and performance algorithms used in this experiment relied heavily on the affected set of measures.

It can be 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 appears that algorithms developed from that data set do not function well when lane errors are controlled to a level more closely reflecting full scale, as they were in the current experiment.

The advisory tone/advisory message combination and both drowsiness countermeasures appeared to have beneficial effects on both driver status and driver performance. However, insufficient data are available to conduct statistical tests with sufficient power to confirm these effects. Inconsistency of data between the previous experiments and the current experiment confounds comparisons as to whether the alarms and countermeasures affected driver performance.

Because of the unanticipated results, data gathering in this experiment was discontinued. A new experiment was then designed in which detections, alarms, and countermeasures were not used, but drivers were required to perform the lane-keeping task in a manner reflecting performance in an actual vehicle. The emphasis was on development of new algorithms having the highest probability of success when used in field trials. The results of the new experiment are presented in Part III of this final report series.

PRESENT RESEARCH

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 present study 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 present study 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. Fairbanks, Lewin, and Wierwille (1995) showed 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 algorithms developed by Fairbanks, Lewin, and Wierwille (1995) for estimation of the definitional measure PERCLOS were used. 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. (See page 7 for an explanation of the A/O task.) 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 (Fairbanks, Lewin, and Wierwille, 1995) are presented in Appendix A of this report. 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 B of this report. 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 C of this report.

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 pressing 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 D. 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 be 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 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 C.

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 pressing 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 pressing 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 pressing 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 pressing 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 as well. 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 as follows:

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).
- THRSHLD: 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.
- LNERRSQ: 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 utilizing 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 E of this report.

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 F. 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 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 and a statistical package.

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, Kim, 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, Kim, and Wierwille, 1993) with data collected in the present 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 below.

		Subjective Ratings		Experimenter Ratings	
Subject	End of Minute	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

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

Experimenter ratings were measured from the scales in Appendix E 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. (There was no videotape of the roadway.)

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 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 G. Charts comparing LANEX with a composite of the two eLANEX algorithms appear in Appendix H. 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 below. 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). R values 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.

Subject	*Composite ePERCLOS R*	*Algorithm F4a-N R*	*Algorithm D4a-N R*	*Composite eLANEX R*
*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

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

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

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		*	*	*

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	*	*		

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		*	*	

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

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 I. Three ANOVAs were calculated on countermeasure type with PERCLOS (Table I1), LANEX (Table I2), and LANDEV (Table I3) as dependent measures. Countermeasure type was shown to have a significant effect on each of the three dependent variables.

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 0.15. For data taken when no countermeasure was operating, ANOVAs were calculated with PERCLOS (Table I4), LANEX (Table I5), and LANDEV (Table I6) as dependent measures. For data taken when the A/O task was operating, ANOVAs were also calculated with PERCLOS (Table I7), LANEX (Table I8), and LANDEV (Table I9) 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 G. 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 G and H 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 G and H 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 I show that the countermeasures had a significant alerting effect on subjects. For cases in which the no countermeasure was operating, and those 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 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.

REFERENCES

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.

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.

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.

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.

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.

APPENDIX A

Regression Summaries and Classification Matrices for Unbaselined PERCLOS Algorithms used in Study

(Fairbanks, Lewin, and Wierwille, 1995)

Unbaselined Regression Summary for Dependent Variable: PERCLOS
$R = 0.78785881$ $R^2 = 0.62072151$ Adjusted $R^2 = 0.61427120$
$F(5,294) = 96.231$ $p < 0.0000$ 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		
	Group	% Correct	Awake	Questionable	Drowsy
Original Observed	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		
	Group	% Correct	Awake	Questionable	Drowsy
New Observed	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 A1: 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 $R^2 = 0.73449877$ Adjusted $R^2 = 0.72906189$
$F(6,293) = 135.10$ $p < 0.0000$ 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

	Group	% Correct	Awake	Questionable	Drowsy
Original Observed	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

	Group	% Correct	Awake	Questionable	Drowsy
New Observed	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 A2: 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
$R = 0.84140802$ $R^2 = 0.70796745$ Adjusted $R^2 = 0.69567134$
$F(4,95) = 57.577$ $p < 0.00000$ 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		
	Group	% Correct	Awake	Questionable	Drowsy
Original Observed	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		
	Group	% Correct	Awake	Questionable	Drowsy
New Observed	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 A3: 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: PERCLOS
$R = 0.87153552$ $R^2 = 0.75957416$ Adjusted $R^2 = 0.74406281$
$F(6,93) = 48.969$ $p < 0.00000$ 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

	Group	% Correct	Awake	Questionable	Drowsy
Original	Awake	95.65	**66**	1	2
Observed	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

	Group	% Correct	Awake	Questionable	Drowsy
New	Awake	91.55	**130**	11	1
Observed	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 A4: Unbaselined Regression Summary and Classification Matrices Showing Accuracy of Algorithm L3a-N When Applied to Original Data and New Data.

APPENDIX B

Regression Summaries and Classification Matrices for Unbaselined LANEX Algorithms used in Study

(Fairbanks, Lewin, and Wierwille, 1995)

Regression Summary for Dependent Variable: LANEX
$R = 0.83026466$ $R^2 = 0.68933941$ Adjusted $R^2 = 0.68677620$
$F(5,606) = 268.94$ $p < 0.0000$ 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

	Group	% Correct	Satisfactory	Borderline	Unsatisfactory
	Satisfactory	65.41	**174**	91	1
Observed	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 B1: 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 R^2 = 0.69307499 Adjusted R^2 = 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		
	Group	% Correct	Satisfactory	Borderline	Unsatisfactory
	Satisfactory	78.23	**97**	27	0
Observed	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 B2: Regression Summary and Classification Matrix Showing Accuracy of Algorithm LDV-4 When Applied to Combined Data Set (24 Subjects).

APPENDIX C

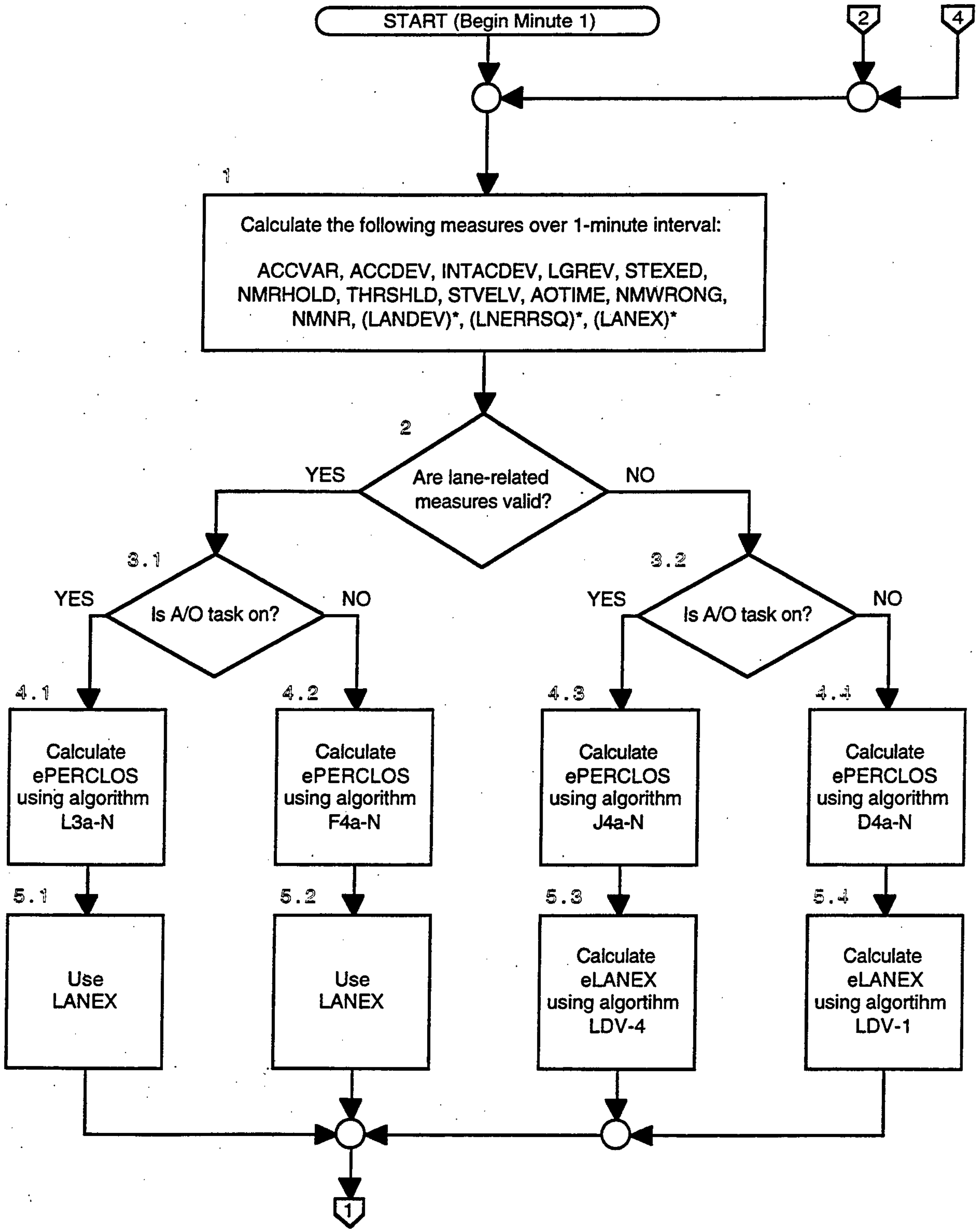
Logic Diagram Representation of Test and Evaluation System Specifications

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, THRSHLD, STVELV) and lateral-accelerometer-related measures (ACCVAR, ACCDEV, INTACDEV) will be calculated (Figure C1, 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 C1, Cell 2), and the status of the A/O task (Figure C1, 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 C1. Logic Diagram for Detection System Computational Procedure

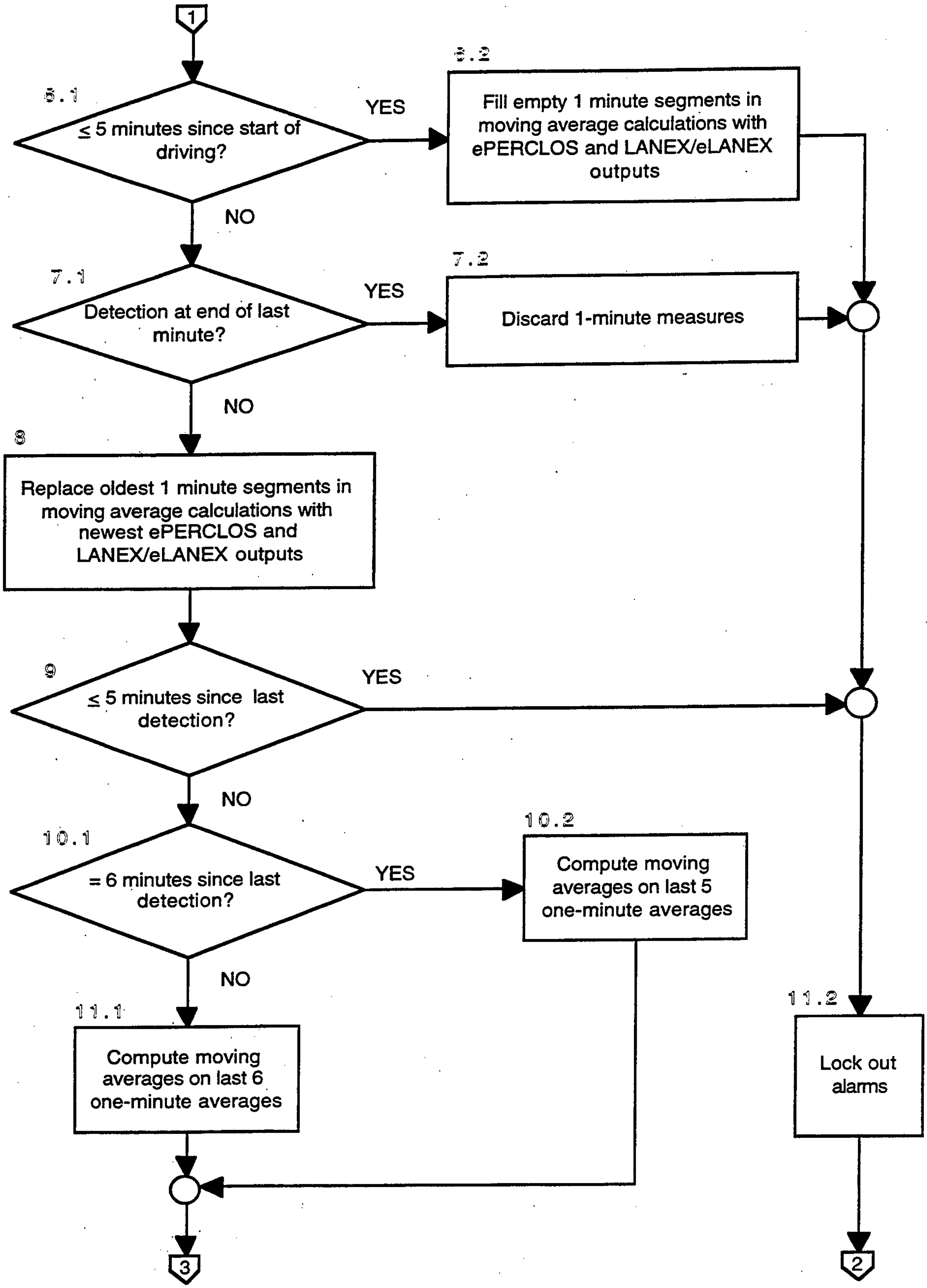


Figure C2. Logic Diagram for Detection System Computational Procedure

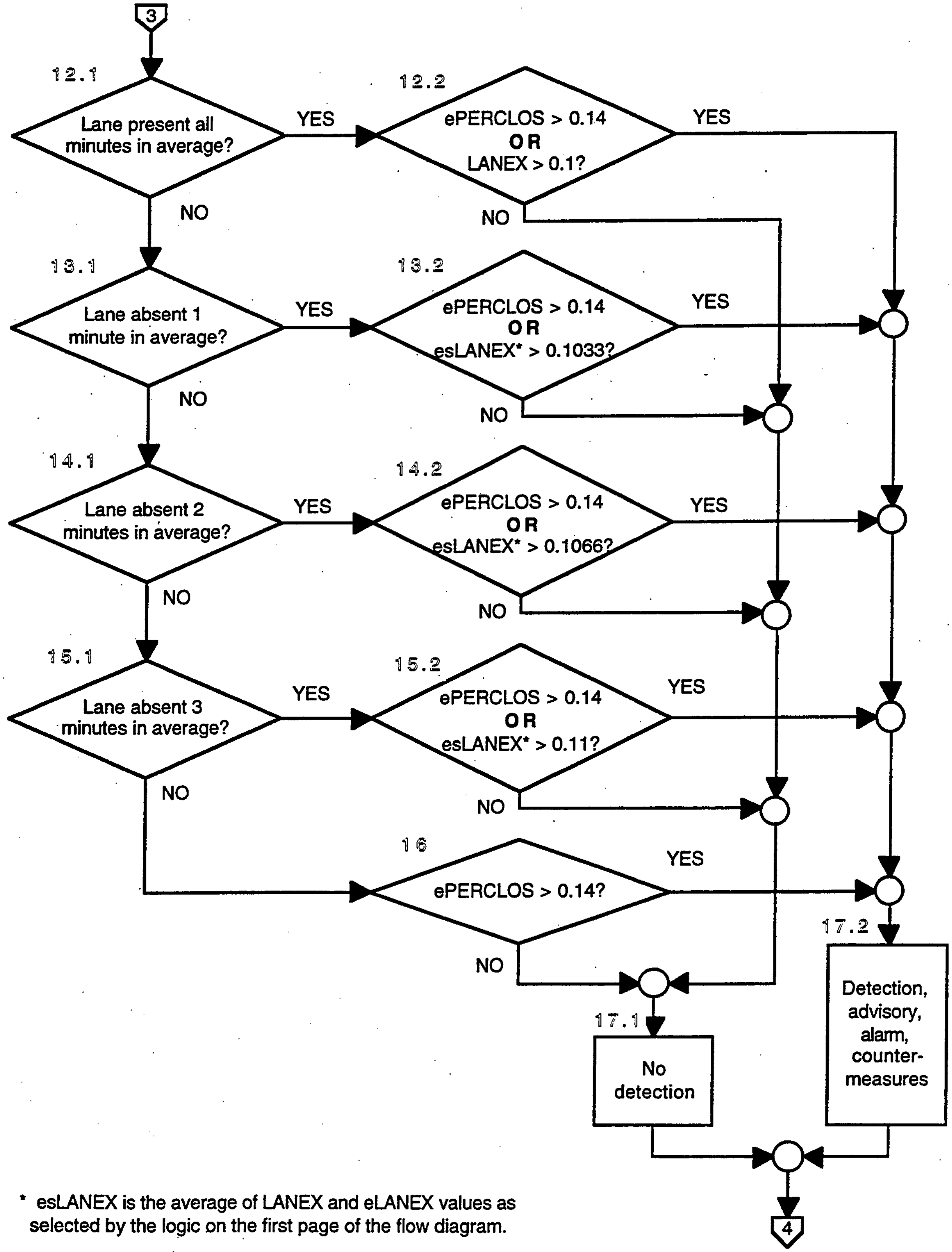
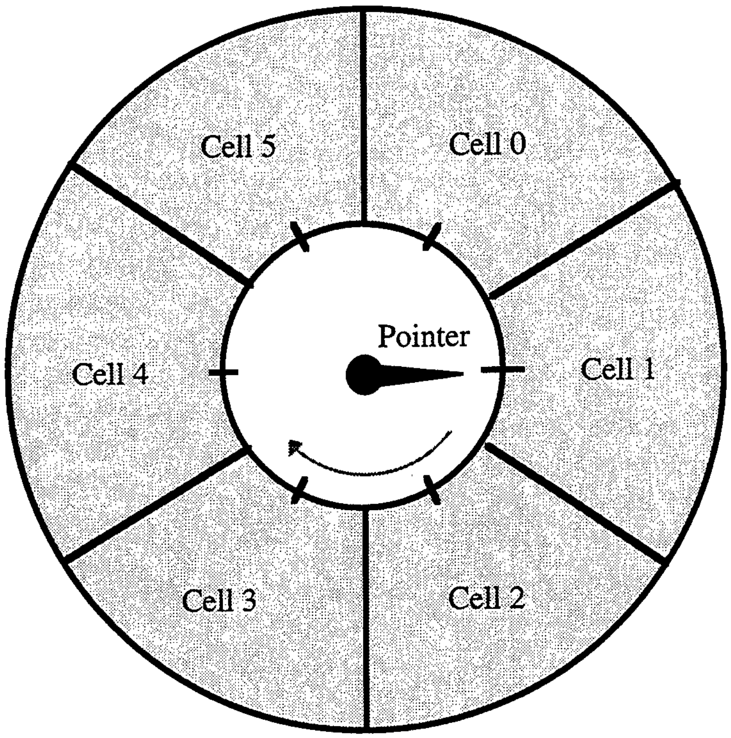


Figure C3. Logic Diagram for Detection System Computational Procedure

- If lane-related measures are *valid* and the A/O task is *on*, ePERCLOS will be calculated using algorithm L3a-N (Figure C1, Cell 4.1) and LANEX will be used (Figure C1, 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 C1, Cell 4.2) and LANEX will be used (Figure C1, 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 C1, Cell 4.3) and eLANEX will be calculated using algorithm LDV-4 (Figure C1, 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 C1, Cell 4.4) and eLANEX will be calculated using algorithm LDV-4 (Figure C1, 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 C2, 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 C4).

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 C2, 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 C2, 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 alarms will be locked out for that minute, since there will be insufficient data to make a detection (Figure C2, Cell 11.2). Also, since no detection can be made, the program will



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 C4. Procedure for Calculation of Moving Averages 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 C2, 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 C2, 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 C2, 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 C2, Cell 10.1). (esLANEX is the average of LANEX and eLANEX values as selected by the logic in Figure C1.) However, since the outputs for the first minute after detection were discarded (Figure C2, Cell 7.2), the moving averages will be five-minute averages (Figure C2, 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 C2, 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 C3, 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 C3, Cells 12.2, 17.2). If neither average exceeds threshold, there will not be a detection (Figure C3, Cell 17.1).

If lane is absent for exactly one minute in the moving averages (Figure C3, Cell 13.1), there will be a detection if either the ePERCLOS average exceeds 0.14 or the esLANEX average exceeds 0.1033 (Figure C3, Cells 13.2, 17.2). If lane is absent for exactly two minutes in the moving averages (Figure C3, Cell 14.1), there will be a detection if either the ePERCLOS average exceeds 0.14 or the esLANEX average exceeds 0.1066 (Figure C3, Cells 14.2, 17.2). If lane is absent for exactly three minutes in the moving averages (Figure C3, Cell 15.1), there will be a detection if either the ePERCLOS average exceeds 0.14 or the esLANEX average exceeds 0.11 (Figure C3, Cells 15.2, 17.2). In each of the three cases, there will not be a detection if neither average exceeds threshold (Figure C3, 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 C3, Cells 16, 17.2).

If a detection occurs, the sequence of advisory tone, alarm, and countermeasures (if applicable) will be initiated (Figure C3, 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 D

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?

never almost never occasionally moderately 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 E

Subject and Experimenter Rating Scales Used In Experiment

Subjective Ratings - Test and Evaluation Experiment

How drowsy were you when the tone and advisory message began?	Not Drowsy
	Slightly Drowsy
	Moderately Drowsy
	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 E1: 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?*

Not Drowsy	Slightly Drowsy	Moderately Drowsy	Very Drowsy	Extremely Drowsy

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

Poor	Fair	Moderate	Good	Excellent

Figure E2: Subjective Rating Scale used by Experimenter at each Detection during Test and Evaluation Experiment.

APPENDIX F

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.

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 G

Line Charts Comparing Measured PERCLOS with ePERCLOS Algorithm Outputs

Figure G1: All Subjects: PERCLOS vs. ePERCLOS (straight 6-min avgs, minutes following detection removed)
R = 0.48391

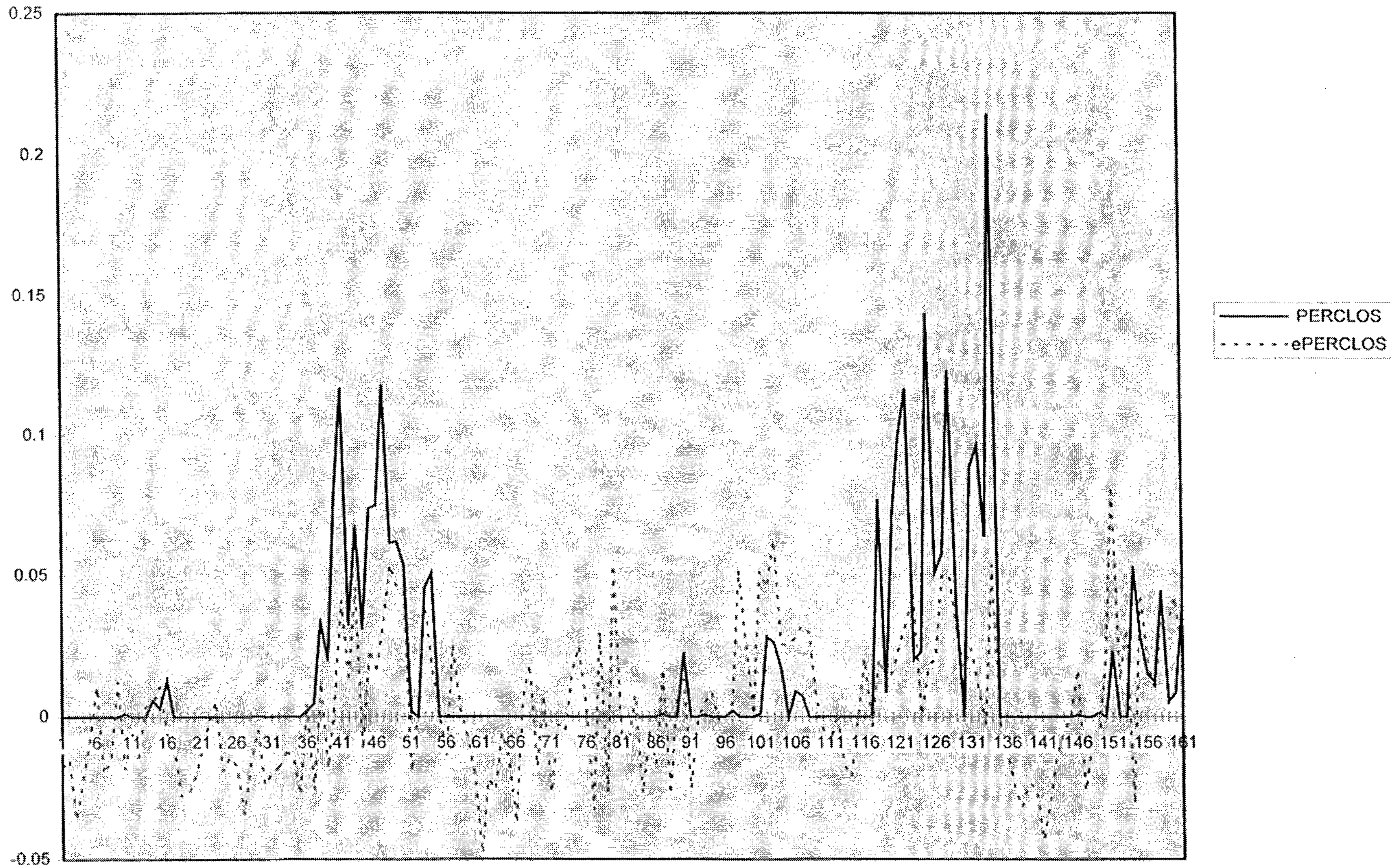


Figure G2: All Subjects: Measured PERCLOS vs. Algorithm F4a-N (straight 6-min avgs, minutes following detection removed)
R = 0.54350

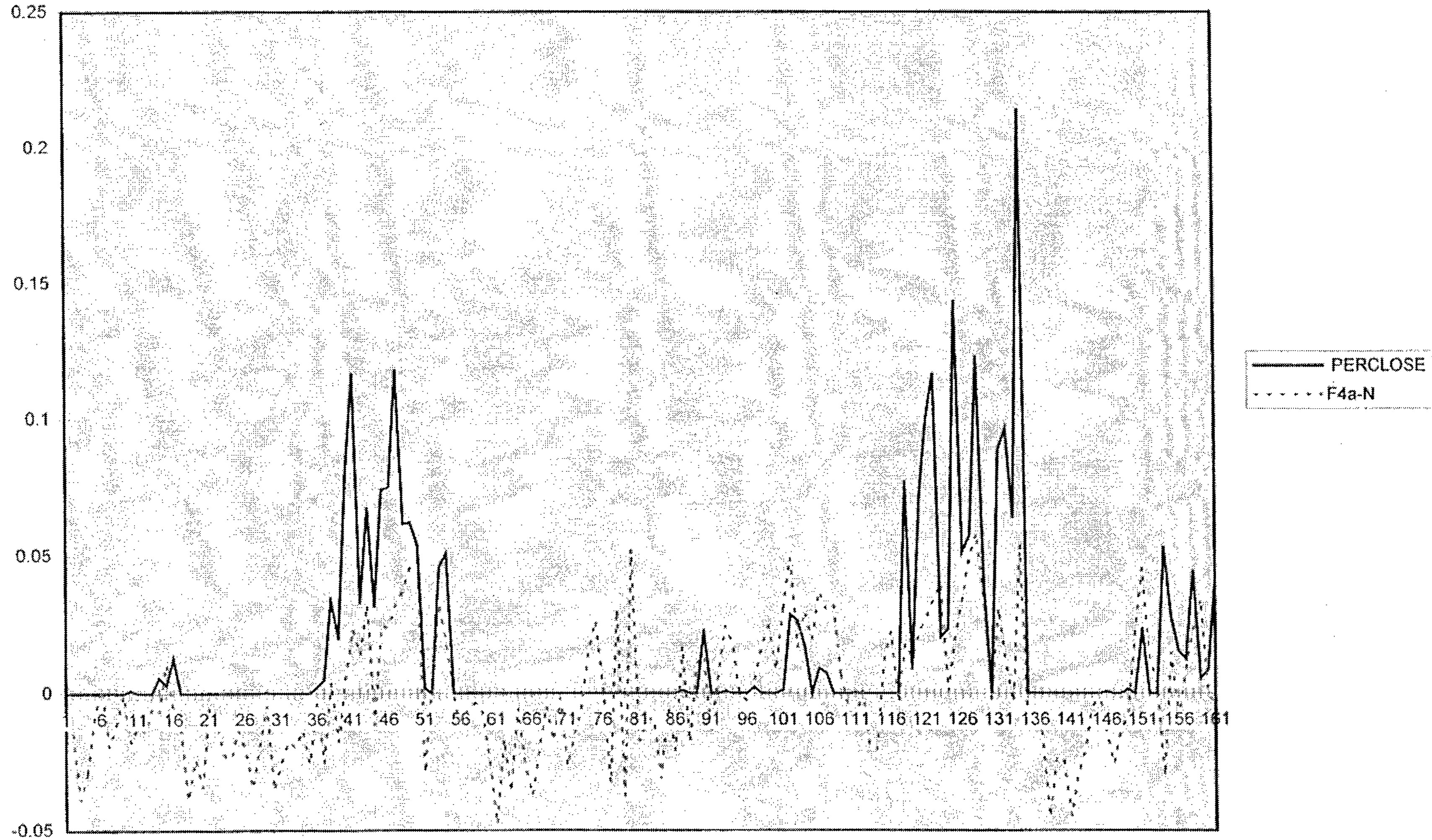


Figure G3: All Subjects: Measured PERCLOS vs. Algorithm D4a-N (straight 6-min avgs, minutes following detection removed)
R = 0.21614

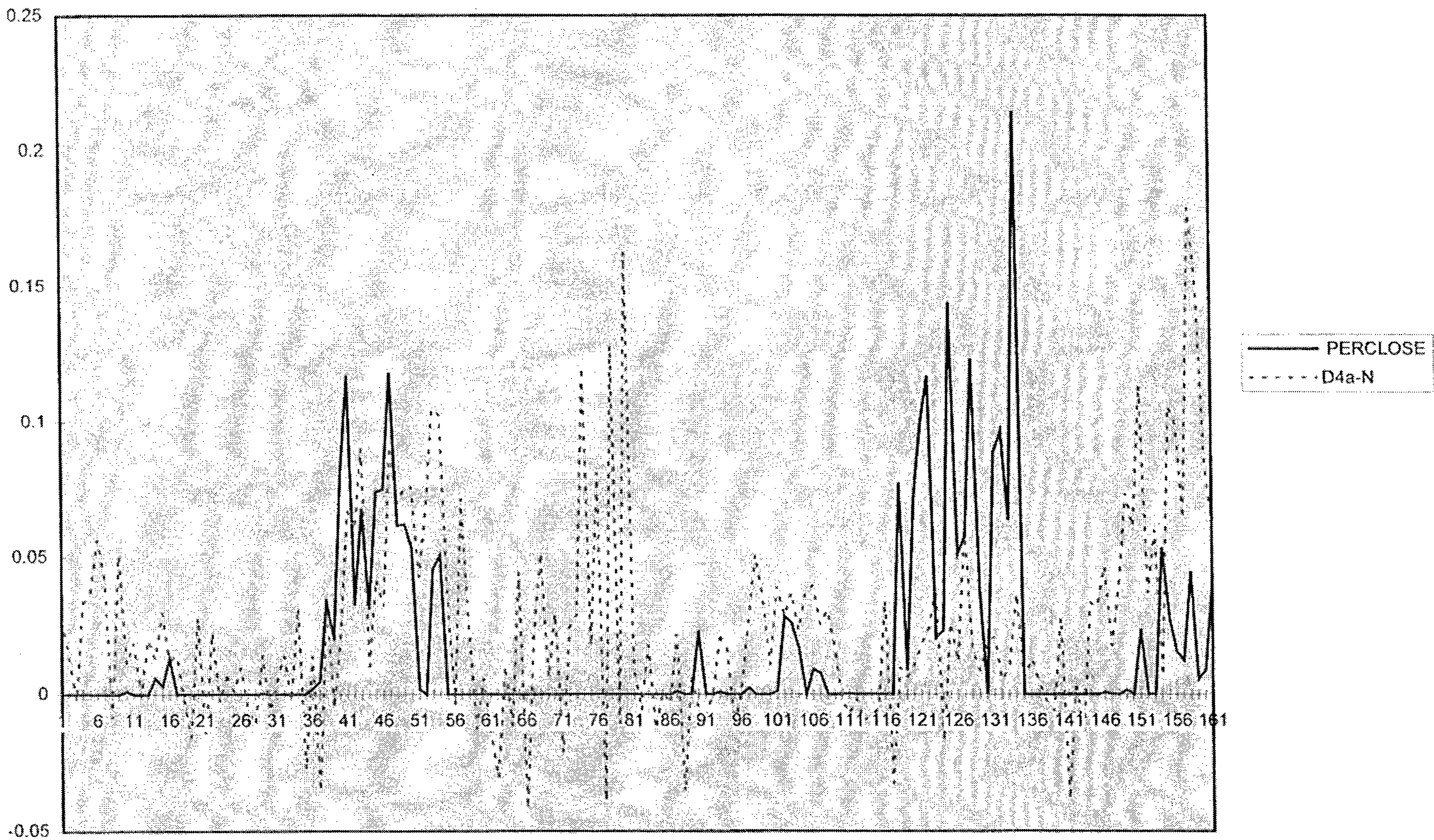


Figure G4: Subject #4: PERCLOS vs. ePERCLOS (straight 6-min avgs, minutes following detection removed)
R = 0.2395

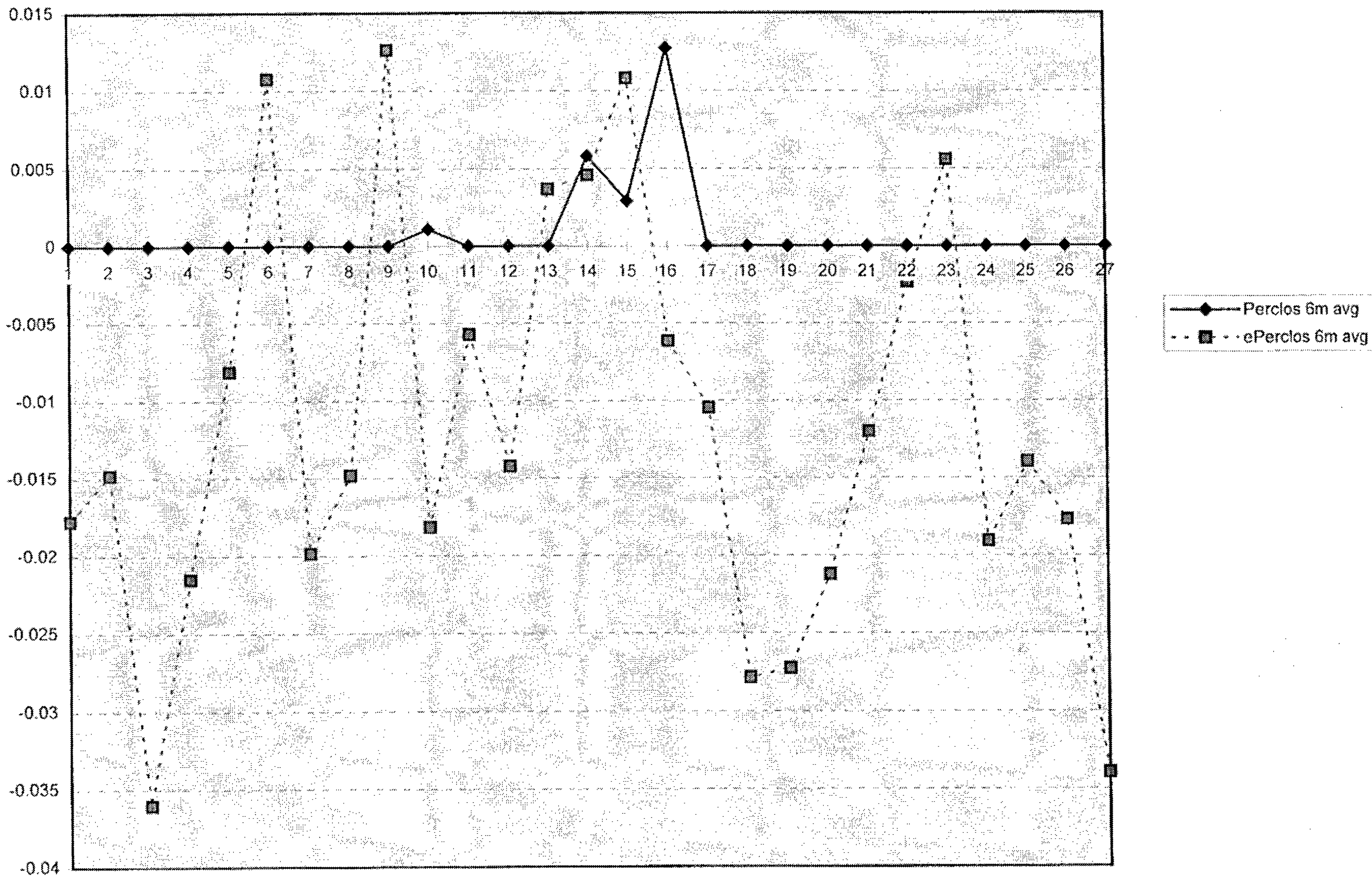


Figure G5: Subject #4: Measured PERCLOS vs. Algs F4a-N and D4a-N (straight 6-min avgs, minutes after detection removed)

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

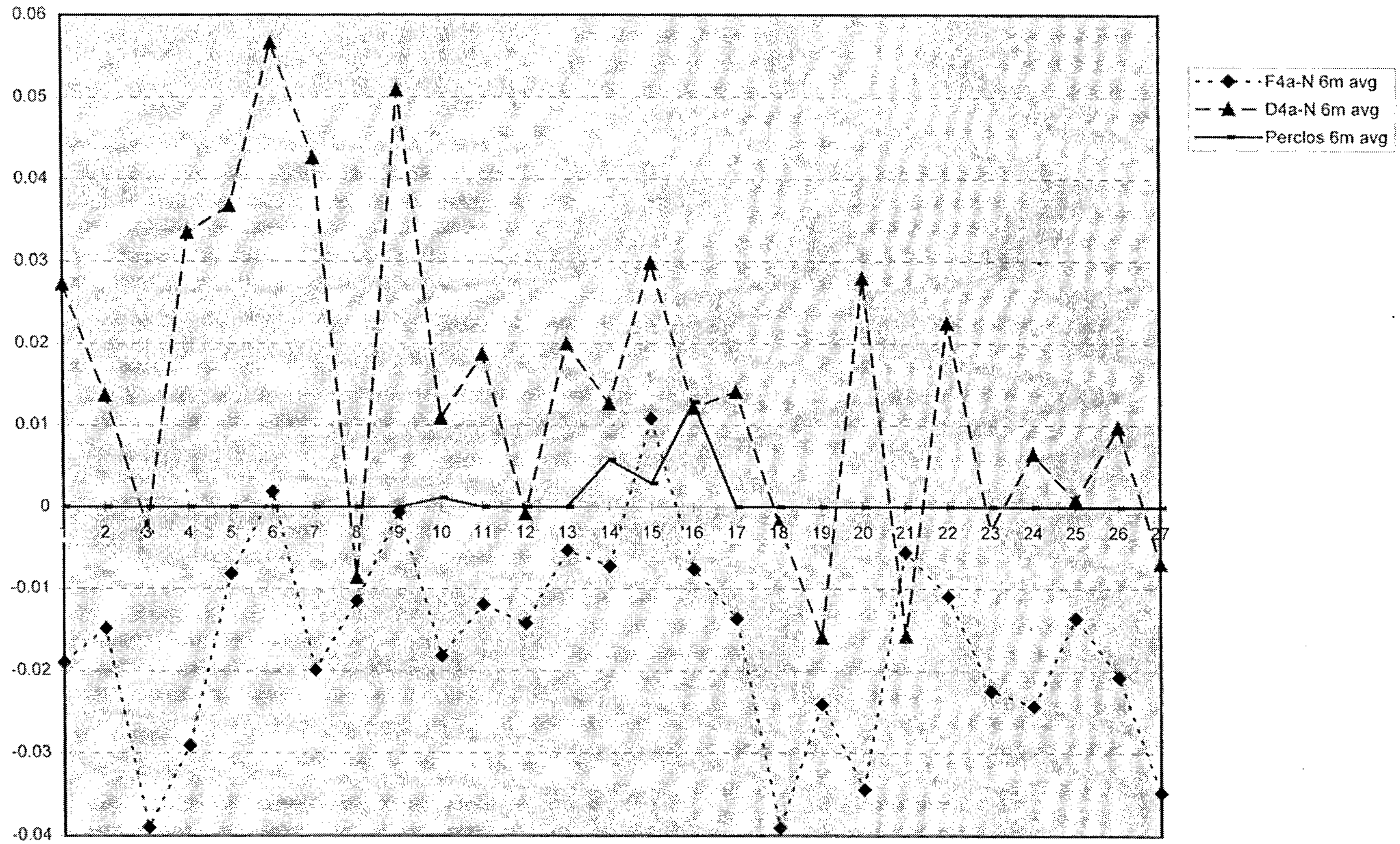


Figure G6: Subject #5: PERCLOS vs. ePERCLOS (Straight 6-min avgs, minutes following detection removed)
R = 0.7843

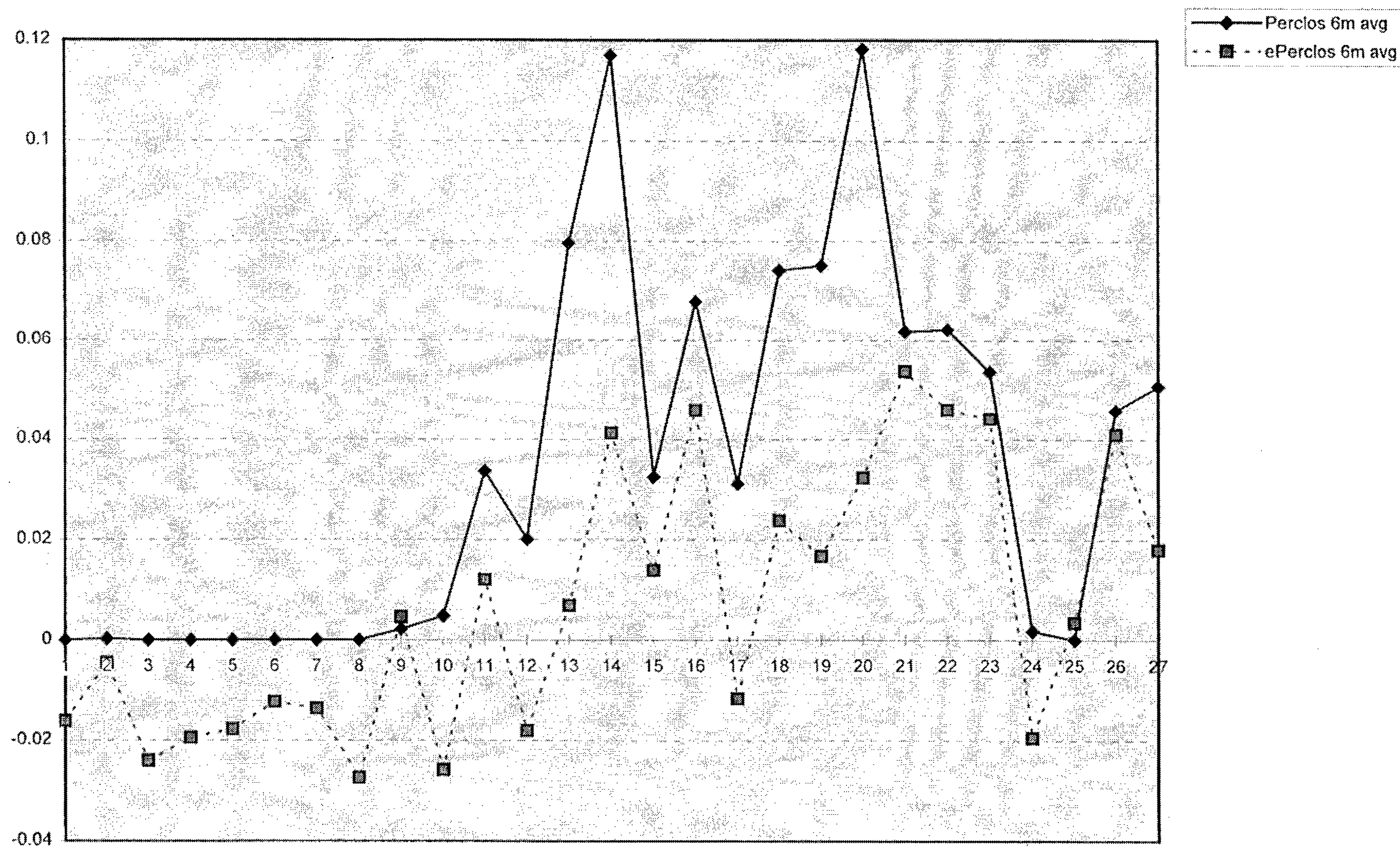


Figure G7: Subject #5: Measured PERCLOS vs. Alg D4a-N and Alg F4a-N (straight 6-min avgs, minutes after detection removed)

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

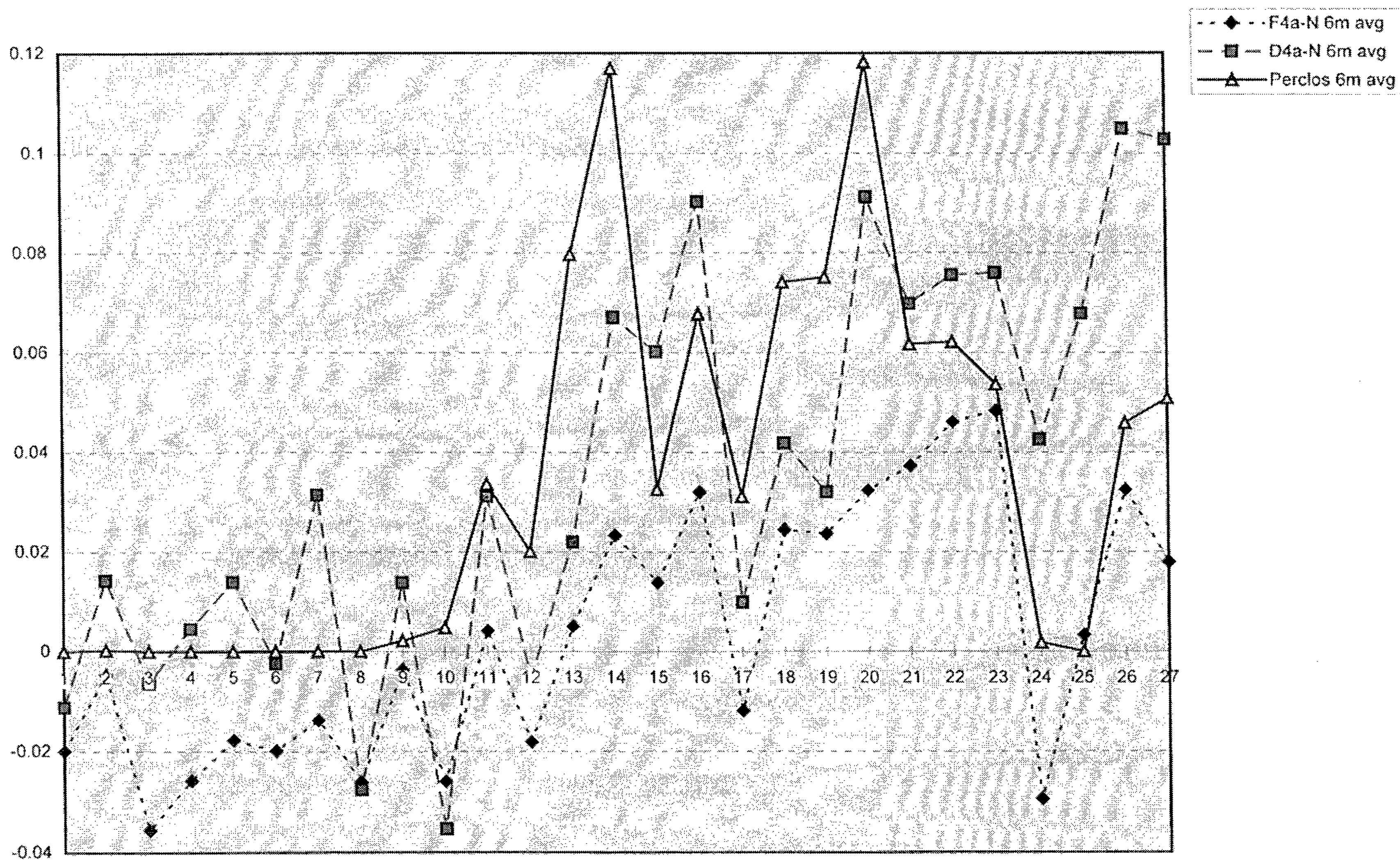


Figure G8: Subject #6: PERCLOS vs. ePERCLOS (straight 6-min avgs, minutes following detection removed)

R value not applicable

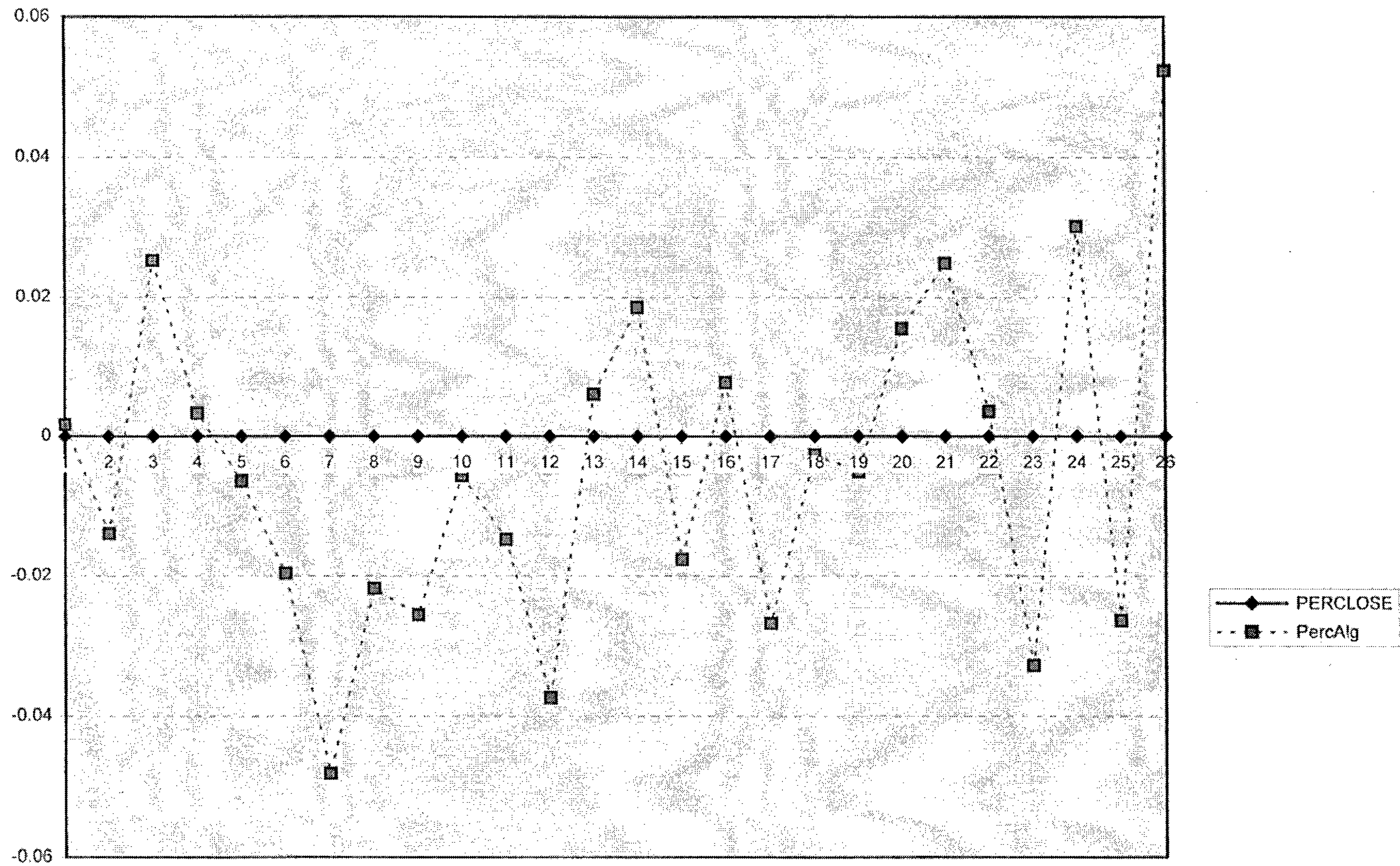


Figure G9: Subject #6: Measured PERCLOS vs AlgD4a-N and Alg F4a-N (straight 6-min avgs, minutes after detection removed)

R values not applicable

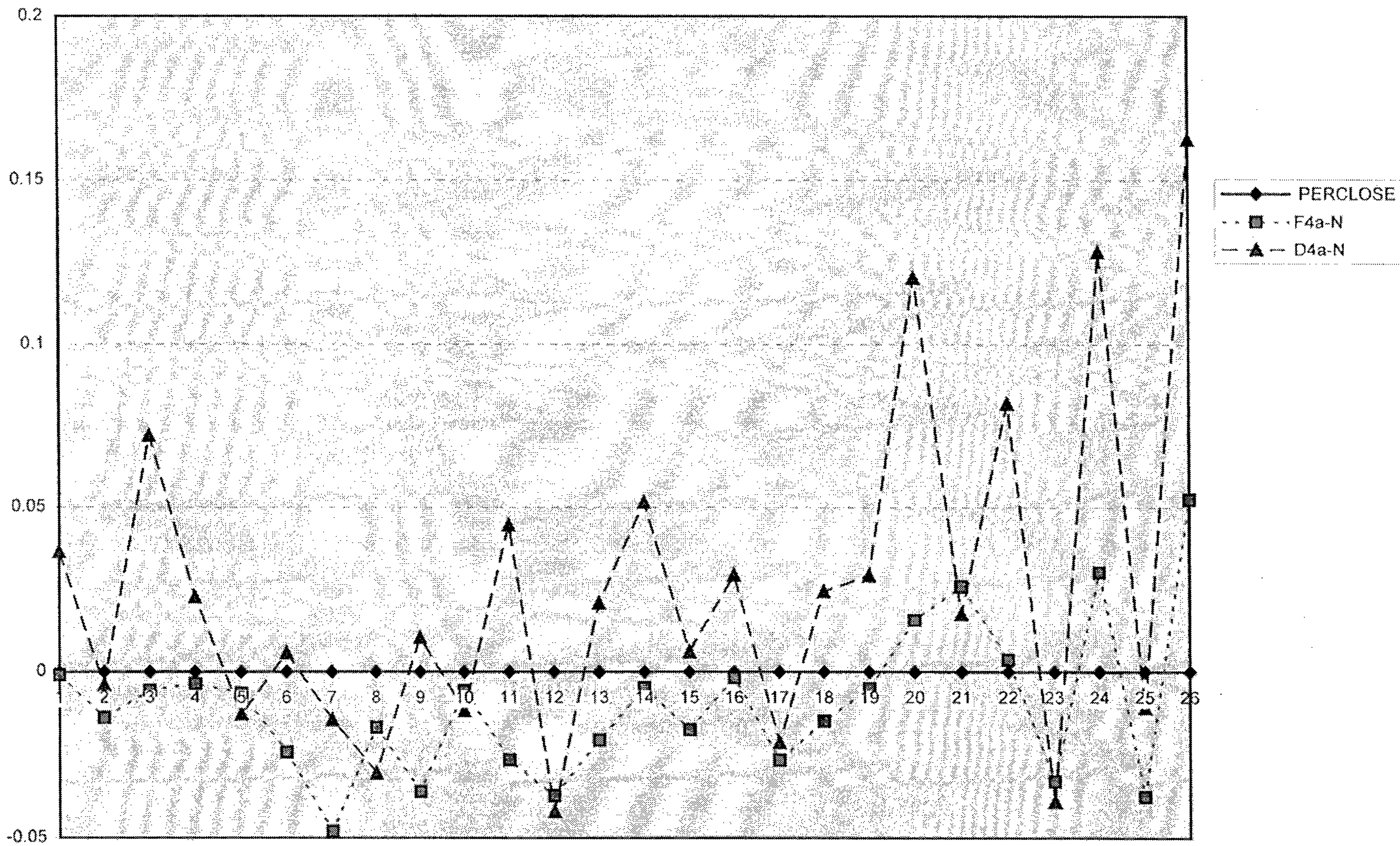


Figure G10: Subject #7: PERCLOS vs. ePERCLOS (straight 6-min avgs, minutes following detection removed)

R = 0.5549

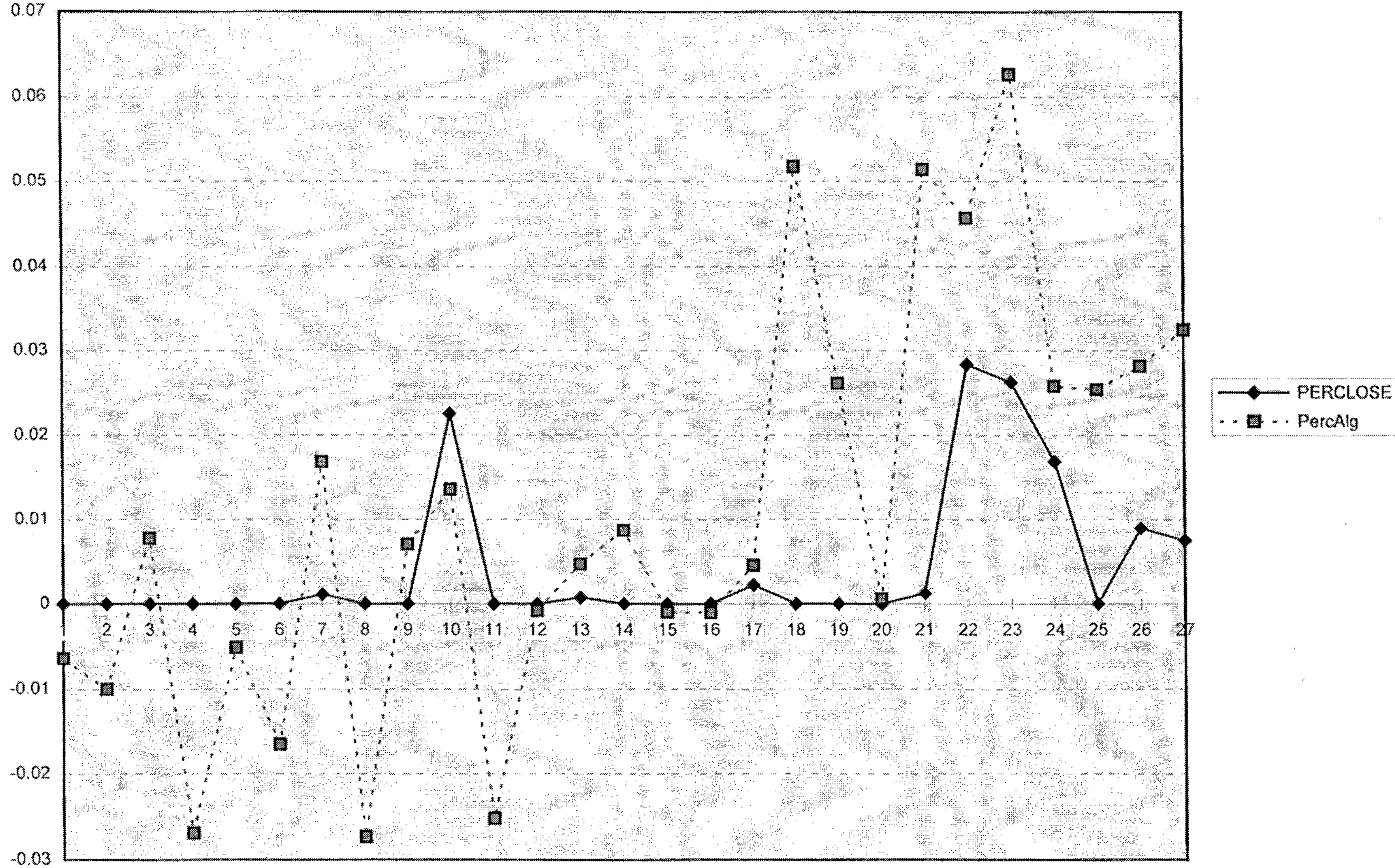


Figure G11: Subject #7: Measured PERCLOS vs. Algs F4a-N and D4a-N (straight 6-min avgs, minutes after detection removed)

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

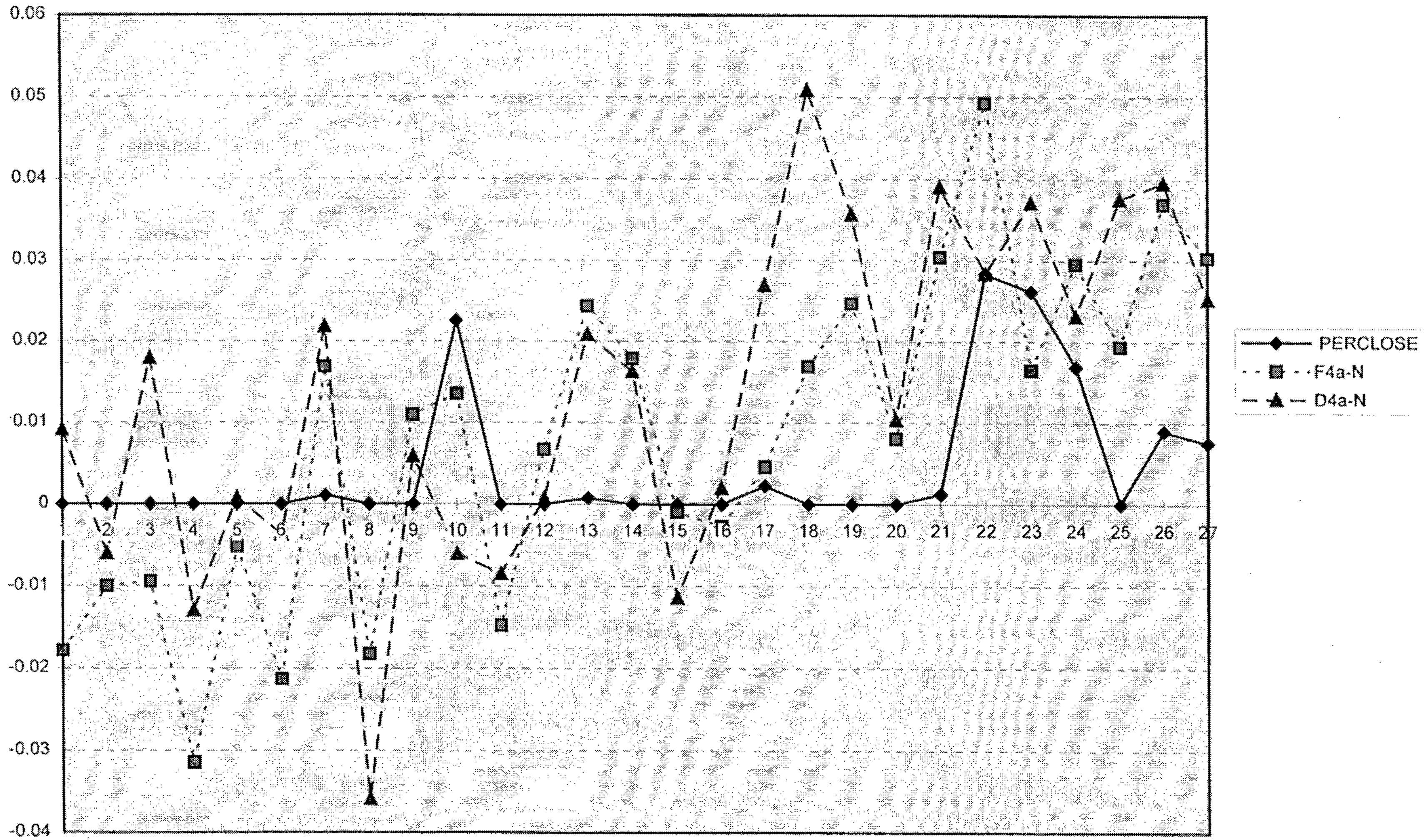


Figure G12: Subject #8: PERCLOS vs. ePERCLOS (straight 6-min avgs, minutes following detection removed)
R = 0.6052

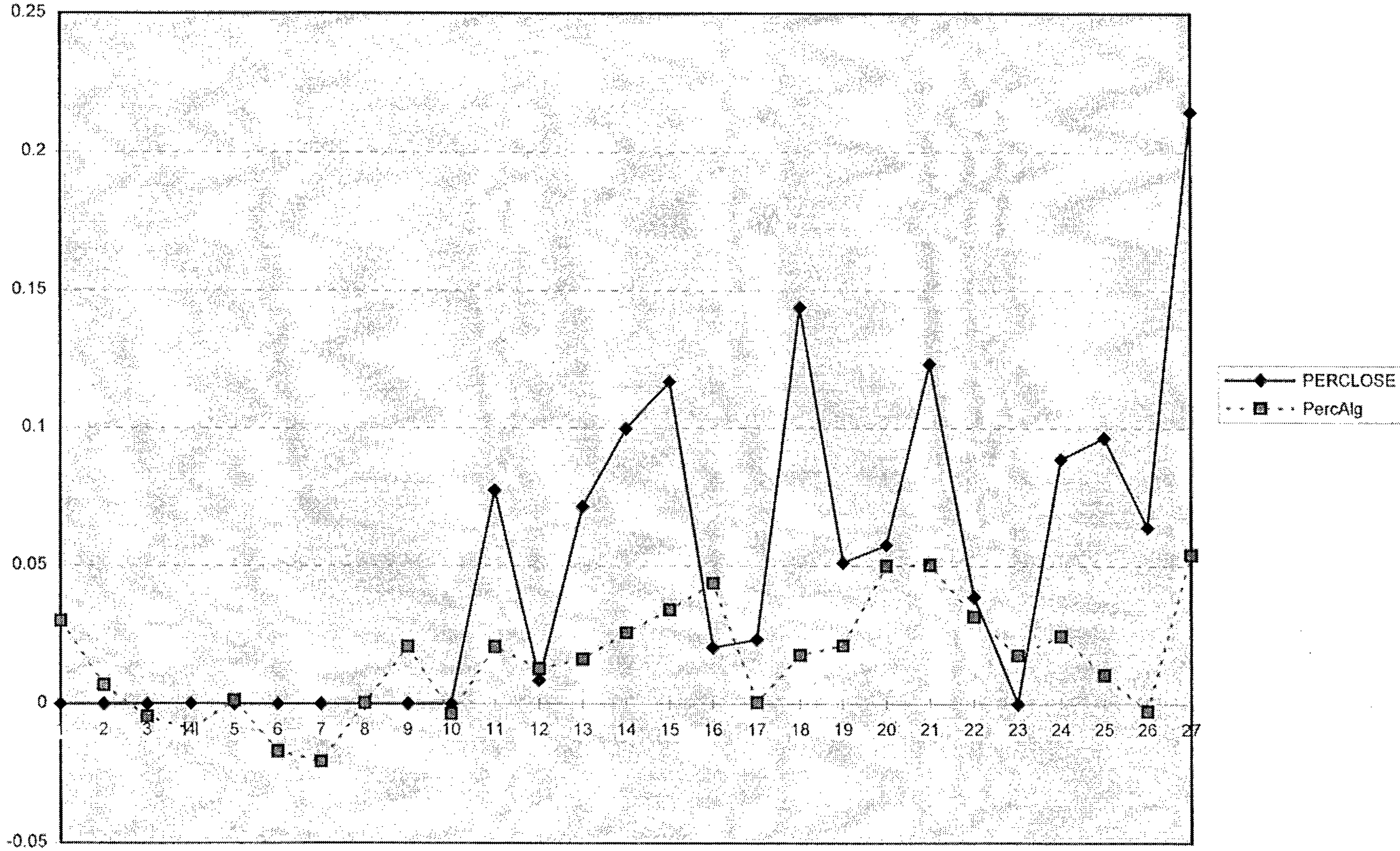


Figure G13: Subject #8: Measured PERCLOS vs. Algs F4a-N and D4a-N (straight 6-min avgs, minutes following detection removed)

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

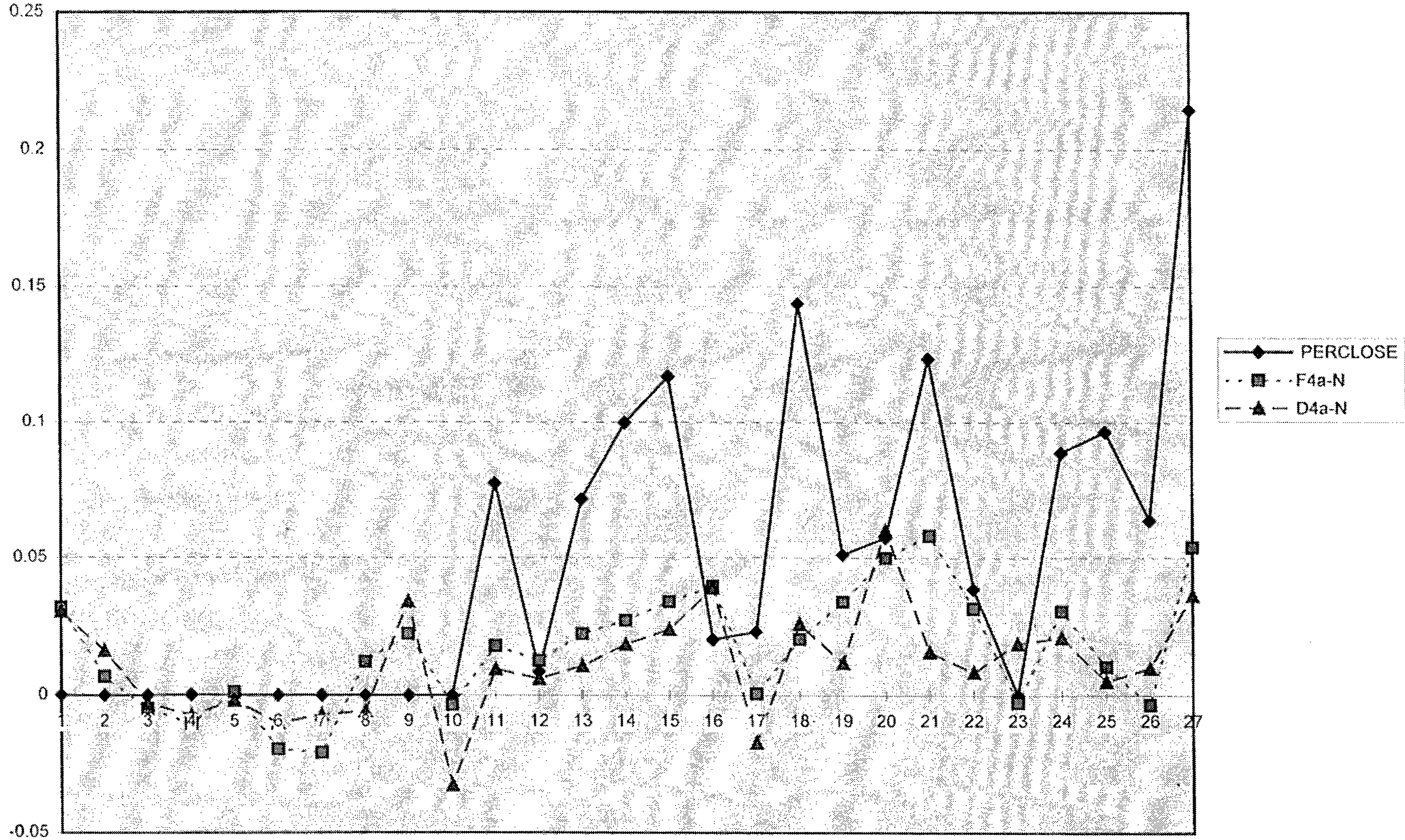


Figure G14: Subject #9: PERCLOS vs. ePERCLOS (straight 6-min avgs, minutes following detection removed)
R = 0.3356

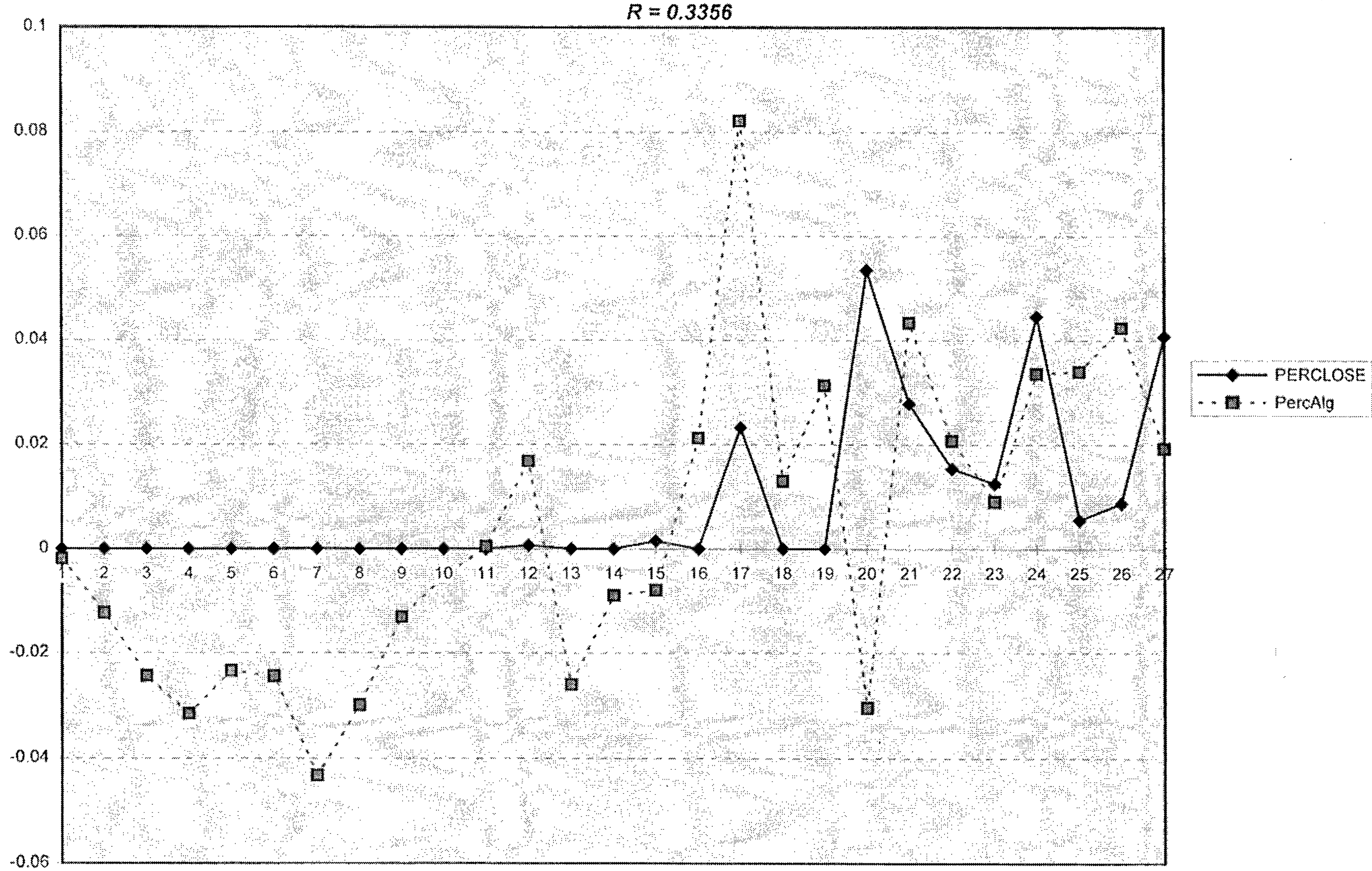
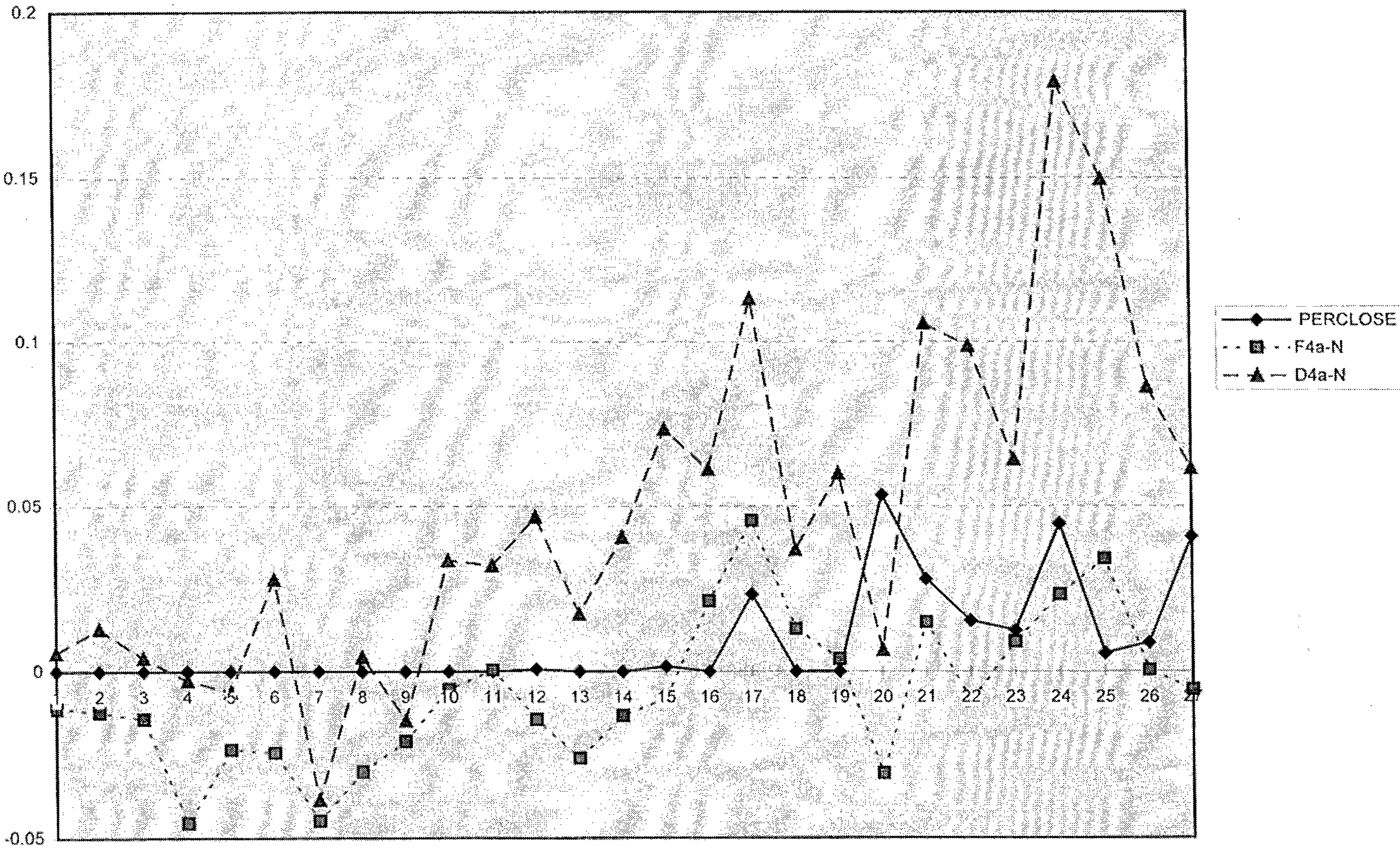


Figure G15: Subject #9: Measured PERCLOS vs. Algs F4a-N and D4a-N (straight 6-min avgs, minutes following detection removed)
F4a-N: R = 0.2558, D4a-N: R = 0.4599



APPENDIX H

Line Charts Comparing Measured LANEX with eLANEX Algorithm Outputs

Figure H1: All Subjects: LANEX vs. eLANEX (straight 6-min avgs, minutes following detection removed)

R = 0.44081

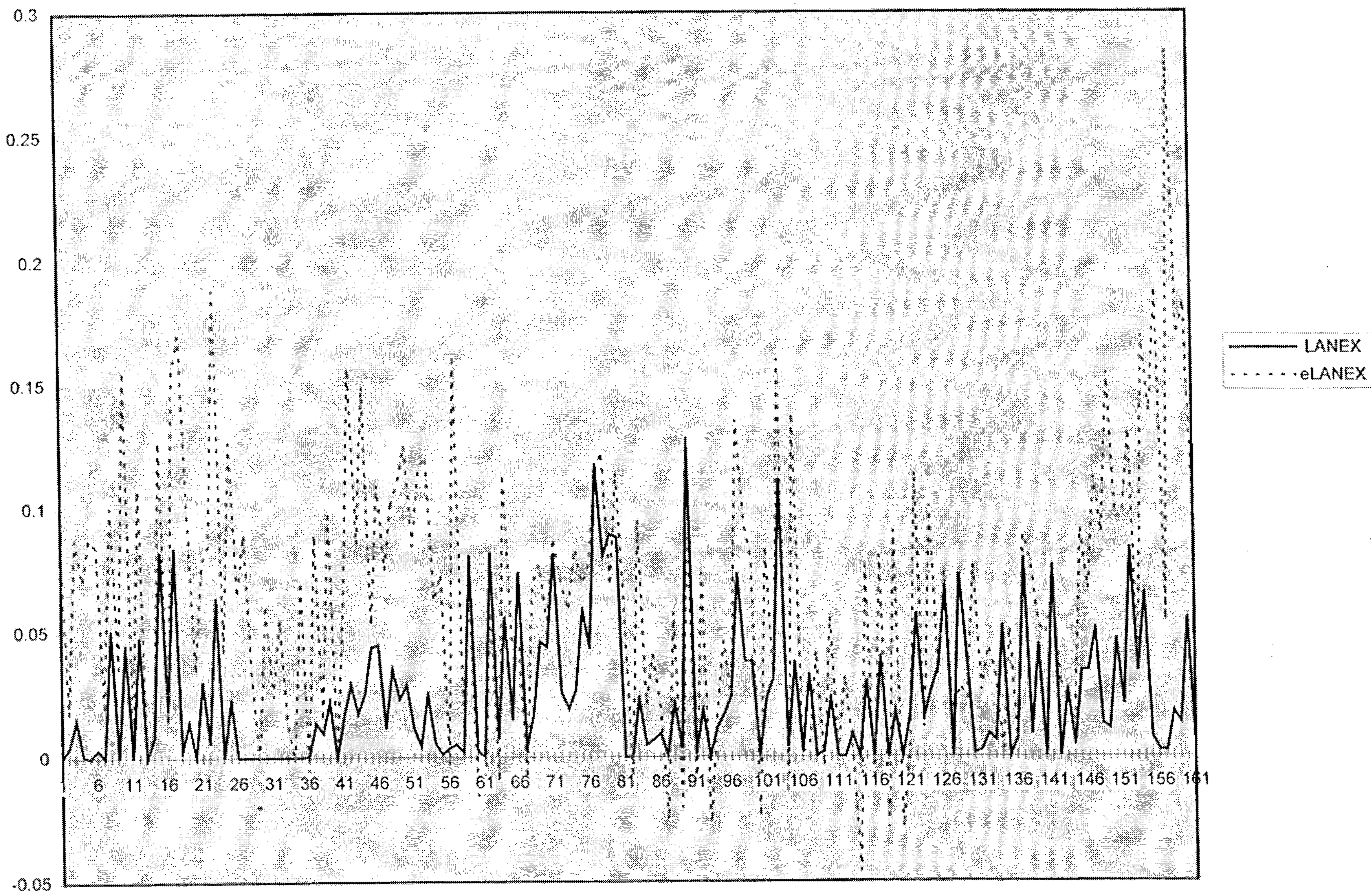


Figure H2: Subject #4: LANEX vs. eLANEX (straight 6-min avgs, minutes following detection removed)
R = 0.6493

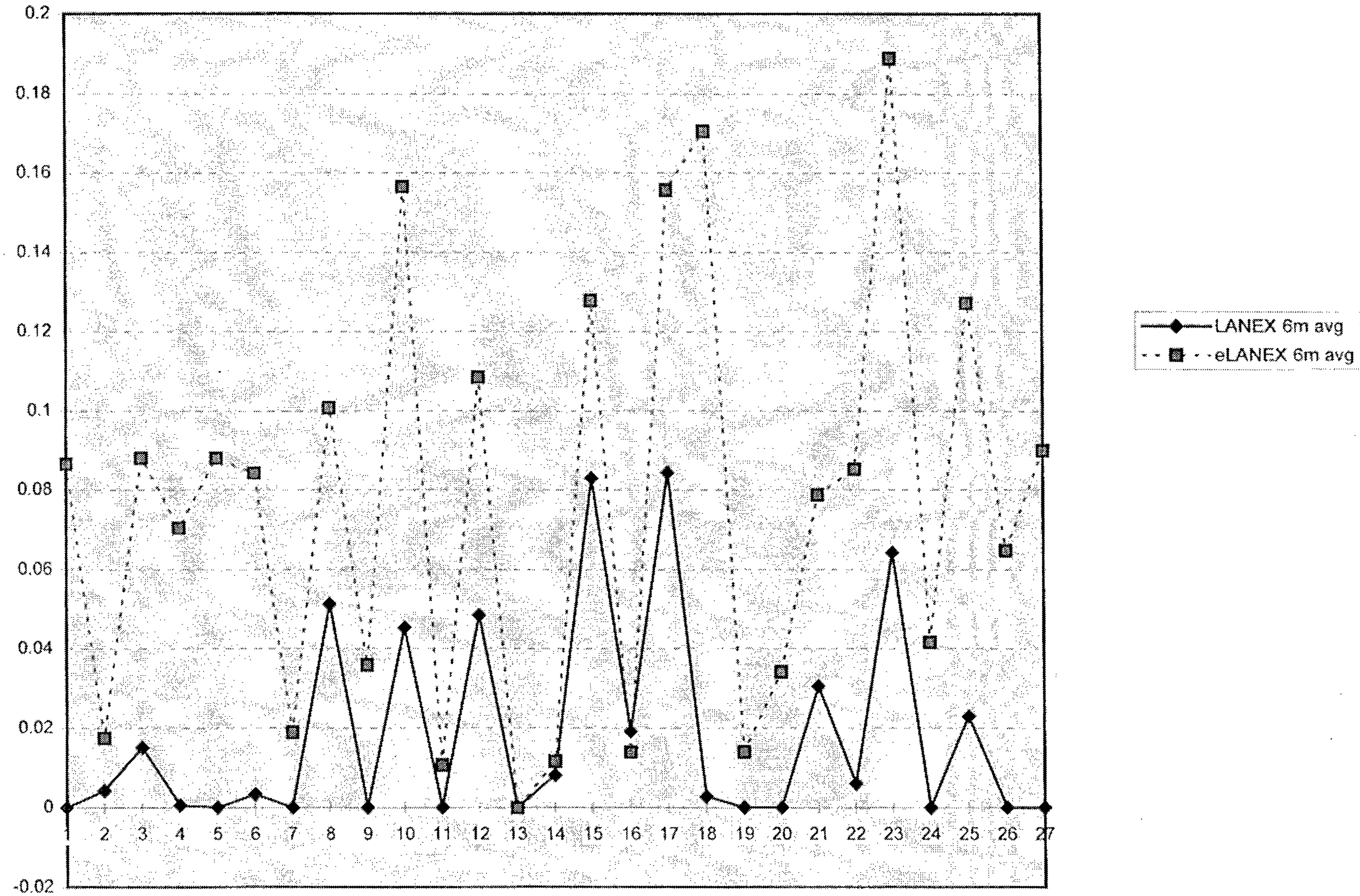


Figure H3: Subject # 5: LANEX vs. eLANEX (straight 6-min avgs, minutes following detection removed)
R = 0.6648

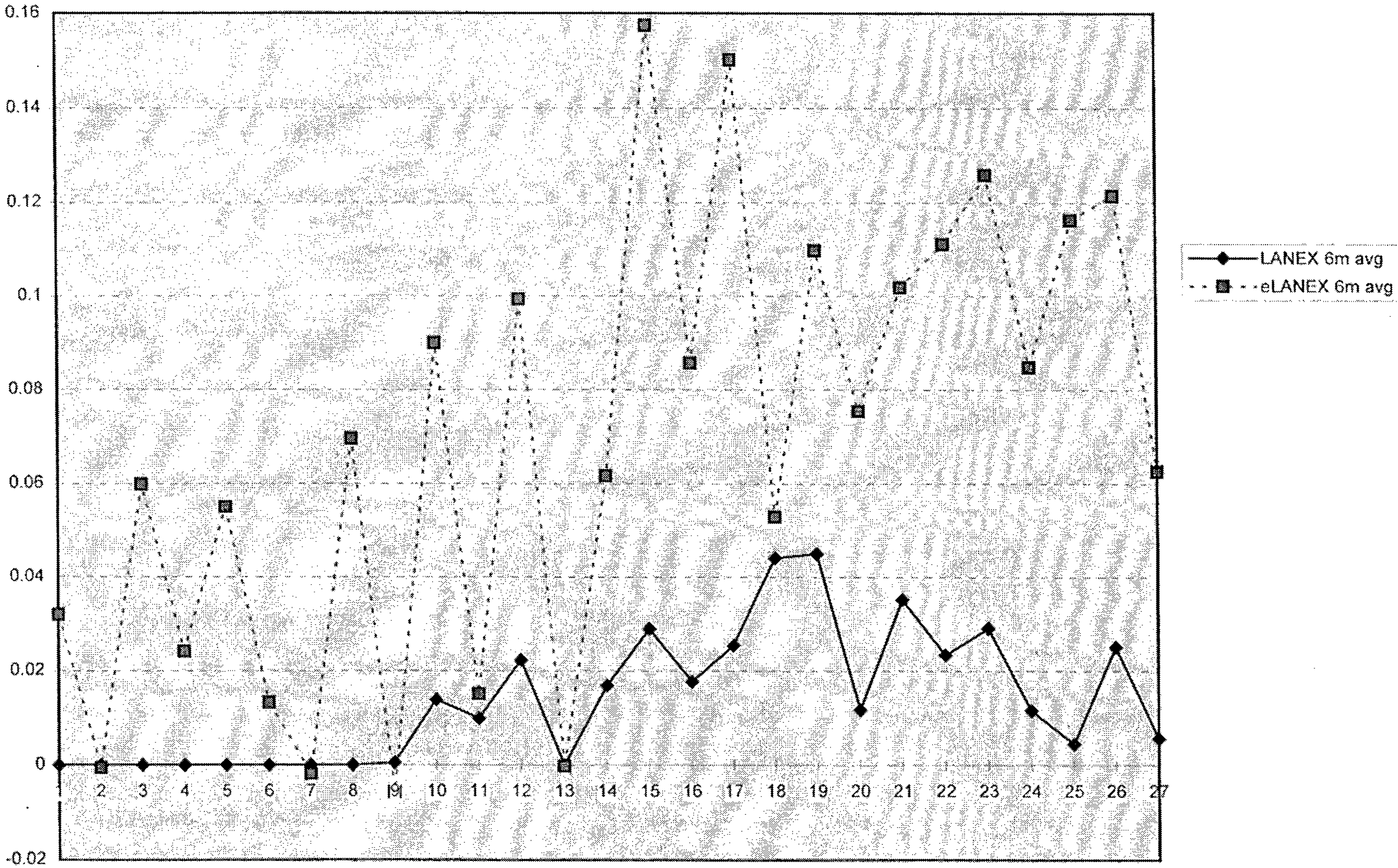


Figure H4: Subject #6: LANEX vs. eLANEX (straight 6-min avgs, minutes following detection removed)
R = 0.5225

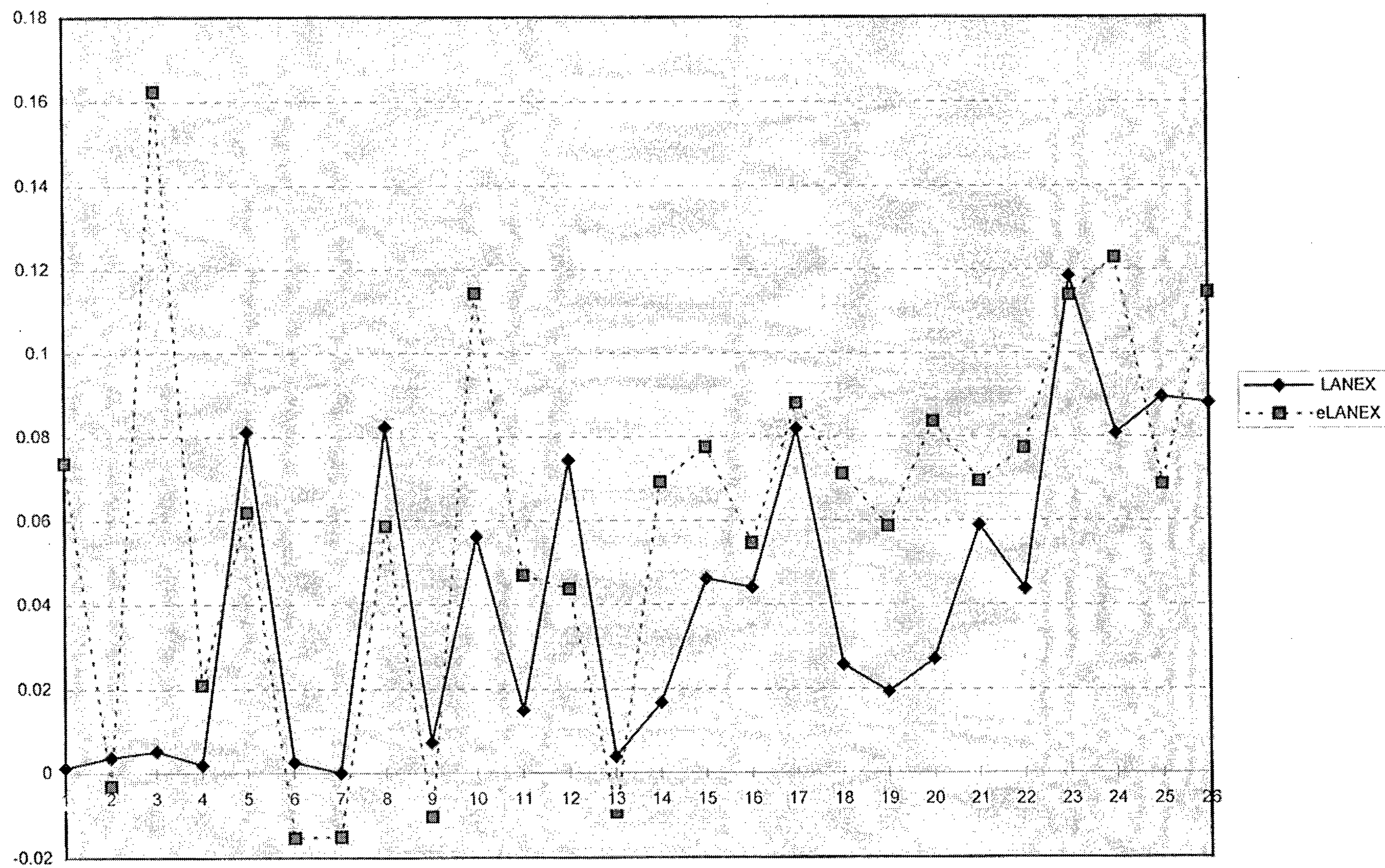


Figure H5: Subject #7: LANEX vs. eLANEX (straight 6-min avgs, minutes following detection removed)
R = 0.7225

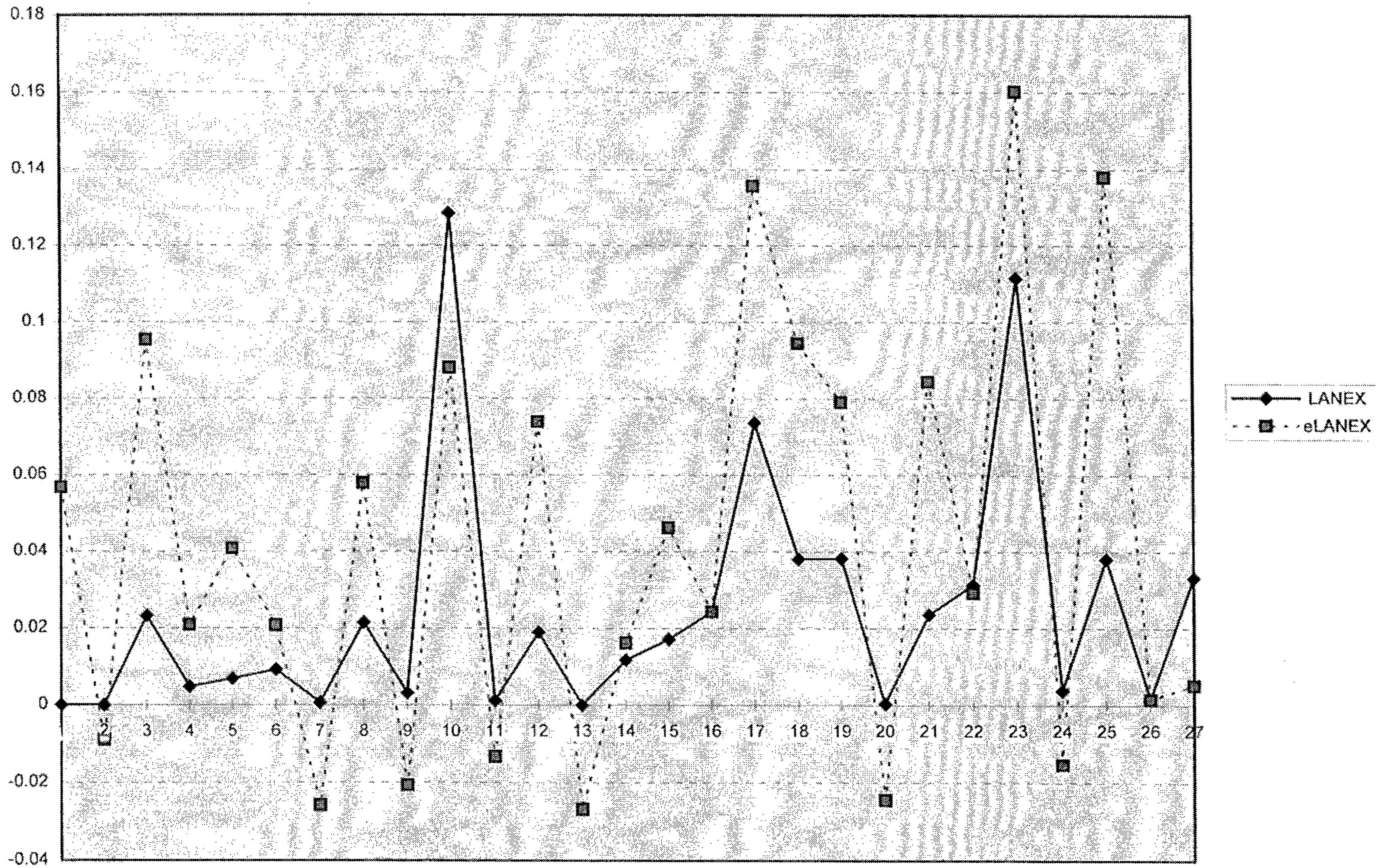


Figure H6: Subject #8: LANEX vs. eLANEX (straight 6-min avgs, minutes following detection removed)

R = 0.4697

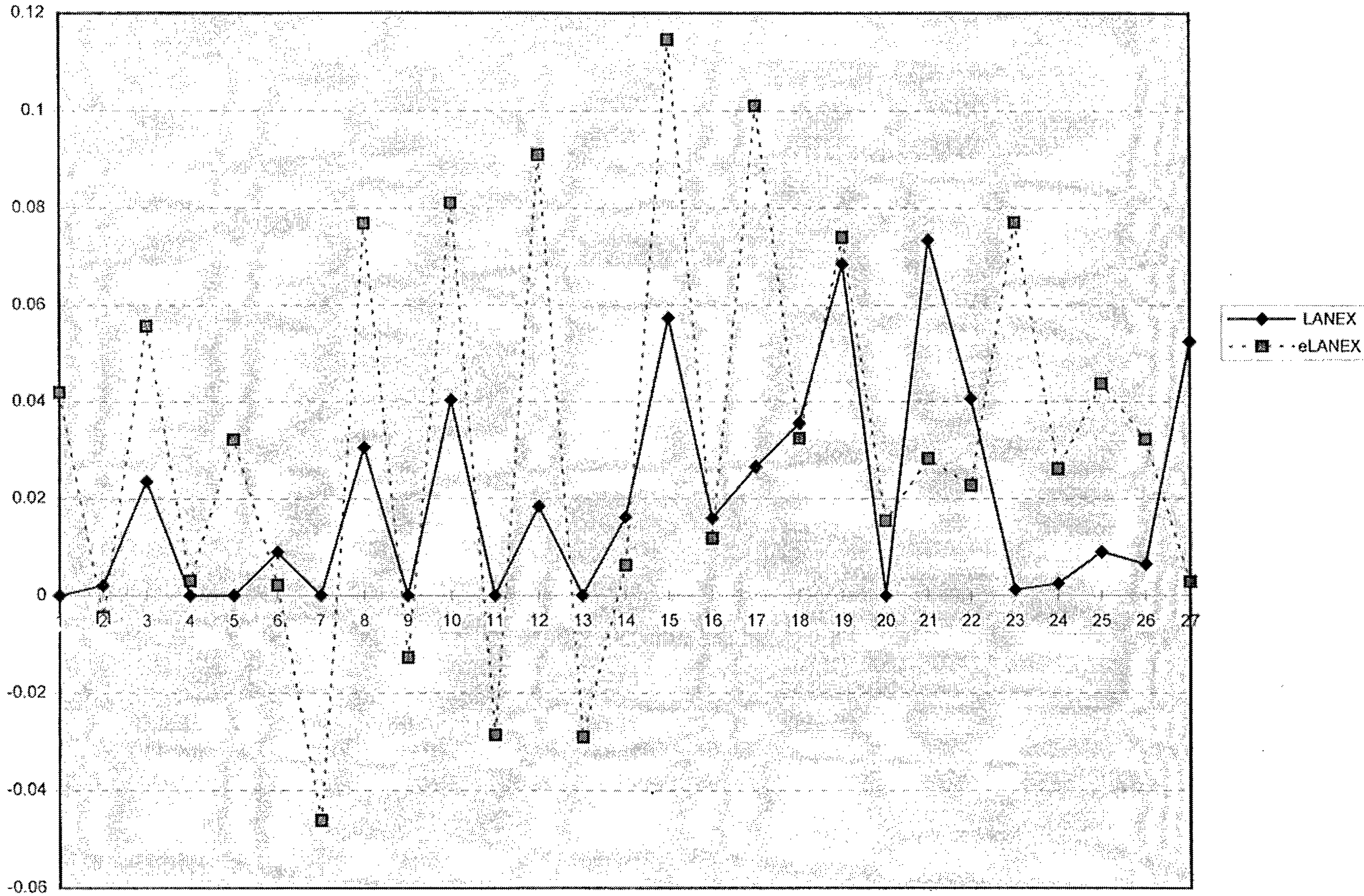
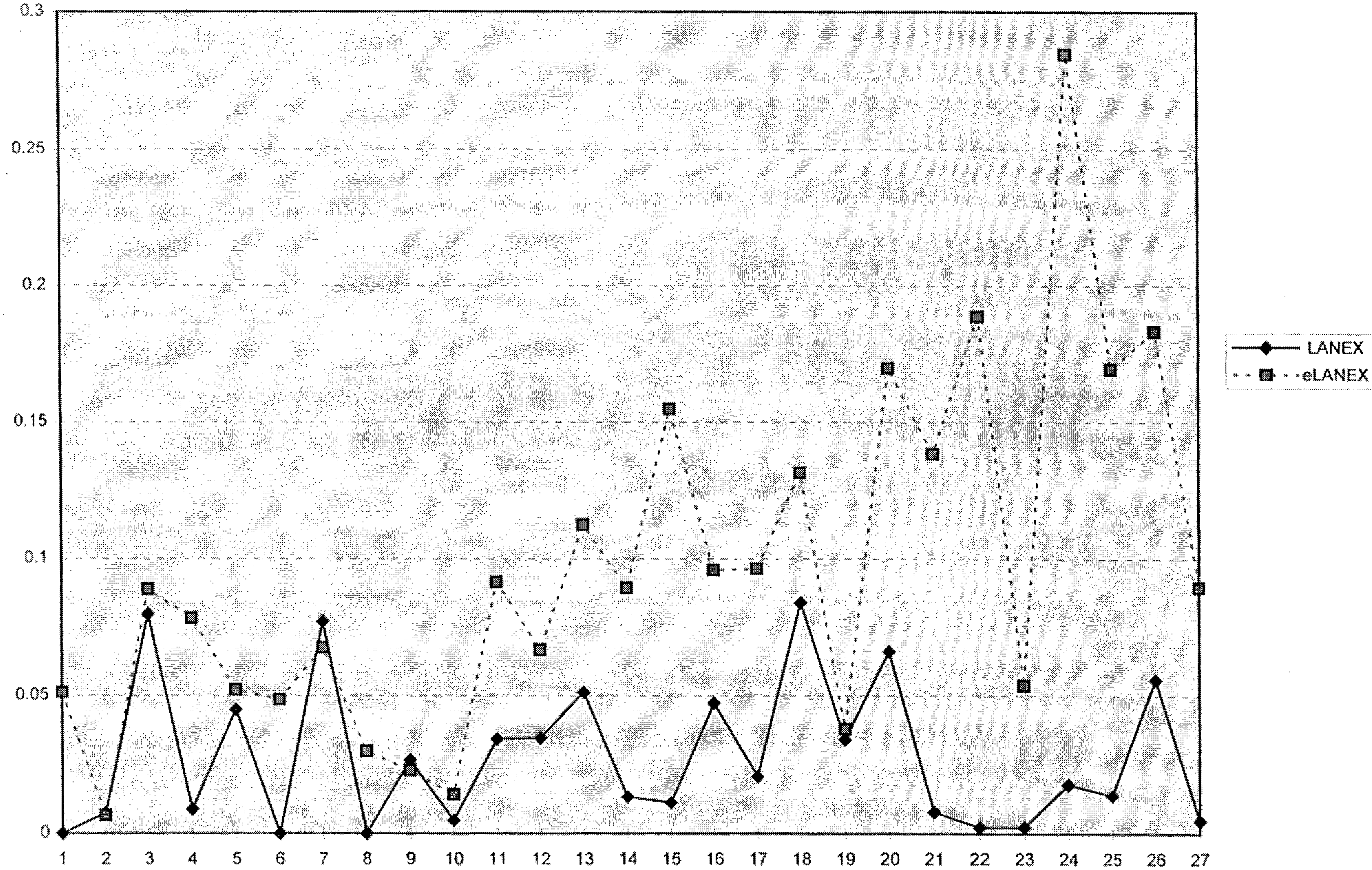


Figure H7: Subject #9: LANEX vs. eLANEX (straight 6-min avgs, minutes following detection removed)
R = 0.1629



APPENDIX I

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

SUMMARY

Groups	Count	PERCLOS Sum	PERCLOS Average	PERCLOS Variance
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

Source of Variation	SS	df	MS	F	P-value	F crit
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 I1: ANOVA Summary Table for Countermeasure Type After Detection
(Dependent Variable: PERCLOS)

SUMMARY

Groups	Count	LANEX Sum	LANEX Average	LANEX Variance
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

Source of Variation	SS	df	MS	F	P-value	F crit
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 I2: ANOVA Summary Table for Countermeasure Type After Detection
(Dependent Variable: LANEX)

SUMMARY

Groups	Count	LANDEV Sum	LANDEV Average	LANDEV Variance
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

Source of Variation	SS	df	MS	F	P-value	F crit
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 I3: ANOVA Summary Table for Countermeasure Type After Detection (Dependent Variable: LANDEV)

SUMMARY

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

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
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 I4: ANOVA Summary Table for Test and Evaluation Experiment vs. Development Experiment, Drowsy Condition, No Countermeasure Operating (Dependent Variable: PERCLOS)

SUMMARY

Groups	Count	Sum	Average	Variance
Test & Eval. LANEX	19	0.779056309	0.041002964	0.001549067
Development LANEX	62	10.21316994	0.164728547	0.011476526

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
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 I5: ANOVA Summary Table for Test and Evaluation Experiment vs. Development Experiment, Drowsy Condition, No Countermeasure Operating (Dependent Variable: LANEX)

SUMMARY

Groups	Count	Sum	Average	Variance
Test & Eval. LANDEV	19	16.06613985	0.845586308	0.043000252
Development LANDEV	62	123.9100933	1.998549893	3.265676166

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
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 I6: ANOVA Summary Table for Test and Evaluation Experiment vs. Development Experiment, Drowsy Condition, No Countermeasure Operating (Dependent Variable: LANDEV)

SUMMARY

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

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
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 I7: ANOVA Summary Table for Test and Evaluation Experiment vs. Development Experiment, Drowsy Condition, A/O Task Operating (Dependent Variable: PERCLOS)

SUMMARY

Groups	Count	Sum	Average	Variance
Test & Eval. LANEX	23	0.41690526	0.018126316	0.000602849
Development LANEX	20	4.891557224	0.244577861	0.026083578

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
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 I8: ANOVA Summary Table for Test and Evaluation Experiment vs. Development Experiment, Drowsy Condition, A/O Task Operating (Dependent Variable: LANEX)

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Test & Eval. LANDEV	23	20.49152547	0.89093589	0.057672623
Development LANDEV	20	47.80940916	2.390470458	1.015238466

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
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				

Table I9: ANOVA Summary Table for Test and Evaluation Experiment vs. Development Experiment, Drowsy Condition, A/O Task Operating (Dependent Variable: LANDEV)

U.S. DEPARTMENT OF COMMERCE
Technology Administration
National Technical Information Service
Springfield, VA 22161 (703) 605-6000