DISCRIMINANT ANALYSIS OF DRIVER BEHAVIOR DURING SAFETY CRITICAL EVENTS

Montasir Abbas, Virginia Tech, abbas@vt.edu

Bryan Higgs, Virginia Tech, Bryan.Higgs@vt.edu

Alejandra Medina, VTTI, ale@vtti.vt.edu

C.Y. David Yang, FHWA, david.yang@dot.gov

Abstract

This research effort aims to shed some light upon the behaviors that drivers show during and immediately before safety critical events. The 100-car Naturalistic Driving Study conducted by the Virginia Tech Transportation Institute (VTTI) collected very useful data in this regard. By instrumenting automobiles and allowing them to be used in normal daily routines, the data collected included normal driving as well as safety critical events. This allows the two to be compared in order to find any differences. A discriminant analysis was used for this task which resulted in interesting results when analyzing the data immediately before safety critical events for two drivers. The discriminant analysis resulted in a way to "predict" events as the discriminant scores of the data immediately before a safety critical event show a deviation from normal car following behavior.

Keywords: Naturalistic Data, Discriminant Analysis, Safety Critical

INTRODUCTION

Many efforts have been put forward in order to increase traffic safety [1-4]. The focus of the past has been on analyzing vehicle crashes through crash testing[5]. The key to improving safety is to understand the behavior of drivers. The 100-car Naturalistic Driving Data [6] offers a new way to view and analyze driver behavior. The 100-Car Naturalistic Driving Study data includes data that was collected during crashes and near crashes. Near crashes are very similar to crashes except that a successful evasive maneuver or action is taken in order to avoid a collision. Whissell and Bigelow created seven driving attitude scales to represent driver behavior and beliefs. The scales were analyzed using discriminant analyses in order to find out that the Speeding Attitude Scale was sufficient in explaining the cause of speeding tickets [7]. Mayer and Treat conducted a study to find the contributing factors to high accident drivers. The study used a discriminant analysis to find the the major contributing factors were that the high accident group scored higher in personal maladjustment, social maladjustment[8].

I. METHODOLOGY

For this research, data from the VTTI 100-car Naturalistic Driving Study was used. As opposed to traditional epidemiological and experimental / empirical approaches, this *in situ* process uses drivers who operate vehicles that have been equipped with specialized sensors along with processing and recording equipment. In effect, the vehicle becomes the data collection device. The drivers operate and interact with these vehicles during their normal driving routines while the data collection equipment is continuously recording numerous items of interest during the entire driving. Naturalistic data collection methods require a sophisticated network of sensor, processing, and recording systems. This system provides a diverse collection of both on-road driving and driver (participant, non-driving) data, including measures such as driver input and performance (e.g., lane position, headway, etc.), four camera video views, and driver activity data. This information may be supplemented by subjective data, such as questionnaire data.

Three forms of data were collected by the 100-car Naturalistic Driving Study DAS: video, dynamic performance, and audio. Approximately 43,000 driving-data hours covering 2,000,000 miles traveled were collected. One hundred cars were instrumented with the DAS.

The following is a typical description of how the data collection is performed, along with accompanying screen shots and information describing how the system works and how data can be used. Four cameras monitor and record the driver's face, forward road view, and left- and right-side of the tractor trailer, which are used to observe the traffic actions of other vehicles around the vehicle. Figure 1 displays the four camera views. Low-level infrared lighting (not visible to the driver) illuminates the vehicle cab so the driver's face and hands could be viewed via the camera during nighttime driving. The sensor data associated with the project were originally collected in a proprietary binary file format. A database schema was devised and the necessary tables were created. The schema preserves the organization of data into modules; i.e., all of the variables associated with a particular module are stored in one table in the database. The import process itself consisted of reading the binary files, writing the data to intermediate comma separated value (CSV) files and "bulk inserting" the CSV files into the database. A stored procedure is available that allows one to query the database using the module and variable names rather than database table and column names.

Figure 1: View of DAS data Collection

The methodology employed in this research effort involves four different steps: the identification and extraction of car following periods, the identification and extraction of safety critical events, discriminant analysis of the previously identified data sets, and validation of the results. The steps are described in detail, but the general idea is to use a discriminant analysis to find a method of classifying a car following behavior as safe or safety critical.

Identification and extraction of car following periods

Car-following situations were automatically extracted from the enormous volume of driving data in the database in order to analyze the car following driver behavior. The filtering process is an iterative process where initial values and conditions are used and after the events are flagged they are reviewed in the video data to adjust the values accordingly in order to obtain minimum noise. Visual inspection of the first subsets created revealed some non car-following events, so additional filtering was thus performed to remove these events from the database.

Specifically, car following periods were extracted automatically according to these conditions:

- Radar Target ID>0

    This eliminates the points in time without a radar target detected
- Radar Range<=120 meters

    This represents four seconds of headway at 70 mph
- -1.9 meters<Range*Sin (Azimuth) <1.9 meters

    This restricts the data to only one lane in front of the lead vehicle
- Speed>=20km/h

    This speed was used in order to minimize the effect of traffic jams, but still leave the influence of congestion in the data
- Rho-inverse <=1/610 meters^{-1}

    This limits the curvature of the roadway such that vehicles are not misidentified as being in the same lane as the subject vehicle when roadway curvature is present.
- Length of car following period >= 30 seconds

The automatic extraction process was verified from a sample of events through video analysis. For the random sample of 50 periods, all 50 were valid car following periods.

~~B.~~A. Identification and extraction of safety critical events

The methodology employed in this research effort involves four different steps: the identification and extraction of car following periods, the identification and extraction of safety critical events, discriminant analysis of the previously identified data sets, and validation of the results. The steps are described in detail, but the general idea is to use a discriminant analysis to find a method of classifying a car following behavior as safe or safety critical.

The safety critical events were identified and analyzed in a previous work by VTTI[9]. The method used to identify the safety critical events were triggers or thresholds on individual variables that were collected. For an event to be flagged, only one of the triggers has to be met. Those triggers are as follows:

- Longitudinal Acceleration greater than or equal to -0.2g
- Forward Time-to-Collision of less than or equal to 2 seconds
- Swerve greater than or equal to 2 rad/sec^2
- Lane Tracker Status equals abort (lane deviation)
- Critical Incident Button
- Analyst Identified

These triggers resulted in a number of potential safety critical events that were analyzed by trained data analysts that verified all of the potential events. This reduction resulted in the number of crashes and near crashes that are shown in Table 1. The near crashes occurring directly in front of the vehicle involving multiple vehicles are the events that are the closest match to car following behavior. Of the numerous near crashes that did not occur directly in front of the vehicle, most occurred to the sides of the vehicle which means no radar data is available for these events. The only means of identifying that a vehicle is beside of the subject vehicle is through the use of the video recording.

Table 1: Enumeration of Crash and Near Crash Data

Type of Event	Crashes	Near Crashes
Animal	2	10
Pedestrian		6
1 Vehicle no objects	23	46
1 Vehicle with objects	12	15
Multiple Vehicles:		
Not Directly in front or behind	5	227
Directly in front or behind	27	457

~~C.~~B. Discriminant analysis

For the discriminant analysis, thirty data points from different car following periods were used for each driver along with the event data for that driver. Thirty car following or normal driving points, selected at random, were used in order to gain a fair representation of normal driving behavior while not overpowering the safety critical event data in the analysis. Seven variables were used for the discriminant analysis which are as follows: Longitudinal Acceleration, Lateral Acceleration, Vehicle Speed, Yaw Angle, Lane Offset, Range, and Range Rate. A discriminant analysis is a statistical method that finds coefficients for the input variables that when summed creates a value that can be used to distinguish between datasets, in this case two datasets. Equation 1 below describes the mathematical form of the resulting discriminant score and how it relates to the coefficients for each variable.

$$Discriminant\ Score = \sum \beta_i * X_i \qquad (1)$$

Where:
β_i is the coefficient for variable i
X_i is variable i

For this analysis, there are seven variable and thus seven corresponding coefficients. When these seven variables and coefficients are combined to create the discriminant score, the score will serve as a way to classify the data points as normal car following behavior or safety critical behavior. Misclassification can occur and needs to be taken into consideration when choosing the best set of coefficients.

~~III.~~II. RESULTS

Table 2 presents the coefficient values resulting from a discriminant analysis at different time steps before the occurrence of safety critical events. The safety critical events were verified as safety critical through video reduction and the time of the event was noted during the video reduction. The time steps up to the event are simply the data points that occur when stepping back in time from the occurrence of the safety critical event. The data used for Driver 103-A is nine near crash events and thirty data points from normal car following periods. The table shows that consistency exists in some of the coefficients over time while the rest of the coefficients appear to be time dependent. Figure 2 shows the percent of misclassification error at each time step with 0 being the time that the safety critical events occurred. The time step at 6 seconds is chosen because it is the highest amount of time that still maintains relatively low error as shown by Figure 2. Figure 3 shows the results of applying the coefficients of the different time steps to the near crash data as well as the normal car following data. The results show some separation between the datasets at each time step with some overlapping occurring. Figure 2 shows that the error weighs on the side of misclassifying safety critical events as normal driving behavior which is not an acceptable error. Time step 6 shows a clear separation between the two datasets which means that the coefficients belonging to that dataset would be the optimal choice for use in further study.

Table 2: Discriminant Coefficients for Driver 103-A

Time	Longitudinal Acceleration	Lateral Acceleration	Speed	Lane Offset	Yaw Angle	RANGE	Range Rate
0	5.4	2.7	0.051	0.019	-0.4	0.013	0.12
1	1.5	-13.1	0.013	0.019	1.6	0.013	0.15
2	-10.3	-10.6	0.030	0.022	-5.4	0.030	0.18
3	-4.0	-5.9	0.057	0.017	2.1	0.024	0.19
4	1.8	-8.7	0.053	0.014	15.3	0.000	0.13
5	-17.9	14.2	-0.005	-0.003	-16.3	-0.004	-0.15
6	16.5	-28.1	-0.026	0.014	38.0	0.001	0.15
7	9.8	-12.2	0.006	0.009	23.2	0.014	0.16
8	22.0	-17.5	-0.009	0.016	37.5	0.008	0.20
9	-14.8	15.3	0.006	-0.007	-3.0	-0.004	-0.15
10	14.3	-12.3	-0.004	0.014	-3.5	0.007	0.14

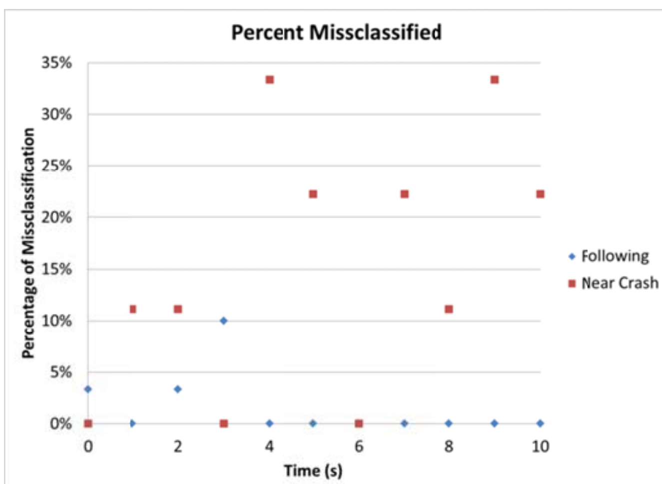

Figure 2: Percent of Misclassified Data Points for Driver 103-A

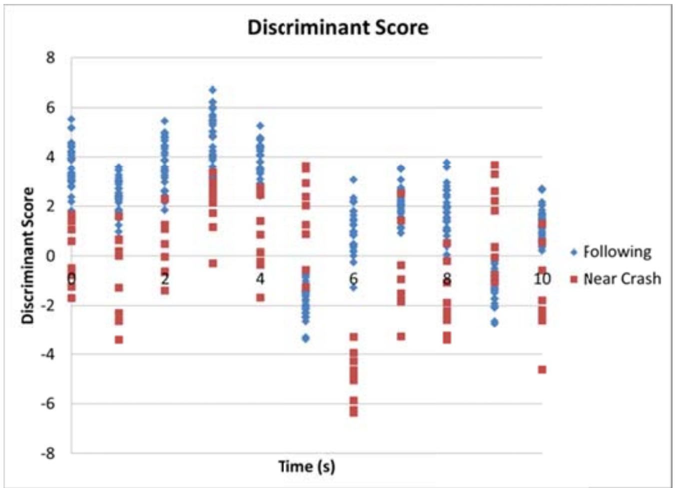

Figure 3: Discriminant Scores for Driver 103-A

Figure 4 shows the resulting discriminant scores result from when the coefficients from the 6-second time step are applied to a full car following period. The figure shows some erratic behavior but the discriminant scores are still above a value of -3 which is the highest point of the near crash data in Figure 3 for the 6-second time step which means that the full car following period would be correctly identified without misclassification occuring. Figure 5 shows the results of applying the same coefficients to a sample near crash event. The discriminant scores seem similar to a normal car following period at the beginning of the episode up until the score suddenly jumps below a value of -5. This jump corresponds to the sudden lane change of a vehicle to the same lane as the subject vehicle with a low distance left between the vehicles. In plainer terms, this refers to one driver "cutting off" another driver.

Figure 6 shows the speed trajectory of the same near crash event along with a marker of the 6-second prediction given by the discriminant coefficients. The 6-second prediction is well before the driver's reaction. Also, the discriminant score plateaus around a value of -5 immediately after the 6-second prediction. This means that the driving conditions remain dangerous, but the driver has not adjusted to the change in driving conditions. When the driver decelerates, as seen in the trajectory, the discriminant score increases indicating a change from dangerous conditions to safe, normal driving conditions.

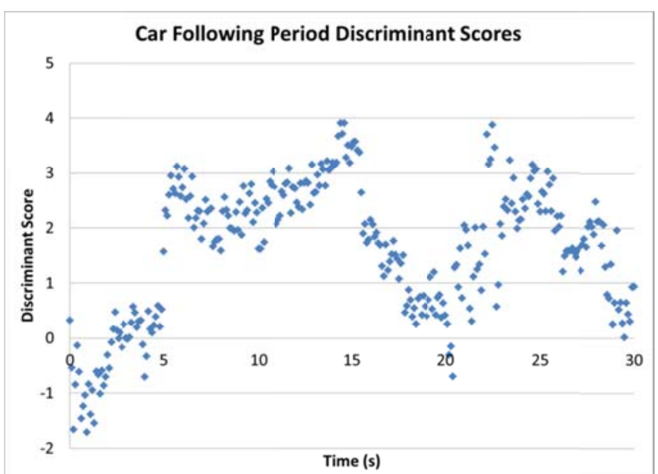

Figure 4: Discriminant Scores of a Sample Car Following Period

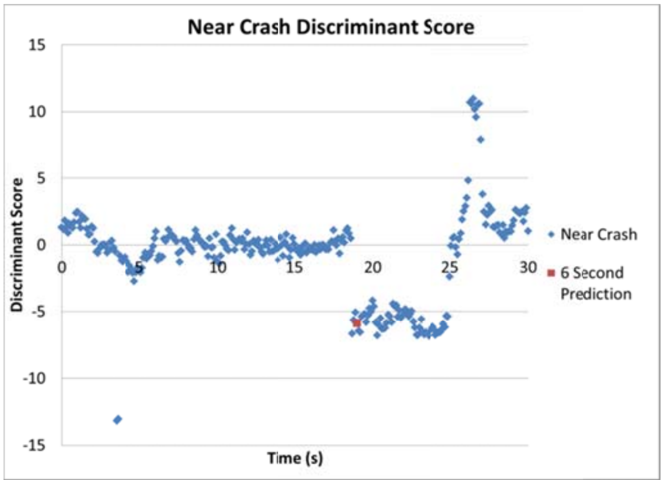

Figure 5: Discriminant Scores of a Sample Near Crash Event

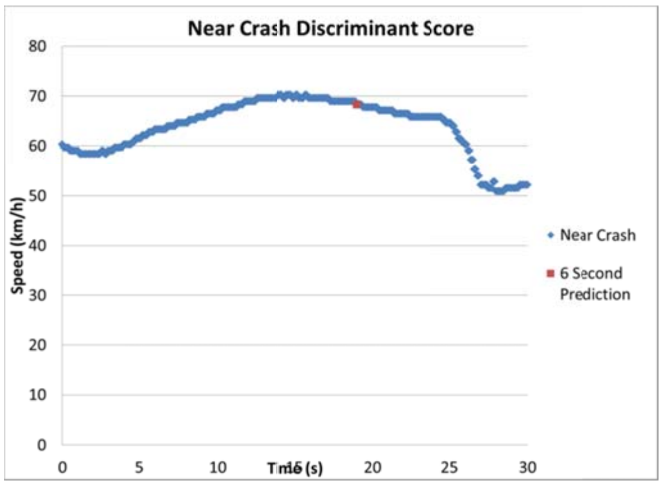

Figure 6: Trajectory of a Sample Near Crash Event

Table 3 shows the discriminant coefficients for repeating the same process as Driver 103-A for Driver 352-A. The acceleration coefficients show that the discriminant score for this driver relies on the longitudinal acceleration more than the lateral acceleration. Driver 103-A showed more of a balance between the two accelerations which serves to show that every person is different and each has unique behaviors. Figure 7 shows the percent misclassification associated with each set of discriminant coefficients.

Table 3: Discriminant Coefficients of Driver 352-A

Time	Longitudinal Acceleration	Lateral Acceleration	Speed	Lane Offset	Yaw Angle	RANGE	Range Rate
0	-15.1	-2.3	0.64	-0.02	14.9	-0.02	0.21
1	12.1	2.8	0.39	0.04	-9.0	0.01	-0.10
2	-26.8	0.0	0.32	0.02	-13.3	-0.01	0.01
3	-32.3	-0.1	0.66	-0.01	2.6	-0.02	0.05
4	-28.7	0.9	0.33	0.00	4.2	-0.01	0.02

5	-30.6	0.2	0.31	0.00	-1.8	-0.02	0.04
6	-27.4	1.0	0.21	0.00	0.7	-0.01	0.03
7	-31.2	0.3	0.29	-0.01	5.8	-0.02	0.04
8	28.2	-0.7	-0.21	0.00	4.2	0.01	-0.01
9	-11.6	2.4	0.14	0.03	-21.4	0.00	-0.07
10	12.3	3.3	0.14	0.02	-14.3	0.02	-0.10
11	3.1	2.8	0.14	0.02	-19.3	0.02	-0.06
12	-10.3	3.3	0.14	0.01	-1.5	0.00	-0.01
13	-7.1	0.0	0.14	0.01	-5.7	0.01	0.05
14	-8.5	-3.4	0.14	0.00	-7.6	0.01	0.08
15	5.1	-1.7	0.15	0.02	-28.9	0.01	-0.02

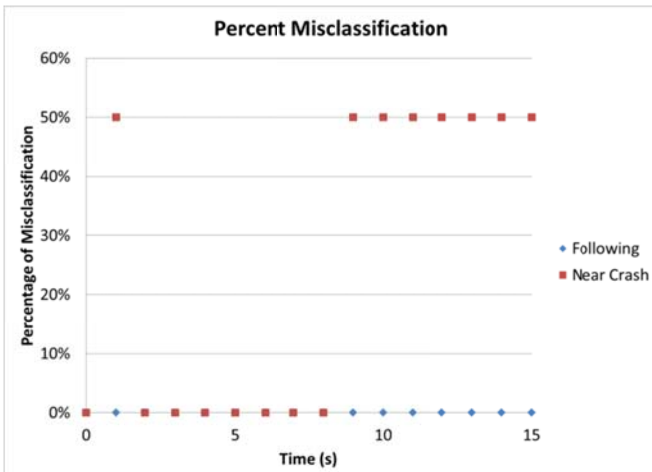

Figure 7: Percent Misclassification for Driver 352-A

Figure 8 shows the discriminant scores for data from two near crash events and normal car following data of Driver 352-A. These scores are calculated by multiplying the discriminant coefficients by the appropriate data values and then taking the summation. There is a separation between the normal car following periods and the near crash events in Driver 352-A at most time steps up until the 9-second time step as shown in Figure 8. At the 9-second time step, the analysis begins to show a misclassification of one of the near crash events as normal driving behavior as shown by Figure 7. To avoid this error and the reversal of the term relationships as seen in the 8-second time step, the coefficients at the 7-second time step should be considered optimal.

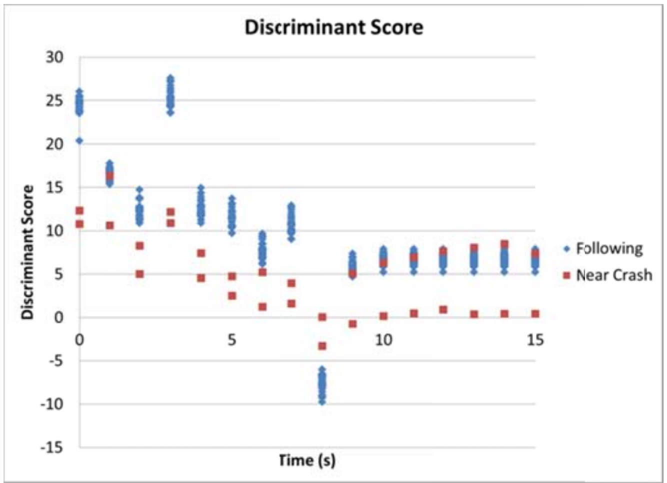

Figure 8: Discriminant Scores for Driver 352-A

Figure 9 shows the resulting discriminant scores when the coefficients of the 7 second time step are applied to a full car following period. Figure 10 shows the results of applying the same coefficients are to the data from a near crash event. In Figure 9 the scores remain relatively between the values of 10 and 12 which suggests that the normal car following behavior of this driver is fairly constant. While in Figure 10, the discriminant scores decrease before the event occurs. Also, the discriminant scores from 10 to 12 seconds in Figure 10 appear to be similar to the scores seen in the normal car following period. This means that a deterioration of the discriminant score can be seen as the data before the near crash event transitions from safe to unsafe driving conditions. Figure 11 shows the trajectory during the near crash event as the driver decelerates suddenly in order to avoid a collision. Figure 11 also shows that the 7 second prediction, given by the discriminant scores, appears well before the reaction of the driver.

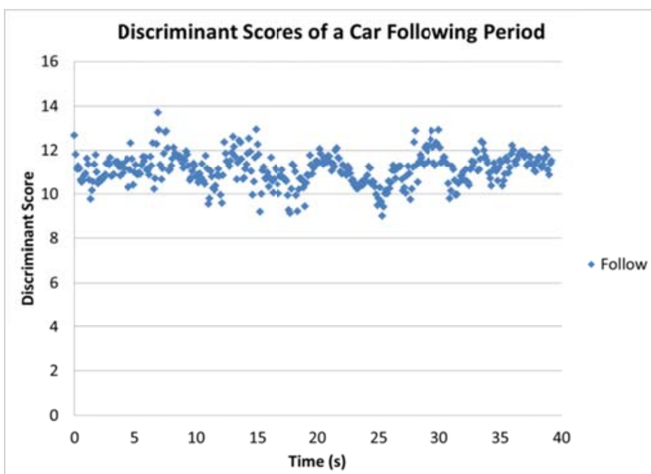

Figure 9: Discriminant Scores of a Sample Car Following Period

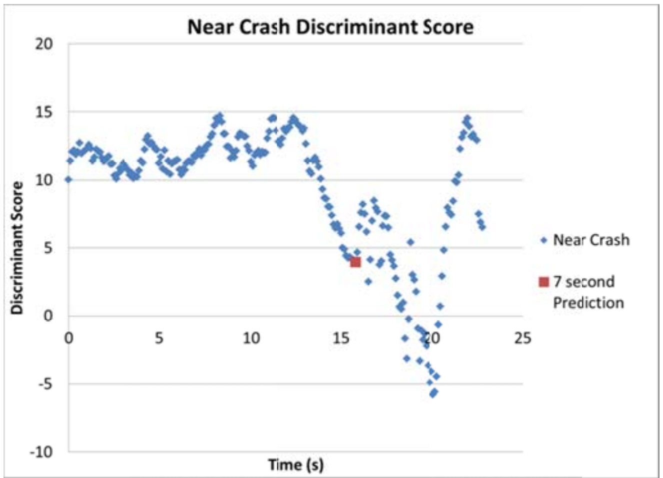

Figure 10: Discriminant Scores of a Sample Near Crash Event

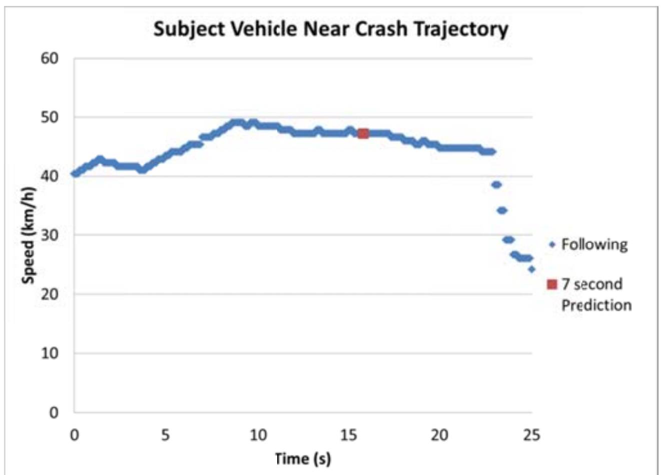

Figure 11: Trajectory of a Sample Near Crash Event with the 7-second Crash Prediction point highlighted

~~IV.~~III. CONCLUSIONS AND FUTURE RECOMMENDATIONS

The results support that using the seven selected variables along with the corresponding coefficients creates a discriminant score that can accurately distinguish between safe and unsafe driving conditions. For Driver 103-A, this distinction can be made 6 seconds before an event occurs. Also, in the case of Driver 352-A, this distinction can be made 7 seconds before an event occurs. The trajectories analyzed also show that the discriminant score increases back to a normal, safe level once a driver has reacted to the unsafe conditions.

The results of the discriminant analysis for both of drivers are obtained by the specific combination of the seven variables selected for analysis: longitudinal acceleration, lateral acceleration, vehicle speed, lane offset, yaw angle, range, and range rate. This means that the warning signs of an impending safety critical event are captured and represented by these seven selected variables. These warning signs are subtle changes in the conditions that go unnoticed when analyzing individual variables, but using the specific combination of the seven variables as defined by the discriminant analysis, results in a noticeable warning sign.

Future recommendations for research are to expand upon this method by adding more drivers. This paper shows that the discriminant analysis results in a driver specific set of parameters, but it would be beneficial to determine if these results apply to all drivers or if the drivers could be divided into groups based upon multiple variables that characterize the behavior of drivers.

ACKNOWLEDGMENT

This material is based upon work supported by the Federal Highway Administration under Agreement No. DTFH61-09-H-00007. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the Author(s) and do not necessarily reflect the view of the Federal Highway Administration.

The authors would like to express thanks to individuals at Virginia Tech and Virginia Tech Transportation Institute who contributed to the study in various ways: Greg Fitch, Shane McLaughlin, Zain Adam, Brian Daily and Rebecca Olson.

REFERENCES

[1] G. M. Fitch, *et al.*, "Analysis of Lane-Change Crashes and near-Crashes," Virginia Tech Transportation Institute, Blacksburg, VA DTNH22-00-C-07007, 2009.
[2] G. M. Fitch, *et al.*, "Safety Benefit Evaluation of a Forward Collision Warning System: Final Report," Virginia Tech Transportation Institute, Blacksburg, VA DTNH22-05-D-01019 Task Order No. 13, 2008.
[3] R. J. Hanowski, *et al.*, "The Drowsy Driver Warning System Field Operational Test: Data Collection Methods," Center for Truck and Bus Safety Virginia Tech Transportation Institute, Blacksburg, VA2008.
[4] S. B. McLaughlin, *et al.*, "Contributing Factors to Run-Off-Road Crashes and Near-Crashes," DOT HS 811 079, 2009.
[5] W. Pawlus, *et al.*, "Development of mathematical models for analysis of a vehicle crash," *WSEAS Transactions on Applied and Theoretical Mechanics*, vol. 5, pp. 156-165, 2010.
[6] T. A. Dingus, *et al.*, "The 100-car naturalistic Driving Study; Phase II - Results of the 100-Car Field Experiment," Virginia Tech Transportation Institute, Blacksburg, VA DOT HS 810 593, 2006.
[7] R. W. Whissell and B. J. Bigelow, "The speeding attitude scale and the role of sensation seeking in profiling young drivers at risk," *Risk Analysis*, vol. 23, pp. 811-820, 2003.
[8] R. E. Mayer and J. R. Treat, "Psychological, social and cognitive characteristics of high risk drivers: a pilot study," *Accident Analysis and Prevention*, vol. 9, pp. 1-8, 1977.
[9] R. Olson, *et al.*, "DRIVER DISTRACTION IN COMMERCIAL VEHICLE OPERATIONS," Center for Truck and Bus Safety Virginia Tech Transportation Institute, Blacksburg VA FMCSA-RRR-09-042, 2009.