UNIVERSITY OF CALIFORNIA, SAN DIEGO

Learning and Inferring Human Intentions: Explorations of Driver Attention and Interactivity

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Electrical Engineering (Signal and Image Processing)

by

Anup Doshi

Committee in charge:

Professor Mohan M. Trivedi, Chair
Professor Serge Belongie
Professor Hal Pashler
Professor Bhaskar Rao
Professor Nuno Vasconcelos

2010

The dissertation of Anup Doshi is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2010

DEDICATION

To my brother, who day by day continues to guide me in the pursuit of happiness.

EPIGRAPH

An 'inner process' stands in need of outward criteria.
—L Wittgenstein

TABLE OF CONTENTS

Signature Page . . . iii

Dedication . . . iv

Epigraph . . . v

Table of Contents . . . vi

List of Figures . . . viii

List of Tables . . . xi

Acknowledgements . . . xii

Vita . . . xiii

Abstract of the Dissertation . . . xv

Chapter 1 Introduction . . . 1
- 1.1 Contributions and Outline . . . 2

Chapter 2 Review of Driver Behavior Predictions . . . 5
- 2.1 Modeling and Understanding Driver Behaviors . . . 6
  - 2.1.1 Operational Maneuvers . . . 7
  - 2.1.2 Tactical Maneuvers . . . 7
  - 2.1.3 Strategic Maneuvers . . . 9
- 2.2 Sources of Variation . . . 9
- 2.3 Toward Real-world Performance . . . 11

Chapter 3 Visual Search and Intent . . . 12
- 3.1 Attention Shifts . . . 12
  - 3.1.1 Introduction . . . 12
  - 3.1.2 Related Research . . . 15
  - 3.1.3 Methods . . . 19
  - 3.1.4 Results and Discussion . . . 24
  - 3.1.5 Investigations into Naturalistic Driving . . . 30
  - 3.1.6 Concluding Remarks . . . 34
- 3.2 Holistic Attention Estimation . . . 35
  - 3.2.1 Introduction . . . 35
  - 3.2.2 Related Research . . . 37
  - 3.2.3 Proposed Model: BRAVVO . . . 40
  - 3.2.4 Experimental Validation . . . 46
  - 3.2.5 Discussion and Concluding Remarks . . . 52
  - 3.2.6 Acknowledgements . . . 53

Chapter 4 Cognitive State and Intent . . . 54
- 4.1 Recency and Sequential Effects . . . 54
  - 4.1.1 Method . . . 55
  - 4.1.2 Results . . . 58
  - 4.1.3 Discussion . . . 64
  - 4.1.4 Acknowledgements . . . 66

Chapter 5 Real-time Systems and Analysis . . . 67
5.1 Comparing Cues to Intent . . . 67
5.1.1 Driver Behavioral Cues . . . 68
5.1.2 Lane Change Intent Prediction . . . 73
5.1.3 Experiments and Analysis . . . 75
5.1.4 Concluding Remarks . . . 81
5.2 Real-time Analysis: ADAS Development . . . 81
5.2.1 Context Capture Framework for Intelligent Vehicles . . . 82
5.2.2 On-Road Intent Prediction Framework . . . 86
5.2.3 On-road Performance Characterization . . . 88
5.2.4 Selection and Performance of Different Sensor Configurations . 93
5.2.5 Concluding Remarks and Future Directions . . . 95
5.2.6 Acknowledgements . . . 98

Chapter 6 Driver Interactivity . . . 100
6.1 Feedback and DAD . . . 100
6.1.1 Background and Related Works . . . 101
6.1.2 LISA-P Testbed and DAD Alerts . . . 104
6.1.3 DAD Evaluation: Speed Compliance Experiment . . . 109
6.1.4 Discussion and Concluding Remarks . . . 117
6.2 Effects of Style . . . 118
6.2.1 Related Works and Driver Style Metrics . . . 120
6.2.2 Experiment 1: Simulator-based Analysis . . . 121
6.2.3 Experiment 2: Real-world Analysis - Predictability and Intent 124
6.2.4 Experiment 3: Real-world - Responsiveness to feedback . . . 125
6.2.5 Discussion and Future Work . . . 128
6.3 Potential of Communicating Intent . . . 128
6.3.1 Cooperative Implementation of Intent-based Advance Driver Assistance Systems using LACASA . . . 130
6.3.2 Quantitative Improvements of Cooperative LACASA-based Systems . . . 132
6.3.3 Concluding Remarks . . . 137
6.3.4 Acknowledgements . . . 137

Chapter 7 Conclusions . . . 139

Bibliography . . . 141

LIST OF FIGURES

Figure 3.1: Motivation for understanding head and eye dynamics in attention shifts. . . 14
Figure 3.2: Flowchart of proposed approach for visual attention shift analysis. 15
Figure 3.3: Examples of various interactions of head and eye movements, with type labels from Zangemeister & Stark, 1982. Note in certain cases eye gaze tends to move first, where in others the head tends to move first. 18
Figure 3.4: Experimental setup of LISA-S testbed. 20
Figure 3.5: Illustrative example of experimental paradigm for distinguishing goal-oriented attention shifts. 21
Figure 3.6: Sample data showing the procedure for detection of the start of the eye saccade. 23
Figure 3.7: Overall distribution of head yaw position and head yaw motion at the time of the eye gaze saccade for each condition, including all examples of all subjects. 25
Figure 3.8: Average Head Yaw prior to Eye Gaze Saccade under each condition of the experiment, aligned to the position of the saccade. 26
Figure 3.9: Distribution of subject-wise median head yaw position and head yaw motions at the time of the eye saccade. Error bars represent standard error of the median. 27
Figure 3.10: Distribution of saccade timings, after the onset of the first cue. 28
Figure 3.11: Distribution of first significant head motions (over a fixed threshold) relative to the gaze saccade. 29
Figure 3.12: Starting Yaw Position for Eye and Head motion under each condition. Error bars represent the standard error of the mean. 30
Figure 3.13: Maximum Yaw Rates for Eye and Head motion under each condition. Error bars represent the standard error of the mean. 31
Figure 3.14: Duration of motion from initial movement until target. Note that in case of eye motion, this is a superset of the saccade duration. Error bars represent the standard error of the mean. 32
Figure 3.15: Maximum Yaw Position for Eye and Head motion under each condition. Error bars represent the standard error of the mean. 33
Figure 3.16: Flowchart of proposed approach for evaluating naturalistic driving data during lane change-associated visual search. 33
Figure 3.17: Graphical model relating Gaze, Attention, and Task. 35
Figure 3.18: Data Flow diagram for BRAVVO model. 41
Figure 3.19: Sample images showing the difficulty involved in gaze estimation in complex environments such as vehicles. 43
Figure 3.20: Three naturalistic driving scenarios under examination. 43
Figure 3.21: Examples of both types of saliency maps used in this approach. 44
Figure 3.22: Sample data and attention estimation components. 46
Figure 3.23: Sample frames comparing approaches for estimating focus of attention. . . . 47
Figure 3.24: Results showing the estimated PDF of focus-of-attention. The blue box marks the estimated point of focus, and the green represents the labeled ground truth. 48
Figure 3.25: Sequence 1: Highway Lane Change - BRAVVO-M approach compared to Baseline Head-Pose and Motion-Saliency Attention Estimators. Raw head pose is shown below, with red indicating no head pose available. 50
Figure 3.26: Sequence 2: Intersection Turn - BRAVVO-M approach compared to Baseline Head-Pose and Motion-Saliency Attention Estimators. Raw head pose is shown below, with red indicating no head pose available. 51
Figure 3.27: Sequence 3: Pedestrian Yield - BRAVVO-M approach compared to Baseline Head-Pose and Motion-Saliency Attention Estimators. Raw head pose is shown below, with red indicating no head pose available. 51

Figure 4.1: Simulator layout and experimental setup for Sequential Effects experiment. . 56
Figure 4.2: Sequential Effects in Pedal Press Response Times (top) and ranking of RT according to the recent sequence (bottom). Error bars indicate one standard error of the mean. 58
Figure 4.3: Sequential effects on response time, broken into foot movement initiation time (top) and time to move to pedal (bottom). 60
Figure 4.4: Mean overall reaction time (ORT) as a function of trial context (circles) and fit of first-order model with exponentially decaying memory (diamonds) . . . 62
Figure 4.5: Lag plots of first- and second-order sequential effects on overall RT (ORT), foot movement initiation Time (FMI) and time to move to pedal (TTM). The error bars represent +/-1 standard error of the mean. 63
Figure 4.6: Sample series in the complex condition of the experiment, showing an example of pedal misapplication. 64

Figure 5.1: Head motion and Eye gaze positions prior to lane changes. 71
Figure 5.2: (Top) Approximate distribution of eye gaze location classifications for labeling purposes. (Bottom) Samples from dataset showing corresponding eye gaze locations. 72
Figure 5.3: Eye glance transition probabilities for lane keeping and lane changing situations used in this study. 72
Figure 5.4: ROC Comparing Different Input Data, 3-sec and 2-sec Decision Time. 77
Figure 5.5: Head motion and Eye gaze positions prior to a lane change. 80
Figure 5.6: LISA-X testbed and sample progression of intent prediction signal. 82
Figure 5.7: Overview of Real-time Intent Detection System. 83
Figure 5.8: Three SWA regions of interest. 86
Figure 5.9: Training-test database split 87
Figure 5.10: Evolution of lane and head signals prior to lane change. 88
Figure 5.11: Plots comparing the $\delta = 2.5$ intent ROC curves with all sensor subsystems for the different test sets. 89
Figure 5.12: Sample image showing multi-suppression technique. 91
Figure 5.13: Distribution of Intent Detections. 92
Figure 5.14: The size of the match window is controlled by the filter detection time δ and the two parameters f, α. f characterizes detection spread and α enables earlier detections. 93
Figure 5.15: Detection performance for different match-window sizes. 93
Figure 5.16: Performance matrix comparing different sensor subsystems, which contribute relevant information at different times before a lane change. 94
Figure 5.17: Real unit ROC curves for competing sensor configurations at different detection timings. (a) $\delta = 1.00$ (b) $\delta = 1.50$ (c) $\delta = 2.50$. 96
Figure 5.18: Real units ROC curve for the `VLAHS` intent classifier at different detection timings. The closer to the lane change the better the performance and prediction horizons beyond 3.0 seconds is completely unreliable. 97
Figure 5.19: Intent detection rate for a fixed false positive per hour rate of 120 FP/hour. 98
Figure 5.20: Examples of on-road intent detection. 99
Figure 5.21: These maneuvers appear very similar to a lane change. (a) Lane ends during a merge. (b) Lane split during a highway exit. 99

Figure 6.1: LISA-P Testbed Setup. 104
Figure 6.2: LISA-P Testbed: DAD Display Capabilities and Limitations, along with approximate location of alerts used for Speed Control Experiment. Within these limitations, the DAD is capable of drawing anywhere on the windshield. . . . 105

Figure 6.3: Samples from NHTSA Backover Trials [147] showing inconsistent performance of automatic rear-vehicle object detectors. 106
Figure 6.4: Demonstration of DAD Backup Aid. 107
Figure 6.5: Samples from a DAD-based Navigation Aid. 108
Figure 6.6: Illustration of the three alerts used. 110
Figure 6.7: Driving Test Path, which includes local roads, main roads, and highways, with speed limits ranging from 15 to 65 mph. 111
Figure 6.8: Results of sample Test Run for Condition 1 - No Display (top) and Condition 2 - Warning sign (bottom). 112
Figure 6.9: Time to slow back down, or the amount of time spent over the speed limit before slowing back down with different alerts. 113
Figure 6.10: Histograms of speeds for each section of road with the given speed limit. . . 114
Figure 6.11: The head pose estimates are classified into three clusters, each corresponding to a certain region in front of the driver: "Up", "Down", or "At DAD" . . . 116
Figure 6.12: Overview of driver-vehicle interaction. 119
Figure 6.13: Evolution of $\sigma_{lat-jerk}$ over time, for each of the 10 drivers in the simulator-based experiment. 122
Figure 6.14: Differences in measured time from lane change initiation until lane crossing (Time-to-lane-crossing), for different driver styles in the simulator experiment. 123
Figure 6.15: Distributions of Intent Prediction Confidences as a function of Driver Style (derived from $\sigma_{lat-jerk}$). ANOVA shows a significant increase in IPCs for "aggressive" drivers ($p < 0.05$). 125
Figure 6.16: Distributions of Intent Prediction Confidences as a function of Driver Style (derived from m_{TTLV}). 126
Figure 6.17: Distributions of "Time-to-slow-back-down" (a measure of responsiveness to the DAD feedback) as a function of Driver Style (derived from $\sigma_{lon-jerk}$). . . 127
Figure 6.18: Layered Architecture for Cooperative Active Safety Applications (LACASA). 129
Figure 6.19: Motivational examples for modular, cooperative LACASA framework. 130
Figure 6.20: Forward Collision Warning - Reduction in risk with earlier alert times. . . . 133
Figure 6.21: Lane Change Simulation Results. 134
Figure 6.22: Intersection Turn Simulation Results: Breakdown of outcome distributions. . 135
Figure 6.23: Intersection Turn Simulation Results: Breakdown of distance distributions. . 136

LIST OF TABLES

Table 2.1: Selected Studies in Single-Target Driver Tactic Prediction (Short-term) 8
Table 2.2: Selected Studies in Short-term Multi-Target Driver Tactic Prediction (Short-term) 9
Table 2.3: Selected Studies in Driver Strategy Prediction (Long-term) 10

Table 3.1: Confusion Matrix for Detecting Endogenous, Goal-Oriented Gaze Shifts (G) versus Exogenous, Stimulus-Oriented Shifts (S) in Simulator Experiment, using Head Yaw Position criteria. Correct classification rate is 69.44%. 29
Table 3.2: Confusion Matrix for Detecting Endogenous, Goal-Oriented Gaze Shifts (G) versus Exogenous, Stimulus-Oriented Shifts (S) in Simulator Experiment, using Head Yaw Motion criteria. Correct classification rate is 65.74%. 30
Table 3.3: Average Intent Prediction Confidences ($\overline{IPC}$) for Each Type of Classifier, where a value of 0 represents chance. 34
Table 3.4: Similar recent research toward the understanding the focus of attention of a human. 38
Table 3.5: Effects of certain tasks on vision-based estimates of gaze, saliency, and orientation given attention. The behavioral influences are noted in parentheses, and discussed in more details in Sections 3.2.3-3.2.3. 39
Table 3.6: Comparison of proposed Bayesian-Attention Estimator with approaches based on Head Pose alone and Saliency alone. 49

Table 4.1: Percent Variance Explained by Three Models For Three Different RT Measures 61

Table 5.1: Glance durations for 3.0sec-ahead and 2.0sec-ahead decision times. Note the lack of a pattern in the eye glances left or right before lane changes, whereas the over-the-shoulder look is more indicative of a lane change. 73
Table 5.2: True Positive and False Positive Rates for a Fixed Threshold (T=0) for the 3 second-ahead Decision Time 76
Table 5.3: True Positive and False Positive Rates for a Fixed Threshold (T=0) for the 2 second-ahead Decision Time 77
Table 5.4: Statistical significance tests comparing outputs of different classifiers. 78
Table 5.5: Lane Change Prediction Datasets 86

Table 6.1: Examples and Related Works involving different display modalities (haptic, audio, visual) and types (static, dynamic, and dynamic+active). 103
Table 6.2: Average time-to-slow-back-down with different alerts over all drivers. 112
Table 6.3: Normalized ratios of time spent “looking” in each direction while above the speed limit, with vs. without alerts, averaged over all drivers. 117
Table 6.4: ANOVA: Time-to-lane-crossing in Non-aggressive vs. Aggressive Drivers, based on each style metric 123
Table 6.5: ANOVA: IPC in Aggressive vs. Non-Aggressive Drivers, based on each style metric 126
Table 6.6: Sample “Levels” of Implementations of LACASA. Future vehicles may have one of these two LACASA-based ADASs, or include some combination or subset of the sensor and communications equipment. 132
Table 6.7: Lane Change Simulation Outcomes (when front-vehicle crosses lane boundary) 134
Table 6.8: Intersection Turn Simulation Outcomes (when turning-vehicle leaves intersection) 135

ACKNOWLEDGEMENTS

All this would not be possible without the wonderful support of my advisor, Prof. Mohan M. Trivedi. His friendship, guidance, mentorship, inspiration, enthusiasm, advocacy, patience, and determination have been limitless and invaluable.

I truly appreciate the contributions of my committee members, Prof. Serge Belongie, Prof. Hal Pashler, Prof. Bhaskar Rao, and Prof. Nuno Vasconcelos. Their insightful questions, criticisms, and guidance have significantly improved the content and direction of this work.

The advice, support, cajoling, inspiration, and knowledge of all my colleagues and predecessors from the CVRR laboratory and beyond will never be forgotten. I'd especially like to acknowledge the collaborations and efforts of my co-authors, Shinko Cheng, Erik Murphy-Chutorian, Brendan Morris, Cuong Tran, Matt Wilder, Prof. Mike Mozer, and Prof. Trivedi.

I am thankful for the support of the Jacobs School Graduate Fellowship from 2005 to 2008, and a Doctoral Dissertation Grant from the University of California Transportation Center, from 2009 to 2010. Parts of the dissertation are also sponsored by the UC Discovery Program, Volkswagen, and the National Science Foundation, all of to whom I am indebted.

Special thanks are required for the many hours donated by colleagues and friends who volunteered to sit through pilot experiments and test prototypes for me. Nothing ever worked the first time, often we were fooled twice, and the third time was rarely charming. I humbly offer this thesis as a consolation prize for all your patience.

Finally, thanks to all my wonderful friends; whether in life, love or laughter, all of whom will never stop believing;
To my San Diego and Southern California families, for taking me in with open hearts;
To my family the world over, for caring, patience, and understanding;
Most importantly, to my dearest parents, for their undying, unquestioning love and support.

Publication acknowledgements: Chapter 3 is in part a reprint of material that is submitted to the Journal of Vision (2010), and IEEE Transactions on Intelligent Transportation Systems (2010), both by Anup Doshi and Mohan M. Trivedi. Chapter 4 is in part a reprint of material that is submitted to Cognitive Science (2010), by Anup Doshi, Cuong Tran, Matthew H. Wilder, Michael C. Mozer, and Mohan M. Trivedi. Chapter 5 is in part a reprint of material that is published in the IEEE Transactions on Intelligent Transportation Systems (2009), by Anup Doshi and Mohan M. Trivedi, and submitted to IEEE Pervasive Computing (2010), by Anup Doshi, Brendan T Morris, and Mohan M. Trivedi. Chapter 6 is in part a reprint of material that is published in the IEEE Transactions on Systems, Man, and Cybernetics - Part B (2009), by Anup Doshi, Shinko Y Cheng, and Mohan M. Trivedi, submitted to IEEE Intelligent Transportation Systems Magazine (2010), by Anup Doshi and Mohan M. Trivedi, and published in the IEEE Intelligent Transportation Systems Conference (2010), by Anup Doshi and Mohan M. Trivedi.

VITA

2005	B. S. and M. S. in Electrical and Computer Engineering, Carnegie Mellon University
2005-2010	Graduate Student Researcher, University of California, San Diego
2010	Ph. D. in Electrical Engineering (Signal and Image Processing), University of California, San Diego

PUBLICATIONS

Anup Doshi and Mohan M. Trivedi, "Driver Behavior and Intent Prediction: Survey and Challenges", *in preparation*, 2010.

Anup Doshi and Mohan M. Trivedi, "Head and Eye Gaze Dynamics during Visual Attention Shifts in Complex Environments", *under review, Journal of Vision*, 2010.

Anup Doshi, Cuong Tran, Matthew H. Wilder, Michael C. Mozer and Mohan M. Trivedi, "Sequential Dependencies in Driving", *under review, Cognitive Science*, 2010.

Anup Doshi, Brendan T. Morris and Mohan M. Trivedi, "Context-Sensitive Prediction of Intentions for Driver Assistance", *under review, IEEE Pervasive Computing*, 2010.

Anup Doshi and Mohan M. Trivedi, "Towards 'Driving Style'-Sensitive Driver Assistance Systems", *under review, IEEE Intelligent Transportation Systems Magazine*, 2010.

Anup Doshi and Mohan M. Trivedi, "A Bayesian Approach for Estimating Attention by Viewer and View Observations", *under review, IEEE Transactions on Intelligent Transportation Systems*, 2010.

Anup Doshi and Mohan M. Trivedi, "On the Roles of Eye Gaze and Head Dynamics in Visual Attention and Context Learning", *IEEE Transactions on Intelligent Transportation Systems, 10(3)*, Sept. 2009.

Anup Doshi, Shinko Y. Cheng, and Mohan M. Trivedi, "A Novel, Active Heads-up Display for Driver Assistance", *IEEE Transactions on Systems, Man and Cybernetics - Part B, 39(1)*, Feb. 2009.

Anup Doshi and Mohan M. Trivedi, "Communicating Driver Intents: A Layered Architecture for Cooperative Active Safety Applications", *IEEE Conference on Intelligent Transportations Systems*, Sept. 2010.

Anup Doshi and Mohan M. Trivedi, "Attention Estimation by Simultaneous Observation of Viewer and View", *IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Workshop on CVPR4HB*, June 2010.

Anup Doshi and Mohan M. Trivedi, "Head and Gaze Dynamics in Visual Attention and Context Learning", *IEEE CVPR Workshop on Visual and Contextual Learning*, June 2009.

Anup Doshi and Mohan M. Trivedi, "Examining the Impact of Driving Style on the Predictability and Responsiveness of the Driver: Real-world and Simulator Analysis", *IEEE Intelligent Vehicles Symposium*, June 2010.

Anup Doshi and Mohan M. Trivedi, "Investigating the Relationships Between Gaze Patterns, Dynamic Vehicle Surround Analysis, and Driver Intentions", *IEEE Intelligent Vehicles Symposium*, June 2009.

Anup Doshi and Mohan M. Trivedi, “A Comparative Exploration of Eye Gaze and Head Pose Cues for Lane Change Intent Prediction”, *IEEE Intelligent Vehicles Symposium*, June 2008.

ABSTRACT OF THE DISSERTATION

Learning and Inferring Human Intentions: Explorations of Driver Attention and Interactivity

by

Anup Doshi

Doctor of Philosophy in Electrical Engineering (Signal and Image Processing)

University of California, San Diego, 2010

Professor Mohan M. Trivedi, Chair

The focus of this research is on the challenges of using multi-sensory, multi-modal data focused on drivers and their environment, to infer the cognitive states and intentions of the driver. We propose several relevant research tasks including behavioral attention and priming analysis, cue selection, and data fusion and model development. Ultimately we arm an Intelligent Driver Assistance System with such information in order to improve decision making, safety, and comfort. Specifically, the objective of this research is to build a holistic driver intent inference system by (1) observing and understanding the most reliable visible behaviors, characteristics, and environmental cues that indicate driver intentions, and (2) analyzing the interactive nature and performance of real-time learning-based driver assistance systems. The thesis makes contributions to, and documents research interactivity between, the fields of electrical engineering, signal processing, cognitive science, human-computer interaction and psychology. The advantage of including human intent information in assistive feedback is found to be substantial, especially

for reducing risk in safety-critical situations in driving and other task-oriented scenarios. The contributions of this research could be extended to applications in Advanced Driver Assistance Systems, Assistive Living, and Smart Meeting Rooms.

Chapter 1

Introduction

For many years research into intelligent systems has been focused on the development of more advanced sensors and networks to perceive, understand, communicate with, and interact with humans and the environment. As these systems have become increasingly adept at performing complex tasks, they have even begun to take control away from humans, in a broad range of arenas from finance to medicine to manufacturing. The resulting increases in safety, productivity, and lifestyle have been a transformative hallmark of the past century.

In certain fields, though, technology continues to lag in its effectiveness at improving the safety and comfort of human agents. Specifically, this includes such tasks as driving, where the human operator continues to play a central role in the task performance. In spite of leaps in automotive technology, every year over 30,000 deaths occurred on U.S. roadways [145], with 1.2 million traffic-related injuries worldwide [158]. Recent estimates have put driver inattention as the underlying cause of up to 80% of these accidents [6]. This would seem to indict the human as the weak link in the driver-vehicle-environment "ecosystem". Nevertheless, it is the cognitive capabilities of human drivers that keep them at the center of this ecosystem, given their abilities to rapidly process and respond to novel, unforeseen circumstances. In a legal sense, every major decision that a vehicle makes is generally sanctioned by the driver, even including such "autonomous" components as adaptive cruise control and parallel-parking assistance systems.

As the human continues to remain at the center of the holistic driving ecosystem, the focus of this research is on the development of new generations of intelligent systems to utilize advanced sensing and computational power to assist the humans. Intelligent Driver Assistance Systems (DASs) have the potential to improve safety by observing and interacting with the driver in a way to decrease risky behaviors and maneuvers [208, 207].

For the most part, however, current DASs are insensitive to the state of the driver and whether the driver needs the feedback, potentially annoying the driver to the point of disregarding or disabling the safety device, defeating its purpose. An opportunity for improvement is the utilization of driver information to improve the prediction of the vehicle's upcoming maneuvers. Prior research has shown that data processed from a set of smart cameras focused on the driver can improve maneuver recognition significantly. Cues such as driver head dynamics, eye gaze, hand position, and foot hovering information, have been shown to reveal a driver's intent to change lanes, brake, and turn [45, 121, 120, 30]. This is made possible because of the way that drivers prepare for the maneuvers moments before the start of the maneuver. A driver may check her blind-spot, adjust her speed and grip the steering differently moments prior to performing the maneuvers.

Along with driver pose information gleamed from driver-facing cameras, data from a number of other embedded sensors measuring environmental obstacles can be used to determine future vehicle maneuvers based on driver intentions. The addition of driver pose as a proxy for driver intent allows for considerable disambiguation between potentially dangerous situations and normal driving behavior. This is a very important factor in human assistance systems design, as reducing the false alarm rate is crucial to driver acceptance of the system.

Our proposed research pertains to the development of a general framework for analyzing driver intent, by analysis and investigations into the attentive and interactive processes surrounding human intentions. Ultimately the research will produce a real-time, interactive Driver Assistance System that incorporates intent inference along with analysis of continuous performance of the system. To improve the performance of the system, and reduce false positives, we also engage in an analysis of several fundamental phenomena, including visual search and attentive processes, causes of unintended maneuvers, and driver style analysis. These studies, inspired by research in cognitive science and psychology, allow the engineering of safer, more comfortable, and more reliable human-machine systems in the real world.

1.1 Contributions and Outline

The contributions will range from basic results in how to detect attentive processes from body language, to include analysis of the most relevant internal (driver) and external (environment) cues for intent inference. We demonstrate for the first time significant sequential effects on driver response times and errors, a novel finding in such a complex environment. Based on this analysis of human behavior, we implement and analyze a cost-effective framework for a holistic (joint human-environment), real-time intent inference system on a vehicular testbed. Finally, we study the interactivity and appropriate design of such style-conscious assistance systems. Among these results, we find that aggressive drivers are more predictable, but less responsive to feedback, than cautious drivers.

As is apparent, the contributions in the thesis are motivated by interactions between the fields of electrical engineering, signal processing, human-computer interaction, cognitive science, and psychology. Throughout this thesis, we will move from basic, controlled laboratory experiments to understand the nature of driver behavior and intention, back and forth to real-world test-beds and real-time algorithms to assess interactive systems in naturalistic settings. The ultimate goal in all of this work is to move towards improved performance of human behavior and intent inference systems.

Among the novel findings in this thesis:

- For the first time, in complex environments, we demonstrate the unique existence of *early head movements* in goal-oriented gaze shifts. (Ch. 3)
- For the first time, in complex environments, we demonstrate the existence of *sequential effects* on driver response times and on driver errors (i.e. unintended behaviors, or pedal misapplications). (Ch. 4)
- We conclude for the first time that head pose is *more predictive* than eye gaze of a drivers intention to change lanes. (Ch. 5)
- We show for the first time that aggressive drivers are actually *more predictable*, but *less responsive*, than cautious drivers. (Ch. 6)

A significant portion of this work included the development and engineering of several test-beds and frameworks:

- Developed the LISA-S driving simulator test-bed for basic, controlled laboratory studies of human behavior, interactivity, and cognition. (Chs. 3 and 4)
- Developed a novel framework, BRAVVO, for driver attention estimation incorporating noisy viewer and view observations. (Ch. 3)
- Developed a real-time on-road intent inference system in the LISA-X test-bed, demonstrating real-time intent inference for the first time in the literature. (Ch. 5)
- Proposed novel metrics for driver style assessment. (Ch. 6)
- Demonstrated how a novel framework for V2V communications, LACASA, could harness intent prediction to improve safety. (Ch. 6)
- Developed a novel heads-up display to assist drivers and improve safety, while minimizing distractions through driver state awareness. (Ch. 6)

The following chapter details a history and review of the literature related to the prediction of human behaviors, specifically in vehicular environments. In particular, we find that the inference of human intent is a useful cue for vehicular trajectory prediction. One of the most useful indicators of intent is the visual search, which we explore further in Chapter 3. We find first that the interaction of head and eye gaze foretells whether a visual search is premeditated and intent-related. Further, we propose a system to tightly integrate environmental sensing with driver sensing to estimate the driver's attentive state and target. In Chapter 4 we take a step deeper into the cognitive state of the driver, to understand how driver behaviors are affected by recent actions. We find in particular that certain sequences of cues or responses tend to prime the driver, significantly altering their response behavior and resulting in unintended actions.

This leads us to ask which cues are the most indicative of a driver's intentions, which we explore in the first part of Chapter 5. We then explore the real-world performance of such a real-time, cost-conscious intent inference system. Indeed, style and interactivity play a big role in these systems, leading us to an analysis of different manners of feedback in Chapter 6. We end with a view towards collaborative vehicles of the future, in which the potential for anticipatory communication between vehicles and infrastructure could lead to greatly improved safety.

Chapter 2

Review of Driver Behavior Predictions

The first automobile accident may have occurred as early as 1770, in what were then very slow-moving vehicles. The first recorded *automobile fatality* did not occur until 1869, when a passenger was thrown from a relatively fast steam-powered carriage, as the driver jolted the vehicle around a turn. *Human* error, as opposed to any mechanical factor, has been cited as a primary cause of this accident. Indeed, human error had even been cited in the first railway fatality as early as 1830. [56]

Nearly two centuries later, human errors are still at the foundation of many accidents in every form of transportation, including vehicles, trains, ships, and planes. As technology has evolved, systems have been developed to help mitigate these errors and dangerous situations. In automobiles, safety systems started out from simply requiring brakes to be installed on automobiles, and have evolved into much more advanced technologies in modern Driver Assistance Systems.

However there are still over 30,000 deaths and 1.2 million injuries yearly on roadways in the U.S.; up to 80% of these are due to driver inattention [6], or as a result of unintended maneuvers [145]. This has motivated researchers to develop new ways to assist drivers and prevent dangerous situations.

Drawing upon fundamental research in *human* behavior prediction, recently there has been a focus on how to predict *driver* behaviors. In this chapter we review the field of driver behavior and intent prediction. The aim of a driver behavior prediction system is to forecast the trajectory of the vehicle prior in real-time, which could allow a DAS to compensate for dangerous or uncomfortable circumstances. Though we present some opportunities for DAS feedback design, we focus mainly on the research in understanding driving behaviors.

This field is heavily influenced by research into driver modeling, which is broad in scope and seeks to characterize all aspects of human drivers, from cognition to operations. Several good reviews of driver modeling research can be found in recent literature [169, 114, 165]. Here we focus on those studies that attempt to understand any and all of those components necessary to identify and predict *patterns* of behaviors.

2.1 Modeling and Understanding Driver Behaviors

To predict the trajectory behavior of a driver, we must first ask whether the behavior is intended or unintended. An unintended behavior may be difficult to predict well before-hand, as in these cases the driver loses control of the situation at some point. In Chapter 4 we will explore some potential causes of unintended behaviors, in which the driver commits to a particular intentional action, but nevertheless performs a different action. Some other commonly cited causes for unintentional behaviors include distractions and workload [6], multi-tasking [109], and fatigue [213], among others.

Intentional maneuvers may be planned on a critical, tactical, or strategic timescale. These timescales have been proposed by prior researchers [169, 165] in driver modeling. The critical, or operational, timescale, on the order of hundreds of milliseconds, is the shortest possible timescale for human interaction. Tactical, or short-term, timescales are on the order of seconds, and encompass many successive critical operations. Finally, the strategic, or long-term, group is associated with minutes or hours of prior planning.

These timescales are widely used in the driver modeling literature [169, 165], where most driver models tend to focus solely on the critical, operational aspect of driving. Very detailed models of operational behaviors have been developed, integrating detailed information about the vehicle and road into understanding how drivers behave [114]. In the following section we will examine more closely the evolution and motivations of a maneuver on all these time scales.

The motivations for any particular driving maneuver can be understood from a driver's desire for (a) *safe* (no accidents) and (b) *comfortable* (e.g. not too slow, not too close) guidance to (c)*a destination*, given the dynamics of the vehicle (model), driver (style), and environment (other vehicles and obstacles, weather). There are a number of studies focused on intent inference of other objects in the environment [68, 87, 7], however we focus on inference of the driver in the ego-vehicle.

For the purposes of this research, we will define a situation as an interaction of any of the three components, the driver, vehicle, and environment. There is certainly feedback amongst these components. We can this explore how planned maneuvers develop with these driver-centric interactions in mind, along each of the time-scales above.

2.1.1 Operational Maneuvers

Operationally-planned intentional maneuvers are a generally a result of a driver's desire to remain safe in following the rules of the road (posted speed limit, road curves, etc), while carefully operating the vehicle within its limits. As mentioned above, many operational models for driver behavior have been proposed, such as [202]; these have been surveyed by Macadam [114] and Plochl [165].

Other instances of operationally-planned intentional maneuvers include unsafe situations, and may involve sudden braking or swerving to avoid any danger. The targeted execution of critically-planned maneuvers is usually in the steering and pedal actuators, where the driver has direct operational control.

Several studies monitor and detail the operational actions of drivers in these critical situations [187, 117, 228], some with an eye toward prediction [167, 125, 37, 185] and taking corrective actions in case of emergencies [2, 20]. Due to the proliferation of good surveys and literature on operational maneuvers [169, 114, 165], we instead focus more on tactical and strategic maneuvers below.

2.1.2 Tactical Maneuvers

Tactically-planned intentional maneuvers are usually motivated by an uncomfortable situation, or occasionally by a recently modified destination goal of the driver. These are of particular interest, as in these cases the driver still has time to react to feedback from an assistance system. The target space of tactical maneuvers are groups of operations; i.e. lane changes, turns, or stops. A driver may change lanes because the surrounding situation is not optimal for the driver, or because a desired exit is coming up soon. In many cases it may be useful to know the upcoming tactics of the driver in order to assist them.

There have been several research thrusts into modeling of driver tactics [159, 168, 181, 186, 126], or in historical analysis of drives to segment individual tactics [105, 80, 33, 184, 64, 124, 194, 127, 93, 203]. These models and algorithms focus on identifying maneuvers after the completion of the maneuver.

In the case of predictive models, we can separate out two broad categories of studies. In the first case, are those which target the prediction of single tactics, shown in Table 2.1. Targets include lane changes, turns, or stopping maneuvers. These studies tend to make use of both discriminative algorithms, such as Support Vector Machines, as well as generative methods, such as Bayesian Nets or Hidden Markov Models. The inputs to the models generally include vehicle data, as well as some measure of the surround. To adequately measure driver intention, we note that there are several studies that incorporate measurements of driver behaviors [120, 150, 30, 121, 49]. These studies include more information useful for driver intent inference, and tend to outperform the studies which attempt to predict maneuvers with no direct measurement of the

Table 2.1: Selected Studies in Single-Target Driver Tactic Prediction (Short-term)

Paper	Target	Inputs	Algorithm	Loc	Dat	Num	TPR	FPR	Time
[71]	S	CAN, ECG	DBN + HMM	S	I	5	-	-	-2.6
[100]	S	CAN	HMM, SLDS	R	N	1	-	-	-
[120]	B	CAN, ACC, Foot, Head	RVM	R	N	28	80	20	-1
[150]	TR, TL, LK	CAN, GPS, Head	Conceptual Fuzzy Sets	R	-	7	89.7	-	0
[30]	TR, TL, LK	CAN, Head, Hands	RVM	R	N	1	76	20	0**
[67]	LC, LK	CAN, Lane	Markov chain, Kinematics model	R	I	1	-	-	0
[121]	LC, LK	CAN, ACC, Head, Lane	RVM	R	N	3	90	4	-2.5
[99]	LC, LK, dLC	CAN	HMM	S	I	10	100	-	0
[183, 180]	LC, LK	CAN, ACC, Lane, SWA+	Clustering	R	N	9	77	5	+1
[35]	LC, LK	CAN, SWA+	Dynamic belief networks	-	-	-	-	-	-1.5
[42]	LC, LK	CAN, Lane	RNN, FFNN, SVM	S	N	10	100	10	-1.5
[49]	LC, LK	CAN, ACC, Lane, Head, SWA	RVM	R	N	15	70	0.2*	-2.5

Targets: (S)stop, (B)brake, (TR)turn right, (TL)turn left, (LC)lane change, (dLC)dangerous lane change, (LK)lane keep.
Inputs: (CAN)vehicle params, (GPS)location, (ECG)heart rate monitor, (ACC)forward-radar, (SWA)side-radar, (SWA+)surround-radar, (Head,Hands,Lane)positions.
Algorithms: (DBN)Dynamic Bayesian Networks, (RVM)Relevance Vector Machine, (HMM)Hidden Markov Model, (SLDS)Switching Linear Dynamic System, (RNN, FFNN)Recurring / Feed-forward Neural Net.
Loc: Location - Real-world or Simulator.
Dat: Data Collection - Naturalistic or Instructed.
Num: Number of Subjects.
TPR: True Positive Rate., *FPR*: False Positive Rate. (percentages)
Time: Time (seconds) from prediction to maneuver (negative=before, positive=after start of maneuver).
* - reported in false positives per second.
** - time before entering intersection, not before maneuver.

driver behavior. A detailed description of one of these systems [49] is included in Chapter 5.

In Table 2.2, we show those studies focused on multi-target prediction of driver behavior. These tend to exclusively rely on generative models, such as Hidden Markov Models, especially since they are better suited to multi-target inference. Of note, several studies [152, 142] include explicit information of driver behavior, and by directly inferring driver intent demonstrate better performance, earlier before the maneuver.

These studies demonstrate the value of incorporating direct measurements of driver behavior, and thereby direct inference of intent, into the prediction of tactically-planned maneuvers. The measurement of behavior allows for the detection of behaviors associated with planning for the maneuver - behaviors such as visual search and preparatory hand or foot movements.

Table 2.2: Selected Studies in Short-term Multi-Target Driver Tactic Prediction (Short-term)

Paper	Target	Inputs	Algorithm	Loc	Dat	Num	TPR	FPR	Time
[36]	A, B, LK, LCL, LCR	CAN, ACC, Lane	Dynamic belief networks	-	-	-	-	-	-
[21]	TR, TL, TU, RO, B	CAN	HMM	R	I	20	-	-	-
[97]	NS	GPS	Markov	R	N	100	90	-	0
[152]	P, TR, TL, LCL, LCR, ST, S	CAN, Eye, Lane	HMM and CHMM	R	N	70	66.5	-	-1
[142]	FL, FC, P	CAN, ACC, Eye, Lane, SWA+	PWARX, Clustering, Grammatical Inference	S	N	5	70	-	-2.5
[15]	LCL, LCR, TL, TR, LK	CAN, GPS, Map*	HMM	R	N	NR	69.6	14.2	+2

Targets: (A)accelerate, (S)stop, (B)brake, (TR)turn right, (TL)turn left, (TU)u-turn, (LCL)lane change left, (LCL)lane change right, (LK)lane keep, (RO)roundabout, (NS)next segment in map, (P)passing, (ST)start, (FL)follow-long, (FC)follow-close, .
Inputs: (CAN)vehicle params, (GPS)location, (ECG)heart rate monitor, (ACC)forward-radar, (SWA)side-radar, (SWA+)surround-radar, (Head,Hands,Lane)positions.
Algorithms: (DBN)Dynamic Bayesian Networks, (RVM)Relevance Vector Machine, (HMM)Hidden Markov Model, (SLDS)Switching Linear Dynamic System, (RNN, FFNN)Recurring / Feed-forward Neural Net.
Loc: Location - Real-world or Simulator.
Dat: Data Collection - Naturalistic or Instructed.
Num: Number of Subjects.
TPR: True Positive Rate, *FPR*: False Positive Rate. (percentages)
Time: Time (seconds) from prediction to maneuver (negative=before, positive=after start of maneuver).
* - Advanced map marked with lanes and intersections.

2.1.3 Strategic Maneuvers

Finally, strategically-planned intentional maneuvers are motivated by the destination goals of the driver, and sometimes by comfort as well (e.g., fastest route, no tolls). Given the starting point, the target space of strategic planning is the desired destination and route of the driver. The route plan gives away important details as to the planned trajectory of the driver.

In Table 2.3 we list several studies related to the prediction of strategically-planned maneuvers. Most of these studies target the route or destination of the driver, and include GPS and Map data as inputs. One issue with these studies is the lack of a common evaluation metric; given that they are evaluated in their own manners it is difficult to compare performance. Ultimately, however, these studies could be used as inputs to tactic and operation prediction models [231].

2.2 Sources of Variation

In order to improve the performance of intended trajectory prediction, it becomes important to characterize the variations in the evolution of maneuvers. A number of variables interact

Table 2.3: Selected Studies in Driver Strategy Prediction (Long-term)

Ref	Target	Inputs	Algorithm	Dataset	Sample Performance
[231], [230]	Destination, Route, Turn	GPS, Context, Map	MDP, etc	*CMU Taxi-driver Set*: 100k miles of Taxi Driver GPS Data	6.7% destination error with 10% observed; 82.6% route distance match; 93.2%turn detection accuracy
[190]	Destination, Route	GPS	HMM	On-road, naturalistic, one-month of data	98% overall-prediction of next segment; 72% in unforced situations
[98]	Destination	GPS	Bayes	Naturalistic	2km error halfway through route
[60]	Route	GPS	Nearest neighbor	*Microsoft Multiperson Location Survey*: 252 subjects, 2+weeks, 10k+ routes	After first mile: 13% of trips predicted, 30% of repeat trips; 70% of repeat trips within top-10 NNs
[38]	Destination	GPS, Map Meta-data	Markov + GMM	Naturalistic	4.29km destination error
[205]	Destination, Route	GPS, Map	String matching	Real, 14 drivers	-
[196]	Destination, Route	GPS, Context, Map	Nearest neighbor + Bayes Net	Real, 1 driver, 4 months	Correct destination predicted during 67% of the trip
[92]	Route	GPS	Hierarchical Tree-based Map Matching	Real, UCR vehicle database	-
[149]	Route	GPS	Markov chain	*INFATI Data-set*: Several cars, One month	50% correct cell, 84.5% with neighbor cell included

with and affect the evolution of intended maneuvers. These variables include the following:

- attentive state and the dynamics visual search
- recency, or the effect of historical actions on current behaviors
- individual styles of drivers, whether aggressive, cautious, or otherwise
- cognitive distraction (effects on intent studied in [227])
- age and experience, especially in older and younger drivers
- drowsiness and fatigue
- emotional state
- multitasking scenarios

- weather, road conditions, and other environmental factors

In this dissertation we focus on improving *tactical* maneuver intent prediction through a deeper understanding of several of these variables: **attention and visual search** in Chapter 3, **recency and sequential effects** in Chapter 4, and **driver style** in Chapter 6.

2.3 Toward Real-world Performance

One goal of research in intelligent vehicles is to develop intelligent advanced driver assistance systems (ADAS), that can use pervasive sensors and interfaces to enhance the driver's safety and comfort. By predicting a driver's intention to perform a particular maneuver, an ADAS could actively engage in avoiding dangerous situations. Prior research studies, such as those listed above, have identified the possibility of predicting a driver's intent to change lanes. Among these research studies, various maneuvers, sensors, algorithms and environments were used to predict intentions. However to our knowledge such a system has not yet been implemented and deployed in a moving vehicle.

Several questions remain unanswered in the existing literature, which we tackle in Chapter 5. Foremost among them is the question of how performance translates from basic laboratory classifier analysis, into real-world driving performance. We also seek to realize the optimal real-world sensor configuration for lane change intent prediction, as the sensor requirements could have a significant cost on any production system. This question is intertwined with an analysis of optimal timing for an intent detection system, as different sensors could provide important information at various times leading up to the lane change. Furthermore, it is important to characterize and understand how feedback from a real-time ADAS could interact with driving behaviors (Chapter 6). This dissertation moves toward answering these questions and providing a basis for future developments in real ADAS systems.

As demonstrated in related literature [58], the applications of these studies extend well beyond vehicular drivers. In many cases such as air traffic management, airplane pilots, and air traffic controllers would also benefit from advanced intent inference-based interactive systems [58]. Additionally, any task-oriented behavior involving a human machine interface, such as smart meeting rooms or assistive living situations, could benefit from intent inference and behavior prediction.

Chapter 3

Visual Search and Intent

3.1 Attention Shifts

Visual search is clearly one of the most important indicators of intent. However, visual searches can be goal-oriented, or stimulus-driven. In the first part of this chapter, we demonstrate that the interaction of head and eye gaze can help foretell whether a visual search is premeditated and intent-related.

3.1.1 Introduction

Analysis and understanding of human behavior, particularly of head and eye gaze behavior, has been a subject of interest for many years [41] in Cognitive Psychology and Neurophysiology; yet a full understanding of the causes, dynamics, and control mechanisms of head and eye movements is still a subject of active research [59]. More recently, researchers in the Computer Vision and Artificial Intelligence community have increasingly sought to incorporate information gleamed from patterns of human behaviors into intelligent Human-Machine Interfaces (HMIs) [207, 161]. The confluence of these two research paradigms allows for a deeper understanding of fundamental human cognition and behavior, as well as how to interpret and modify that behavior, if necessary.

Active perception of human operators by "intelligent environments" can allow assistance systems to improve performance and even help to avoid dangerous circumstances. Every year traffic accidents result in over one million fatalities worldwide [158], and around 40,000 in the U.S. alone. Of those, an estimated 26,000 are due to some form of driver inattention [106, 6]. Recent advances in active vision and machine intelligence have resulted in incorporation of camera-based driver analysis into Driver Assistance Systems [207, 208], to predict future behaviors or inattention, and thereby counteract poor driving behavior. By detecting patterns of body language in critical situations, such as the time prior to a lane change, these systems are able to predict the context of the situation.

In other interactive environments such as intelligent Command-and-Control centers or intelligent meeting rooms, systems monitoring the participants or operators, could provide assistance based upon the subjects' body language. This may help reduce distractions and help improve performance of whatever task is being performed. Landry et al. [103] discovered that certain patterns, or "gestalts", of aircraft on a radar screen drew the attention of the air traffic controllers due to their location, though they were not relevant to the task. An air traffic control training manual from the FAA [26] states that "even in low-workload conditions, distractions can clobber short-term or working memory." An assistance system could mitigate such dangerous situations by detecting the context of the attention shift and providing warnings when the attention shift is not task-related.

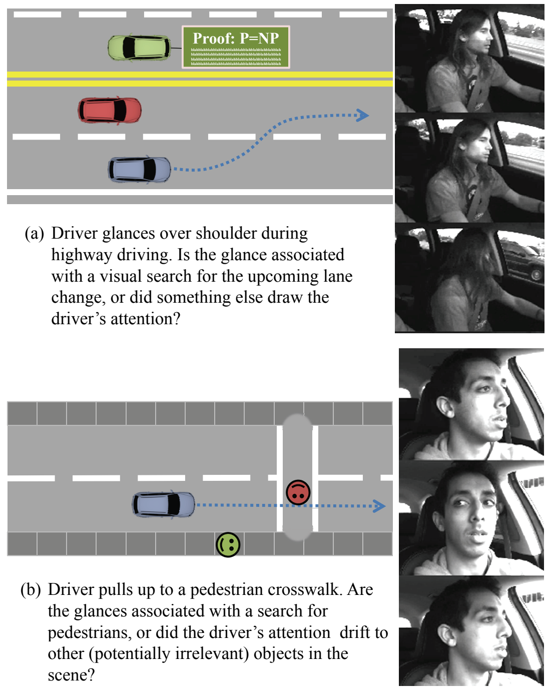
Those Intelligent Environments that must assist humans in time- and safety-critical situations would clearly benefit from knowledge of the human's cognitive state. Such information could allow the system to detect, for example, whether the driver is distracted or actively engaged in a relevant task. Information about salient properties of the environment [82] could help to understand attention, however such systems are not good at predicting the actual focus of attention [11], nor do they give insight into the cognitive state of the driver.

Studies into the dynamics of head and eye gaze behavior tend to hint that attention is linked to the types of movements of head and eye. A number of studies have shown that when presented with a stimulus in the field of view, the head tends to lag behind the eye during the gaze shift [12, 65]. Others have found that early head motions, with respect to the saccade, are associated with target predictability, location, propensity to move the head, and timing [223, 128, 176, 59, 61, 94]. These studies occurred in very controlled, arguably unnatural environments; more recent studies into "natural" tasks involving grasping, tapping, or sitting in a lecture [76, 160, 122, 47] have suggested that task-oriented gaze shifts may be associated with early head motions. These studies suggest the hypothesis that the interactive dynamics of head and eye movements may encode information about the type of attention shift, and thus the cognitive state, of the human subject.

These preliminary investigations led to two related questions in more complex environments such as a driving scenario:

1. Is it possible to observe differences in human eye-head dynamics during different styles of attention shifts?
2. Is it possible to extract and use those cues to help identify or learn information about the subject's cognitive state?

In the following sections we detail novel findings regarding the interaction of eye gaze and head pose dynamics under various attention-switching conditions in more complex environments, and while engaged in safety-critical tasks such as driving. In particular, we find that sudden,



(a) Driver glances over shoulder during highway driving. Is the glance associated with a visual search for the upcoming lane change, or did something else draw the driver's attention?

(b) Driver pulls up to a pedestrian crosswalk. Are the glances associated with a search for pedestrians, or did the driver's attention drift to other (potentially irrelevant) objects in the scene?

Figure 3.1: Motivation for understanding head and eye dynamics in attention shifts. In safety-critical situations such as driving, knowledge of the driver's attentive states would allow an Intelligent Driver Assistance System to help the driver avoid crashes or dangerous situations. In both cases, the driver's glances could be associated with a task-relevant visual search, or with a stimulus-driven potential distraction. We propose that by detecting the relative latencies of eye-head dynamics, an intelligent driver assistance system could distinguish the two cases and assist the driver if necessary.

bottom-up visual cues in the periphery evoke a different pattern of eye-head yaw movement latencies as opposed to those during top-down, task-oriented attention shifts. In laboratory vehicle simulator experiments, a unique and significant ($p < 0.05$) pattern of preparatory head motions, prior to the gaze saccade, emerges in the top-down case. In contrast to the findings of Land [102], we also observe the early head motion patterns during task-oriented behavior in real naturalistic driving scenarios.

We ultimately aim to understand whether one can detect whether the human subject had prior knowledge of the scene or the task upon which they are focusing. That contextual knowledge may be useful to real-time interactive assistance systems. Specifically, it could provide semantic hints such as whether there is a distraction or unplanned stimulus in a scene, or what kinds of goals the human subjects are pursuing. We validate our simulator-based findings in qualitative analysis of naturalistic real-world driving data. These results show that measurements of eye-head dynamics are useful data for detecting driver distractions, as well as in classifying human attentive states in time-critical and safety-critical environments.

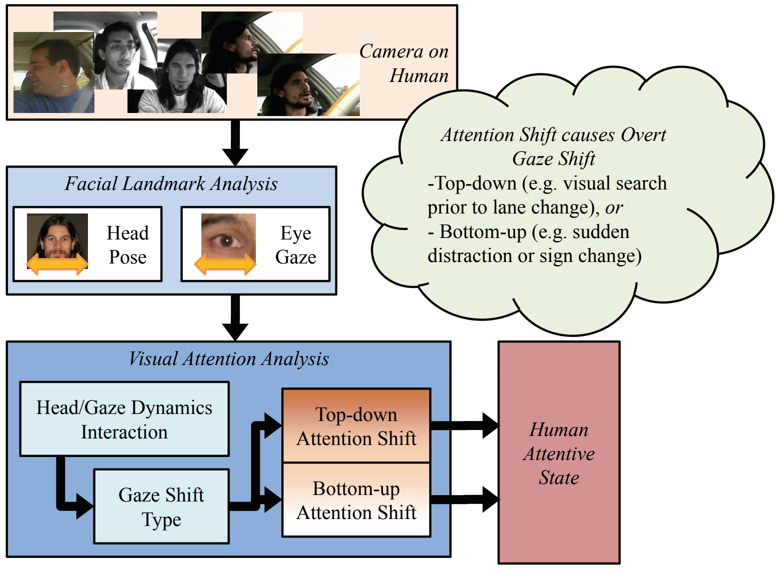


Figure 3.2: Flowchart of proposed approach for visual attention shift analysis. Upper body and facial landmark analysis is used to detect relative latencies of eye and head motion after an attention shift. This can be used to classify the gaze behavior, giving important clues to the attentive state of the human driver.

3.1.2 Related Research

Human Behavior Analysis and Prediction

A great amount of recent research has been in the detection of human behaviors. Many of these examples use patterns of human behavior to learn about the scene or predict future behaviors [179, 154, 162]. Such predictive systems could be used in many different human-machine interfaces, from air traffic control [103] to assistive living [161]. For the following research we are motivated by time- and safety-critical situations in driving scenarios, specifically in behavior analysis and prediction in vehicles.

Recent research has incorporated sensors looking inside the vehicle to observe driver behavior and infer intent [207, 208]. Bayesian learning has been used to interpret various cues and predict maneuvers including lane changes, intersection turns, and brake assistance systems [30, 121, 120]. More recently head motion has been shown to be a more useful cue than eye gaze for discerning lane change intentions [48].

The assumption made in all of these systems has been that head motion or eye gaze is a proxy for visual attention. In other words the system tries to measure head motion given that the driver is likely paying attention to whatever they are looking at, in whichever direction they are looking. The system then infers that because their attention is in a certain direction, they must have goals associated with that direction. For example, a driver may look left prior to changing lanes, as a direct result of their need to be attentive of vehicles in the adjacent lane.

Gaze Behavior in Complex Environments

A gaze shift may or may not be associated with a particular goal. The broad question of "why people look where they look" is a subject of research for cognitive psychologists, neuroscientists, and computer vision researchers alike.

A significant amount of research in psychology has examined whether such visual searches are guided by goals (such as the goal of changing lanes), or by external stimuli [82, 146, 163, 91, 178]. These stimuli may include visual distractions, which could pop up in a scene and thereby attract the attention of the observer. For the most part any visual search is presumed to be guided by some combination of a goal-driven and stimulus-driven approach, depending on the situation [221].

Itti et al. [82, 146, 163] and others [225, 115] have made inroads in developing saliency maps that model and predict focus of attention on a visual scene. Initial models were based on "bottom-up" based cues, such as edge and color features. However it was found that "top-down" cues may be more influential; in other words the *context* of the scene and the goals of the human subject are crucial to determining where they look. Hayhoe et al. [91] determined that even in the presence of potentially dangerous distractions, in complex environments gaze is tightly coupled with the task. Several other works have similarly concluded that in natural environments, saliency does not account for gaze, but task and context determine gaze behavior [102, 178, 74, 224].

It is clear that in certain critical environments, distractions play an important role in attracting attention. Carmi et al. [27] show that dynamic visual cues play a causal role in attracting visual attention. In fact, perceptual decisions after a visual search are driven not only by visual information at the point of eye fixation but also by attended information in the visual periphery [53]. In certain cases these stimuli may affect task performance; Landry et al. [103] found certain unrelated gestalt motion patterns on radar screens drew the attention of air traffic controllers away from the task at hand. In the driving context there are many well-known cognitive and visual distractions that can draw the driver's attention [6, 44]. Recarte et al. [170] measured the number of glances to the mirror during a lane change, noting that visual distractions decrease the glance durations from by 70-85%. This result is well-aligned with more recent results indicating the limitations of drivers' multi-tasking abilities [109, 108]. Moreover, some suggest that visual distractions may even increase likelihood of "change blindness", a phenomena whereby a subject may look in a certain area and not see or comprehend the objects in front of them [52, 107]. In these cases, it would be useful to know whether a gaze shift is attributable more to irrelevant visual stimuli or to a specific goal or context.

Several studies have used eye gaze or head pose to detect the attention of the subject [13], or estimate user state or gestures [8, 118]. In this study, we instead use the interaction of eye gaze and head pose to determine the attentional state of the subject, and proceed to use that information as contextual input to event detection and criticality assessment systems. In the following sections we describe the recent research in eye-gaze interactivity analysis in overt attention shifts, and in the later sections we describe some supporting experiments.

Overt Attention Shifts

Attention shifts can be of two kinds, or some combination of the two [221]. Top-down attention shifts occur when the observer has a particular task in mind. This task may necessitate a visual search of a potentially predetermined location, or a search of likely relevant locations. An example may be as simple as shifting one's attention from a television to a newspaper, after having turned the television off. There may also be learned tasks, such as the search for oncoming cars when crossing a road, as a driver in a vehicle or as a pedestrian at a crosswalk.

Bottom-up attention shifts are caused by interesting stimuli in the environment. These stimuli may include distractions or salient regions of a scene. For example, flashing police lights on the highway may draw unnecessary attention, as may an instant chat message popping up on the screen during a technical presentation.

Generally speaking, Pashler et al.[157] observes that "whereas exogenous attention control is characterized as stimulus driven, endogenous control is typically characterized as cognitively driven." Others have encoded such notions into computational models of attention [135].

In many cases an object in the scene may easily be a distraction in one instance, and part of the task at hand at another time. For example, in a classroom, the person standing in front of the blackboard may be the teacher during a lesson, to whom the student should be paying attention. On the other hand, the person in front of the blackboard may just be someone walking by, who is distracting the student from the task at hand. By classifying the student's interactive behaviors, we may be able to set the context for the attention shift and understand the state of the subject, whether cognitively-driven (top-down) or stimulus-driven (bottom-up).

Head and Eye Gaze during Overt Attention Shifts

Zangemeister and Stark [223] performed a controlled study of eye-head interactions and posited various conditions of different styles of eye-head movements. In their paper they found several styles of movements, depicted in Figure 3.3. Among the most pertinent of movements are those labeled "Type III," which include early or anticipatory head movement with respect to the gaze shift. They theorized that this behavior is associated with a repetitive, predetermined, or premeditated attentional shift, as is the case for any goal-directed attentional shift.

Morasso et al. [128] examined control strategies in the eye-head system, and observed that "The head response [to a visual target], which for random stimuli lags slightly behind the eyes, anticipates instead for periodical movements [of the target]." The implication is once again that for trained or predetermined gaze shifts, the head movement anticipates the gaze.

Indeed eye shifts can occur much faster than head movements and thus extremely fast

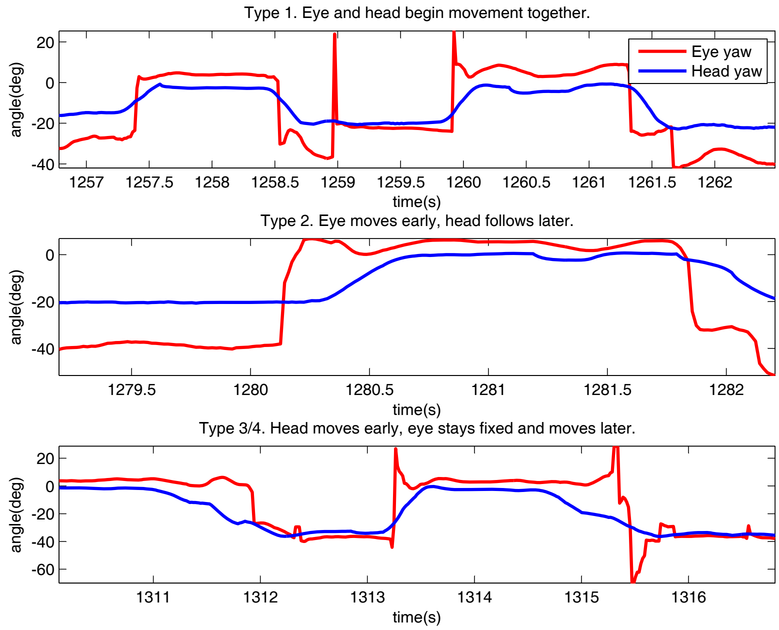


Figure 3.3: Examples of various interactions of head and eye movements, with type labels from Zangemeister & Stark, 1982. Note in certain cases eye gaze tends to move first, where in others the head tends to move first.

gaze changes can be made with eye shifts alone. Shifts of larger amplitudes would necessitate a head movement to accommodate the limited field of view of the eyes. Most likely, when a large visual attentional shift is about to occur, and the observer has prior knowledge of the impending shift, these studies imply that there may be some amount of preparatory head motion [160].

Fuller [61] and Freedman [59] each included thorough reviews of eye-head coordination and verified the same results seen above. Several variables including initial eye, head, and target positions, along with predictability, seem to affect the latencies of head movements [176]. A few studies have touched the possibility that attention and task can influence the dynamics of eye-head movements [34, 76, 123, 94]. These studies are well-controlled in laboratory environments but limited in their generalizability to natural environments and more complicated tasks.

In this study we venture to examine directly the effects of goal- vs. stimulus-driven attentional shifts on eye-head coordination. Further, by classifying the type of shift, we are able to propose a novel model to determine the cognitive state of the subject, which may prove useful to assistive human-machine interfaces such as Driver Assistance Systems.

In the following sections, we show that by extracting the yaw dynamics of eye gaze and head pose, it may be possible to identify those gaze shifts which are associated with premeditated or task-oriented attentional shifts, driven by "endogenous" cues. In each case, we find that a majority of endogenous, task-related shifts occur with an anticipatory Type III gaze shift. Based on these results and the studies listed above, we might further hypothesize that a Type III gaze shift could imply a task-related shift, and Types I or II are more likely to occur in conjunction with stimulus-related gaze shifts, associated with "exogenous" cues.

3.1.3 Methods

Participants

Ten volunteers consented to participate in the experiment, 9 male and 1 female. The participants were mostly in their 20s, with two in the 30s, and one over 40; with the female subject was near the median age of 25. Every subject had a valid driver's license, with varying driving experience level from novice to decades of driving experience.

Apparatus

The experiment was conducted in a driving simulator, with several additional features to facilitate the various conditions of the experiment. The simulator was configured as shown in Figure 3.4. A 52" screen was placed 3 feet in front of the subject, with a Logitech steering wheel mounted at a comfortable position near the driver. An additional secondary screen was placed to the left of the main monitor, in the peripheral vision of the subject.

The main monitor was configured to show a PC-based interactive open source "racing" simulator, TORCS [204]. The software was modified in several ways to make it a more appropriate driving simulator. The test track was chosen to be a two-lane highway/main road running through a city. Several turns on the driving track necessitated a significant slowdown in speed to maneuver through, keeping the driver actively engaged in the driving task. Additionally, the maximum speed of the "vehicle" was limited to appropriate highway speeds, in order to discourage excessive speeding.

In this experiment the track contained no other vehicles, to limit the complexity of interactions. The ego-vehicle and road parameters such as friction, vehicle weight, and others were fixed to approach real-world conditions. However they were also constrained in order to facilitate an easy learning process, as some subjects had never used driving simulators before.

A stereo-camera-based non-intrusive commercial eye gaze tracker, faceLAB by Seeing Machines [188], was set up just in front of and below the main monitor. It was appropriately placed not to obstruct the field of view of the driver. The system required calibration for each subject, which was performed prior to any data collection. Once calibrated, it output a real-time estimate of gaze location (pitch and yaw) and head rotation (pitch, yaw, and roll). These quantities were calculated and transmitted concurrently in "Minimum Latency" mode to the PC running the driving simulator. The gaze and head data, along with all the driving parameters such as distance traveled and lateral position, were automatically timestamped and logged to disk every 10 milliseconds.

Figure 3.4: Experimental setup of LISA-S testbed.

The secondary monitor showed various text messages as described below, depending on the condition. This display was controlled synchronously with the same PC which ran the simulator.

Design

Each subject was run in two conditions, *endogenous* and *exogenous*. In each condition the subject is cued in a different way to check the secondary monitor to find out whether to change lanes. These are demonstrated in the illustrative example in Figure 3.5.

The "endogenous" condition *(Condition 1)* of the experiment was intended to stimulate "goal-oriented" attention switching, such as the planned visual search of a driver checking mirrors prior to a lane change. Large overhead signs appeared at several constant, predefined locations around the track. After noticing these signs coming up in the distance, the subject should glance over at the secondary monitor. This monitor would be displaying a message, "Left Lane" or

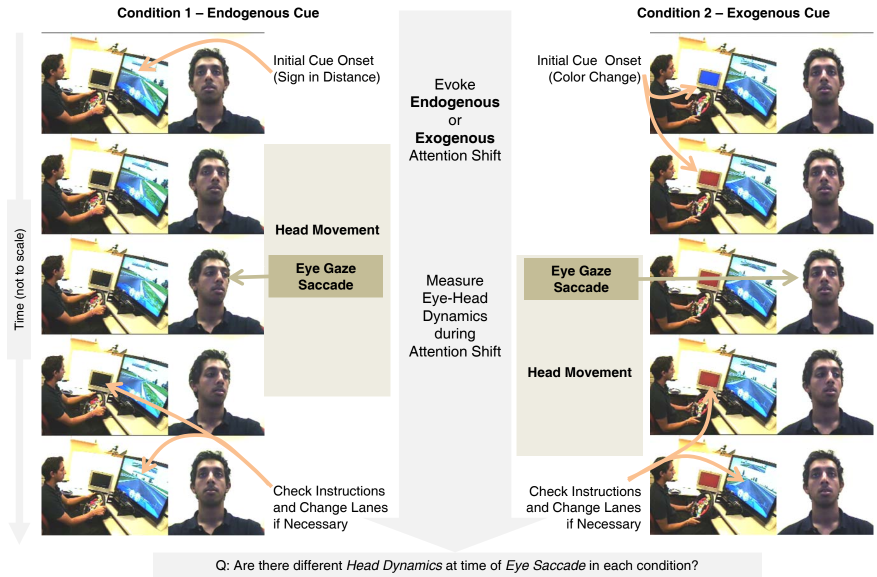


Figure 3.5: Illustrative (staged) example of experimental paradigm. In each cueing condition, we measure the differences in eye-head interactions during attention shifts to a secondary monitor, which the driver is required to check for instructions. In the "endogenous" condition, the driver is presented with a cue in the primary monitor and allowed to make a goal-oriented or pre-planned attention shift. In the stimulus-oriented "exogenous" cueing condition, the secondary monitor displays a sudden change in color, drawing the driver's attention to the target.

"Right Lane," indicating in which lane the driver should be. The subject was told to move to that lane by the time the overhead sign passed by. The duration of the cue appearance varied from 5 to 20 seconds. In this manner the subject was allowed time enough to *plan and initiate* the attention switch by herself; we label this as the "endogenous" cue condition (*Condition 1*).

The second condition was designed to evoke an unplanned "stimulus-oriented" attention switching response, as if the driver was suddenly distracted. In this condition the driver was also told to maintain the lane as best as possible. The cue to change lanes would come from the secondary monitor, whose entire screen would change color suddenly, at a set of predefined times unknown to the subject. This color change occurred in concert with a potential change of the text message, once again to either "Left Lane" or "Right Lane." Upon noticing the change, the driver was tasked with maneuvering to the appropriate lane as soon as was safe to do so. The subject was told that the colors were random, not correlated with the text, and would occur at random times. Thus the subject's attention switch was hypothetically initiated by some external cue in the peripheral field of view; we label this as the "exogenous" cue condition (*Condition 2*).

The "exogenous" cue would ideally have been a stimulus not associated with the task of driving. However it is difficult to collect enough data where the subject freely decides to attend to an irrelevant stimulus. By associating the stimulus-style cue with the secondary task of lane selection, it becomes possible to gather consistent data about how the subject responds to the stimulus-oriented cues such as sudden flashes, motion, or other unplanned distractions. This can then be directly compared to driver behaviors under the "endogenous" condition described above.

Each condition consisted of 10 minutes of driving, corresponding to 12 to 15 lane changes per condition. The order of the conditions was presented randomly, and to ensure comfort participants were offered breaks between conditions.

Procedure

Participants were told that the experiment pertains to naturalistic lane-keeping behavior. After calibrating each subject on the gaze tracker, the subject was given 5-10 minutes to acclimatize herself to the simulator. She was queried once she felt comfortable with the simulator, and her comfort level was verified by subsequently asking her to keep the vehicle in a single lane for at least 60 seconds.

For the remainder of the experiment, the subject was tasked with maintaining her current lane to the best of her ability. This allowed the subject to be actively engaged in the driving process throughout the experiment. They were instructed to respond to the cues in each condition, and if necessary, change lanes when safe to do so. This required them to check the secondary screen, or "side mirror," in order to decide which lane to move into. Though this is not entirely naturalistic, it is reminiscent of glancing at the mirror prior to lane changes to scan for obstacles, and participants had little difficulty following the instructions.

Analysis

The automatically logged datasets from each experiment were then processed to analyze the dynamics of head and eye movements leading up to the attention shift. For each condition, the appearance time of the cue was determined and the subsequent gaze shift was determined to be an example of a shift of interest. The secondary monitor was fixed at an angle of approximately 55° from the subject, so only gaze shifts which resulted in glances of that magnitude were considered.

In 34% of the cases in the entire experiment, cues appeared which were not followed by a tracked gaze shift of sufficient magnitude. This was caused either by a lack of response from the subject, or a lack of tracking confidence from the head and eye tracker. Occasionally the tracker would lose an accurate representation of the subject's head or eye movements, and this would be represented by a "validity" signal, output by the tracker. Whenever there was no valid gaze shift of sufficient magnitude following the appearance of a cue, the example was discarded from further analysis.

Fewer than 5% of the examples in the "exogenous" condition were discarded due to the unforeseen effects of high cognitive load. Occasionally a color change would appear while the driver was actively engaged in a sharp turn, either causing the driver to lose control of

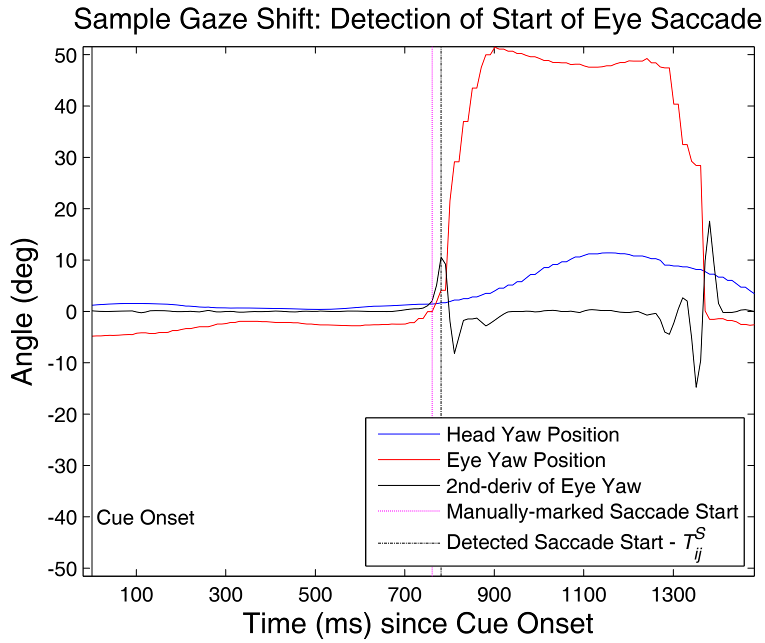


Figure 3.6: Sample data showing the procedure for detection of the start of the eye saccade. A local area around a manually marked point is searched for the maximum eye yaw acceleration. This point is labeled as the start of the saccade, T_{ij}^{S}. The dynamics of the head yaw, i.e. *position* and *motion*, at this point T_{ij}^{S} are extracted for further analysis.

the vehicle, or to shift their glance slightly without paying much attention to the color change. Where these effects were observed the examples were discarded, as in these cases drivers behaved inconsistently, most likely due to a task overload. Such effects could be the topic of further investigations, but examples were too few to discuss in detail in this study.

Each subject j had approximately 10 examples of each condition after the pruning step described above. For the remaining examples in both conditions, the exact location of the first gaze saccade to the secondary monitor was found in an iterative manner. This was done to avoid any false positives in an automatic saccade-detection procedure, given the somewhat noisy nature of the gaze data.

In the first step, the example was manually annotated as to the approximate time of the initiation of the gaze saccade, during the first gaze shift of approximately 55° in yaw (i.e., to the target monitor). Subsequently the point of maximum yaw acceleration in a local 50-millisecond window W around the annotated point T_L was found. This was calculated using a 5-tap Derivative-of-Gaussian filter to temporally smooth and find the second derivative of the gaze rotation signal. The maximum point of the second derivative was fixed as the location of the gaze saccade:

$$T_{ij}^{S} = argmax_{t \in W}(e''_Y[t])$$

where $e_Y[t]$ represents the yaw position of the eye at time t. An example of this detection procedure can be seen in Figure 3.6; we found that the local point of maximum eye yaw acceleration was a consistent labeling method for the following analysis.

Given the exact time of the gaze saccade, T_{ij}^S, two features based on head dynamics were calculated for each example i of subject j. The first was simply the value of the head yaw at the point of the saccade:

$$P_{ij} = h_Y[T_{ij}^S]$$

where $h_Y[t]$ represents the yaw position of the head at time t.

The second feature was the average yaw velocity in the 100ms prior to the saccade,

$$M_{ij} = \frac{1}{10} \sum_{t=T_{ij}^S-100ms}^{T_{ij}^S} h'_Y[t]$$

P_{ij} and M_{ij} thus represent the yaw position and motion features of each example i of subject j. These features were chosen to capture whether the head was moving in a preparatory motion prior to the saccade, and by how much.

Based on the features above, several additional statistics were calculated:

$$\widetilde{P_j^1} = median_i(P_{ij}|\text{Condition1})$$

$$\widetilde{P_j^2} = median_i(P_{ij}|\text{Condition2})$$

$$\widetilde{M_j^1} = median_i(M_{ij}|\text{Condition1})$$

$$\widetilde{M_j^2} = median_i(M_{ij}|\text{Condition2})$$

3.1.4 Results and Discussion

Each trial was characterized by one main independent variable: the *condition*, either 1-endogenous or 2-exogenous. Two related variables were measured, namely the head yaw P_{ij} and head motion M_{ij} at the point of the gaze saccade.

The *order* of presentation was also randomized, but there is no interaction between order and condition ($p > 0.5$), so we collapse across the order groups.

A main effect of the condition is found both in yaw (mean(P_{ij}^1) $= 6.68°$, mean(P_{ij}^2) $=$ $1.54°$, $F(1,9) = 9.44, p < 0.05$) as well as in motion (mean(M_{ij}^1) $= 41.36°/sec$, mean(M_{ij}^2) $=$ $22.12°/sec$, $F(1,9) = 11.23, p < 0.05$). Figure 3.7 shows the distribution of all the data at the point of the saccade, clearly demonstrating that the yaw and motion both tend to be greater at the point of saccade in the endogenous condition.

Figure 3.8 shows the average head yaw over all the examples time-aligned to the saccade location. The head yaw in Condition 1, shown in red, demonstrates a clear increase in head motion as much as 500 ms prior to the saccade, with the greatest motion occurring around 200 ms prior. Head yaw in condition 2 begins clearly increasing in earnest only 50-100ms prior to the saccade.

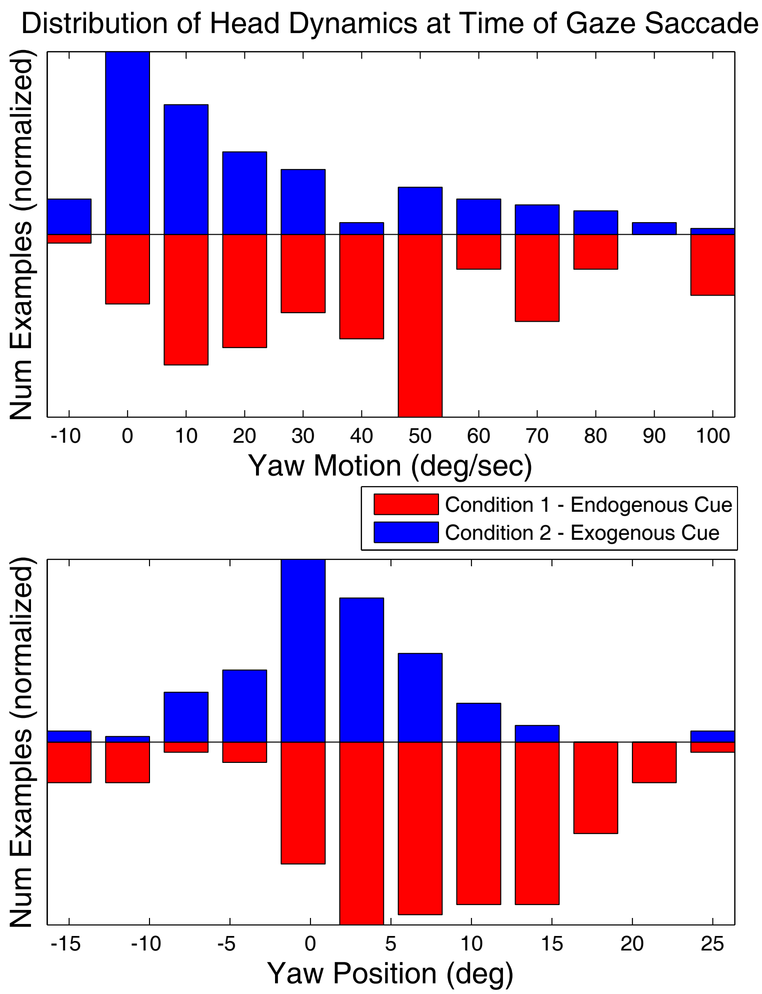


Figure 3.7: Overall distribution of head yaw position and head yaw motion at the time of the eye gaze saccade for each condition, including all examples of all subjects.

Prior studies [223, 160, 59] have noticed early head motions up to 200 ms in advance of the saccade. The results of these experiments are thus in mild agreement, although it is apparent that in preparation for task-oriented gaze shifts in the driving situation, the head movement may begin as much as 500 ms in advance of the saccade.

To discount the effect of each subject's outliers, the subject-wise medians of these data, $\widetilde{P_j^{[1,2]}}$ and $\widetilde{M_j^{[1,2]}}$, were also calculated to analyze in further detail. Given the limited dataset size, the two pairs of statistics were then analyzed using the Wilcoxon Signed-Rank test, a non-parametric extension of the t-test. This analysis showed significant differences in both cases; $\widetilde{P_j^1} = 6.2° > \widetilde{P_j^2} = 1.1°$ $(Z = 2.80, p = 0.0051 < 0.01)$, and $\widetilde{M_j^1} = 45.3°/sec > \widetilde{M_j^2} = 13.9°/sec$ $(Z = 2.50, p = 0.0125 < 0.02)$. Figure 3.9 shows the differences in the median statistics computed above, for $\widetilde{P_j}$ and $\widetilde{M_j}$. The endogenous cueing condition, corresponding to a "task-oriented"

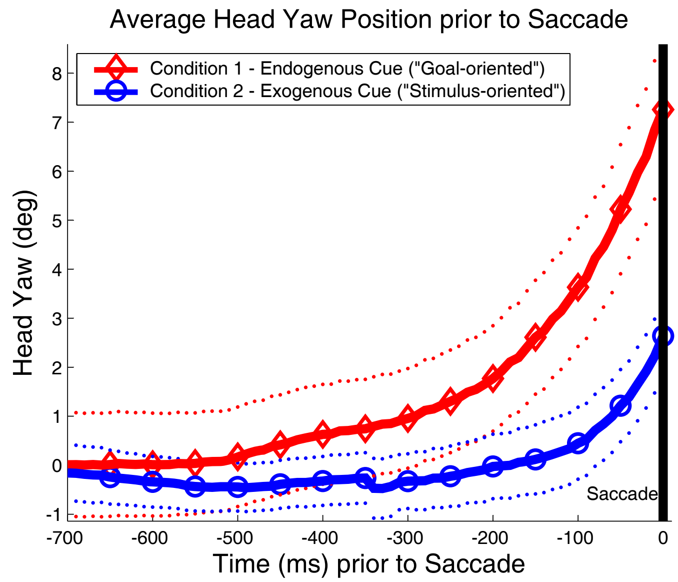


Figure 3.8: Average Head Yaw prior to Eye Gaze Saccade under each condition of the experiment, aligned to the position of the saccade. Dotted lines show the variance of the overall data. In condition 1, a clear pattern emerges of early head movement, beginning 0.5 seconds prior to the actual gaze shift. This early head movement is much less evident in condition 2.

attention shift, demonstrates a clear pattern of greater and earlier head motion just prior to the saccade.

Figure 3.10 demonstrates the relative timings of the gaze saccade after the appearance of the cue. The exogenous cue tends to attract attention very quickly, and most gaze shifts are made within 500 ms of the cue. This is in clear contrast to the endogenous cue, which as expected, results in the subject shifting gaze in a pre-planned manner. This separation shows the experimental conditions elicited the two different styles of gaze shifts appropriately.

In order to characterize the time course of the head movements with respect to the gaze shift, we can measure the timing of the first "significant" head motion. In Figure 3.11, the histograms of the first head motion (where head motion goes above a fixed threshold of $17°/sec$) is shown for each condition. This is found by searching in both directions from the point of the gaze saccade, to determine where the head motion first exceeds the threshold. The endogenous cueing condition can be observed eliciting a greater portion of early head motions.

Kinematics of Eye and Head Movements

Here we measure a number of other variables that could have influenced the onset of early head motion due to the kinematics of the eye-head motion.

Prior studies in controlled conditions have determined that the amplitude of the gaze shift had a significant impact on the amount of head movement prior to the saccade [59], as

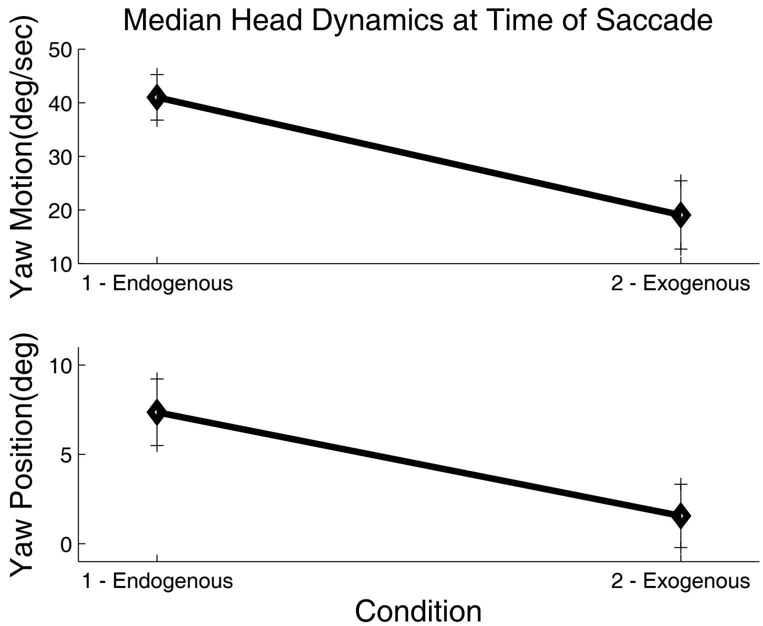


Figure 3.9: Distribution of subject-wise median head yaw position and head yaw motions at the time of the eye saccade. Error bars represent standard error of the median.

larger movements tended to correlate with early head motions. Another critical variable was found to be the predictability of the target location, which also positively influenced the early movement of the head. However in this study we have fixed the target location, and the amplitude as determined by initial position is generally constant as well. To verify this, we compared the starting yaw positions in Figure 3.12 and found no reliable differences in starting position ($EyeInitialYaw : F(1,9) = 2.84, p = 0.09, HeadInitialYaw : F(1,9) = 0.95, p = 0.33$). The subject is always aware of the location of the target, in both the endogenous and exogenous cueing cases. In spite of the constant target and shift amplitude, we still find variations in the amount of early head movement, correlating with the type of cueing condition; whether driven endogenously by top-down motor commands, or exogenously by bottom-up stimulus responses.

Bizzi et al.[18] also report in controlled environments that predictive movement conditions include lower *peak velocities* and *movement durations* than triggered conditions. To analyze these effects, we measured the peak velocities and saccade durations for both eye and head movements in Figures 3.13 and 3.14. In contrast to the earlier studies, the peak eye velocities in the endogenous condition actually trend toward being significantly greater than in the exogenous condition ($F(1,9) = 5.16, p = 0.02$). In all other cases, there were no significant differences ($PeakHeadVelocity : F(1,9) = 0.14, p = 0.71$, $EyeMovementDuration : F(1,9) = 1.24, p = 0.26$, $HeadMovementDuration : F(1,9) = 1.10, p = 0.30$). We are thus not able to observe slower peak velocities or movement durations in the endogenous cueing case.

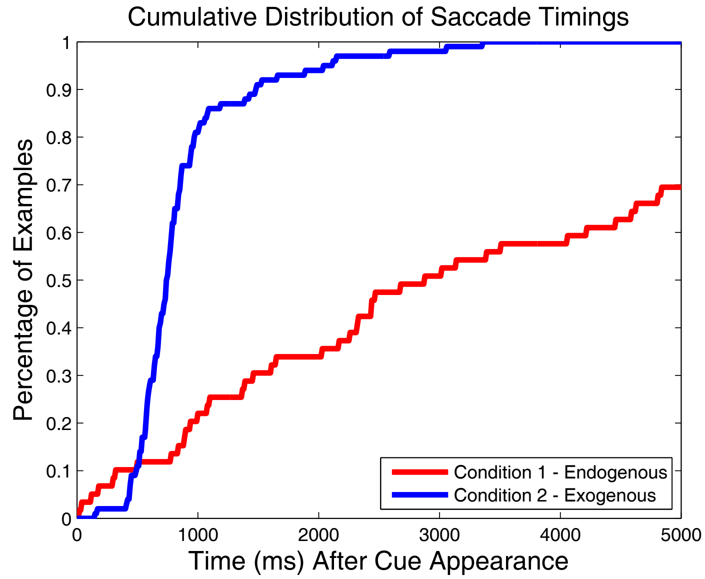


Figure 3.10: Distribution of saccade timings, after the onset of the first cue. The delay in Condition 1 is as expected as subjects take time to detect the cue and plan the saccade. Condition 2 follows the pattern of unplanned saccades, most occurring around 500ms after cue onset.

Finally, during predictive movements, it has been reported that the *head contribution* may be larger than during triggered movements [59]. If it were the case that the head contribution were larger in predictive movements, we would observe a greater maximum in the head yaw in the endogenous case. As demonstrated in Figure 3.15, we find no reliable differences between the maximum yaw positions of either the head or the eye, between the two conditions ($EyeMaxYaw$: $F(1,9) = 2.12, p = 0.15, HeadMaxYaw : F(1,9) = 0.15, p = 0.70$).

Some of these results have been repeated in other controlled environments [223, 61, 94]. In this more naturalistic study, we found almost no variations between the conditions. This implies that by approximating more natural conditions, along with a more complex task of driving, variables such as initial gaze position and gaze shift duration seem uncorrelated with either style of gaze shift. This could further imply that under these conditions, such variables are irrelevant to to onset of early head motion, in contrast to the conclusions of earlier studies, such as those reviewed in [59]. However more experiments should be done in these cases to verify these claims.

Feasibility of Classification

Given the significant effects of the condition on head yaw dynamics at the saccade time, it may be possible to generate a classifier to determine in real-time whether the individual example is more similar to Condition 1 or Condition 2. The statistical basis for such a classifier is provided by the ANOVA results in the previous section, giving credence to the possibility of automatically

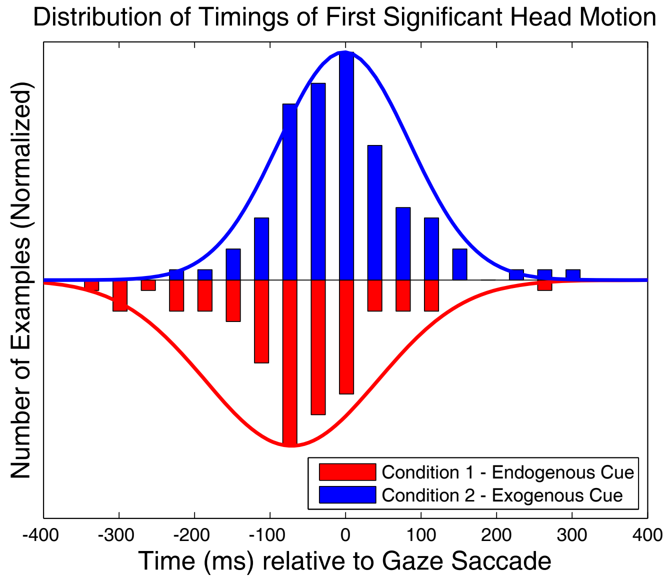


Figure 3.11: Distribution of first significant head motions (over a fixed threshold) relative to the gaze saccade. The histograms represent the actual measurements, and the solid lines represent a fitted Gaussian. The endogenous, "goal-oriented" condition shows a marked difference, with a majority of head motions occurring prior to the saccade.

Table 3.1: Confusion Matrix for Detecting Endogenous, Goal-Oriented Gaze Shifts (G) versus Exogenous, Stimulus-Oriented Shifts (S) in Simulator Experiment, using Head Yaw Position criteria. Correct classification rate is 69.44%.

Number of examples...		
	Actually G	*Actually S*
Predicted G	63	41
Predicted S	25	87

classifying the examples.

The most basic classifier consists of thresholding the test statistic (either P_{ij} or M_{ij}). Results of these simple classifiers are shown in Tables 3.1 and 3.2. Results approach 70% classification rates using the Head Yaw Position criteria. To our knowledge, the potential for real-time classification of attentive states is a unique proposition in the literature. Such a classifier could then directly be used to improve advanced Human-Machine Interfaces in task-oriented environments.

The results in these prior sections serve to validate the hypothesis that humans exhibit different behaviors when attention-switching in various contexts. Of importance is the attention switch in time- and safety-critical situations such as driving, where distractions pose significant dangers. Knowledge of a planned attention shift could give hints to whether the driver is actively engaged with a task. In the next section we discuss an application of this eye-head dynamics knowledge in real-world task-oriented situations, in order to design classifiers to detect driver's

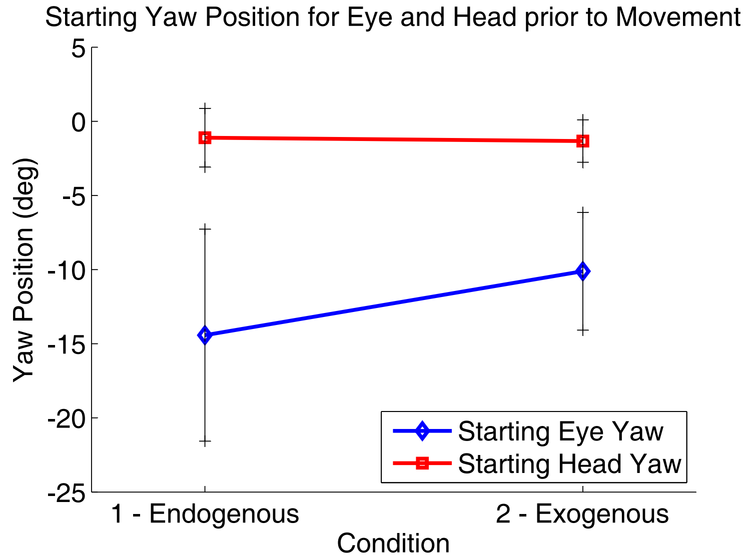


Figure 3.12: Starting Yaw Position for Eye and Head motion under each condition. Error bars represent the standard error of the mean.

Table 3.2: Confusion Matrix for Detecting Endogenous, Goal-Oriented Gaze Shifts (G) versus Exogenous, Stimulus-Oriented Shifts (S) in Simulator Experiment, using Head Yaw Motion criteria. Correct classification rate is 65.74%.

Number of examples...		
	Actually G	*Actually S*
Predicted G	65	39
Predicted S	35	77

intentions.

3.1.5 Investigations into Naturalistic Driving

McCall et al. [121] demonstrated the ability to detect a driver's intent to change lanes up to 3 seconds ahead of time, by analyzing driver head motion patterns. Doshi et al. [48] extended this study and found that head motion was in fact an *earlier* predictor of lane change intentions than eye gaze. However the reasons for this interesting finding were not clear.

We propose that the visual search that occurs prior to lane changes, and potentially in other similar common driving maneuvers, is initiated by a top-down process in the driver's mind. The driver has a goal in mind, and thus is trained to initiate a search in particular locations such as the mirrors and over the shoulders, for obstacles. Here we present a deeper analysis into real driving data to support this hypothesis, that the visual search prior to lane changes is a Type III search. The ability to detect this type of behavior is crucial in being able to identify the context of the situation, and then to assess its criticality or determine if objects around the vehicle are

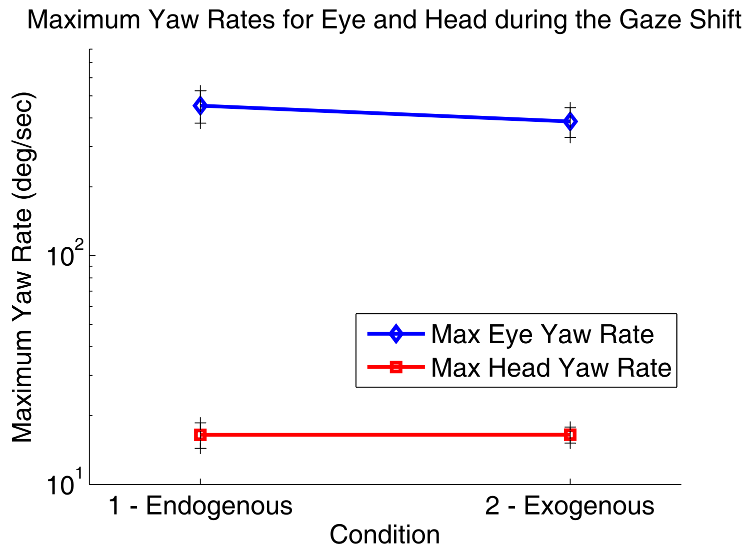


Figure 3.13: Maximum Yaw Rates for Eye and Head motion under each condition. Error bars represent the standard error of the mean.

of interest.

Naturalistic Driving Data Collection

For this research, data was collected in a driving experiment with an intelligent vehicle testbed outfitted with a number of sensors detecting the environment, vehicle dynamics, and driver behavior. This data is drawn from the same data as was used in the lane change intent work by McCall et al [121]. A camera-based lane position detector and CAN-Bus interface provided most of the data related to the vehicle and surrounding environment.

The main driver-focused sensor was a rectilinear color camera mounted above the center console facing toward the driver, providing 640x480 resolution video at 30 frames per second. To calculate head motion, optical flow vectors were compiled and averaged in several windows over the driver's face (detected with the Viola-Jones [211] face detector). This method was found to be stable and robust across different harsh driving conditions and various drivers. Other methods could be used for these purposes [139, 137]. Various automatic eye gaze detectors exist (e.g., Wu et al.[220]), however to ensure accuracy and reliability, eye gaze was labeled into one of 9 categories using a manual reduction technique similar to several recent NHTSA studies on workload and lane changes [106, 6] (Details on this procedure can be found in Chapter 5.1 and in [48]). While the technique did not capture the subtle variations in eye movements, it was useful to capture the timing of the eye saccades to the mirrors and side windows.

The dataset was collected from a naturalistic ethnographic driving experiment in which the subjects were not told that the objective was related to lane change situations. Eight drivers

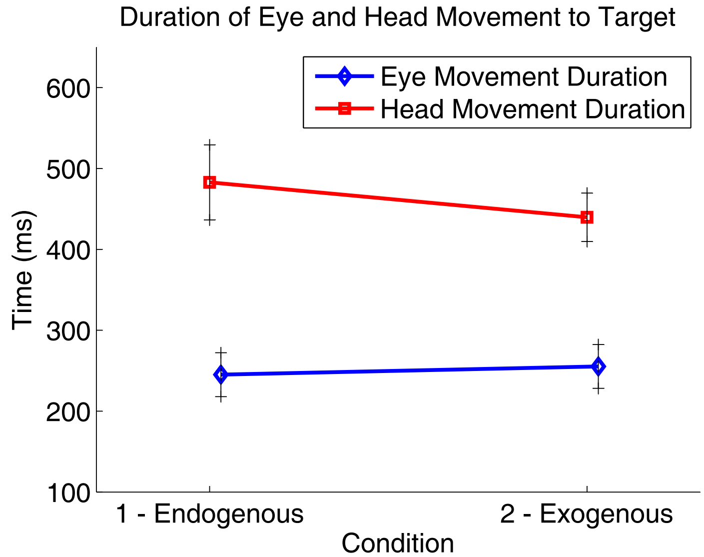


Figure 3.14: Duration of motion from initial movement until target. Note that in case of eye motion, this is a superset of the saccade duration. Error bars represent the standard error of the mean.

of varying age, sex, and experience drove for several hours each on a predetermined route. A total of 151 lane changes were found on highway situations with minimal traffic. 753 negative samples were collected, corresponding to highway "lane keeping" situations.

Analysis

These examples were used with the features described above to train a classifier to predict 3 seconds in advance of a lane change, whether the driver would intend to change lanes. Another classifier was trained for 2 seconds ahead of the lane change. In similar studies an extension of SVM, namely Relevance Vector Machines, was found to be a good classifier (see [201, 121] for more details), and was thus used here as well. In a comparative analysis, similar to that in Chapter 5.1, we find that such a classifier based on head motion has significantly more predictive power than one based on eye gaze 3 seconds ahead of the lane change, but not 2 seconds ahead of time.

By looking at the outputs of each classifier, we can get a sense of the performance of eye gaze and head pose over time. Specifically, the RVM classifier outputs a class membership likelihood, ranging from -1 (for negative examples) to 1 (for positive examples); thus the more positive the value, the more confident it is in its predictions of a true intention. Figure 3.16 shows the processing framework for this scenario. By looking at the average over all positive examples of these "Intent Prediction Confidences," in Table 3.3, we can tell that the Eye-Gaze-based classifier is hardly better than chance 3 seconds before the lane change, but improves significantly in the

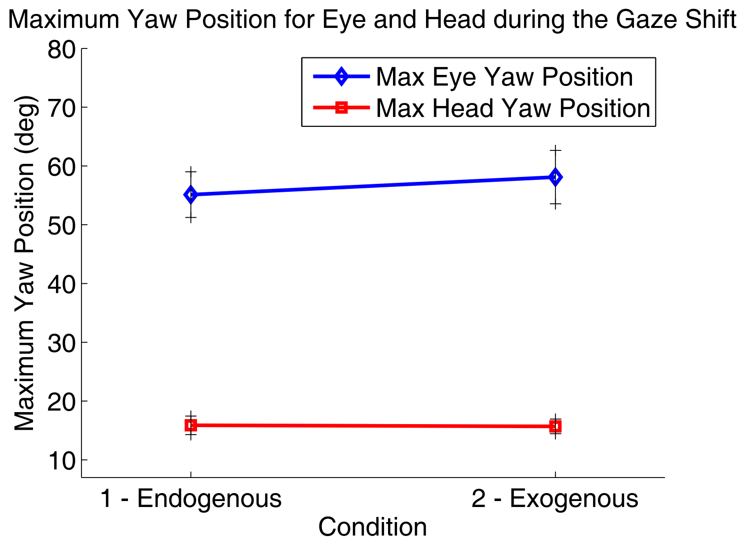


Figure 3.15: Maximum Yaw Position for Eye and Head motion under each condition. Error bars represent the standard error of the mean.

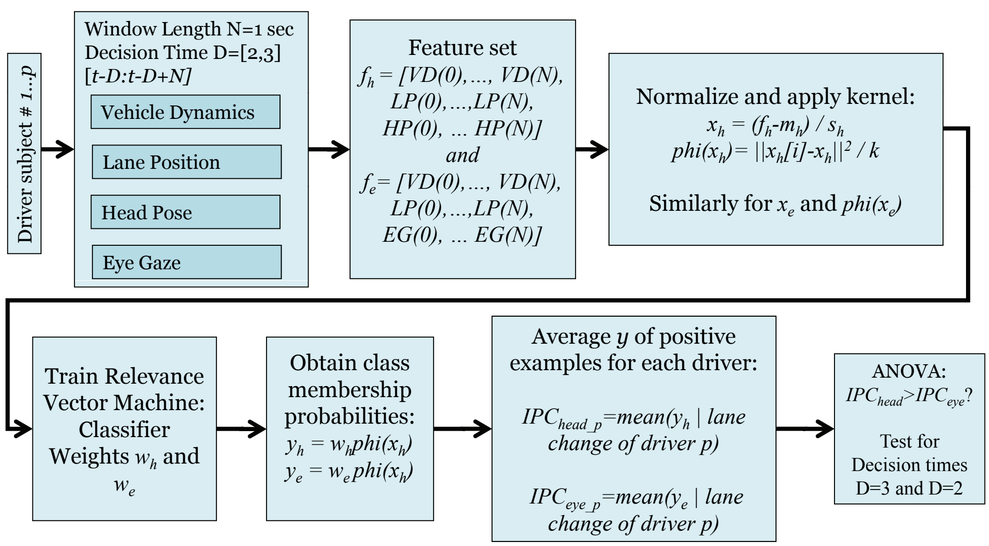


Figure 3.16: Flowchart of proposed approach for evaluating naturalistic driving data during lane change-associated visual search.

2 second case. On the other hand, the Head-Motion-based classifier works very well even in the 3 second case.

The results indicate that drivers engage in an earlier preparatory head motion, before shifting their gaze to check mirrors or blind spot.

ANOVA significance tests comparing the population of Intent Prediction Confidences ($\overline{IPC}$) demonstrate quantitatively that the preparatory head motion prior to the eye gaze shift,

Table 3.3: Average Intent Prediction Confidences ($\overline{IPC}$) for Each Type of Classifier, where a value of 0 represents chance.

Seconds before Lane Change:	*3 Sec*	*2 Sec*
Eye-Gaze Classifier ($\overline{IPC_{eye}}$)	0.0027	0.4691
Head-Pose Classifier ($\overline{IPC_{head}}$)	0.4411	0.6639
ANOVA: $\overline{IPC_{head}} > \overline{IPC_{eye}}$	$p < .01$	$p > .05$

is a significant trend 3 seconds prior to the lane change: The head-pose classifier is significantly more predictive than the eye-gaze classifier ($F(1,7) = 24.4, p < 0.01$). We can conclude that the early head motions begin to indicate a driver's intentions prior to the actual gaze saccade. This implies that the detection of a goal-oriented gaze shift may be quite useful in future Advanced Driver Assistance Systems, by helping to determine the context of the drive and whether the driver is indeed paying attention to task-relevant objects.

3.1.6 Concluding Remarks

We have demonstrated how the dynamics of overt visual attention shifts evoke certain patterns of responses in eye and head movements, in particular the interaction of eye gaze and head pose dynamics under various attention-switching conditions. Sudden, bottom-up visual cues in the periphery evoke a different pattern of eye-head yaw dynamics as opposed to those during top-down, task-oriented attention shifts. In laboratory vehicle simulator experiments, a unique and significant ($p < 0.05$) pattern of preparatory head motions, prior to the gaze saccade, emerges in the top-down case.

This finding is validated in qualitative analysis of naturalistic real-world driving data. In examining the time-course of visual searches prior to lane changes, the same significant pattern of early head motions appeared, indicating a "task-oriented" attention shift. Though it would be dangerous to collect data regarding stimulus-oriented attention shifts in real driving, the simulator experiments presented here seem to demonstrate that sudden stimuli attract eye motions as early as, or earlier than, head motions.

One important question arises regarding the nature of the stimulus presented in this experiment, specifically whether the "exogenous" cueing condition actually corresponds to a distraction-driven attention shift. Pashler et al. [157] notes that "abrupt onset" cues may not be inherently attractive in a bottom-up sense, but are only attractive if those cues also happen to be useful for whatever top-down tasks the user may be engaged in. This "cognitive penetration" of bottom-up cueing could imply that the "exogenous" cueing case presented in these experiments may not correlate to a real bottom-up, reflexive "distraction". However it could be argued that in real driving scenarios, users are primed to detect abrupt onset stimuli because they may be relevant to the safety of the driving situation. In this context, any abrupt onset stimulus which is then irrelevant to the task at hand, can be considered a "distraction," and the results of this study shows that the visible dynamics of attention shifts differ in this case. More work could be done to characterize the nature of distractions in complex environments, as they relate to top-down goals of the subject, and then to include different forms of distractions in these experiments.

The results presented here indicate that measurement of eye-head dynamics may prove to be useful data for classifying human attentive states in time- and safety-critical environments. By detecting a "predictive" or early head movement prior to a gaze shift, an intelligent assistance system could improve its estimate of whether the subject is actively engaged in the task at hand. In potentially dangerous situations such as driving and flying, active perception of humans thus has a large role to play in improving comfort and saving lives.

3.2 Holistic Attention Estimation

Where eye-head interaction is useful in understanding attention, in many cases in automotive environments, we are unable to measure driver behaviors, such as eye movements, accurately. Even with noisy measurements of driver behavior, below we propose a method to incorporate environmental measurements of salient objects, to better understand attention and visual search.

3.2.1 Introduction

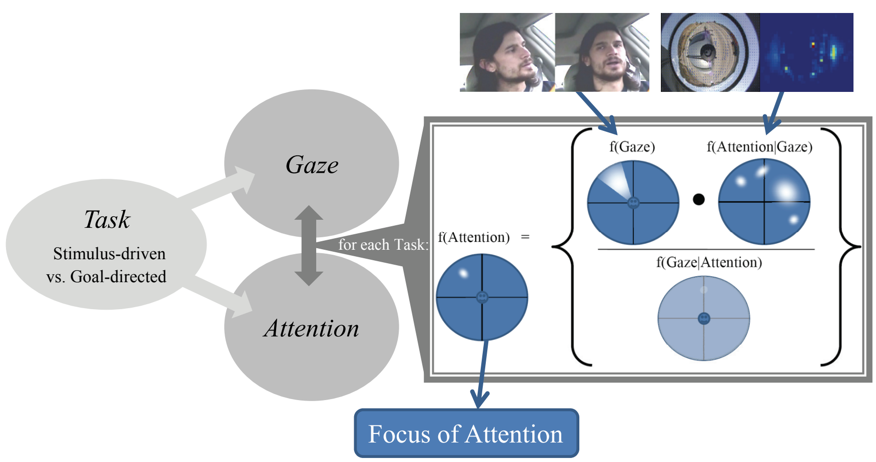


Figure 3.17: Graphical model relating Gaze, Attention, and Task. This leads to a Task-dependent Bayesian treatment of the joint densities of Gaze location and Attention location. Sample probability distributions for each term are shown below the terms. Vision-based **Gaze** estimation and **Saliency** maps are used to define the numerator terms, and the denominator encodes gaze inhibitors such as "inhibition of return," multi-tasking, and other cognitive distractions affecting Gaze **Orientation given Attention**.

In prior research associated with detecting human behavior in complex environments such as driving, it has sometimes been assumed that *attention* is indistinguishable from *gaze* [177]. However, early in the 20th century Helmholtz concluded that it is possible to attend to locations in the field of view without resorting to eye movements [73]. More recent research into visual attention has brought out the notion that attention and gaze are non-trivially interdependent. While attention generally must precede gaze shifts [77], a number of cognitive processes and distractions are known to affect the relationship between gaze and attention [177].

Computational models of visual attention look to simulate human attentive mechanisms in order to understand the most salient regions of a scene [82]. To estimate the attentive state of a human interacting with a complex, dynamic scene, it becomes necessary to observe the observer, to observe the scene, and to maintain some knowledge of the observer's tasks or cognitive state.

In the following research we introduce a new integrated model to estimate the attentive state of a human subject engaged in a complex task such as driving. We start out by applying a Bayesian treatment to the joint probability distributions of gaze and attention, as shown in Figure 3.17. A marginalization of the "ongoing task" allows us to break the model down into semantically meaningful and computable terms.

We demonstrate it is possible to encode many of the real-world cognitive phenomena that affect the relationship between gaze and attention, into the proposed model. This provides an elegant and principled way to estimate the subject's attentional state in dynamic environments, by incorporating vision-based estimates of both gaze and saliency.

The proposed system, labeled BRAVVO (Bayesian appRoach for estimating Attention by Viewer and View Observations) demonstrates a 45% improvement in Focus-of-attention estimation over a baseline system based on head pose alone, and a 61% improvement over a saliency-based attention estimation system. These results are generalizable to any human-machine interface in complex environments where the sensors monitor the subject and surroundings simultaneously.

Motivation

In complex environments where intelligent machines monitor and assist humans, such as driving or flying, it is very important for the assistance system to understand the attentional state of the human. Camera-based systems are becoming increasingly practical in real-time environments. We are especially interested in camera systems focused on a human and the surrounding environment; such information could provide valuable cues to infer the attention of the human subject. The proposed framework generalizes to any human-machine interface in complex environments, such as vehicles or control rooms, where the sensors monitor the subject and surroundings simultaneously.

State of the art research into estimating attention assumes precise knowledge of eye gaze (along with the notion that gaze corresponds to attention) [163], or that the environment is limited and controlled by the experimenter [9]. Those conditions are hard to come by "in the wild," where humans have free reign to look wherever they please, the environment is uncontrolled, and neither the humans nor the environment can be observed accurately. Such environmental constraints are important barriers to overcome in developing and realizing robust and useful human-computer interfaces [156].

Common sources of “noise” in observing humans in naturalistic, complex situations include distance and inability to use intrusive technologies (such as head mounted eye-tracking or EEG measurements). Eye gaze measurements are more accurate in controlled situations, where head pose measurements are more robust but less precise.

Simultaneous observation of a naturalistic, complex environment may also be hampered by several factors. Of primary concern is the difficulty in determining the most “salient” regions of the scene, especially when for different tasks, humans may find different parts of the same scene to be salient. Computer vision techniques have been developed to tackle these issues and are discussed in more detail below.

In order to combine these noisy measurements in a principled way, we propose a Bayesian approach to attention estimation. The fundamental motivation for this approach is the notion that gaze and attention are non-trivially interdependent. An approach to estimate the joint location of attention and gaze must thus be derived without ignoring the subtleties of each variable. Further, each component is dependent on the on-going task. This Bayesian approach allows us to divide the problem into semantically meaningful terms and brings out the importance of considering other cognitive factors, such as inhibition of return (IOR) and center bias.

In time-critical environments, an assistive system would do well to understand the fundamental attentive state of a human subject or controller. Such a system must use all available means to understand the attentive state of the human. Further, it should incorporate a sound understanding of the cognitive factors affecting humans as well as the sources of noise in its own underlying observations. As shown in Table 3.4, the proposed Bayesian appRoach for estimating Attention by Viewer and View Observations, or BRAVVO, is the first such system to demonstrate these qualities.

The remainder of this section is organized as follows, with a specific focus on evaluation of BRAVVO in driving-related situations. We review related research and concepts in Section 3.2.2. In Section 3.2.3, we develop our proposed model and independently consider each relevant concept. The remainder of the chapter contains experimental validation in the context of several driving situations in Section 3.2.4, and finally concluding remarks in Section 3.2.5.

3.2.2 Related Research

While it may not be possible to measure attention directly, prior studies show that we can observe gaze behavior and also identify salient features of the environment to help identify focus-of-attention.

Table 3.4: Similar recent research toward the understanding the focus of attention of a human. Saliency studies, benchmarked by the work of Itti et al.([82, 163]), do not operate with noisy, unconstrained measurements of the human. Other more recent studies([9, 138]) seek to control parts of the experiment, such as regions of interest (ROIs), to make attention estimation easier.

	Uncontrolled Environment	Uncontrolled Observer	Noisy Measurement of Environment	Noisy Measurement of Observer	Task Information
Itti et al. [82, 163]	Yes	Partial (videos)	Partial (videos)	No	Yes
Ba et al. [9]	Partial (meeting room)	Partial (predefined ROIs)	No	Yes	No
Murphy-Chutorian et al. [138]	Partial (meeting room)	Partial (predefined ROIs)	No	Yes	No
BRAVVO (proposed)	Yes	Yes	Yes	Yes	Yes

Gaze Detection in Harsh Conditions

Gaze detection in vehicles shares many of the challenges of other complex environments, such as airplane cockpits and surveillance situations. To detect where the driver is looking, robust monocular in-vehicle head pose estimation systems have been developed [139, 140, 138, 10, 229], though head pose is generally not a sufficient estimate of true gaze. More precise gaze estimates can be derived from eye gaze detectors [191]. NHTSA has most recently conducted studies of Driver Workload Metrics [6], including eye gaze as a proxy for driver workload. As discussed in Section 3.2.3, eye gaze tracking in vehicles is quite challenging. It is especially compounded when considering the reluctance of subjects in natural environments to submit to intrusive systems such as head-mounted eye gaze trackers.

As discussed above, there have been a few studies that examine gaze behavior in specific contexts such as lane changes on highways. Several of these studies demonstrate that gaze behavior is heavily influenced by task, especially during lane changes [199, 133]. The experiments of Land [102] in a real automotive setting lead to the conclusion that driver scan patterns during visual searches are predictable and task-oriented.

Numerous other studies have shown that fatigue [6, 69], traffic [175], and other cognitive and visual distractions [170] may also have an effect on gaze patterns of the driver. Cognitive distractions involve mental tasks that increase the workload of the driver without necessarily adding visual clutter, whereas visual distractions draw the driver's focus of attention away from the road. This result is well-aligned with more recent results indicating the limitations of drivers' multi-tasking abilities [109].

Vision-based gaze estimation has thus helped to demonstrate that drivers' gaze behavior is predictable and task-oriented ("top down"), especially during lane changes [102]. However cognitive and visual ("bottom up") distractions [6] have a significant effect on the gaze patterns of drivers. The proposed model incorporates both "top down" and "bottom up" influences in vision-based estimation of gaze.

Surround Analysis and Saliency

Visual saliency detectors have been developed to determine the relative attractiveness of objects in a scene [82, 84, 104, 70, 225, 115]. The potential structure of these saliency maps vary based on the motivation and context of the scene. A primitive saliency map could include features based on intensity, color, and orientation. Itti extended this saliency map by examining the eye gaze patterns of human subjects on several scenes to build up a prior for that particular type of scene [82]. Recent studies, however, show that such "bottom-up" saliency maps can not explain the fixations of goal-oriented observers [91, 178]. More recent versions of saliency maps have thus incorporated top-down goals [146, 163, 84], however much of their development has been motivated by images with relatively static backgrounds, or with significant amounts of training data.

Ultimately it is the interaction between an observer's goals, and the salient properties of the scene, that guide the observer's attention [221, 224]. Itti found that motion cues are much stronger predictors of gaze changes than any other cue in complex scenes [83]. Therefore given that we are in the specific context of driving, with highly dynamic background scenes, we ultimately choose to use motion-based features to build up a "bottom-up" saliency map, as detailed in Section 3.2.3. There have been similar motion-based approaches to scene analysis using omnidirectional cameras, in order to identify and track interesting objects in a scene [62, 63].

This motion-based saliency map will also be compared with a biologically-plausible saliency map developed by Itti et al [82]. We combine these with "top-down" maps dependent upon the task.

Focus of Attention in Complex Environments

Table 3.5: Effects of certain tasks on vision-based estimates of gaze, saliency, and orientation given attention. The behavioral influences are noted in parentheses, and discussed in more details in Sections 3.2.3-3.2.3.

	Gaze Estimate	**Saliency Estimate**	**Orientation given Attention**
Lane Keeping task	Field of view to 10 degrees *(Parafoveal vision)*	Goal-directed scan pattern - *front (Top-down)*	Focus on scanning road. *(Higher IOR, Center-bias)*
Secondary task (Lane Change / Turn / Yield)	Field of view to 10 degrees *(Parafoveal vision)*	Goal-directed scan pattern - *sides (Top-down)*	Likely paying attention to gaze location *(High IOR, No Center Bias)*
Distracted - No task	Field of view to 60+ degrees *(Peripheral vision)*	Motion, color change cues more salient *(Bottom-up)*	Gaze location and attention not necessarily linked. *(Low IOR)*

Evidence has shown that in some cases behavior is influenced by objects that have not been explicitly attended to [177]. Moreover, "cognitive distractions" which take the subject's mind off the task at hand would inherently imply a de-coupling of gaze and attention. Cognitive psychologists have reported other relevant phenomena which could inhibit the relationship between gaze and attention, including "Inhibition of Return" and "Change blindness" [166]; these are discussed in part in more detail in Section 3.2.3. It is important to consider these factors in any computational model of attention, and especially in complex environments such as vehicles where these distractions are quite dangerous [107].

Several studies have proposed models to estimate the Focus of Attention in various environments [79, 9, 138]. The current study builds upon previous preliminary work [51], by significantly expanding the context and analysis, including a comparison of many different driving situations. In addition, this study includes a comparison of the proposed motion based-saliency to other state-of-the-art saliency maps.

To the knowledge of the authors, this is the first study in which visual gaze and salient, task-dependent targets are dynamically computed, and attentional likelihood is then modeled in a Bayesian manner. This allows a simple yet principled incorporation of many attention-related phenomena, along with noisy measurements of gaze and saliency, to provide a robust estimate of attention. The system quantitatively outperforms baseline attention estimators based either on gaze or saliency alone, as well as naive combinations of the two. In the following section we propose and discuss the framework of this model.

3.2.3 Proposed Model: BRAVVO

The basic calculations for the BRAVVO model are visualized in Figure 3.18, with the key emphasis on a principled Bayesian combination of simultaneous observations of the viewer (Gaze estimation) and the view (Saliency estimation).

Let G represent the Gaze Estimate, A represent the Focus of Attention Estimate, and T represent the ongoing task. G, A are drawn from continuous 2-D distributions in polar world coordinates (d, θ).

We can consider the joint probability distribution functions of gaze and attention, without assuming anything about the dependencies between them. As we have alluded to, the location of a subject's gaze may or may not indicate that attention is being paid to that particular location. Therefore by Bayes' rule for probability densities,

$$f_A(a) = \frac{f_A(a|G = g)f_G(g)}{f_G(g|A = a)}. \tag{3.1}$$

For the purposes of this work T is a discrete random variable drawn from the space of all tasks, $T \in \{T_1, T_2, \ldots T_n\}$. By applying the law of total probability, conditioning on the task

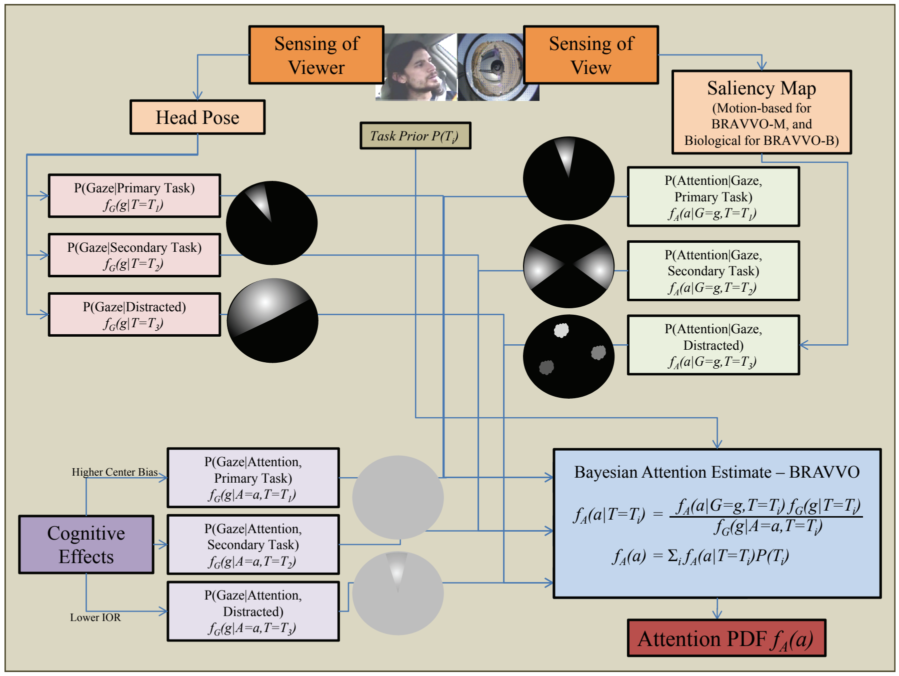


Figure 3.18: Data Flow diagram for BRAVVO model. The measurements of the viewer and the view are used as inputs to several terms in the model; other terms are predefined based on the tasks or based on modeled cognitive effects. Sample distributions are shown for each of the components. These components are combined in a Bayesian framework to produce a final distribution of attention. Note that two models can be generated, one based on Motion-Saliency (*BRAVVO-M*), and one based on Biological Saliency (*BRAVVO-B*).

T,

$$\begin{aligned} f_A(a) &= \sum_{i=1}^{n} f_A(a|T=T_i)P(T_i) \\ &= \sum_{i=1}^{n} \frac{f_A(a|G=g,T=T_i) f_G(g|T=T_i) P(T_i)}{f_G(g|A=a,T=T_i)}. \end{aligned} \tag{3.2}$$

We can consider each term in Equation 3.2 separately. First we consider the effect of the task on gaze and attention, in Section 3.2.3. Here we define the "prior" $P(T_i)$ as well as the space of "tasks" for this particular application context.

Each of the other terms are task-dependent; in other words for each task $T_i, i \in \{1...n\}$, there will be a different probability density for those terms. The first term, $f_G(g|T=T_i)$, corresponds to the probability density function of the gaze location. As described in Section 3.2.3 below, we can encode any uncertainty in the gaze estimate in this term.

We will consider the numerator term $f_A(a|G=g,T=T_i)$, as the likelihood that attention is focused on a particular location given gaze location and task, in Section 3.2.3. Finally, we will consider the distribution of gaze given attention and task, $f_G(g|A=a,T=T_i)$, in Section 3.2.3. This section will incorporate such concepts as "center bias" and "inhibition of return" which can affect the link between gaze and attention.

Influence of Tasks: $P(T_i)$

The particular task which the human is performing has a significant influence on attention [178]. When engaged in a task even in complex scenes and environments gaze follows a predictable pattern, focusing on the most pertinent subjects [102]. It is even evident that gaze can be used to predict whether a driver is distracted from any particular task [6].

Without loss of generality, for the purposes of this study we find it useful to define three "tasks," corresponding to normal driving; other environments may incorporate a different set of tasks into the same model. A primary task T_1 is defined as the task upon which the subject should be focused in most situations; in the case of driving that task involves maintaining the vehicle heading and speed within appropriate boundaries while avoiding obstacles. A secondary task T_2 arises when the driver chooses to change lanes, make an intersection turn, or stop to yield to other traffic or pedestrians. The secondary task includes the setup to these maneuvers and the maneuvers themselves. Finally, whenever the driver is distracted from either of these tasks, we label that state as T_3.

The prior probability $P(T_i)$ can be defined in a straightforward manner using a data-driven approach. Here for simplicity we assume it is uniform, however we acknowledge that in various environments, certain tasks are more likely than others. It is useful to treat the task as a nuisance variable, since it will be easier to analyze the rest of the equation assuming knowledge of the ongoing task. The direct estimation of task ([121, 48, 208]) could assist the model here by adjusting the prior, however we leave that beyond the scope of this work.

In Table 3.5, we demonstrate the effect of various tasks in the driving context, on the estimates of each of the terms in Equation 3.2. Each of these effects stems from biological and psychological considerations, noted in the italics and discussed in more detail in sections below.

Gaze Estimation: $f_G(d, \theta)$

Practical means for gaze estimation include monocular vision-based systems, which are capable of reducing cost and impact on the environment. Monocular *eye gaze* estimation is a difficult problem, in light of occlusions, shadows, and other tough situations. Figure 3.19 demonstrates some of the tough conditions under which gaze estimation must occur in vehicle-based environments. Monocular *head pose* estimation can prove to be more robust, though less precise than eye gaze estimation. In either case, the gaze estimate will be a noisy approximation of the true gaze.

In this study we use a commercially-available monocular head-pose estimation system. The system outputs the head yaw, synchronized with the rest of the system. For estimating

Figure 3.19: Sample images showing the difficulty involved in gaze estimation in complex environments such as vehicles. Even with controlled lighting and reasonable camera locations, there are still issues due to shadows, lighting changes, occlusions, appearance changes, and other challenges. Direct eye gaze estimation would not be possible in many of these cases, where head pose estimation may be possible but not as precise. Either way there is bound to be some degree of error in the gaze estimate.

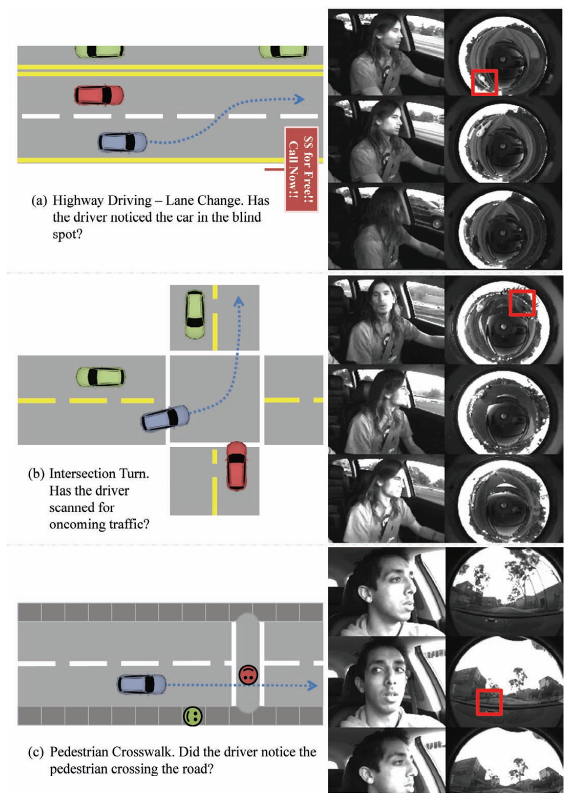


Figure 3.20: Three naturalistic driving scenarios under examination in this research. Each scenario requires the subject to shift focus of attention; the goal of the proposed system should be to answer whether the subject has actually paid attention to relevant objects in the scene. Sample images from the scene are shown on the right. The first two sequences, (a) Highway Lane Change and (b) Intersection Turn, use an omni-directional camera to observe the environment. The third sequence, (c) Pedestrian Yield, uses a wide-angle camera for forward-looking vision.

attention in task-oriented behavior, this rough approximation of the true gaze may be sufficient. In primary tasks such as general scanning patterns, research has shown that gaze patterns tend to follow a "center bias" [70, 225] whereby most glances occur toward the center of the field of view, in correspondence with the head pose. On the other hand for secondary tasks involving large visual searches, there has been recent support for the notion that head pose moves with eye gaze, making it easier to measure gaze using just head pose [47]. Nevertheless it is still important to account for gaze movements independent of head pose, especially in a "distracted" state as discussed below.

To deal with uncertainty in the gaze measurement, we take the output $\widehat{\theta}$ of a vision-based gaze tracker, and model uncertainty in the estimate with a 1-D Gaussian density function in θ (as an approximation to a Von Mises distribution). To be specific,

$$f_G(d,\theta) = \frac{1}{\sqrt{2\pi\sigma_{T_i}^2}} exp\left\{-\frac{(\theta-\widehat{\theta})^2}{2\sigma_{T_i}^2}\right\} \tag{3.3}$$

Note that we modify the variance of the Gaussian to be task dependent σ_{T_i}. The effect of task directly on the estimation of raw gaze is largely evident in the case of distractions. In those cases peripheral vision plays a larger role (motion and color cues affect periphery more). In the distracted state it then becomes necessary to increase the variance of the gaze estimate to include a wider field of view. In the primary and secondary-task oriented states, we can model the field of view after the typical human parafoveal field of view, which is approximately 10° [173].

Visual Saliency: $f_A(a|G=g,T=T_i)$

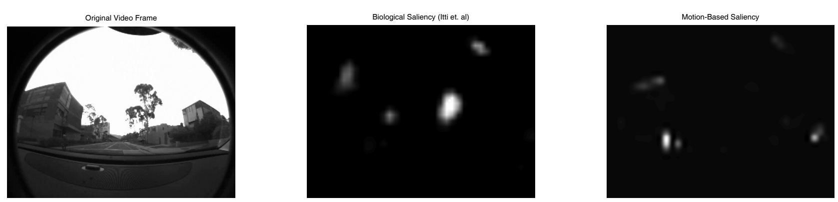


Figure 3.21: Examples of both types of saliency maps used in this approach. Original frame (left). The "biological" approach (center) captures interesting objects such as the tree. The "motion-based" approach (right) captures independent motion such as the pedestrian crossing toward the left.

The term $f_A(a|G=g,T=T_i)$ in Equation 3.2 corresponds to the likelihood of paying attention to a particular location, given that the subject is looking at that direction and engaged in a given task. This most directly evokes the notion of visual saliency, that is, the regions of a scene that are most likely to draw attention. Given a gaze estimate in some direction, there are likely going to be locations in that field of view that are more salient than others. In this section we model those effects.

As discussed above, the derivation of visual saliency of a scene is a well-studied problem. Recent studies have demonstrated that task and context are the most significant factors in predicting gaze patterns and attention [178]. However the relationship is not clear, especially when subjects are cognitively or visually distracted. In driving situations, distraction is an extremely consequential issue [6]. There have been many attempts to model glance behavior based on interest point detection in a “bottom-up” approach [82]; here we can make use of the context of driving to develop more relevant saliency maps correlating to driver distraction [50].

In the case of the goal-oriented behavior in the lane-keeping and secondary driving tasks, we presume that the most salient regions in the scene, are those where the driver should be scanning in normal situations. Namely, we set the front of the vehicle as the most salient region in lane-keeping, and the sides of the vehicle as more salient under lane-changing, turning, and yielding to traffic. Saliency in these situations could be modified to include detection of objects relevant to the task as well [28], however we leave that beyond the scope of this work. These regions would correspond to “top-down” processing of saliency in a scene.

In the “No-task” or “Distracted” state, the drivers are more likely to be paying attention to abnormal events in the scene. In the case of normal driving, motions and color changes tend to be most salient cues [83]; these are the cues most likely picked up by peripheral vision [173]. Therefore we make use of a motion-based saliency map in this case, where normal “background” scene motions are calculated using optical flow, and then “foreground” motions are subtracted out [50].

$$f_A(d, \theta | g = \widehat{\theta}, T = T_3) = \frac{m(d,\theta) - \overline{m}(d,\theta)}{\sum_{d,\theta} m(d,\theta) - \overline{m}(d,\theta)} \tag{3.4}$$

The current frame motion is m, and $\overline{m}$ represents an average “background” motion. Examples of this sort motion-based saliency can be seen in Figures 3.22 and 3.21.

We also compare this “motion-based” saliency approach with a biologically-inspired (“biological”) approach proposed by Itti and Koch [82] which remains the benchmark for visual saliency. This comparison demonstrates the versatility of the proposed framework. Multiple versions of saliency could be implemented as inputs to the system depending on the situation.

In the generated results, the “biological” saliency map is incorporated as the saliency map in a version labeled BRAVVO-B. The framework incorporating the “motion-based” saliency map is labeled as BRAVVO-M. Examples of each version of the saliency map can be seen in Figure 3.21.

Orienting of Attention: $f_G(g|A = a, T = T_i)$

The final term in Equation 3.2, $f_G(g|A = a, T = T_i)$, corresponds to the distribution of gaze given the location of attention. This term could encode any style of gaze inhibitor, or those cognitive processes which prevent gaze from moving to the focus of attention.

Of primary concern is “Inhibition of Return”, the phenomena by which human subjects tend not to glance at objects which they have already seen [166]. This response could be due to the subject already having paid attention to those objects; by keeping track of them the subject does not need to look at that particular area again.

The task also plays a role in linking gaze to attention. In a goal-oriented state, a subject may be more focused on the task at hand. However in a distracted state, the subject may be more likely to be paying attention to something unrelated to the visual field of view. Such "cognitive distractions" have been found to significantly alter the safety of driving situations in particular [6].

In order to account for such notions, we define a task-oriented PDF, $f_G(g|A = a, T = \{T_1, T_2\}) = C$, which amounts to a uniform distribution. That would indicate that no particular area is more likely to be glanced at. The distraction-driven PDF, $f_G(g|A = a, T = T_3)$ places slightly more emphasis on the gaze angle, which corresponds to a decreased likelihood of paying attention to the area where gaze is located and a lower inhibition of return (IOR). Such a model could be modified or trained to account for more subtle variations in these gaze inhibitors, given more detailed or informative training data.

3.2.4 Experimental Validation

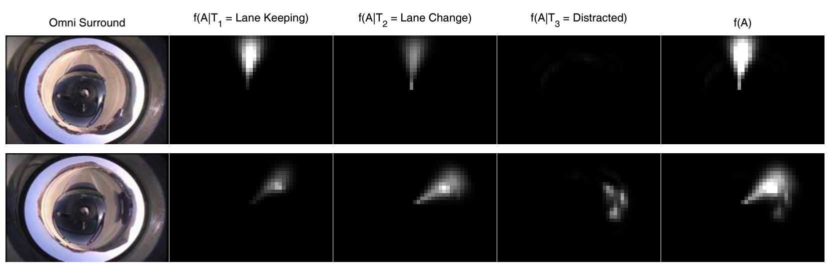


Figure 3.22: Sample data of the omni surround view, shown on the left. The task-dependent pdf's of attention corresponding to Equation 3.2 are in columns 2-4, and the last column shows the final density of the attention estimate. The top row corresponds to a normal lane-keeping state, and the bottom row shows data prior to a lane change. Notice in the second row how the vehicle on the side draws more attention based on the "lane-change" and "distracted" states, thus heavily affecting the final attention density.

In order to validate the model we take examples from real-world, naturalistic driving data. The vehicular testbed is outfitted with a commercial head tracking unit, which involves a monocular camera pointed at the driver. This system outputs head pose information; eye gaze trackers were implemented but deemed too unreliable in many situations.

Additionally, an omni-directional camera is mounted above the vehicle directly above the driver's seat. In this manner, the camera is able to see most of the driver's field of view. Sample images from this camera can be seen in the first column of Figure 3.22.

Finally, for situations involving yielding at intersections, a wide-angle camera is mounted on the front of the dashboard looking forward. Sample images from this camera can be seen in Figure 3.21.

A sample series of calculations is shown in Figure 3.22. For every frame in the sequence, and every task, the terms in Equation 3.2 are computed and stored. The component of attention given each task is shown in columns 3-5 of Figure 3.22. The effects of the tasks on attention are clear; for a “lane keeping” task T_1 the model presumes that the attention is mostly on the forward areas. The attention given a “lane change” task T_2 are less focused on the front and more on the side.

However in the second sample row, the “distracted” state T_3 causes the attention to shift to the vehicle driving by. This effect is caused by the wide gaze estimate along with the motion-based saliency map. Assuming the driver is distracted, they would more likely notice the vehicle in their peripheral vision and then turn to pay attention to it. The vehicle is seen in both the tasks T_2 and T_3, so it would be a bit difficult to presume the actual underlying task.

By marginalizing the task, we arrive at a distribution function of the attention, $f_A(a)$, seen in the right-most column of Figure 3.22. The resulting estimate of attention can be obtained in a maximum likelihood manner, or the attentional distribution can be used as a prior for some temporal attention tracking scheme.

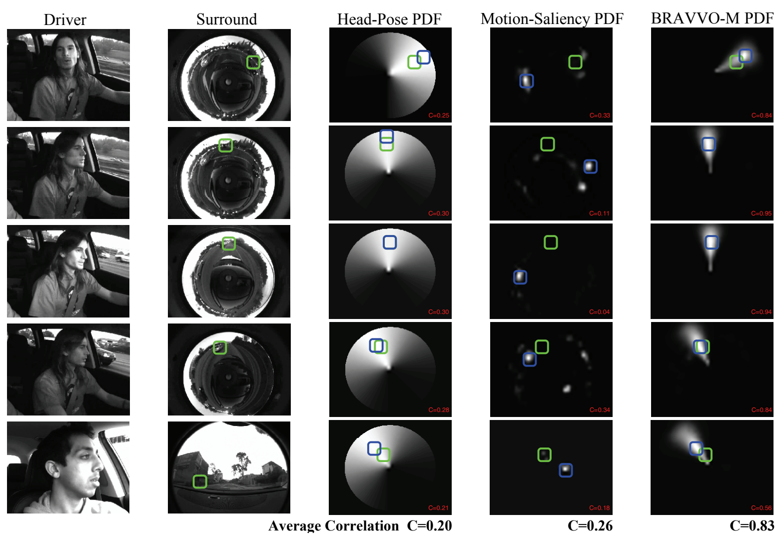


Figure 3.23: Sample frames comparing approaches for estimating focus of attention. The left column shows the head pose, followed by the surround frame. For each of the approaches, the blue square shows the estimated attention, and the green square shows the labeled ground truth. The average correlation (C) between the pdfs and the ground truth is shown; the average correlation in these frames for the BRAVVO-M approach is 0.83, compared to 0.20 using just Head Pose, and 0.26 using just Motion-Saliency.

Quantitative Evaluation

In this section we discuss the results of an experimental evaluation of over a thousand images of manually labeled data. Three video sequences were chosen for analysis, all in naturalistic driving scenarios as seen in Figure 3.20.

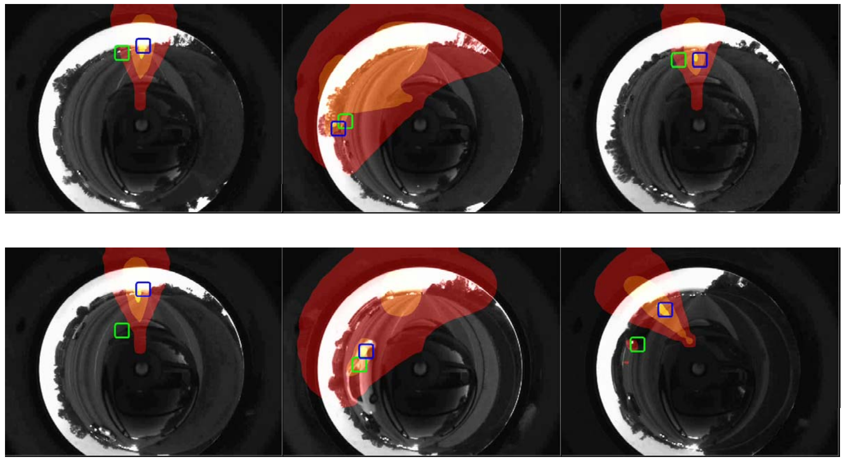


Figure 3.24: Results showing the estimated PDF of focus-of-attention. The blue box marks the estimated point of focus, and the green represents the labeled ground truth.

The first sequence includes normal highway driving leading up to a lane change on a highway, with vehicles passing by at high speeds. The second situation consists of the lead up to an intersection turn, and the turn itself. Finally the last sequence includes mostly local street driving with a single stop, yielding to a pedestrian at a crosswalk.

Ground Truth

In these cases the external environment images were shown to several annotators along with the images of the driver. After cross-validating the manually labeled data, ground truth for the attentional location was determined, as discussed below.

The validation of the system should ideally include some notion of the ground truth of the driver's attention patterns. However it is difficult to obtain actual ground truth, without an extremely detailed understanding and measurement of the attentional processes in the driver's neurological system. Several simulation-based studies have included a questionnaire asking the driver to annotate their own data.

In this case we would like to emulate the performance of an expert assistance system; generally speaking a passenger sitting next to the driver has a good awareness of where the driver is paying attention. It has been shown that having a conversation with a passenger is much safer (than using a cell phone) because the passenger can modulate the conversation based on the driver's attentional and stress patterns [193]. An assistance system that approaches the performance of a passenger or human onlooker would be fundamentally useful. In this case we ask a human labeler to examine video of the driver and of the environment, acting as a passenger. Using those videos, the "passenger" then marks their interpretation of the focus of attention of the driver.

Figure 3.23 demonstrates the ground truth labeling as well as sample results for some

Table 3.6: Comparison of proposed Bayesian-Attention Estimator with approaches based on Head Pose alone and Saliency alone. Avg. Correlation corresponds to the correlation between the estimated PDFs and the labeled Focus of Attention. Also shown is the error (in pixels) between the estimated and labeled Focus of Attention. BRAVVO-B incorporates "biological" saliency, whereas BRAVVO-M incorporates "motion-based" saliency.

	Seq 1: Lane Change		Seq 2: Intersection Turn		Seq 3: Pedestrian Yield	
Estimator	*Avg. Correlation*	*Avg. Error*	*Avg. Correlation*	*Avg. Error*	*Avg. Correlation*	*Avg. Error*
Head-Pose alone	0.24	114.20	0.23	135.82	0.20	126.44
Biological-Saliency	0.26	159.04	0.11	246.66	0.17	147.38
Motion-Saliency	0.17	191.04	0.17	216.96	0.25	134.00
BRAVVO-B	0.63	68.31	**0.55**	**77.29**	0.43	88.57
BRAVVO-M	**0.63**	**62.15**	**0.55**	85.55	**0.44**	**88.14**

frames. The left column shows the head pose, followed by the surround frame. Three approaches are compared; for each of the approaches, the blue square shows the estimated attention, and the green square shows the labeled ground truth.

The following tables and graphs demonstrate the experimental accuracy of the proposed system based on this ground truth.

Results

As points of comparison for the proposed Bayesian approach (*BRAVVO*), we also include results of several simple attention estimators, one using only head pose (*HeadPose*), one using only a raw motion-based saliency map (*Motion-Saliency*), and one using only a raw biological saliency map (*Biological-Saliency*). The head pose-based attention estimator places the focus of attention on the central field of view of the driver's head. The saliency-based detectors uses the saliency maps derived above, and places the estimated attention at the point of greatest saliency.

As mentioned above, we are able to generate two distinct estimators, one based on each of the saliency maps. Results for both BRAVVO-B (biological) and BRAVVO-M (motion-based) are shown, and it is apparent that either approach constitutes significant improvement over the baseline. Table 3.6 shows comparative results using several metrics. In the first metric, we correlate a box around the labeled focus of attention with the PDF of the estimated attention. This metric demonstrates the ability of the estimation system to capture the most relevant point in the scene. The BRAVVO approaches not only outperform the baseline approaches, but it is also apparent that a simple combination (by addition) of the two baseline approaches would still not reach the level of the proposed system. Figure 3.23 demonstrates some of these correlation values; the average correlation in the frame of Figure 3.23 for the BRAVVO-M approach is 0.83, compared to 0.20 using just Head Pose, and 0.26 using just Motion-Saliency.

As seen in Figure 3.23, an approach based upon head pose alone is not precise enough, whereas motion-based saliency is more precise, but places emphasis on inappropriate locations. Through a principled Bayesian combination of these inputs, the BRAVVO-M approach achieves a much greater level of accuracy and precision in estimating the focus of attention.

We also calculate the error distance between the labeled and estimated focus of attention. Average results over the data run can be seen in the final column of Table 3.6. Figures 3.25, 3.26, and 3.27 show the performance improvement of BRAVVO-M over the baseline estimators, for a sequence of data.

It is interesting to note that in the cases when no head pose was detected due to the head being out of range of the head pose detection system (highlighted in red in Figures 3.25, 3.26, and 3.27), the BRAVVO system still performs quite well. It is generally able to do so because it can revert to the saliency map as an estimate of what the driver should be looking at, while the head pose is out of range. Overall, the BRAVVO system clearly outperforms both baseline systems and more closely approaches the performance of an ideal agent.

The following three sections a closer look at the comparisons between BRAVVO-M approach and the baseline approaches for each of the three sequences, as well as the source of errors.

Sequence 1 - Highway Lane Change

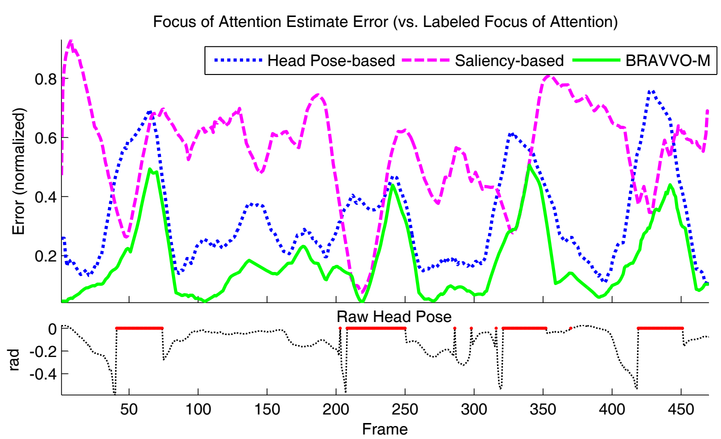


Figure 3.25: Sequence 1: Highway Lane Change - BRAVVO-M approach compared to Baseline Head-Pose and Motion-Saliency Attention Estimators. Raw head pose is shown below, with red indicating no head pose available.

In the first sequence, a driver approaches a lane change on a highway, and glances several times at the surrounding environment. In several of these glances, the head pose detection system loses track of the head; thus error based on the head pose alone increases significantly as seen in Figure 3.25. However the BRAVVO-M approach still maintains relatively low error throughout. Upon loss of the head pose, the system places less emphasis on the head-related measurements and correspondingly the saliency-based component helps improve the performance.

Sequence 2 - Intersection Turn

Similar results can be seen in the second sequence, in which the driver passes through an intersection turn and turns, ostensibly to check for oncoming traffic. The portions of lost head pose, as seen in Figure 3.26 again demonstrate the power of the system to utilize combinations of information to improve attention estimation. Saliency-based approaches offer more precision, but an estimate based on the maximum likelihood of the saliency pdf is often times misplaced.

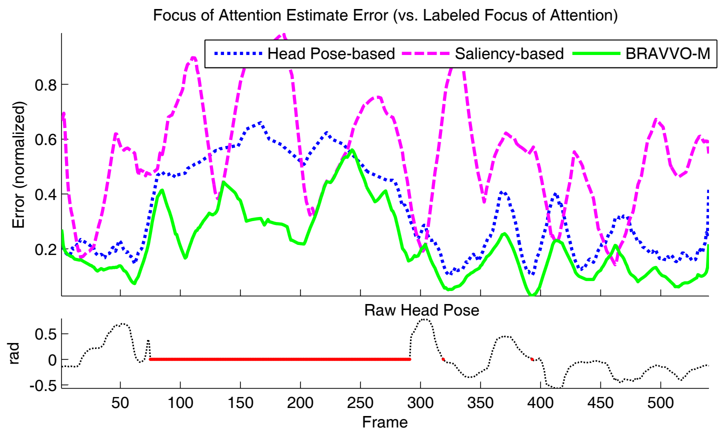


Figure 3.26: Sequence 2: Intersection Turn - BRAVVO-M approach compared to Baseline Head-Pose and Motion-Saliency Attention Estimators. Raw head pose is shown below, with red indicating no head pose available.

Sequence 3 - Pedestrian Yield

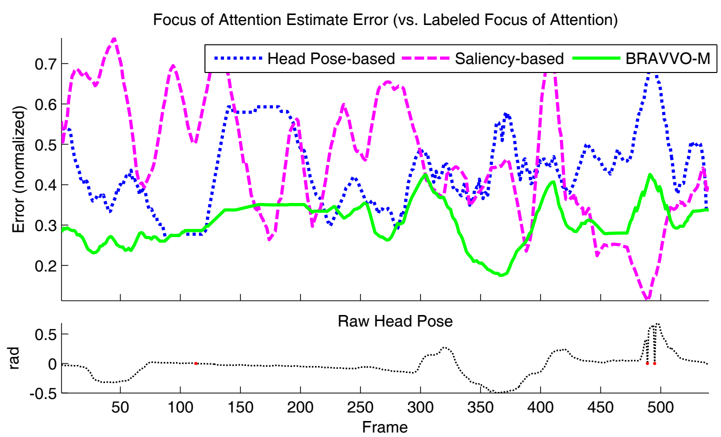


Figure 3.27: Sequence 3: Pedestrian Yield - BRAVVO-M approach compared to Baseline Head-Pose and Motion-Saliency Attention Estimators. Raw head pose is shown below, with red indicating no head pose available.

Finally, we demonstrate the ability of the BRAVVO approach on different camera setups in the third sequence. Instead of an omnidirectional camera mounted on top of the car, a camera is mounted in front of the rearview mirror, with a wide angle lens. A basic calibration allows the transformation of object locations from the front-view camera to a surround image. Given this, we are able to compare the approaches in similar metrics as above.

In this sequence a pedestrian crosses the street, and the driver glances to the right, possibly at another pedestrian, before glancing at the pedestrian. In such a situation, it is critical for the system to understand whether the driver is completely aware of the relevant surroundings. As shown in Figure 3.27, the error from the BRAVVO-M approach is overall the least of the three compared approaches.

3.2.5 Discussion and Concluding Remarks

We have introduced a new approach to analyzing the attentive state of a human subject, given cameras focused on the subject and their environment. In particular, we are motivated by and focus on the task of analyzing the focus of attention of a human driver. We have developed a new Bayesian paradigm for estimating human attention specifically addressing the problems arising in complex, dynamic situations. The model incorporates vision-based gaze estimation and several types of visual saliency maps, with cognitive considerations including "inhibition of return" and "center bias" that affect the relationship between attention and gaze.

It is important for the system to have a fundamental knowledge of the tasks at hand. Different task sets could significantly affect the design of the priors as well as the choice of distributions, as described above. To demonstrate some generality, we have shown performance on various kinds of tasks, using several different sensor setups as well. In each of the cases, the system requires sensors which cover at least the relevant field of view of the human subject.

The results demonstrate the potential of the model. We are able to demonstrate a 45% improvement in Focus-of-attention estimation over a baseline system based on head pose alone, and a 61% improvement over a saliency-based attention estimation system. These results show the capability of the system to more closely approximate an ideal assistive agent who is aware of the attentive state, based only on observations of the subject and surrounding environment. We demonstrate the model's ability to use various kinds of saliency maps (e.g., [225, 115, 84]) with other sets of cognitive factors, by comparing to a state-of-the-art saliency approach [163].

Future work includes, for example, a more explicit representation of time dependency in the model. The current calculations for head pose and saliency inherently track features over time and thus implicitly encode some time dependency. However a number of cognitive factors could affect the performance of humans over time, and these could be explicitly modeled in the process. Other factors, such as age and experience, also give rise to different performance on critical tasks [46]. Expert pilots would have different performance, including scan patterns and attentive states, than novices. Such information could also be coded into the model in the form of different priors for various model components.

Recent research has shown that up to 80% of vehicle-related crashes are due to driver inattention, resulting in over 25,000 fatalities each year in the U.S. alone [6]. Intelligent assistance systems that interface with the driver could help to reduce the number and severity of these accidents. Recently IDASs have even incorporated the inference of driver intentions to brake and change lanes [207] in order to improve safety. It is also important to consider the role of appropriate feedback to the human in these critical situations [44, 132].

The BRAVVO approach to estimating attention could help to mitigate and prevent many inattention-related crashes by simultaneous observation and understanding of both the viewer and the view. Similar situations arise not only in driving, but also in other safety-critical events, such as airplane cockpits, as well as command-and-control centers and control towers. Landry et al. [103] found certain unrelated gestalt motion patterns on radar screens drew the attention of air traffic controllers away from the task at hand. BRAVVO could help in such situations in its ability to holistically determine critical events to assist the human operators. The proposed approach is general enough to be applicable to any human-machine interface in complex environments.

3.2.6 Acknowledgements

Chapter 3 is in part a reprint of material that is submitted to the Journal of Vision (2010), and IEEE Transactions on Intelligent Transportation Systems, by Anup Doshi and Mohan M. Trivedi. The dissertation author was the primary investigator and author of these papers.

This research was supported by research grants from the University of California Transportation Center (US Department of Transportation Center of Excellence), the UC Discovery Program, Volkswagen, and the National Science Foundation. The author is grateful for the advice and support of colleagues from the CVRR laboratory.

Chapter 4

Cognitive State and Intent

In the following chapter we take a step deeper into the cognitive state of the driver, to understand how driver behaviors are affected by recent actions. We find in particular that certain sequences of cues or responses tend to prime the driver, significantly altering their response behavior.

Basic laboratory experiments in cognitive science have shown that there are significant performance effects based on recent experience. We emulate these studies in the more complex driving environment, and for the first time demonstrate the existence of sequential effects outside simple laboratory tasks.

Further, we are able to demonstrate a potential cause for unintentional behaviors: Sequential priming may be one cause of driver mistakes, in which the intend to hit one pedal, but mistakenly step on the other. This phenomena is more commonly known as "pedal misapplication."

4.1 Recency and Sequential Effects

In many naturalistic tasks, it is critically important for an individual to respond quickly to a sequence of cues in a rapidly changing environment. For example, drivers on highways worldwide are increasingly getting stuck in stop-and-go traffic [29], and are forced to engage in a sequence of braking and accelerating actions as dictated by cues from other vehicles. These situations sometimes take a turn for the worse: over 3,500 vehicles were involved in fatal rear-end collisions on highways in the U.S. in 2008 [145].

In this research, we suggest that the recent sequence of stimuli and responses—here, cues to brake and accelerate—have a profound effect on driver response times. Such *sequential dependencies* have been studied in the psychological literature for over half a century [5, 54, 171]. A sequential dependency is an influence of one incidental experience on subsequent experience.

Sequential dependencies arise when individuals perform a task repeatedly or perform a series of tasks, and performing one trial influences behavior on subsequent trials. Diverse behavioral measures are affected, including response latency, accuracy, type of errors produced, and interpretation of ambiguous stimuli. Recency effects occur across all components of the cognitive architecture, including perception [116], selective attention [96], language [19], decision making and judgment [88], and motor control [39, 40, 215].

Sequential dependencies suggest that what might appear to be intrinsic uncertainty in human behavior can instead be understood in terms of a systematic response to a changing world. Recent theoretical perspectives characterize sequential dependencies in terms of an individual's adaptation to the statistical structure of a nonstationary environment [134, 216, 222]. Nonstationarity implies that recent events provide a stronger predictor of the future than events further back in time.

Although reliable patterns of sequential dependencies have been demonstrated in controlled experimental tasks, it is not known how robust these effects are outside the laboratory. Typical controlled experiments involve two alternative force choice (2AFC) button presses, neatly parceled into discrete trials. In contrast, driving requires continuous, ongoing behavior, and graded responses that include pedal depression and steering adjustments. In 2AFC, the response-to-stimulus interval is very brief, typically on the order of 500-1000 ms. Recency effects may be short lived and not persist in more naturalistic environments where the events occur irregularly and less frequently. Finally, in 2AFC, participants are focused on a single task, quite different than a situation such as driving, where multiple goals and extraneous stimuli compete to control an individual's attention, and where multiple responses (steering, pedal depression) are required simultaneously.

Here, we examine sequential dependencies in a more complex, naturalistic, time critical task. In our experiment, participants navigated a vehicle in a driving simulator, and were occasionally required to brake or accelerate, as indicated by a traffic-light style red or green cue on the windshield. We demonstrate the existence of sequential dependencies in both pedal-press latencies and errors.

4.1.1 Method

Participants

Eighteen volunteers participated in the experiment, 6 female and 12 male. The participants were in their 20s and 30s and varied in driving experience levels from beginner to decades of driving experience.

Figure 4.1: Simulator layout and experimental setup. The screen is shows a typical driver view in the *complex* condition; the left inset shows a typical view in the *simple* condition. The right inset shows a sample frame of the camera used for initial foot motion detection.

Apparatus

The experiment was controlled by a PC running an interactive open source simulator, TORCS [204], depicted in Figure 4.1. The ego-vehicle and road parameters such as friction and vehicle weight were fixed to approach real-world conditions. However they were also constrained in order to facilitate an easy learning process, as some subjects had never used driving simulators before. In particular, the maximum speed of the vehicle was limited, and the dynamics of the vehicle were set such that the car tended to travel at a slow, constant velocity when the pedal was not pressed.

A 52" screen was placed 3 feet in front of the subject, with a Logitech steering wheel mounted at a comfortable position near the driver, and accelerator and brake pedals placed accordingly. Mounted onto the toy steering wheel was a full-size steering wheel (15") typical of those found in compact cars. The brake and accelerator pedals, positioned and sized similar to realistic automobile configurations, registered continuous deflection.

All the driving parameters such as pedal position, lateral position of the vehicle in the road, and steering wheel data were automatically timestamped and logged to disk every 10 milliseconds. Additionally, a camera synchronously captured views of the driver's foot movements at 30 frames per second using an NTSC color camera.

Design

Each participant was run in two conditions: *simple* and *complex*. In the simple condition, the simulator was configured with a wide, flat, straight track that required little steering effort to keep the car on the track. No other vehicles were present, and lane markings, background scenery, and sky features were eliminated. In the complex condition, the track was designed to be a constant S-curve, i.e., a series of switchbacks; the environment was populated with buildings, trees, and other scenery such as mountains in the distance; and the track consisted of three lanes, though no other vehicles were present.

Each condition was run as two continuous driving episodes, with a short break between conditions and between episodes. The simple condition always preceded the complex condition, to give participants practice with the simulator. Trials were embedded within a driving episode and consisted of a cue (red or green disk) and the corresponding driver response (braking or accelerating, respectively). In the simple condition, the interval from response to the next stimulus cue (RSI) was two seconds. In the complex condition, the RSI varied from one to four seconds, drawn from a uniform distribution.

Each driving episode consisted of seven practice trials followed by 128 experimental trials. Experimental trials were counterbalanced such that all stimulus sequences of length five (e.g., red, green, red, red, green) occurred equally often, resulting in eight replications of each five-back sequence per episode.

Procedure

Participants were told that the experiment examines driver behavior in stop-and-go traffic. They were instructed to respond to a series red and green cues which appear as circular traffic-light style disks in the central field of view; red and green cues required participants to "pump" the brake and accelerator, respectively, with their right foot. Participants were told not to hold their foot on the brake or accelerator position and that the default foot position should be a neutral one in which neither pedal was depressed. Participants were allowed to hover their driving foot over either pedal. Most hovered over the accelerator, as is typical in naturalistic driving, but some hovered over the last pedal that was depressed. Although this hovering requirement is not entirely naturalistic, it is reminiscent of stop-and-go driving and participants had no difficulty following the instruction.

The cues remained visible until a response pedal was pressed. If neither response pedal was pressed within one second of stimulus onset, the trial was terminated and the next stimulus appeared following the RSI.

Data Analysis

A pedal press was defined to occur at the time when either of the pedals was depressed 20% or more of its maximum. Several participants tended to braise one pedal briefly along a trajectory to pressing the other pedal, and we determined criteria—consistent with video observations—that considered the second pedal press as the response.

The video data coming from the foot camera was also analyzed to detect the time of the initial foot movement, and to allow us to decompose the pedal press latency into a foot movement initiation time and a time to move the foot to the pedal. The video processing algorithm used a moving pixel-wise median background to distinguish novel changes in the current frame. Due to the camera perspective, these changes were caused mostly by foot motion; thus any major foot motion could be detected by placing a threshold on the amount of motion in the frame. Once a foot movement event was detected in this manner, the foot movement signal was traced back to the initial inflection point of the motion, to determine the time of the beginning of the foot motion.

4.1.2 Results

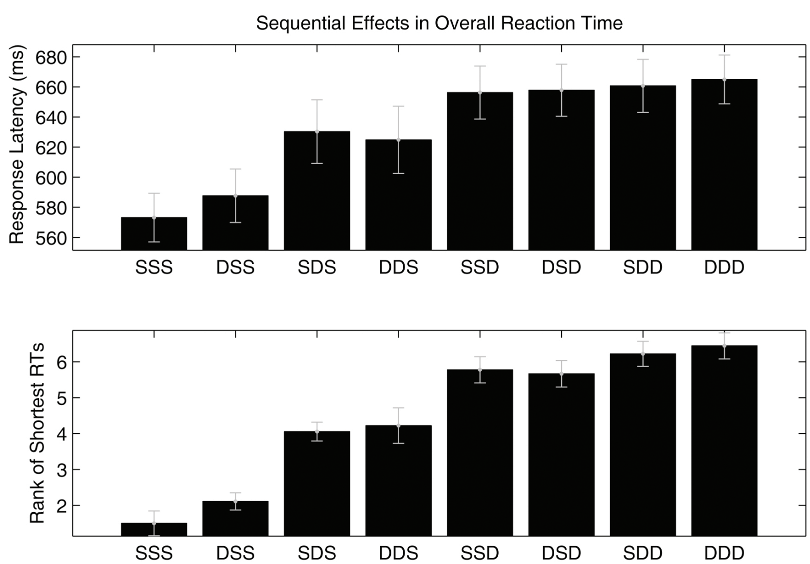


Figure 4.2: Sequential Effects in Pedal Press Response Times (top) and ranking of RT according to the recent sequence (bottom). Error bars indicate one standard error of the mean.

Each trial is characterized by three independent variables: (1) the driving environment (simple versus complex), (2) the target response (accelerate versus brake), and (3) the context arising from the three most recent trials. The context is specified in terms of whether the stimulus (or target response—the two are confounded in the present study) on the current trial, trial n, is the same as (S) or different than (D) the stimulus on trials n–3, n–2, and n–1. For example, the context SDD represents a sequence in which trial n–3 is the same as trial n, but trials n–2 and n–1 are different, which would correspond to the cue sequence red-green-green-red or green-red-red-green (ordering is always with the oldest first).

Participants made relatively few response errors. We discuss errors later, but first present an analysis of response times (RTs). For this analysis, trial n is discarded if an error has occurred on any of trials from n–3 to n, or if the RT is not within three standard deviations of the individual participant's mean RT.

A main effect of driving environment is found: RTs are slower in the complex environment than in the simple (654 ms versus 612 ms, $F(1, 17) = 27.8, p < .001$), expected given the additional distractors and variability present in the complex environment. Participants are slower to brake than to accelerate (659 ms vs. 607 ms, $F(1, 17) = 6.97, p < .02$), not a big surprise because in normal driving the foot tends to hover over the accelerator pedal. Finally, there is a main effect of context ($F(7, 119) = 26.99, p < .001$). However, there are no higher order interactions involving any of the independent variables (all $p > .275$). Consequently, we collapse across driving environment and response pedal in all subsequent analyses.

Sequential Dependencies on Response Time

The mean RTs across subjects for each of three-back context are shown in Figure 4.2. The contexts are ordered by similarity to the current trial, with recent trials being same response on the left and different response on the right. Similarity has a robust effect on RT. For the most extreme context contrast—SSS versus DDD—there is a reliable difference of 91 ms ($F(1, 17) = 63.46, p < 0.01$).

To remove the variability due to differences in mean RT and sequential effect magnitudes across participants, we transformed the eight context-specific RTs for each participant to a rank score from 1 (shortest RT) to 8 (longest RT), shown in the bottom half of Figure 4.2. The mean rank obtains smaller error bars. The pattern in Figure 4.2 is consistent with a simple response (or stimulus) priming model that posits that recent trials leave behind a rapidly decaying trace, and the response on the current trial is faster to the extent that the trace is consistent with the current response (or stimulus). This model predicts a monotonically increasing pattern of RTs from left to right in Figure 4.2.

Decomposing the Response Time

Using automated video-based analysis of the camera aimed at the participant's foot, the overall response time (ORT) can be decomposed into two stages: the time from stimulus onset to initiation of foot movement (FMI), and the time to move from initiation to pedal press (TTM). Trials for which the analysis technique failed (7.87% of the overall trials) were discarded. These

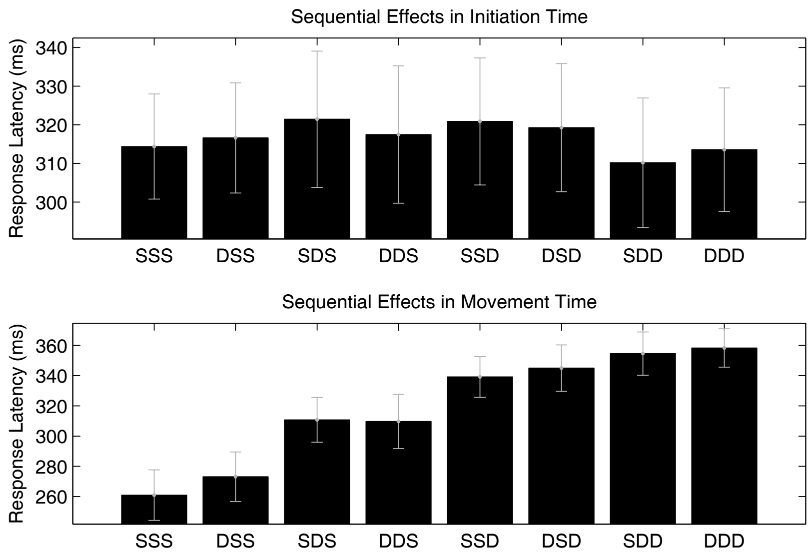


Figure 4.3: Sequential effects on response time, broken into foot movement initiation time (top) and time to move to pedal (bottom).

cases were caused either by foot motion which began before the stimulus appeared (causing ambiguity in determining response initiation), or by the lack of significant motion prior to the pedal press.

Figure 4.3 shows sequential effects for the two components of ORT. Systematic variation is observed in TTM but not in FMI. Indeed, the sequential dependencies in TTM appear to be as robust as those in ORT. For example, contrasting the extreme contexts of SSS and DDD, the TTM difference is 89 ms in the simple condition and 104 ms in the complex condition, comparable to the difference in ORT. However, no SSS/DDD context X simple/complex interaction was observed for TTM ($F(1, 17) = 1.4, p > .25$).

Whereas FMI is not susceptible to sequential effects, it is affected by the interval from the preceding response to the current stimulus (RSI). In the complex condition, we divided trials into shorter and longer RSIs, with the dividing line being the median RSI of 2.52 seconds. FMI was reliably slower following short RSIs (FMI of 343.8 and 365.7 ms for short and long RSI, respectively; $F(1, 17) = 14.55, p < 0.01$). However, RSI had no systematic effect on TTM (298.6 and 301.1 ms for short and long RSI; $F(1, 17) = 0.28, p > 0.25$), suggesting a double dissociation between FMI and TTM. A stimulus readiness effect is observed in the early stages of processing that determine FMI, whereas sequential dependencies are observed in later stages reflected in the TTM.

Table 4.1: Percent Variance Explained by Three Models For Three Different RT Measures

	Model		
RT Measure	1st Order	2nd Order	1st+2nd Order
FMI	9.81	8.23	12.61
TTM	94.32	17.56	95.32
ORT	87.91	16.97	89.68

First- And Second-Order Sequential Effects

Theoretical accounts of sequential effects in two-alternative forced-choice tasks [32, 216], have obtained exquisite fits to data—accounting for over 98% of the variance in mean RT conditioned on the recent context—by assuming two distinct, additive mechanisms. Following the terminology of Wilder et al. [216], we refer to these effects as first and second order. First order effects depend on the recent history of stimulus/response identities, i.e., the S-D context depicted in Figures 4.2 and 4.3. Second order effects depend on the history of repetitions (R) and alternations (A) of the stimulus/response. For example, the trial sequence red-green-red-green corresponds to the first-order sequence DSD and the second-order sequence AAA. Although the first and second order sequences are in one-to-one correspondence, they can be dissociated in terms of the priming that they predict. For example, the sequence red-green-red-green produces first-order priming for green because two recent trials have been green, but strong second-order priming for red because the three recent trials have been alternations. In fitting data from 2AFC tasks, first- and second-order priming both contribute roughly equally to explaining RT variability [216].

To investigate the relative contribution of first- and second-order effects in the current experiment, the data were fit to models that assume that priming from recent trials decays exponentially. In addition to testing pure first- and second-order models, we also tested a combined model in which both forms of priming were allowed. The pure models have one free parameter—the memory decay rate; the combined model has three—two decay parameters and one additional parameter for a mixture weight. (See [216], [222], for details of these models.) Parameters for each model are obtained by minimizing the squared prediction error.

Table 4.1 reports the percent of variance explained by the three different models for the three different RT measures—ORT, FMI, and TTM. As expected, no model does a good job of fitting the FMI, which has no systematic relationship with the recent history. The first-order model does a good job of fitting the patterns of ORT and TTM, consistent with what one observes in Figures 4.2 and 4.3.

Although the second-order model obtains a poor fit, second-order effects may be present but dominated by first-order effects, as suggested by the fact that the combined model shows an improvement in fit over the first-order model for all RT measures. However, the improvement could be due only to the combined model's additional degrees of freedom. Comparing the first-

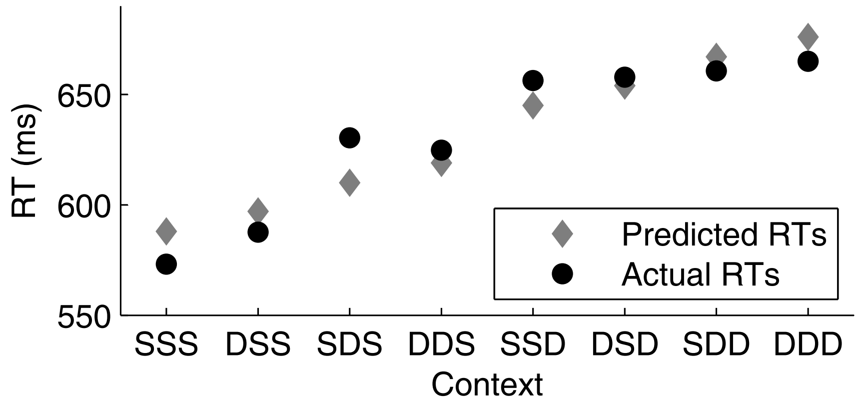


Figure 4.4: Mean overall reaction time (ORT) as a function of trial context (circles) and fit of first-order model with exponentially decaying memory (diamonds)

order and combined models, a likelihood-ratio test fails to reject the null hypothesis that the two models obtain a comparable fit (for all RT measures, $p > 0.25$). Thus, reaction times exhibit primarily first-order priming effects that can be well explained by a first-order model with exponential memory decay (Figure 4.4).

Another way of assessing the magnitude of sequential effects is to construct a *lag profile* that shows the influence of the nth trial back on the current trial. To construct a first-order lag profile, the RT data is partitioned by whether the current trial is the same or different as the nth-back trial. The difference in mean RT between these two conditions reflects the extent of priming. The upper graph of Figure 4.5 shows the first-order lag profile for the three response time measures.

The lag profile confirms that for ORT and TTM, first-order effects are statistically reliable at lags 1, 2, and 4 ($p < 0.01$ by one-tailed t-test), but not at lag 3 (ORT lag 3 $t(17) = 1.20, p = 0.248$; TTM lag 3 $t(17) = 1.79, p = 0.092$). There is no significant first-order priming of FMI ($p > 0.10$ for all lags).

The second-order lag profile (lower graph of Figure 4.5) is similar except that each trial is characterized as a repetition or alternation (of the previous), and the lag n priming is the speed up that occurs when the current trial matches the nth trial back. Second-order priming at lag 1 is identical to first-order priming at lag 2. Consequently, evidence for second-order effects comes from reliable priming at lags 2 or 3. Second-order priming at lags 2 and 3 is statistically reliable for ORT (lag 2: $t(17) = 3.73, p < 0.01$; lag 3: $t(17) = 2.70, p = 0.015$), albeit small in magnitude. For TTM, second-order priming is not statistically reliable (lag 2: $t(17) = 1.27, p = 0.221$; lag 3: $t(17) = 1.83, p = 0.086$).

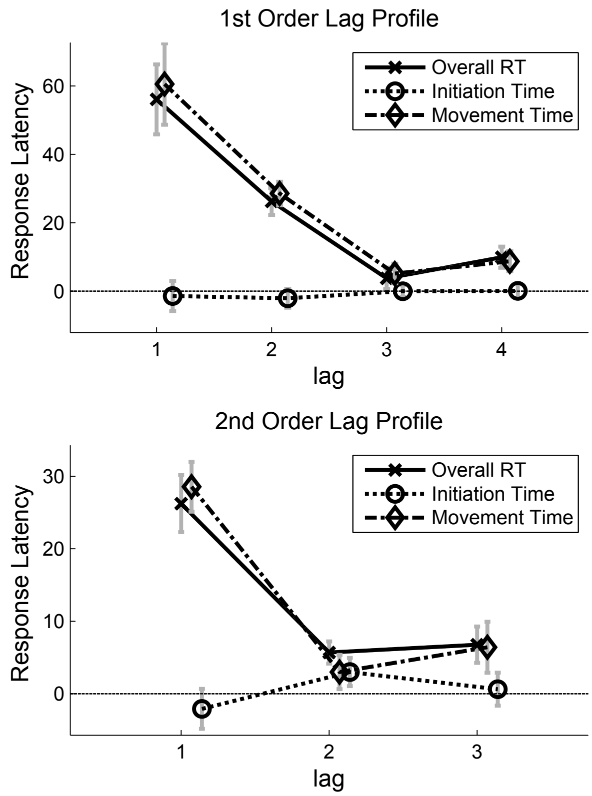


Figure 4.5: Lag plots of first- and second-order sequential effects on overall RT (ORT), foot movement initiation Time (FMI) and time to move to pedal (TTM). The error bars represent +/-1 standard error of the mean.

Pedal Misapplications

Sequential effects typically arise in error rates as well as in latencies [32], with the slower responses being more error prone, indicating that sequential effects are not merely a speed-accuracy trade off.

Trials in which the participant initially fully pressed the wrong pedal were rare in the experiment, occurring on only 2.33% of trials. In nearly every case, participants detected their error and corrected it afterward (e.g., Figure 4.6). Due to the small number of errors, instead of comparing error patterns in three-back contexts, we examined errors based only on the previous trial context.

Pedal misapplications occurred on 3.30% of trials when the previous stimulus differed from the current, but only on 1.46% when the previous stimulus was the same as the current (reporting a subjectwise median). The difference is significant by a Wilcoxon signed-rank test ($Z = 2.46, p < 0.05$). On 0.39% of trials, participants made no response. When these no-response trials are included, the effect of the previous trial is still evident ($Z = 2.16, p < 0.05$).

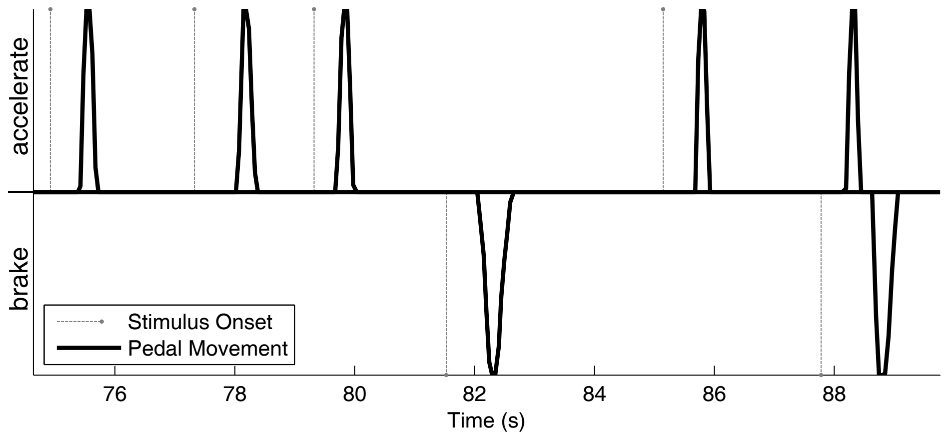


Figure 4.6: Sample series in the complex condition of the experiment, showing an example of pedal misapplication. After a number of accelerate cues in the recent history, at the 88-second mark, the driver sees a brake cue but accidentally steps on the accelerator instead.

4.1.3 Discussion

In a simulated driving task, we observe a robust effect of the recent sequence of pedal presses on both response latencies and error rates. The task is perhaps the most realistic, naturalistic scenario in which recency effects have been explored. The effects extend many trials back, certainly beyond the immediately preceding trial. The effects are what we characterize as first order, corresponding to the traditional conception of priming of the response (or stimulus, as the two are confounded). In contrast to laboratory 2AFC tasks, second order effects—effects of the repetition and alternation sequence—are not observed.

Our experiment extends the canonical laboratory 2AFC task toward a more naturalistic scenario in three respects. First, the visual environment is cluttered, heterogeneous, constantly changing, and contains elements that distract from the pedal pressing task, elements such as buildings, trees, mountains, and grass. Second, driving requires multitasking: in addition to braking and accelerating as demanded by the pedal pressing task, the driver needs to constantly monitor and adjust steering to remain in the center lane of a regularly curved road. Third, the response-to-stimulus intervals are variable, ranging from 1 to 4 seconds, and are larger than the typical interstimulus intervals of less than a second.

Even the simple condition in our experiment—which offered little in the way of distractor stimuli and required minimal multitasking—moves in the direction of ecological validity over the canonical 2AFC task. However, the finding that sequential effects are robust and comparable in the complex condition supports the notion that environmental and task complexity—via additional variability of the stimulus onset, additional visual distractors, and additional difficulty of the secondary (driving) task—do not modulate the nature of sequential effects.

Although it's not altogether unexpected to find priming in a complex, multitasking environment, a further analysis of behavior offers an intriguing suggestion concerning the stage of processing at which priming occurs. We decomposed the overall response latency into the time from stimulus onset to foot movement initiation (FMI), and the time from movement initiation to pedal press (TTM). We find a double dissociation: only TTM is affected by the trial sequence, and only FMI is affected by the response-to-stimulus interval.

Our motivation for the RT decomposition is a related decomposition that has been performed via event-related potentials to analyze sequential dependencies. In a 2AFC task with the two responses mapped to the two hands, the start of response selection can be inferred as the onset of the lateralized readiness potential (LRP). Jenztsch et al. [86] and Jones et al. [90] have analyzed sequential effects on RT, decomposing the overall RT into stimulus processing (pre-LRP) and response processing (post-LRP) stages. Wilder et al. [216] have shown that the stimulus processing component of RT is well fit by a second-order model of sequential dependencies, and the response processing component is well fit by a first-order model. That is, sequential dependencies are based on frequency of repetitions and alternations of the *stimuli*, and on the frequency of the two alternative *responses*.

The presence of first-order effects in the TTM in the driving task is predicted by the dual priming theory, because TTM reflects a late stage of processing. However, on its face, the theory predicts second-order dependencies at an early stage of processing, which are not observed in the FMI. It remains to be determined whether the absence of second-order effects is a failure of the theory or is attributable to another factor that differentiates the driving task from canonical 2AFC tasks, e.g., the complexity of the visual environment, multitasking, or the asymmetry of pedal presses.

What are the practical implications of recency effects in driving? Consistent with previous studies using simple 2AFC tasks, we find that recent events can result in response delays of nearly 100 ms. Because sequential dependencies are found in complex laboratory conditions that approximate real-world driving, it seems altogether plausible that they will be observed in situations such as stop-and-go traffic, or a series of traffic lights in an urban environment.

At 65mph, a delay of 100 ms corresponds to an increased stopping distance of nearly three meters, which could be the difference between running into a leading vehicle and stopping safely. Even smaller differences in response times have been shown to have severe consequences in both the probability and severity of vehicle crashes [55].

Beyond delays in responding, we have shown that recent events can affect the likelihood of a pedal misapplication. In our study, the probability of a pedal misapplication more than doubled if the target pedal differs from the previous pedal versus when the two are the same.

Many recent unintended-acceleration-related accidents in the U.S. may have been a result of pedal misapplications [72]. The context-dependent pattern of driver response delays and errors could potentially be exploited in a holistic Advanced Driver Assistance System, or ADAS [44, 207]. Given a particular observed history of pedal presses and familiarity with a particular driver, the ADAS could predict real-time performance. These predictions could be used to provide additional or earlier alerts to drivers in situations where delays/errors were likely. In more critical circumstances, the vehicle could take action to reduce impending accident severity by priming the brake system. In less critical situations, if the upcoming context is ripe for higher response times (e.g., more than 600 ms), an urgent alert could help to refocus the driver to the potentially appropriate response. An adaptive intelligent ADAS design could thus utilize sequential context to help the driver and mitigate dangerous or uncomfortable circumstances.

4.1.4 Acknowledgements

Chapter 4 is in part a reprint of material that is submitted to Cognitive Science (2010), by Anup Doshi, Cuong Tran, Matthew H. Wilder, Michael C. Mozer, and Mohan M. Trivedi. The dissertation author was the primary investigator and author of this paper.

This research was supported by research grants from the University of California Transportation Center (US Department of Transportation Center of Excellence), the UC Discovery Program, Volkswagen, and the National Science Foundation. The author is especially grateful for the insight and support of Prof. Michael Mozer, and Cuong Tran, along with the support of Matthew Wilder, Colter Cederlof, and other colleagues from the CVRR laboratory.

Chapter 5

Real-time Systems and Analysis

We are now impelled to ask which cues are the most indicative of a driver's intentions, given that visual search and recency are demonstrably important considerations. In order to implement a real world system, we must remain conscious of systems that are cost effective, along with the robustness of both sensors and cues. In the first part of this chapter, we compare the most useful cues in a comparative framework for intent prediction. We then explore the real-world performance of such a real-time intent inference system, implemented as part of an on-road driver assistance system.

5.1 Comparing Cues to Intent

The design of robust and practical intelligent assistance systems is an active field of research and development. In particular, the type and placement of sensors to achieve maximum performance has yet to be well understood. Several recent systems, including collision warning and brake support, backup warning systems, and lane departure warning systems, use very specific environmental sensors associated with their application to augment the driver's awareness. However the basis of all these systems are sensors detecting the environment outside the vehicle, along with the vehicle dynamics.

Recent research has supported the incorporation of sensors looking inside the vehicle into these systems, in a holistic manner [207, 208]. A major advantage of monitoring drivers is the ability to observe driver behavior and potentially infer driver intent. Such data could inform an ADAS in several manners: (1) *Awareness*: Is the driver aware of the pertinent surroundings in their environment? What is the focus of attention and distraction level? (2) *Intent*: Is the driver planning to move into a dangerous situation? What is the planned trajectory or possible movements?

Given the driver's attention patterns, it may be possible to infer whether or not the system should warn the driver and reduce false alarms.

In particular we are interested in determining the important driver cues for distinguishing intent, in order to support future ADAS designs. In prior intent prediction research [30, 121, 120], *head dynamics*, a derivative of head pose, has been proposed as a pertinent cue. While robust monocular in-vehicle head pose estimation systems have been developed [137, 10, 229], it may be argued that head pose is not a sufficient estimate of true gaze. In order to derive precise gaze estimates, it follows that *eye gaze* should be included [191]. Unfortunately there are several drawbacks with modern eye-gaze estimators in vehicles, including the need to overcome lighting changes, shadows, occlusions, and potentially cumbersome stereo rigs or intrusive head-mounted cameras. Therefore we are motivated to determine if eye gaze and head motion are useful intent predictors, and furthermore which one (or combination) is the more informative cue.

In this section, we use a lane change intent prediction system [121] to determine the relative usefulness of eye gaze and head dynamics data. Our comparative experiments are designed to distinguish the merits of the two cues and compare their importance. By determining the better cue, we hope to provide the basis for appropriate future designs of lane change intent systems, as well as a foundation for interactive driver assistance systems in general.

5.1.1 Driver Behavioral Cues

The analysis of driver behavior has long been a popular field of research in light of the potential for safety improvements. With respect to the particular maneuver of lane changes, the analysis of driver behaviors dates back at least 30 years.

Here we present a summary of relevant research. We then present our methodology for determining driver behavior, in preparation for our comparative experiments.

Related Research in Lane Change Behavior Analysis

According to early research in the field, there is significant reason to believe that behavior analysis of the driver can lead to reliable predictions about lane change intent. As described below, in order to safely decide to change lanes, a driver should have recently given some *attention* to the occupancy and state of neighboring objective lane. According to Hoffman [77], "attention is free to move independent of the eyes, but eye movements require visual attention to precede them to their goal". Thus by measuring the driver behavior corresponding to a visual search, we can hope to capture obvious shifts in attention and thereby deduce lane change intentions.

The time period three seconds ahead of the actual lane change was determined to be a critical time period during which the driver engages in a visual search to determine feasibility of lane change [133]. In the period several seconds prior to the lane change maneuver, certain patterns emerge in driver' visual search behaviors.

In fact according to Tijerina et al. [199], there are specific eye glance patterns which take place in the period before a lane change. It was determined that during left lane changes, there were between 65-85% chance of looking at the left mirror, and 56-67% chance of looking at the rearview mirror. Correspondingly, during right lane changes drivers looked at the right mirror with 36-53% probability, and the rearview mirror with 82-92% probability. Moreover the mirror glance duration before lane change manuevers lasted on average 1.1 seconds, varying between .8 and 1.6 seconds [106]. Mourant and Donohue observed that lengthy blind spot checks occured only in conjunction with lane change maneuvers; in lane keeping situations no such checks were performed by the drivers [133].

Bhise et al. [16] studied a series of naturalistic lane changes in real world settings, and discovered that most visual searches prior to lane changes involve head motions, and relatively very few (5.4% in their study) involve eye glance alone. Furthermore, Robinson [175] found a remarkably stable relationship between eye glance behavior and head movement behavior: In an experiment where a visual fixation was placed at 60 degrees, the eyes moved first and the head followed approximately 50 ms later. The relationship that head movement immediately follows eye movement was found to be stable across all individuals in the experiment.

The experiments of Land [102] corroborate these results in a real automotive setting. Upon examining the head movements and eye behavior of several drivers approaching intersections, the author found some remarkable tendencies in the drivers' behavioral patterns. When a decision to change gaze has been generated subconsciously or "unthinkingly", eye and head movements by "default" begin at nearly the same time. The eyes move quicker than the head; however the velocity of the head movement is a direct correlation of the magnitude of the gaze change (i.e., the head moves faster for a larger gaze change). These results indicate that head and eye movements are correlated under an unguided visual search, in a situation similar to the search prior to a lane change.

These results lead to the hypothesis that eye gaze and head pose can be reliable indicators of a driver's intent to initiate a lane change. Furthermore, it might be posited that head pose alone could be good enough, given that fixations tend to reliably draw head movements along with eye gaze changes.

Other studies more recently have included eye gaze measurements as a part of laboratory tests of driver fatigue [110, 89] or of simulated lane change events [182, 226]. Simulators, though, do not capture all the dynamics and variability of real-world environments [6]. Some real-world studies of driver behavior during lane changes have measured eye gaze by manually reducing data [106, 6, 199, 155, 81]. By doing so they can ensure the reliability and accuracy of the eye gaze data; these studies have shown some promise of using eye gaze as a cue for driver intent. There have also been real-world studies that relied on automatically detecting eye gaze, but their results were limited due to robustness issues, especially with regards to occlusions from sunglasses and harsh lighting conditions [111, 192, 13, 14]. Finally, there have been several studies that achieved promising results using just head motion as a cue for behavior prediction [121, 120, 30, 31].

As they involve testing more dangerous situations, we leave the in-depth study of the lane change behaviors in the presence of of fatigue, distractions, and heavy traffic to future research; here we assume they play no more than a minor role in the process of changing lanes. In order to fully study the effects of these variables, simulator studies could be developed. In the following research we investigate highway situations without heavy traffic or abnormal fatigue levels and distractions, in the interest of driver safety.

Heming et al. [75] examined the predictive power of "glances to the left mirror" prior to lane changes. They found the left mirror glances are a good predictor of intention to change lanes, but should be combined with other indicators to reduce false alarms. The participants in the study were told about the goals of the study, however, potentially changing their behavior. In fact the participants' near-ubiquitous usage of turn signals to indicate lane changes was significantly higher than turn indicator usage from naturalistic studies of drivers' lane change behaviors [106]. As described below the data used in our study comes from experiments in which drivers were not told about the goals of the study, increasing the likelihood of naturalistic behaviors.

Prior studies have suggested that head pose may be better suited than eye gaze to infer lane change intent [45] without providing any statistical evidence. This research significantly expands upon the analysis and extends those preliminary results, including a larger sample size and quantified significance tests which finally allow concrete conclusions to be drawn. No other study has been found that quantifiably compares the predictive power of each of these cues in determining a driver's intention to change lanes. Additionally, we will propose an biological explanation for why head pose consistently has more predictive power than eye gaze, earlier in time prior to a lane change.

Experimental Data Collection and Reduction

For this research, data was collected in a driving experiment with an intelligent vehicle testbed outfitted with a number of sensors detecting the environment, vehicle dynamics, and driver behavior. This data is the same data as was used in the lane change intent work by McCall et al [121]. A lane position detector and CAN-Bus interface provided most of the data related to the vehicle and surrounding environment. The lane detector was a camera-based lane detector based on the VioLET lane tracker [119], with the camera mounted on the top right of the windshield.

The main driver-focused sensor was a rectilinear color camera mounted above the radio facing toward the driver, providing 30fps at 640x480 resolution. This data from this camera was collected and post-processed to extract behavioral cues based on head pose and eye gaze, as described below.

The dataset was collected from a naturalistic ethnographic driving experiment in which the subjects were not told that the objective was related to lane change situations. Eight drivers of varying age, sex, and experience drove for several hours each on a predetermined route. A total of 151 lane changes were found on highway situations with minimal traffic, at speeds in the range of 55-75mph. 753 negative samples were collected, corresponding to highway "lane keeping" situations.

Head Dynamics

In order to be invariant to illumination changes and independent of driver identity, head movement is estimated using optical flow and block matching, as described below.

In order to capture the essence of the head motion, optical flow vectors are calculated for each of the regions falling within the detected face region, which is found using the Viola/Jones face detector [211]. These vectors are calculated for each of the frames, within the window specified. The vectors are integrated over time, and separately over space; the integrated values are input as features to the classifier. In this manner any sort of rapid head movements will be captured, and the length of time and extent to which the head moved left or right will also be recorded. This methodology is based on the one developed by McCall et al. [121] and proves to be a robust estimator of head motion. Other methods could also be used to estimate and derive dynamical cues from the true pitch and yaw of the driver's head [137].

Figure 5.1 shows the head motion of a driver plotted along with the eye gaze patterns prior to lane changes. It can be seen that the driver's head motion increases before and during lane change maneuvers.

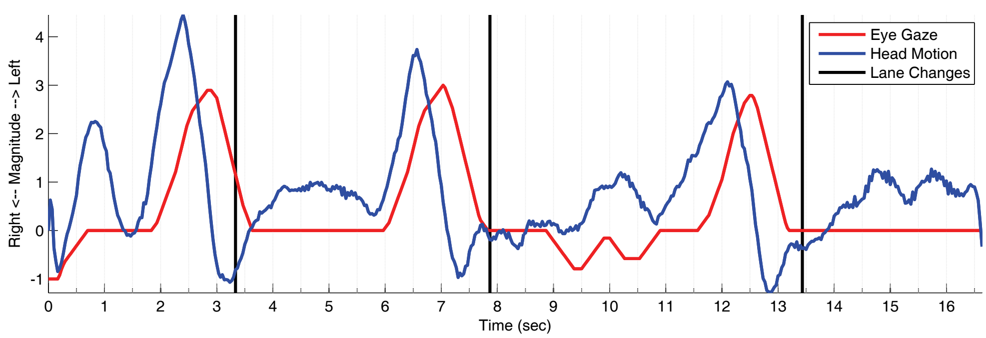


Figure 5.1: Head motion and Eye gaze positions prior to lane changes. Black lines denote lane changes. The red line represents the magnitude of the eye gaze shift left or right, and the blue line represents side-to-side head motion.

Eye Gaze

Because of the nature of the camera positioning, which was placed to minimize occlusion and distraction, automatic eye gaze measurements were deemed too cumbersome and unreliable to collect. The data came from a monocular camera mounted near the radio controls, in the center of the dashboard looking at an angle toward the driver. This angle was too obtuse for monocular eye gaze estimators such as [220] to work reliably. In an ideal world, a properly designed stereo or monocular eye gaze system would provide robust data. Therefore in order to

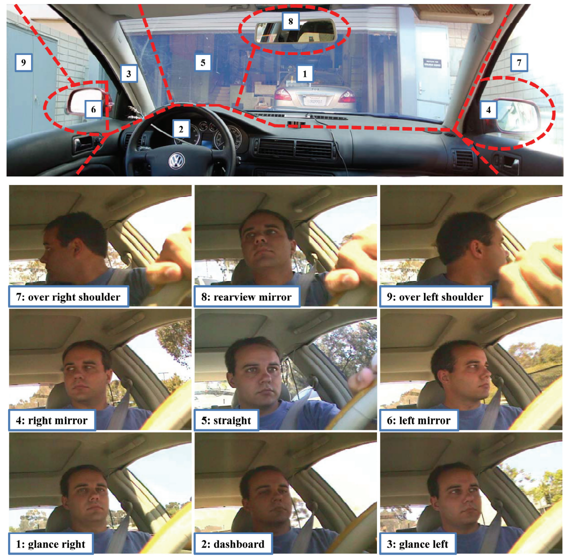


Figure 5.2: (Top) Approximate distribution of eye gaze location classifications for labeling purposes. (Bottom) Samples from dataset showing corresponding eye gaze locations.

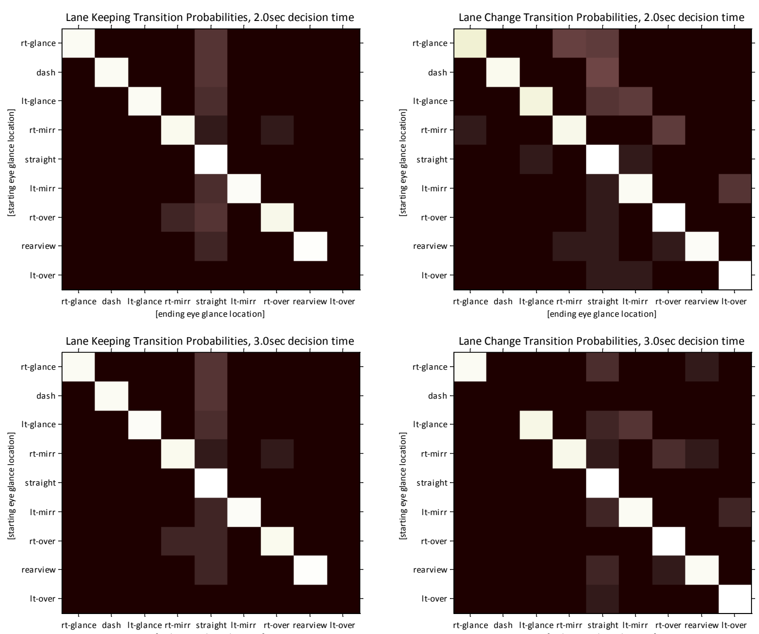


Figure 5.3: Eye glance transition probabilities for lane keeping and lane changing situations used in this study (brighter squares indicate greater likelihoods; black indicates no such data was found). During regular lane keeping situations, the driver is likely to be looking straight; any other glances will likely also end up looking straight. Prior to lane changes, it is much more likely for a left or right mirror glance to transition to a corresponding over-shoulder glance.

Table 5.1: Glance durations for 3.0sec-ahead and 2.0sec-ahead decision times. Note the lack of a pattern in the eye glances left or right before lane changes, whereas the over-the-shoulder look is more indicative of a lane change.

Time(ms) looking...		*Straight*	*Glance Lt*	*Glance Rt*	*Lt Mirror*	*Rt Mirror*	*Over Lt*	*Over Rt*
Time window between 4.0sec and 3.0sec	Before Lane Change	573.5	49.1	51.1	129.0	23.3	68.1	44.4
	Before Lane Keeping	840.3	21.8	53.2	12.5	3.4	0.0	1.2
Time window between 3.0sec and 2.0sec	Before Lane Change	416.2	51.1	20.5	158.3	26.3	182.9	96.6
	Before Lane Keeping	842.6	22.3	53.4	11.7	3.4	0.0	1.3

approximate the ideal case and retain the best possible chance of getting reliable and accurate eye gaze estimates, the data was manually reduced. The procedure followed was similar to those followed in the NHTSA lane-change and workload experiments [106, 6], to produce output that a real-world eye gaze tracker would output in an optimal setting.

Nine different gaze locations were derived from the procedure described in [6] as relevant to the task at hand: Looking Straight, Glancing Left or Right, Looking at Dash or Rearview Mirror, Looking at Left or Right Mirrors, and Looking Over Left or Right Shoulders. Sample images from each of these cases can be seen in Figure 5.2.

As described in Section 5.1.2, data over a span of one second prior to the decision time is input to the lane change intent system. The dynamics of the eye gaze over these time windows can be seen in Table 5.1 and Figure 5.3. Table 5.1 shows the average amount of time spent looking in certain directions prior to lane changes, compared to normal lane keeping situations. A significant difference is seen between lane change and lane keeping situations, especially in mirror and over-the-shoulder glances. Glance transition probabilities for each situation are shown in Figure 5.3. For each decision time, the eye glance behavior does not change significantly during lane keeping situations, as expected. However prior to lane changes, it is much more likely for mirror glances to end up in over-shoulder glances. Thus to robustly represent the eye gaze over this time period, several different formats were input to the system, including the raw eye gaze classifications as well as the histogram of glances over the time window.

5.1.2 Lane Change Intent Prediction

Driver intent inference is a challenging problem, given the wide variety of potential behaviors and intents. To limit the scope of the problem, we examine simply the driver's intent to change lanes, at a particular time in the near future. We base our experiments on the lane change intent system developed by McCall et al. [121], labeled the Driver Intent Inference System (DIIS). It is important to note that we are using this system as a basis for our comparative study about various input cues. There are a number of other works in this field [99, 153, 183, 81, 75]; our current research will hopefully help inform the future design of these and other intent predictors when considering which inputs to include.

Driver Intent Inference System

The DIIS is a discriminative classifier to distinguish between two events: lane changing (either right or left) and lane keeping. The following classes of variables are available to the classifier: *Vehicle State Variables*, *Environment Variables*, and *Driver State Variables.*

Vehicle State Variables include gas pedal position, brake pedal depression, longitudinal acceleration, vehicle speed, steering angle, yaw rate, and lateral acceleration; these are derived from the vehicle's CANBus network and are henceforth referred to as **Vehicle** Data. *Environment Variables* collected include road curvature metric, heading, lateral lane position, lateral lane position 10m ahead, and lateral lane position 20m ahead; and are referred to as **Lane** Data. For more information on the process and methodology for acquiring the lane data, please see McCall et al [121]. Finally, *Driver State Variables*, including the variables of particular interest in this research, namely head or **Head** dynamics and **Eye** gaze measurements, collected and preprocessed as described above.

Each of these variables, as a time series, is windowed to a length of 1 second prior to the chosen decision time. They are then concatenated into a large feature vector, from which a small subset of useful features should be chosen to determine the intent. In order to find these important features and their relative weightings, the Relevance Vector Machine is employed as described below. This classifier outputs a class membership probability, which can then be thresholded to determine a true positive and false positive rate for the predicted lane change intent.

Relevance Vector Machine

The Relevance Vector Machine (RVM) classifier used to train the DIIS is based on Sparse Bayesian Learning (SBL), developed by Tipping [200, 201] and implemented in [121]. The algorithm is a Bayesian counterpart to the popular Support Vector Machines (SVM); it is used to train a classifier that translates a given feature vector into a class membership probability. RVMs in particular use a parameterized prior to prune large feature vectors and facilitate a sparse representation of the data.

The basic form of the RVM classifier is given as follows:

$$y(\mathbf{x}) = \sum_{i=1}^{M} w_i K(\mathbf{x}, \mathbf{x}_i) \tag{5.1}$$

where $\mathbf{x}$ is the input feature vector, w_i is the learned model weight, and $K(\cdot, \cdot)$ is a kernel function. The output $y(\mathbf{x})$ then represents the probability that $\mathbf{x}$ belongs to a particular class.

For our purposes, the feature vector for each example $\mathbf{x}_i$ includes temporal blocks of each of the input cues described above. For example, at time t, the feature vector looks like

$$\begin{aligned}\mathbf{x}(t) \quad = \quad & [\text{LateralPos}(t), \ldots, \text{LateralPos}(t-N+1); \\ & \text{SteeringAngle}(t), \ldots, \text{SteeringAngle}(t-N+1); \\ & \text{EyeHistogram}(1), \ldots, \text{EyeHistogram}(9); \\ & \text{etc. }], \end{aligned} \tag{5.2}$$

where N represents the number of past values of each variable that have been stored internally; we selected N such that the feature vector represented a one second long sliding window of data.

A detailed description of the RVM algorithm can be found in [200]; the specific algorithm used in these experiments is described in more detail in [201]. The RVM has been shown to be quite effective in predicting lane change intent [121], and so for our comparative study we have selected it as a baseline, though other methods could be used.

Indeed several advantages of this methodology motivate the use of RVMs over other algorithms such as SVMs and Hidden Markov Models. The RVM is designed to sift through large feature sets and obtain a sparse representation of the data, which is especially useful in this application in identifying a small set of useful features. Multi-modal data from various sets of sensors can thereby be combined easily, with the discriminating cues from each modality automatically chosen by the RVM. The resulting sparse representation allows for quick computation and classification in real-time and real-world conditions with limited hardware.

The SBL methodology is general enough to consider cases, such as in our experiment, where there are relatively few training examples as compared with the number of features. As opposed to Support Vector Machines, RVMs also provide a theoretical framework for determining class membership probabilities. This allows the user to tune the decision boundary to achieve desired performance, in a principled manner. Finally, by including the windowed time series of cues in the final feature vector, and applying the kernel function, the RVM is capable of determining non-linear temporal relationships between features, eliminating the need for HMMs.

5.1.3 Experiments and Analysis

In this section we describe the analysis procedures and results from our experiments.

In order to predict lane change intent, the classifier needs to be trained for a particular decision time, with a given window of data prior to that time. Based on the prior research as well as the optimal results in [121], we decided to obtain results for two decision times of 2 seconds and 3 seconds prior to the lane change. By detecting a lane change this far in advance, an ADAS would be able to warn the driver in time for the driver to be able to react safely [182, 183, 121].

Table 5.2: True Positive and False Positive Rates for a Fixed Threshold (T=0) for the 3 second-ahead Decision Time

				TPR	FPR
Lane	Vehicle			47.97%	2.27%
Lane	Vehicle	Head		79.46%	0.66%
Lane	Vehicle		Eye	61.08%	1.82%
Lane	Vehicle	Head	Eye	75.68%	0.86%
		Head	Eye	71.08%	0.97%
		Head		70.95%	0.99%
			Eye	46.76%	1.88%

In each case data in a window of one second prior to that decision time was used to make the decision. The data was formatted as described above into a feature vector.

In order to counter the effects of scale in feature selection, each feature was renormalized to be zero mean and unit variance, where the mean and variance were estimated using the training data. The data was then sent through a Radial Basis Function kernel as described in the SBL algorithm, with a kernel width of 0.5.

Data was split into training and testing datasets, in a ratio of 80%-20% for a 5-fold cross validation. Five such randomized trials were conducted. Since the output of the classifier was a class membership probability, the decision threshold was varied across the range of probabilities to obtain an ROC curve for the set of trials.

In order to judge the relative effects of Head and Eye data, various combinations of input features were tested, by including or excluding some subset of Head, Eye, Lane, or Vehicle data from the feature vector. The results of these comparative experiments for the 3 and 2 second-ahead decision times are presented below.

The data for the 3-second-ahead decision time was collected in a window between 4 to 3 seconds in advance of the lane change. As can be seen in Table 5.2, as well as the comparison ROC curves using different input cues in the top left of Figure 5.4, we can make some general observations. It turns out "Eye" data basically has no effect on the performance of the detector. In fact the best performance occurs by using "Lane Vehicle and Head" data, with "Lane Vehicle Head Eye" barely below that. As discussed below, the reason for this dip may be noisiness in the eye data; especially since people glance around much more than they move their head. However in all cases, the performance of the 2.0-sec detector is much better.

Data for the 2-second-ahead decision time was collected in a window between 3 to 2 seconds in advance of the lane change. Once again the relative effects of Eye data are not great, as seen in Table 5.3 and the right half of Figure 5.4. We can note that adding Head data to Lane and Vehicle improves performance better than by adding Eye data to Lane and Vehicle.

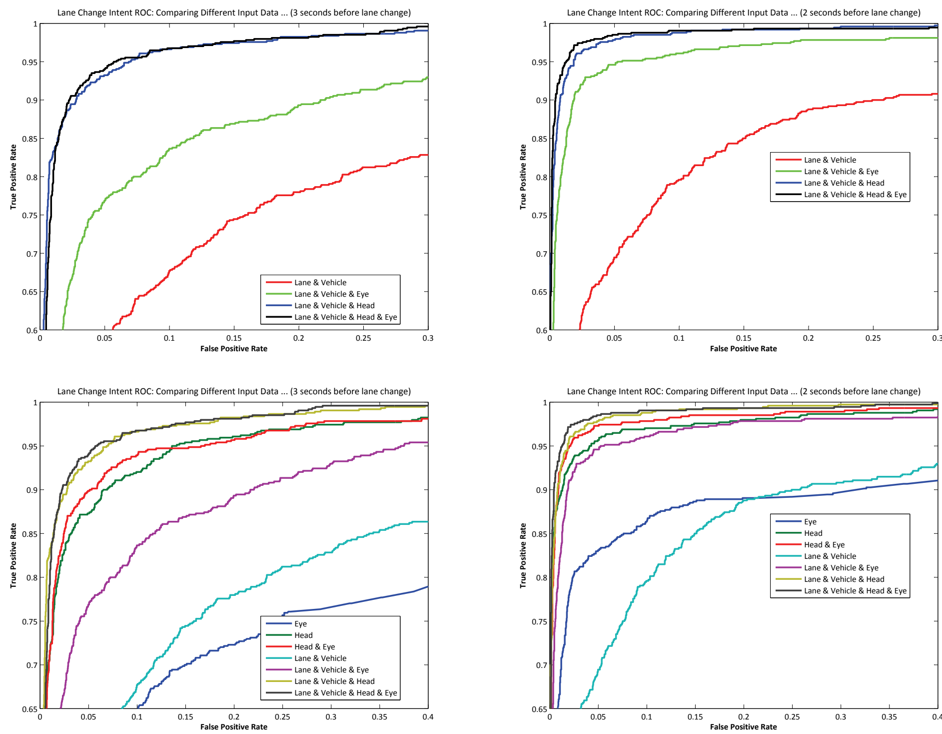


Figure 5.4: ROC Comparing Different Input Data, 3-sec and 2-sec Decision Time. The figures represent the same data, comparing the output of the classifier using various sets of inputs and times. The top figures shows that the addition of Eye data to Lane & Vehicle improves performance, but not as much as the addition of Head data. When using all four sets of inputs, the results are more or less the same as without the Eye data. Similar patterns are seen between the 3-sec (left) and 2-sec (right) Decision times. All the data are shown for comparison in the bottom figures.

Table 5.3: True Positive and False Positive Rates for a Fixed Threshold (T=0) for the 2 second-ahead Decision Time

				TPR	FPR
Lane	Vehicle			58.38%	2.18%
Lane	Vehicle	Head		88.51%	0.72%
Lane	Vehicle		Eye	78.92%	0.83%
Lane	Vehicle	Head	Eye	87.30%	0.39%
		Head	Eye	81.35%	0.39%
		Head		83.38%	0.33%
			Eye	75.14%	1.74%

Table 5.4: Statistical significance tests (1-way ANOVA). A p-value of less than 0.05 indicates that the two populations under test have significantly different means. Various combinations of Lane (L), Vehicle (V), Head (H), and Eye (E) cues are compared against each other. Δ*IPC-Means* shows the difference between the averages of each group's Intent Prediction Confidences. p-values from ANOVA and the Wilcoxon Signed Rank (WSR) Test are shown.

3 second Decision Time				
Comparing...		Δ*IPC-Means*	*ANOVA p-value*	*WSR p-value*
LVE	LVH	-0.438	0.00492	0.00781
LVE	LVHE	-0.423	0.00518	0.01563
LVH	LVHE	0.016	0.87432	0.54688
2 second Decision Time				
Comparing...		Δ*IPC Means*	*ANOVA p-value*	*WSR p-value*
LVE	LVH	-0.195	0.10465	0.03906
LVE	LVHE	-0.135	0.22351	0.19531
LVH	LVHE	0.060	0.52869	0.46094

However, the best performance is achieved by using all four sources of data.

In this case, as opposed to the 3-second-ahead case, the addition of eye data does improve performance of the overall detector. The improvement is slight but noticeable, whereas in the 3-second-ahead data the eye data had negligible effect. The progression of a lane change attempt, then, could be tipped off first by head movements 3 seconds before the lane change, and then closer to the 2-second-ahead threshold, eye movement data would become a useful additional input. This result is corroborated by the statistical significance tests presented below.

Statistical Significance

This study is intended to build upon the results of McCall et al. [121], hence the data used in this study comprises of the usable data obtained from that study, as described above. As this strategy involved a limited population of eight drivers, it is important to determine the statistical significance of the results before drawing conclusions.

To determine the statistical significance of the results we performed an analysis of variance (ANOVA). The data under test are the intent prediction confidences, output as the result of three different lane change intent classifiers. For each driver, the output confidences for every lane change example of that driver are averaged together. We compare this sample population of average prediction confidences across three different classifiers: Lane & Vehicle & Eye (LVE), Lane & Vehicle & Head (LVH), and Lane & Vehicle & Head & Eye (LVHE). The analysis of the confidence outputs allows a judgment on whether the results were significantly different between the eye-gaze-based classifier and head-pose-based classifier, as well as when both cues are included.

Using this analysis we can conclude in a statistically significant way ($p < 0.01$) that 3 seconds before a lane change, the head-pose-based classifier is more confident than the eye-gaze-based classifier. Table 5.4 shows the p values for various pairs of one-way ANOVA calculations. In the case of 3 seconds prior to the lane change, both LVHE and LVH are significantly more confident than LVE.

Two seconds prior to the lane change, the results are not as significant, though the trend ($p = .10$) is similar. This might indicate that the eye gaze cue provides better information about upcoming lane change intentions closer to the actual event. The head dynamics, though, seem to achieve a better result overall, and are certainly significantly better than eye dynamics at an earlier point.

One may call into question the assumption under ANOVA that the data is drawn from a Gaussian distribution. We thus additionally employ a non-parametric test that requires no assumptions about the distribution of the data. We present the results of the paired two-sided Wilcoxon signed rank test in Table 5.4. The resulting p-values are more or less in line with the results of the ANOVA test, so we can reasonably draw similar conclusions based on either of these tests.

So while it may not seem to be a large population, the conclusion may be drawn that the patterns are so consistent and dramatic that the trends for the 3-second decision time are statistically significant. We can reliably conclude based on the ANOVA analysis that for most drivers, head dynamics have significantly more predictive power than eye dynamics 3 seconds prior to an intended lane change.

Discussion

In some sense it is surprising that the addition of head dynamics by itself does as well as or better than eye gaze, since eye gaze would inherently seem to include the motion of the head in its estimate. After observing the data there seem to be two major causes for the lack of influence held by the eye gaze data.

Closer analysis of the patterns of head movement and eye gaze movement from Figure 5.1 can be seen in Figure 5.5. The data suggests that in this movement pattern, and many others in the data set, head motion actually starts before the eye gaze movement. In other words, the gaze shift began with an initial head motion and then the eye shift followed later. This behavior may seem counterintuitive, however such a pattern turns out to be consistent with a specific biological model of attention shifts.

According to an experimental study analyzing the relationship between eye and head movements by Zangemeister and Stark [223], such early head movement with respect to the overall gaze shift occurred mainly in gaze shifts of large amplitude, gaze shifts with predictable targets, and/or very rapid shifts. The study identified various other models of eye-head movement, including early eye movements in situations with small amplitude shifts or unknown target location.

The task-oriented gaze shift associated with lane change glances certainly falls into the former category, where the driver has a premeditated gaze target and thus prepares the gaze shift with a preliminary head movement. By capturing this movement, we can predict the intention of the driver earlier than if we wait for the eye gaze shift.

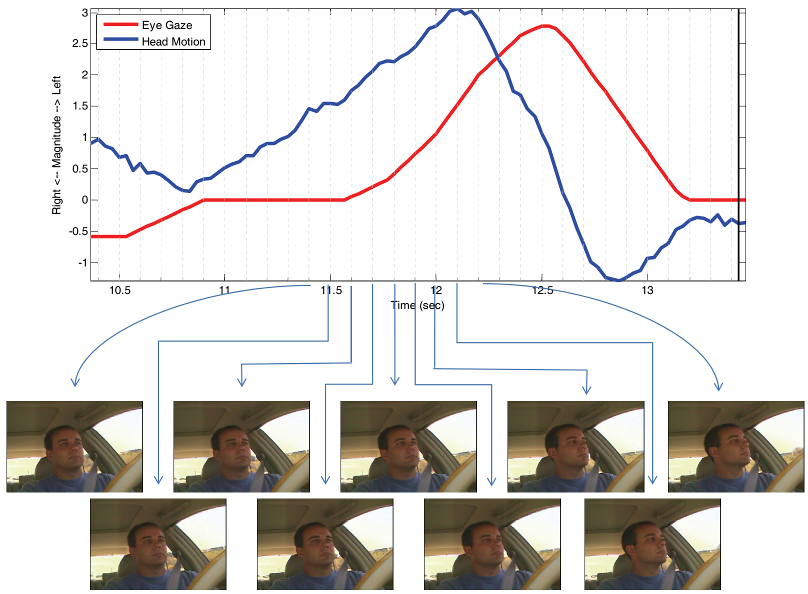


Figure 5.5: Head motion and Eye gaze positions prior to a lane change. The black line demarcates the lane change, red represents the magnitude of the eye gaze shift left or right, and blue represents side-to-side head motion. The driver's image from various points is shown at the corresponding time. Note the slight head motion visibly starts before the eye gaze shift occurs.

This leads us to the conclusion that, in situations where there are tasks requiring large gaze shifts to predetermined locations prior to the execution of the task, head motion can be used to predict the onset of the gaze shift earlier than can eye motion. The data presented above indeed supports this; head motion is shown to have significantly more predictive power than eye gaze 3 seconds prior to the task, but the difference is not as significant 2 seconds prior to the task.

A secondary factor in the performance of the eye gaze data is that the amount of eye movement also varies between drivers, whereas head motion may be more consistent. With a limited data set it is difficult to train a classifier to adapt to each driver's own style of eye-glancing prior to lane changes. Head pose movements, however, occur in a more telling manner across the population; this pattern extends from the general results of Bhise [16]. This property makes head dynamics a more reliable metric for inferring driver intent.

The NHTSA Lane Change Study [106] and the IVHS review [57] both led to the hypothesis that measuring head dynamics may be enough to detect distinctive behavioral cues prior to a lane change. Having confirmed that hypothesis, it can be argued that the relative ease of capturing head motion information as compared to eye gaze in vehicles, outweighs the advantages of adding eye data to head, lane, and vehicle data. Given the choice between the two cues, head pose can be considered as a better and earlier indicator to use for lane change intent inference.

Eye gaze may still be useful, however, for driver workload and distraction studies. For intent analysis, behavioral information derived from head motion is more important than eye gaze data, and robust systems may be designed using just measurements of driver head dynamics along with lane and vehicle data.

5.1.4 Concluding Remarks

We have presented a comparative study of the influence of eye gaze and head movement dynamics on driver behavior and intent prediction with respect to lane change maneuvers. Intent prediction was carried out using a discriminative classifier based on Sparse Bayesian Learning, where various combinations of features were used to train a classifier given labeled naturalistic driving data. We found that in general eye gaze was significantly not as informative as head motion ($p < 0.01$) in helping determine the correct prediction of whether a driver would change lanes, 3 seconds prior to the lane change. Head motion together with lane and vehicle data serves as a very good indicator of lane change intent, and we discuss a biological reason why head pose is actually an earlier indicator than eye gaze.

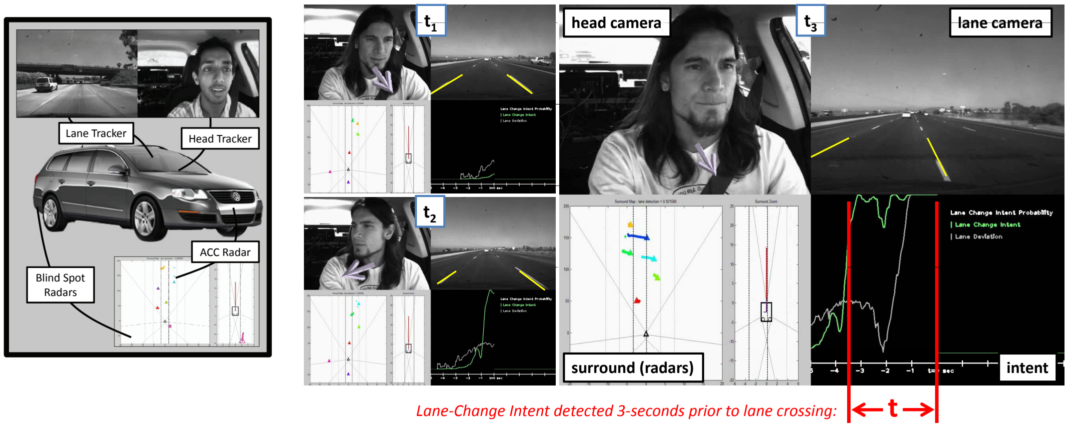
With the design of simple, robust Intelligent Driver Assistance Systems in mind, we have thus attempted to determine the important driver cues for distinguishing driver intent. The addition of eye gaze is relatively cumbersome and potentially unreliable in harsh conditions, and does not improve performance as early nor as robustly as do simpler head dynamics cues.

Future studies could further examine the effects of distractions and fatigue on these behavioral cues prior to lane change events, potentially in a driving simulator or controlled environment, along with the effects of external vehicles on the driver's motivations to change lanes.

5.2 Real-time Analysis: ADAS Development

The analysis of intent detection systems in recent literature has been on limited datasets which used offline implementations of the systems. These analyses are not sufficient to determine performance in real-world conditions. In this article we develop and implement a real-time implementation of a lane-change intent detection system, and propose several analytical methods to more realistically characterize performance in general, real-world situations.

The implications of a system capable of recognizing and predicting the driver's intent to change lanes are numerous. In a safety-critical role, preceding a lane-change maneuver, the vehicle could warn the driver of blind-spot obstacles only when the driver shows intent to change lanes. In contrast to current warning systems, which may often become a distraction or annoyance, by implementing the intent recognition system, warnings are provided only when the driver exhibits intent to perform an unsafe maneuver.



(a) (b)

Figure 5.6: (a) LISA-X is equipped with production grade sensors to maintain the look and feel of a stock Volkswagen Passat to ensure naturalistic vehicle operation of the both the driver in and around the testbed. (b) Real-time intent prediction on LISA-X. The prediction algorithm utilizes the most indicative sensory signals to infer the intention of the driver to change lanes.

An intent-based IDAS could also function in a comfort and convenience role. An attempt to overtake a slower lead vehicle with the automatic cruise control (ACC) system activated is much less responsive than during natural driving since the ACC does not allow throttling after the lane change. Given an accurate IDAS that predicts the driver's intent to change lanes, the system could throttle the engine earlier to mitigate the performance gap between ACC-assisted lane changes and naturalistic driving.

We are primarily interested in answering several important questions in relation to intent prediction systems: (1) How do offline classifier performance evaluations translate to real-world, real-time on-road performance, especially with respect to the false positive rates? (2) What is the performance of various sensors and sensor configurations in detecting intents? Are more advanced, costly sensors necessary to achieve better performance? (3) For a given sensor configuration, what is the earliest time prior to the lane change when intent can be detected?

An overview of the lane change prediction system output is presented in Fig. 5.6.

5.2.1 Context Capture Framework for Intelligent Vehicles

The Laboratory for Intelligent and Safe Automobiles (LISA) is dedicated to exploring multi-disciplinary approaches to making automobiles safer and more intelligent. We take a holistic approach to understand the driving experience by analyzing vehicle dynamics, the vehicle surroundings, as well as the driver and occupants. Unlike previous LISA intelligent vehicle test beds [208, 30], the LISA-X vehicle used in this research is not a reconfigurable research testbed. Instead, it utilizes production-grade sensors, in order to understand what can be accomplished

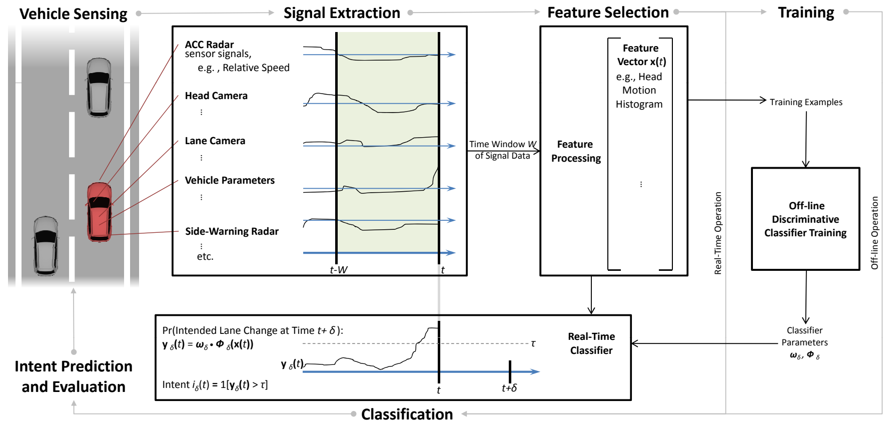


Figure 5.7: Overview of Real-time Intent Detection System.

with current technology with minimal cost.

The LISA-X vehicle is a 2008 Volkswagen Passat Variant 3.6L, modified to include the following sensors: (See Fig. 5.6a): (a) Adaptive Cruise Control (ACC) Radar; (b) Head (HEAD) Tracking Camera; (c) Lane Departure Warning (LDW) Camera; and (d) Side Warning Assist (SWA) Radars.

The ACC and SWA Radar systems are used to get surrounding obstacle information. The ACC Radar utilizes a narrow beam in order to detect and track vehicles in the front while the SWA Radar system is able to track vehicles in the rear. The LDW Camera system tracks the lane markings on the road to determine lane level positioning. The monocular head camera system monitors the driver's head position and orientation. These sensor subsystems are tightly integrated into the vehicle body to ensure minimal distraction during driving.

We outfitted LISA-X with sufficient computational resources to ensure live capture, recording, and processing of all the sensory subsystem data, as well as any other signals derived from the vehicle's Controller Area Network (CAN) Bus. During a typical drive the system captures over 200 sensory signals, synchronized at 30 Hz and timestamped, to provide a rich description of the complete driving experience.

Every 33 msec (30 times a second), a time window of $W = 2$ second's worth of measurements are collected from each of the individual sensor subsystems and processed, to form a contextual descriptor of the driving environment. This descriptor should include the most discriminative set of features available so that classification algorithms can distinguish between different behaviors, e.g. intended lane-changing vs. lane keeping. In this case the descriptor is a feature vector drawing from all the different sensor subsystems:

$$x(t) = \begin{bmatrix} \bar{\alpha} \\ \bar{\beta} \\ \bar{\nu} \\ \bar{\lambda} \\ \bar{\varsigma} \end{bmatrix} = \begin{bmatrix} \text{Vehicle} \\ \text{ACC} \\ \text{HEAD} \\ \text{LDW} \\ \text{SWA} \end{bmatrix}. \tag{5.3}$$

There are a significant number of raw signals from each sensor at any given time. For example, the SWA system outputs detailed information of the location and speed of all tracked objects. In their raw format, a vector of these numbers would be extremely difficult for any pattern recognition algorithm to decipher without prior information about what each number means. The raw data coming out of the sensors is, in many cases like this, too noisy or sparse. Without enough training data for specific situations, it becomes necessary to aid the system by adding "expert" knowledge.

For example, in lane change and overtake situations, it is critical to know when there are cars in or approaching the blind spot. Therefore, we can preprocess the raw SWA data to determine how many critical objects there are, and whether they may be in critical zones. By doing this feature processing step, we give the machine learning system a jump start to use the existing training data to decipher the most useful thresholds for situation criticality.

Below we describe the components of the context descriptor feature vector. In some cases we translate a number of the raw signals into time-series features, where for example

$$x_{j...j+W}(t) = [HeadYaw(t-W)...HeadYaw(t)]$$

. For other features, as described above, we need to use some higher-level processing of the time-series data, for example,

$$x_{j+W+1}(t) = \sum_{n=t-W}^{t} 1[HeadYaw(n) > 30]$$

.

Vehicle Signals $\bar{\alpha}$: The vehicle CAN Bus provides measurements of the vehicle's dynamic state. This supplies several time series features which include the steering wheel angle, yaw rate, and blinker state signals, as well as indicators of blinker direction and length of time it is active.

ACC Signals $\bar{\beta}$: The ACC system is able to track vehicles using RADAR in the front of the vehicle. The ACC RADAR is designed with a narrow FOV cone such that only the vehicle directly in front (in the same lane usually) can be reliably tracked. The 4 ACC features relate to this track and account for the distance to the ACC vehicle, the relative speed of the ACC vehicle, the measured time gap with the ACC vehicle in seconds, and the difference between the current vehicle speed and the desired speed (ACC set speed).

HEAD Signals $\bar{\nu}$: Unlike the other sensor subsystems, the HEAD system does not monitor the driving environment, instead it focuses on the driver. The intentions of a driver can be better inferred when directly measuring driver actions. The HEAD features are a measure of recent driver head motion, both head rotation (yaw) and pitch. The features include the time series of head yaw position, yaw motion (derivative of yaw position), a histogram of head yaw values, a histogram of yaw motion values, a histogram of head pitch position, and an indicator signal of significant yaw rotation. Since preparatory glances are a major indicator of a lane change maneuver [106] there are a large number of features generated for the HEAD system.

LDW Signals $\bar{\lambda}$: The LDW system provides measurements of the vehicle position with respect to the road and the road geometry. The LDW features correspond to the recent time series W of vehicle lateral deviation (position within the lane), lane curvature, and vehicle yaw angle with respect to the lane.

SWA Signals $\bar{\varsigma}$: The SWA system is able to track surround vehicles in the rear and sides using a two RADAR system and delivers the position and relative velocity of each obstacle. Unlike the ACC RADAR, the SWA system is able to track a large number of vehicles simultaneously because of a much larger FOV. The large rear area is quantized into smaller critical zones. The critical zones correspond to the blind spots in the rear of the vehicle as shown in Fig. 5.8. The SWA blind spot features indicate the occupancy and speed state within the critical zones.

$$\varsigma_1 = \max_{N_s} \sum_{t=1}^{W} s_v(t) \quad \forall -5 < s_x(t) < -1.65 \tag{5.4}$$

$$\varsigma_2 = \max_{N_s} \sum_{t=1}^{W} s_v(t) \quad \forall -1.65 < s_x(t) < 1.65 \tag{5.5}$$

$$\varsigma_3 = \max_{N_s} \sum_{t=1}^{W} s_v(t) \quad \forall\, 1.65 < s_x(t) < 5 \tag{5.6}$$

with N_s the number of vehicle surround tracks within a particular zone during the time window W and $s_x(t), s_y(t), s_u(t), s_v(t)$ the relative (x, y) coordinate and velocity (u, v) of a surround vehicle at a given time t. The features ς_i and are intended to represent motivating or deterring factors for a lane change because the presence of a vehicle in the adjacent lane would impede a lane change.

Once these signals are processed into features, they are concatenated into a large vector, as seen in Figure 5.7. Each of these vectors is used as a training example (positive or negative) for the discriminative classifier. In the case of live evaluation, we pass these vectors into the discriminative classifier (as trained below) to obtain the prediction of intent.

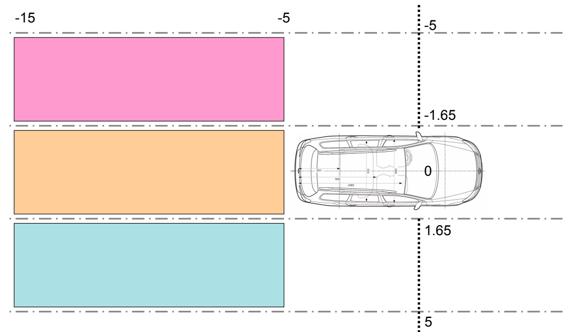


Figure 5.8: Three SWA regions of interest: driver-side lane in pink, ego lane in orange, and passenger-side lane in blue. The SWA features consist of the average longitudinal speed in each of these regions and indicate motivators and deterrents to a lane change.

Table 5.5: Lane Change Prediction Datasets

Label		Data Set	N_{people}	N_{runs}	Hours	Description
D		Long runs for Training	15	24	14.5	782189 frames
	D_{train}	Training Set				266+/2606- exs.
	D_{test}	Cross Validation Test Set				101+/879- exs.
T		Long runs for Testing	5	18	7.9	427497 frames

5.2.2 On-Road Intent Prediction Framework

Training/Testing Databases

In order to train the lane change intent classifier, we proceeded to collect a database of naturalistic driving. The training data collection lasted several months with 15 participants from UCSD as well as VW-ERL in Palo Alto.

For each driver, a data collection run consisted of two phases, designed to capture a variety of behaviors. In the first phase, a driver was instructed to operate with the ACC function active. The desired speed was set greater than the posted highway speed limit and drivers were instructed to stay in a lane and pass when necessary (Lane changes could occur on either the driver or passenger side).

In the second collection phase the driver was free to drive normally without the ACC function engaged in order to obtain completely natural driving behaviors. Drivers (especially those not familiar with ACC operation) had a tendency to behave differently when the ACC was active than in their normal, day-to-day, driving behavior.

The two phases lasted approximately 1 hour in total with 30 minutes dedicated to each of the two collection phases, allowing us to capture sufficient data to find patterns through a wide variety of conditions.

All the lane changes in a collection run were automatically detected using the lane tracking information. Lane changes can be detected by observing the point when lateral lane

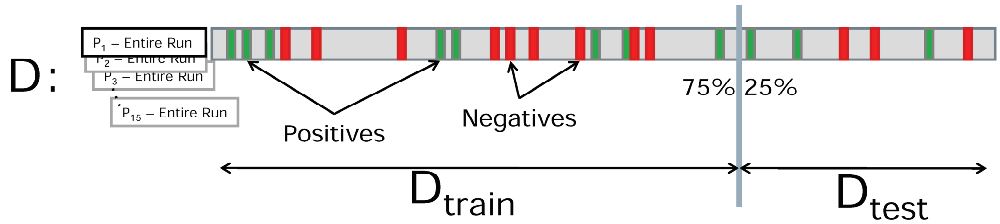


Figure 5.9: Training Database Split: The training database which consisted of 15 subjects was split into a training and cross validation test set in order to train the lane change classifier. 75% of the lane changes were randomly chosen from each subject and used for training and the remaining 25% was for cross validation during training.

position switches sign (from the right to left side of the lane or vice versa). The automatic detection scheme worked well when lane tracking was reliable, but a number of lane changes had to be manually detected.

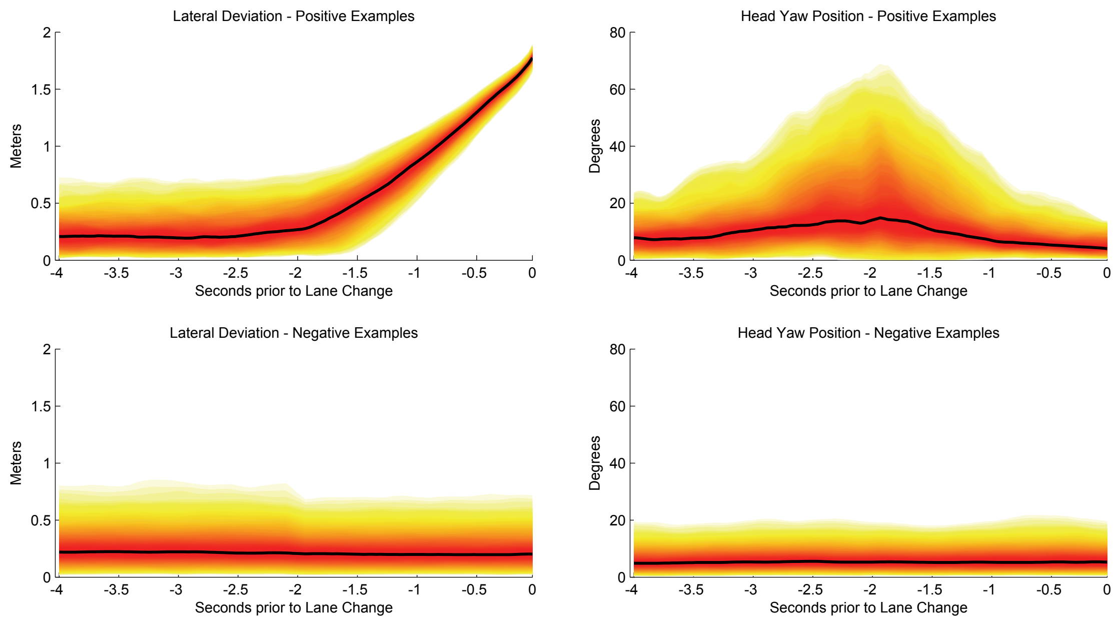
To ensure that the training data was valid, several criteria were used to discard any data in which the sensors might have been performing poorly. (1) **Vehicle** - The ego-vehicle must be at highway speeds. (2) **ACC** - The cruise control must be activated (but this is not strict). (3) **HEAD** - The visual head tracking must be consistent by having limited head rotation and face located within the center of the camera field of view. Finally, there cannot be too many peaks in the head rotation time series. (4) **LDW** - The lane tracking should be consistent (with high confidence) over the data window.

The final validated dataset D included around 370 instances of naturalistic lane changes. This was drawn from nearly 15 hours of driving data, from 15 drivers (12 m, 3 f), with varying nationality, age, experience, and familiarity with the ACC functionality. We broke the training dataset into two separate subsets for cross-validation training. The set D_{train} was for training the classifier while D_{test} was for assessing the performance during training. The data split methodology is depicted in Fig. 5.9.

We also collected another independent testing data set T of long drives, consisting of 5 different drivers and 7.9 hours of completely uncontrolled driving, with 229 lane changes. With this dataset T, we evaluated over 400,000 instances of data, with results seen below. Table 5.5 summarizes the lane change prediction datasets.

Sparse Classifier for On-Road Intent Prediction

We based our lane change intent inference algorithm on one developed by McCall et al.[121], labeled the Driver Intent Inference System (DIIS). There are a number of other works in this field [99, 153, 183, 81, 75]; our experience with developing discriminative classifiers for intent prediction [121, 30] has shown considerable success warranting this investigation into a real-time prototype.



(a) (b)

Figure 5.10: (a) Evolution of lateral lane position before a lane change when lane deviation is maximal. (b) A driver's head yaw (rotation) when scanning tends to occur between 2.5 and 1.5 seconds before a lane change, much later than was previously reported [106].

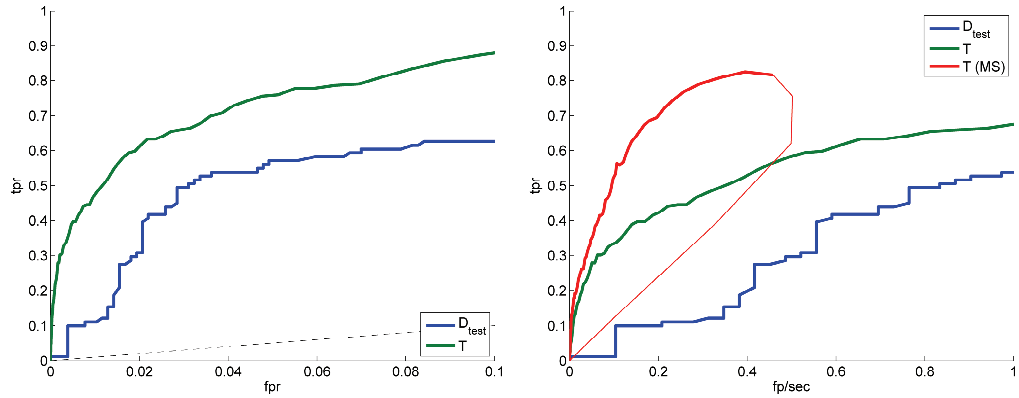
The DIIS is a discriminative classifier to distinguish between two events: lane changing (either right or left) and lane keeping. The general flow of the system can be seen in Figure 5.7.

To train the DIIS, we used the discriminative Relevance Vector Machine (RVM) classifier, which is based on Sparse Bayesian Learning (SBL), developed by Tipping [201] and implemented in [121]. This is the same as the previous section, where more details on the classifier can be found.

5.2.3 On-road Performance Characterization

From Laboratory to Roads

The performance of pattern recognition based classifiers in laboratories and "real" use can be significantly different. In the laboratory setting, pre-specified examples are fed to the classifier to determine performance but in real-world settings, the classifier must continually evaluate data as in arrives. This continuous operation has direct impact on performance because the amount of data to evaluate is much greater, consecutive evaluations will perform similarly, and decisions must be made in real-time. Practically this results in much lower false positive rates because of the significantly larger number of evaluations and necessitates the development of techniques to perform consistent labeling.



(a) (b)

Figure 5.11: Plots comparing the $\delta = 2.5$ intent ROC curves with all sensor subsystems for the different test sets. (a) The traditional TPR vs FPR ROC curve used to asses classifier performance with results on T higher due to the much larger number of negative examples in the set. (b) ROC curve utilizing real units demonstrates the laboratory evaluation does not translate well on-road. The on-road performance can be greatly improved by lowering the number of false positives using a multi-suppression technique (MS) as shown in red.

During the intent classifier training, performance is evaluated with a small independent testset D_{test}. The ROC curve generated using this set of labeled positive and negative examples provides the traditional performance evaluation in the laboratory setting, and is similar to those seen in prior work [121].

As one of the design requirements, we aimed to create an algorithm that is able to predict behaviors across a wide population of drivers. This includes drivers of different styles, some of who may engage in slightly different behaviors prior to lane changes than others. Given a wide variety of styles, we find it necessary to analyze performance of the classifier on individual drivers to understand whether performance degrades between drivers. We compared the class membership probabilities between the positively and negatively labeled data (i.e., lane changes vs. lane keeping situations). Using an analysis of variance, treating the subjects as random factors, we found a consistent pattern across all drivers of separability between classes ($F(1, 14) = 674, p < 0.001$). Across the range of drivers, the classifier has quite consistent performance, and could be expected to perform just as well on a larger population.

We then moved onto a dataset T of fewer drivers, but much longer and continuous drives, to analyze the on-road performance of the predictor response during continuous operation. The two ROC curves, in blue and green, are shown in Fig. 5.11a for detection time of $\delta = 2.5$ using all the sensor subsystems. The ROC from the on-road set T is significantly better than for D_{test} because of the greater rate of negative examples causing a very low false positive rate.

While the traditional ROC curve analysis makes the classifier seem quite strong, viewing the results on a different scale more meaningful for real-world usage a different story unfolds.

Fig. 5.11b presents the same ROC but using false positives per second [FP/sec] rather than the traditional false positive rate. Rather than utilizing academic measures of performance, this allows us to interpret concrete meaning in terms of "real" units. For example, a classifier with a 70% TPR and 5% FPR may seem reasonable in an academic sense, but during real on-road operation this might translate to over 1 false positive every second. Therefore, seemingly "good" classifier performance in the laboratory does not necessarily translate well on-road. In order to have good on-road performance we must significantly lower the classifier false positive rate.

Lowering Classifier FPR

The intent classifier performance analysis using real units [FP/sec] shows that false detections are the main source of error. These false positives are mainly the result of two phenomena. First, on-road evaluation presents data sequentially and continually, which results in non-independent outputs. Second, a classifier is designed to detect a lane change δ seconds before its occurrence, but the classifier is not always so precise in its prediction.

Time Series Dependency and Multi-suppression

In classical pattern classification setups, data is selected, either positives or negatives examples, and fed to a classifier to test its response. The implicit assumption is that individual examples are independent. In the on-road case, a sliding window of data is used on the time series of feature data which results in non-independent samples. Two consecutive windows will contain almost identical data and therefore will have similar responses. The plot in Fig. 5.12 shows the evolution of intent probability of the $\delta = 2.5$ classifier in black for a number of seconds before a lane change.

Ideally, a single large value would spike at -2.5 but in this case there are actually two major responses at -3.5 and -1.25 seconds. As the lane change approaches, the probability of lane change increases and a number of time instances are above the detection threshold. Each of the red deltas indicate a detection but there is only one lane change which means most of them are considered false positives. With the assumption that consecutive detections arise from the same intent, a single detection can be extracted on the rising edge of the intent signal. This multiple detection suppression technique logically handles the time-series effect but may change the detection time away from δ. After this "multi-suppression" only 3 detections remain in this segment. The detection that matches the lane change is the closest to δ as shown in green while the only the remaining 2 detections are false detections which is significantly less than the red.

The on-road performance after multi-suppression (MS) is shown in red in Fig. 5.11b. The MS scheme results in significant performance gain over traditional evaluation but introduces an interesting looping behavior in the ROC curve. At low threshold values many samples are considered continuous detections and then suppressed. Therefore, the false detection rate will be very low but in addition true detections are low as well because the multi-suppression response will occur at times outside the match window. Since this effect occurs only at low thresholds it

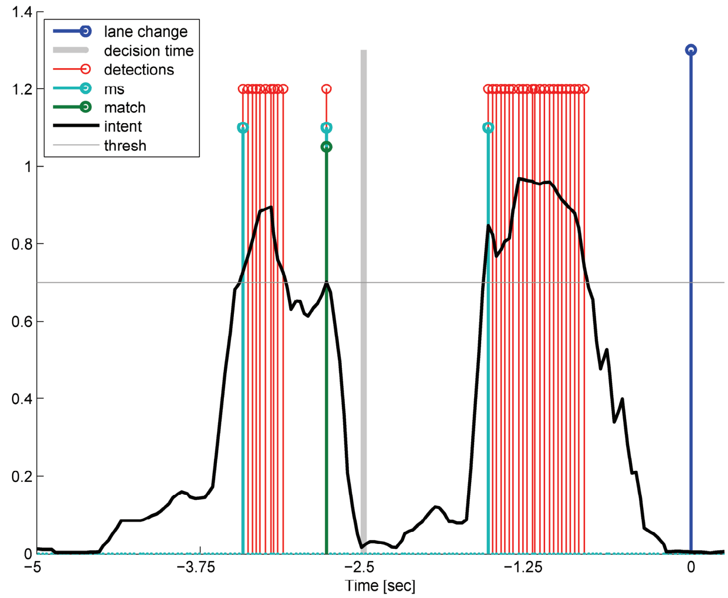


Figure 5.12: Multi-Suppression: The blue indicates the time of a lane change while the red denote all the times where the prediction is above the threshold. Since only one lane change occurs, a single match, in green, is considered as a true positive. The remaining red are false positives. The number of false positives can be managed by suppressing multiple hits and only counting a single response from each of the two groups of red (as shown in light cyan).

is not much of a concern in practice as the selected detection threshold would be high to have high confidence in the prediction and to limit false positives.

Imprecise Prediction Timing and Match Windows

Even though we trained an intent classifier to infer the lane change δ seconds before its occurrence, it is merely attempting to match the current data to a pattern seen during training. The cues that signify an oncoming lane change may not always happen at the exact same time, due to the variability in human behaviors. The timing response of an intent classifier is plotted in Fig. 5.13 for different detection times. These plots were generated by finding all time instances where the prediction probability value was above a low detection threshold and determining the elapsed time to the next lane change. Although a detector is designed for a specific time δ, the detections are not localized and can occur before or after this time. Therefore, it is not sufficient to only consider the sample δ seconds before the lane change as a positive example, there must be flexibility such that examples within a small time window around δ could be considered a positive.

In order to determine if the classifier was able to correctly infer the lane change δ seconds early, a small match window is used to soften the δ constraint to consider a close detection as a

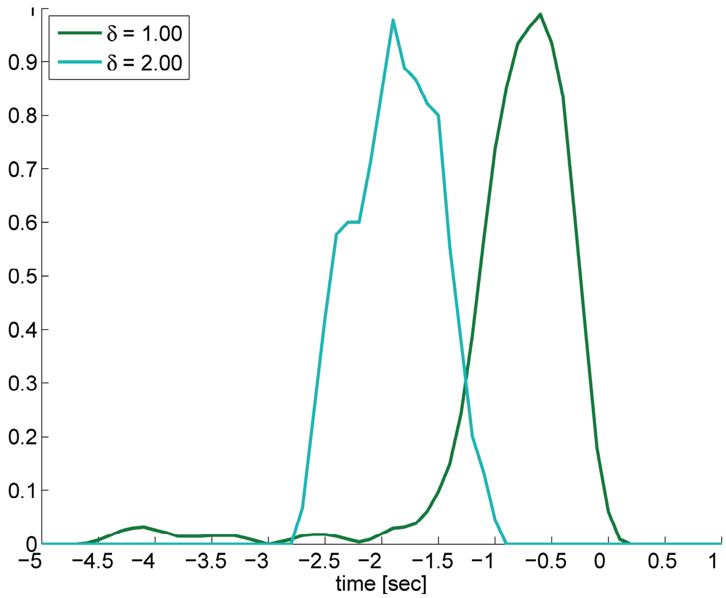


Figure 5.13: The distribution of intent detections is is fairly wide because consecutive intent classifications will result in a succession of positive results. A positive detection does not always occur exactly at the detection time but can occur some time before and after.

match (green in Fig. 5.12). The match window size is parameterized by two values, α and f,

$$t^{+}_{mw} = f\delta \tag{5.7}$$

$$t^{-}_{mw} = \alpha f\delta \tag{5.8}$$

as shown in Fig. 5.14. The match window is not symmetric to bias matching for earlier detection. The early bias is a result of multi-suppression which has the tendency of marking one detection before the maximal response. The f term characterizes the timing of detection spread while α enables earlier detections. An $\alpha = 2.0$ is used for all further analysis since performance improvements was limited for larger values.

In Fig. 5.15a we show the detection rates for intent classifiers with different detection times all increasing with larger match window sizes. Of course, larger windows will have a better chance of collecting a high response but the cost is an increased number of false positives within the window as shown in Fig. 5.15b. The FPR dips down after increasing initially; this is only because the number of negative examples in a window is growing faster than the number of new false positives. This peak gives a good indication of the natural window size for a particular detection time δ.

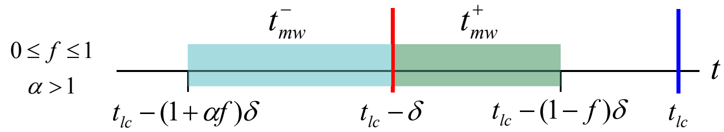
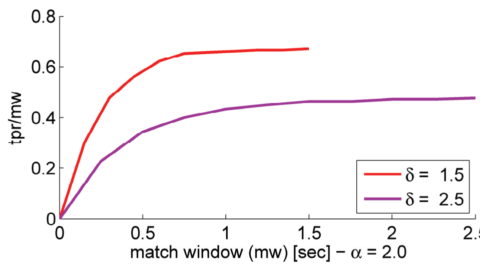
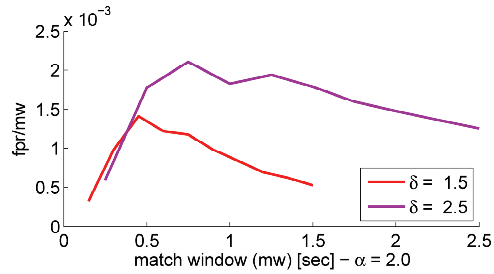


Figure 5.14: The size of the match window is controlled by the filter detection time δ and the two parameters f, α. f characterizes detection spread and α enables earlier detections.



(a)



(b)

Figure 5.15: (a) The true positive rate obviously increases for different match window sizes. The strong classifiers are able to increase rapidly with small window increase (between 0 and 1 second). (b) The cost of better detection is a larger number of false positives within the match window.

5.2.4 Selection and Performance of Different Sensor Configurations

Sub-System Timing Exploration

Each new sensor that is added to an automobile provides both new functionality and a more complete understanding of the driving situation. But, these sensors only provide measurements of the physical world, and it is not immediately clear when they are most useful for lane change intent recognition. It is important to know what time period before a lane change are each of the cues from a particular subsystem relevant. By knowing these times it is possible to design a classifier that utilizes the sensors appropriately.

Fig. 5.16 indicates the performance of each individual sensor subsystem over a range of detection times. The color indicates how important a sensor system is in comparison with the other sensors the area under the ROC curve (AUC) criterion. The LDW lane information is most informative between 0-1.5 seconds prior to a lane change, as demonstrated in Fig. 5.10a. The HEAD information is most relevant between 2-3 seconds, corroborated by the histogram of raw head movements in Figure 5.10b. Interestingly, the SWA features do not seem to be much information through 5 seconds and ACC apparently is the most informative sensor at 5 seconds. By training intent classifiers which utilize only measurements from a particular sensor system, we are thus able to determine its relevant time horizon without deep analysis or understanding of the signals themselves.

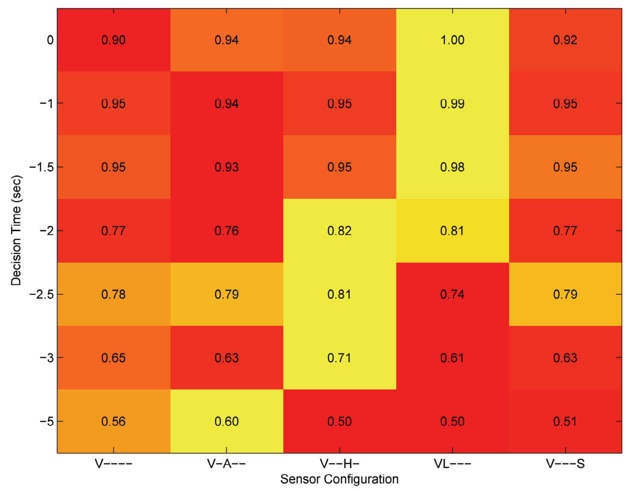


Figure 5.16: Different sensor subsystems contribute relevant information at different times before a lane change. The color indicates at a given decision time how important a sensor system is as compared to the others using the AUC. The lane information (LDW) is most informative close to the lane change and as the prediction happens further away from the lane change the HEAD information counts the most. The quality of a intent prediction is assessed using the area under the ROC curve (AUC) metric where a larger value indicates a better performing classifier. As expected the performance is generally better when making a prediction closer to the actual lane change. The lane information is very relevant at the 0, 1, and 1.5 sec decision times while the head information is most informative at 1.5.

Sensor Configuration Exploration

We might find it reasonable to assume that having more sensors will automatically provide better information. While more data may provide a more complete picture of the driving situation it also becomes more difficult to manage and effectively utilize all the information (a.k.a. the "curse of dimensionality"). In addition, these automotive sensors are not standard equipment on vehicles but instead are added options, which add additional costs to the vehicle. Therefore, we would like to know what is the relative performance given different sensor configurations and whether all the advanced sensor systems are really necessary for accurate intent prediction.

A number of different sensor configurations were examined corresponding to different equipment options {(`V`)ehicle, (`L`)DW, (`A`)CC, (`H`)EAD, (`S`)WA}. For lane change detection, at a minimum we require the Vehicle information and the LDW lane information. The following sensor combinations were evaluated:

- `VLAHS` (Vehicle, LDW, ACC, HEAD, SWA)
- `VLAH-` (Vehicle, LDW, ACC, HEAD)
- `VL-HS` (Vehicle, LDW, HEAD, SWA)
- `VL-H-` (Vehicle, LDW, HEAD)

- `VLA-S` (Vehicle, LDW, ACC, SWA)
- `VLA--` (Vehicle, LDW, ACC)
- `VL--S` (Vehicle, LDW, SWA)

In Fig. 5.17 we compare the real-unit ROC curves for detection times $\delta = 1.0, 1.5$, and -2.5 seconds. For the $\delta = 1.5$ classifier in 5.17a the detection rate for all the classifiers is able to reach 95% after 200 FP/hour. The `VLA--` and `VL--S` perform the worst of the bunch. At this time the use of all the sensors in `VLAHS` actually degrades performance slightly. In 5.17b there is a clear separation between the worst classifier `VL--S`, the non-HEAD (`VLA-S` and `VLA--`), and the HEAD based classifiers. At $\delta = 1.5$ there is significant improvement in performance through utilization of the head viewing camera. Similarly, at $\delta = 2.5$ the HEAD based predictors are the best performing, but this time the SWA only (`VL--S`) classifier out performs the ACC based systems. These results are consistent with prior research that found direct observation of the driver improved lane change intent prediction by detection of preparatory scanning.

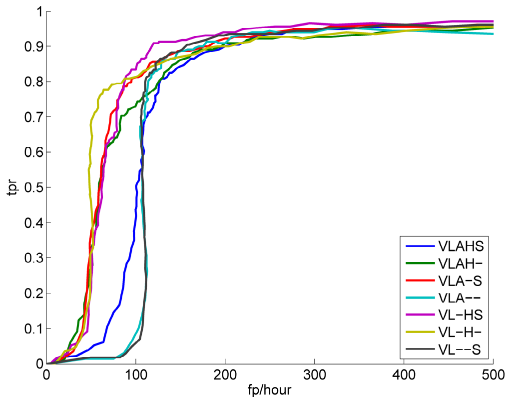
As expected, the quality of the intent classification improves the closer the detection time is to the lane change. Fig. 5.18 shows strong performance up to $\delta = 1.5$ and that intent prediction seems to be unreliable beyond $\delta = 3.0$ seconds. This indicates that intention manifests itself in behavioral movements only in the 3 seconds leading up to the lane change.

It is interesting to note for the $\delta = 0.0$ that the ROC curve bends toward the top left corner during the increasing TPR. This is an artifact of the multi-suppression algorithm, not normally present in ROC plots. At high thresholds the multi-suppression algorithm is not able to conglomerate detections because of noisy peaks. Once the threshold falls below this noise level, more consecutive frames are merged through the multi-suppression procedure, decreasing the number of false positives.

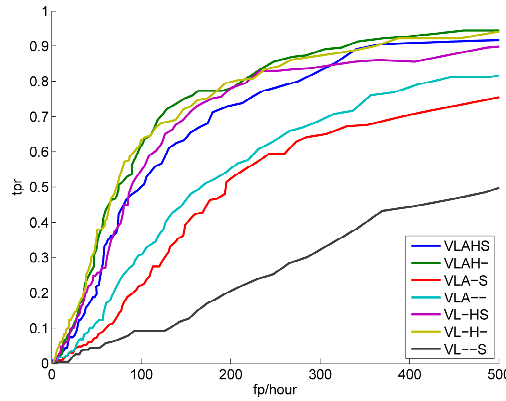
Finally, a complete comparison between all the sensor configurations and detection timings is presented in Fig. 5.19. This matrix indicates the detection rate for a fixed false positive rate of 120 FP/hour. Generally speaking, the performance for all the sensor configurations decreases further away from the lane change except for `VL--S` (maximum at $\delta = 2.5$). The color scale indicates the relative strength of a sensor configuration at a given detection time. The coloring shows that the sensor configurations utilizing the HEAD sensor have significant gains between $\delta = 1.5$ and 2.5 seconds, making it an attractive advanced sensor.

5.2.5 Concluding Remarks and Future Directions

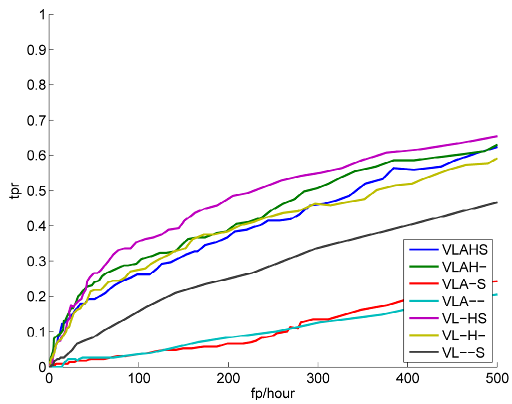
The intent prediction system we presented is able to consistently detect lane changes seconds before its occurrence as demonstrated in Fig. 5.20. While it is certainly important to be able to accurately predict when lane changes will occur, the number of lane changes is relatively rare compared to the lane-keeping instances, implying it is more important to eliminate the



(a)



(b)



(c)

Figure 5.17: Real unit ROC curves for competing sensor configurations at different detection timings. (a) $\delta = 1.00$ (b) $\delta = 1.50$ (c) $\delta = 2.50$.

number of false detections. If intent information is provided as feedback to the driver, either through modification of the vehicle behavior or through an alert system such as a beep or light, the false positive rate is closely tied to driver acceptance of the system. A driver will not want or use a system that is constantly incorrect.

By introducing Match Windows and Multi-suppression, we have shown that it is possible to reduce the number of false positives in on-road performance. However there are still a variety of other reasons for a false positive, among them events that look similar to lane changes (merges, exits); lane changes close to times of poor sensor readings; and an intent without a corresponding maneuver, i.e. an aborted lane change.

Several examples of events that appear like lane changes can be seen in Figure 5.21. These (a) merge and (b) lane split during exit occurrences could be considered lane changes by relaxing the lane change definition. Fig. 5.20d shows an example of an arguably “true” false positive, which may have been an aborted lane change. In this example, the driver spends a

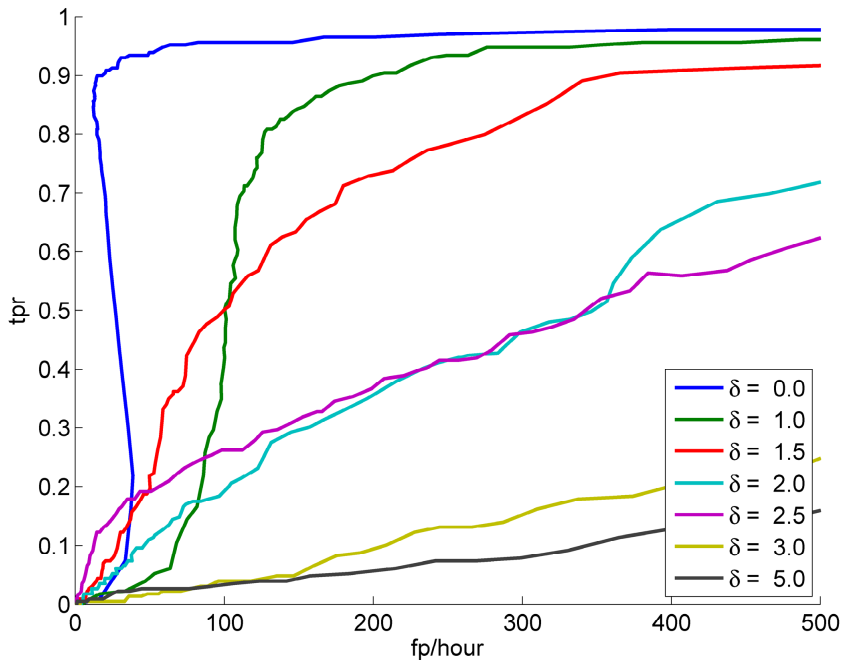


Figure 5.18: Real units ROC curve for the `VLAHS` intent classifier at different detection timings. The closer to the lane change the better the performance and prediction horizons beyond 3.0 seconds is completely unreliable.

significant time looking over his shoulder into the adjacent lane but does not perform a lane change maneuver. By taking into account higher level information about the surround (i.e., if the lane was closed due to construction, or if there was a car moving into the blind spot), we could move toward eliminating these false positives as well.

When implementing a real-time driver intention prediction engine, we have shown that the traditional laboratory methods for training and evaluating performance do not correlate well with the actual on-road results. The difference between independent samples in a lab environment and the time-correlated data received while driving necessitate new evaluation schemes. The use of multi-suppression and a match window improved performance but further improvements are necessary for commercial deployment. New training methods might be applied to account for the time series (correlated) data. This could mean better choice of negative examples which are very close to the detection time or to explicitly model the prediction output as a time-varying process. In addition, the driving context can be further exploited for real system design. For example, the lane change classifier might not use the ACC system when it is not engaged or not following a lead vehicle.

Despite the difficulties on-road, this study showed that lane change prediction can be reliable up to 3 seconds before the actual lane change. Any prediction further out had very low correlation with an actual maneuver. In addition, the value of a camera specifically designed to

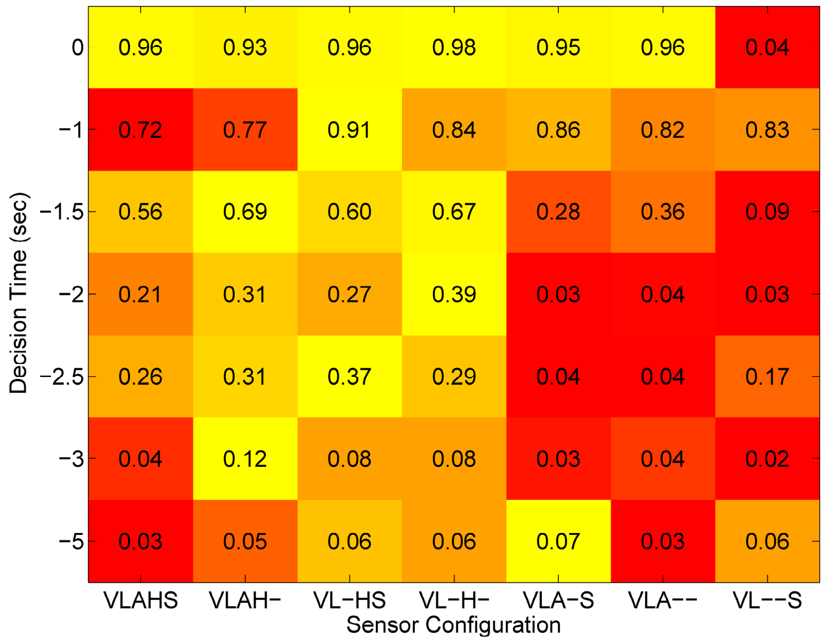


Figure 5.19: Intent detection rate for a fixed false positive per hour rate of 120 FP/hour. Similar to the individual sensor study, lane change intent is improved between 1.5 and 2.5 seconds by utilizing the head sensor to observe the driver actions.

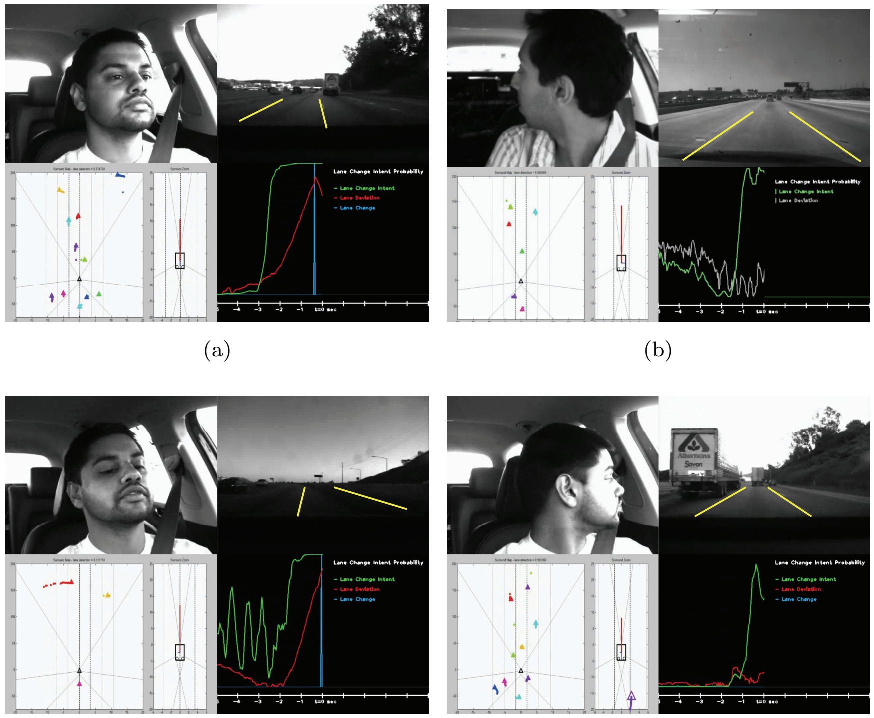
observe the driver has tremendous benefit for detecting lane change intent earlier. This camera could also be used to monitor the driver state and condition to test if distracted or sleepy to further improve road safety.

Automotive systems operate under especially safety and time-critical constraints, where every millisecond helps to save lives. Intelligent driver assistance systems, along with more general pervasive computing systems, gain precious moments and thus stand to benefit greatly from proactively understanding and predicting human behaviors.

5.2.6 Acknowledgements

Chapter 5 is in part a reprint of material that is published in the IEEE Transactions on Intelligent Transportation Systems (2009), by Anup Doshi and Mohan M. Trivedi, and submitted to IEEE Pervasive Computing (2010), by Anup Doshi, Brendan T Morris, and Mohan M. Trivedi. The dissertation author was the primary investigator and author of these papers.

This research was supported by research grants from the UC Discovery Program, Volkswagen, and the National Science Foundation. The author is especially grateful for the insight and support of Dr. Brendan T Morris, along with the support of Mr. Jaime Camhi, and advice and support of colleagues from the CVRR laboratory.



(a) (b) (c) (d)

Figure 5.20: Examples of on-road intent detection. (a) and (b) show examples of lane change intents successfully detected 2-3 seconds prior to the lane change. (c) is also a true detection, although the intent probability is quite noisy leading up to the detection, possibly due to early lateral movement within the lane. (d) shows an example of an aborted intent, in which vehicles in the blind spot most likely caused the driver to abstain from the lane change.

(a) (b)

Figure 5.21: These maneuvers appear very similar to a lane change. (a) Lane ends during a merge. (b) Lane split during a highway exit.

Chapter 6

Driver Interactivity

Style and interactivity play a big role in intelligent human-machine systems, especially driver assistance systems, leading us to an analysis of different manners of feedback in the following chapter. This leads us to an analysis of driver style, and how it influences driver responsiveness and predictability. We end with a view towards collaborative vehicles of the future, in which the potential for anticipatory communication between vehicles and infrastructure could lead to greatly improved safety.

6.1 Feedback and DAD

Safety is a major cause for innovation in the automobile industry. Many new technologies are emerging to meet the increasing desire for safer vehicles and roadways.

One factor that plays a big role in road safety is excessive speed. Speeding is a significant cause of accidents and motor vehicle infractions. According to the National Highway Transportation Safety Administration (NHTSA)[148], 13,113 lives were lost in 2005 in crashes involving speeding. Indeed 30% of all crashes involved speeding, involving a cost of approximately $40.4 billion each year in the U.S. alone. Furthermore, most of those crashes are at speeds under 55mph, off of the highways.

Another important cause of fatalities is "backover" accidents. These occur when a driver is in reverse and thus has a limited field of view of their path. According to NHTSA[147], "backovers"' annually cause 183 deaths and between 6700 and 7419 injuries, with the majority of these young children in driveways.

For these reasons and more, auto manufacturers have begun implementing technology based safety systems. Some of these state-of-the-art systems include collision warning and brake support (Volvo), intelligent night vision with pedestrian detection (Honda), backup warnings and cameras (Nissan, BMW, Toyota, etc.), and lane departure warning systems (Infiniti). Despite this progress it is not clear whether these systems can urgently and reliably warn the driver with out distracting them from the road conditions. Fixed displays will not grab driver attention if they are looking away. Auditory (and haptic) displays may not be able to convey the same amount of information quickly and succinctly. Drivers may already know about impending danger, and they may be annoyed by systems that provide redundant warnings. Side-screen displays might take the driver's attention off the road.

We introduce and evaluate the Dynamic Active Display (DAD), which is a unique *large-area windshield display* designed to actively alert driver in critical situations. As a novel prototype developed in conjunction with Volkswagen, the DAD is able to use a laser to display dynamic visual icons nearly anywhere on the windshield. Within the author's knowledge, the implementation of such a wide-area windshield display in a real test-bed is unprecedented. It allows for the real-world experimentation of concepts previously only testable in laboratory settings, thus moving beyond the realm of simulated studies.

The DAD can be used to provide context-specific alerts to the driver, and it is capable of changing the position and intensity of an alert depending on the attention of the driver. This "active" alert thus makes decisions based upon the state and pose of the driver, vehicle, and environment.

The display is presented as an effective medium to communicate important information to the driver, while minimizing distraction. As part of a quantitative evaluation of this system, we find that DAD demonstrates significant improvements in both effectiveness and minimizing distractions as part of a driver assistance system.

For this research, background and related works are discussed in Section 6.1.1. Section 6.1.2 describes the instrumented testbed, LISA-P, along with some potential capabilities of DAD. A comprehensive evaluation of DAD in a speed control experiment is described in Section 6.1.3. Conclusions and future research are discussed in Section 6.1.4.

6.1.1 Background and Related Works

Background

Modalities of displays. The driver can be informed of critical situations via several different modalities.

Haptic interfaces include any sort of interface that will use force feedback or touch-sensitivity. Examples could include resistance or shaking of the steering wheel, or resistance in the brake and accelerator pedals. These have the advantage of being intuitive; they also can inform the driver quickly even if the driver is distracted, since the driver is usually in contact with the interface.

Audio interfaces may include voiced commands, such as those given by Navigation systems, or even simple beeps and sounds. Most vehicles are equipped with warning beeps to indicate if the driver has left their headlights on, or if the driver is not wearing a seatbelt. The beeps are useful if the driver knows what the sound means, or in conjunction with a visual cue. To convey more information in the auditory channel, as is done with voiced navigation directions, for example, requires more time.

Visual interfaces abound in the vehicular environment. Examples include the dashboard of the vehicle, showing the speedometer and tachometer, as well as side-screen monitors for navigation systems. Visual cues have the advantage of being able to convey a wealth of information to the driver quickly. However there is a great deal of visual distraction in a vehicular environment, so to inform a driver using this modality could be a challenge.

Several cars now also use Heads-up Displays (HUDs) to convey information to the driver. As a subset of Visual interfaces, HUDs are designed to present information to the driver closer to the field of view of the driver. Thus, the driver does not have to look down to see information, but can spend more time looking at the road. It is important to note, however, that drivers will not be able to focus on the HUD and the road at the same time, due to the effects of parallax. Nevertheless by presenting information closer the normal field of view, HUDs require less effort on the part of the driver than other kinds of visual displays [113], so may present an effective choice of conveying visual information.

Types of displays. Moreover, independent of the modality, there can be several types of displays, depending on the criticality of the situation. A “static” display will constantly display information, regardless of the situation. This may be useful if the driver should be constantly aware of the information, as in speed or engine temperature. “Dynamic” displays monitor the state of the environment and the vehicle and change accordingly, potentially alerting the driver only if there is an impending event to be aware of. Navigation systems are certainly dynamic, as well as dashboard warning lights for “engine check” and others. Finally, an “active” display monitors the state of the driver, and displays information in response to the state. The active display could infer intent or focus of the driver, and determine whether displaying information would be useful or distracting. This also allows to the display to change position or even modality depending on the driver state.

Table 6.1 shows examples of each modality and type of display. Many of the displays are mature technologies and are thereby widely implemented in vehicles. However design of a “dynamic and active” display is an active research subject, especially in terms of investigating effectiveness and robustness of the active display. These references are discussed in more detail in the Related Works below.

Related works

As seen in Table 6.1, a majority of research into “active” interfaces has mainly focused on simulator-based experiments. Experiments conducted such environments have evaluated various aspects of the active displays, including appropriate display timings [141, 3], positions [218, 113],

Table 6.1: Examples and Related Works involving different display modalities (haptic, audio, visual) and types (static, dynamic, and dynamic+active).

	Static	Dynamic	Dynamic Active
Haptic	Speed bumps, Rumble strips	[66]	[189]
Audio	Keys-in-ignition alarm, Headlights-still-on alarm	Navigation systems	[189, 195], [43, 3]-Simulator
Visual	Speedometer, Tachometer, Oil/Temp gauges	Navigation systems, In-car computers	[151, 208, 218]-Simulator, [4, 164], [44]-Proposed

as well as driver responses [151] and attitudes [43]. These have demonstrated the potential positive effects of the active heads-up display. However it is difficult to draw conclusions for real-world situations based on simulated experiments, as there are many more variables present in real-world conditions. According to a NHTSA Driver Workload Metrics report [6], "Some effects were observed in the laboratory that were not observed during driving. Until this is better understood, judgments on task effects should reflect a comprehensive evaluation approach that includes more than just laboratory testing." This motivates the need for an implemented testbed to experiment in real-world conditions.

Sharon et al. [189] implemented an active interface for the purpose of "coaching" or giving feedback to driving students. The interfaces are auditory and haptic, however, so very limited information can be communicated in urgent situations. Takemura et al. [195] experimented with a testbed actively sensing driver and environment state using cameras, and potentially advising the driver with voice synthesis. Amditis et al. [4] proposed a testbed solution for active visual displays using a side screen and dashboard displays. Petersson et al. [164] conducted experiments with camera-based systems to detect driver and environment conditions, with an active visual display in a side screen. Unfortunately the side screen requires the user to look away from the road to gauge the information.

Furthermore, Liu and Wen [113] observed in that truck drivers in simulators were able to control their speed better with speed information from a HUD rather than a HDD (heads-down display). The study had shown that a savings of 0.8s to 1.0s in driver reaction time can be achieved with the use of HUDs to display warning information over conventional heads-down-displays. We present results of a similar comparison on actual road-ways with a novel experimental laser-based HUD.

Our contributions are the motivation and introduction of a novel laser-based wide-area heads-up windshield display, which is capable of actively interfacing with a human as part of a Driver Assistance System. The Dynamic Active Display (DAD) is a unique prototype interface that presents safety-critical visual icons to the driver in a manner that minimizes deviation of his or her gaze direction without adding to unnecessary visual clutter. Furthermore, as part of an active safety system, the DAD presents alerts in the field-of-view of the driver "actively" - only

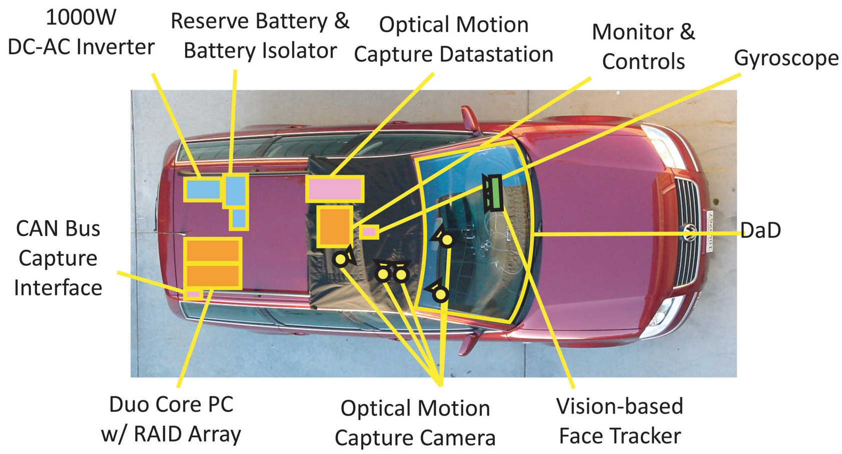


Figure 6.1: LISA-P Testbed Setup.

if necessary, based upon the state and pose of the driver, vehicle, and environment. We examine the effectiveness of DAD through a comprehensive comparative experimental evaluation of a speed-compliance driver assistance system, implemented on a vehicular testbed. Three different types of display protocols for assisting a driver to comply with speed limits are tested on actual roadways, and compared to the conventional dashboard speedometer. Given the inclination, drivers who are given an over-speed warning alert reduced the amount of "time-to-slow-back-down" to speed limit by 38% ($p < .01$) as compared to drivers not given the alert. Additionally, certain alerts decreased distraction levels by reducing the time spent looking away from the road by 63% ($p < .01$). Ultimately, each of these alerts exhibit strengths in complementing ways, demonstrating the utility and promise of the DAD system. Furthermore, the active capabilities of DAD are shown to be useful in Backup Alert Aid and Navigation Aid settings.

6.1.2 LISA-P Testbed and DAD Alerts

The Laboratory for Intelligent and Safe Automobiles - Passat (LISA-P) testbed setup can be seen in Figure 6.1. It is instrumented with a novel laser-based large area windshield display whose capabilities are demonstrated in Figure 6.2. The Dynamic Active Display, or DAD, is capable of displaying alerts anywhere on the windshield via a blue-colored laser. Additionally, the LISA-P is outfitted with an optical motion capture system, a vision-based eye gaze tracker, and GPS and CANBUS sensors to determine the state of the occupant and vehicle. A more detailed description of the LISA-P can be found in [30].

Applications of DAD

The full capabilities of the DAD system are here presented in terms of 3 demonstrative modules. Below each module is described and motivated. It is a unique ability of DAD to be able

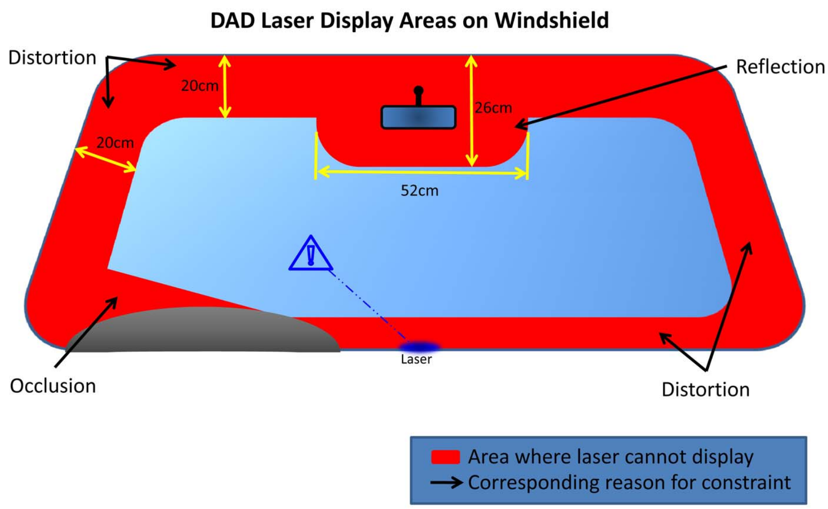


Figure 6.2: LISA-P Testbed: DAD Display Capabilities and Limitations, along with approximate location of alerts used for Speed Control Experiment. Within these limitations, the DAD is capable of drawing anywhere on the windshield.

to simultaneously run each of these using a single active display. However this study is just the beginning of this new display paradigm, and many more uses can be enumerated for a dynamic active large-area windshield display.

The first two proposed systems below, an Active Backup Alert Aid and a Navigation Aid, utilize the DAD to actively update the display based on the driver and environment. The Active Backup Alert demonstrates the ability of the DAD to interface with the driver and adjust its output based on the driver's state. The Navigation Aid integrates the DAD with environmental sensors to proactively inform the driver of upcoming directions.

Finally, a description of design of the Speed Compliance aid is presented. Specifically, its three alert designs stretch the limit of dynamic displays, and it furthermore analyzes safety using driver state information. The discussion of the Speed Compliance module is continued in Section 6.1.3, where it is used to quantitatively analyze whether the DAD System is safe and effective.

Active Backup Alert Aid

As mentioned above, there are a significant number of fatalities and injuries due to vehicle "backovers", when the driver does not take note of someone behind the vehicle. There are several systems that are capable of detecting objects behind the vehicle, but according to a recent NHTSA study [147], their performance is unreliable in certain situations. Example screenshots from the NHTSA experiments can be seen in Fig. 6.3. Additionally, many vehicles have rear-view cameras to show the driver what is in the back; in this case the driver should remember to check the camera before backing up.

Driver-centric safety systems would prove to be a safer manner for avoiding backovers.

Figure 6.3: Samples from NHTSA Backover Trials [147] showing inconsistent performance of automatic rear-vehicle object detectors.

By sensing the state of the driver and assisting him to conduct maneuvers which involve blind spots, one could significantly improve safety rates.

We have implemented an backup alert system, which actively responds to driver head pose to improve safety. Head pose is measured using the motion capture system, but could ultimately be based on vision-based approaches [31]. The system works as shown in Fig. 6.4. In this case the driver is reminded to look back, only if they are in reverse and have not recently looked back. If they happen to look at the rear-view mirror (or rear-view camera), we decrease the urgency of this alert. Finally, as soon as they look back and check the blind spot, we remove the alert from the windshield. Should the driver continue in reverse for a significant period without looking back, we display the alert again. Once the driver is out of reverse, the alert mechanism is turned off.

The advantage of this alert is that it gives useful information to the driver, in an accessible manner. Moreover the alert will not show up when the driver does not need it to, so there will not be as great of an annoyance factor. The simple addition of such an item to safety systems could thereby potentially save lives. Future research can utilize DAD to examine the effectiveness of active blind spot alerts.

Navigation Aid

Another potential application of the DAD is an active Navigation Aid, which could increase the safety of current navigation systems by actively sensing driver state and environment state. With such information it may be possible to assist the driver in making smarter decisions about controlling the vehicle. For example, a primary benefit would be to overlay directional arrows in the driver's field of view. This would allow the driver to quickly and safely assimilate the directions without being distracted.

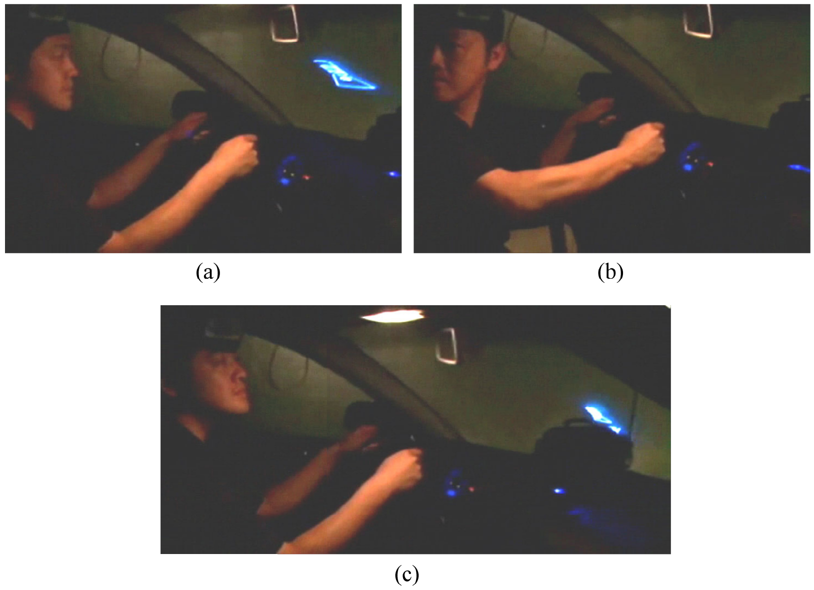


Figure 6.4: Demonstration of DAD Backup Aid. (a) If the driver is in reverse and looking forward, display an alert aligned with their gaze reminding them to look back. (b) Once they look back, remove the alert. (c) If they look at the rear-view mirror, decrease the urgency of the alert.

A sample navigation aid can be seen in action in Fig. 6.5. Future experiments can be conducted on optimal placement and timing of navigation directions, based on driver attention and distraction levels.

Speed Compliance Aid

Since speeding is a leading cause of crashes, any manner of safely reducing speed from over the speed limits may be useful. In this light, a DAD-based Speed Compliance Aid is presented and used to quantitatively analyze the safety and effectiveness of the DAD over having normal dashboard-based displays.

There are three proposed alert modes for this particular aid, whose designs were motivated by the strengths and weaknesses of human vision. The human eye can be divided into three regions based on acuity to different visual cues: the macula, which contains both the fovea and parafovea regions that subtend about 10°, and the peripheral vision region, extending to 180°. Vision within the macula has the highest visual acuity, which is necessary for reading, watching television, driving and any activity where visual detail is of primary importance. The peripheral vision extends beyond it and has good motion detection and temporal resolution [174].

For critically important situations, a visual alert presented directly in the driver's central visual field should be able to catch the driver's attention immediately, but this runs the risk of competing directly with the driver's view of the road and surroundings. For the particular case of a speed limit or current speed alert, the more appropriate placement would be a secondary

Figure 6.5: Samples from a DAD-based Navigation Aid.

location where a driver has the option of taking notice if the situation does not demand complete concentration. Watanabe et al. [214] observed that the fastest response times to HUD warnings presented during videos of drives occurred when the warning was placed 5° to the right of the center of line of sight. Because of the fast response times and the secondary importance of the speed alerts, all alerts in our speed control experiments were placed approximately 5° diagonally to the bottom-right from center of the line of sight.

There is also a need to display alerts that grab the attention of the driver from the secondary location, particularly when the speed limit has been exceeded. Since peripheral visual field is most sensitive to motion cues, we animate the alerts with zooming and bouncing effects for this purpose of attracting attention. The zooming enlarges the alert every other second, and the bouncing consists of a vertical location change that is similar to the motion of a rubber ball bouncing off the ground. Both take into account the apparent need for the driver to fixate upon the alert for a moment in order to recognize its meaning. The zoom consists of two sizes with one second separating the times between changing sizes. The bounce starts with high bounces for 0.5 seconds, but for 1.0 seconds the icon bounces only subtly until it finally comes to rest at the base location. Both allow for some time in which the icon is not or barely moving for the eyes to fixate upon. Furthermore, the alerts were designed to be approximately 2 inches tall, such that they were big enough to clearly understand and yet small enough to prevent occlusions. The zoom doubles the size, whereas the bounce moves the icon approximately 2 inches.

As it is easy to measure amount of time spent over speed limit (given the current speed and the speed limit), the Speed Compliance module is chosen over the other two modules for a quantitative evaluation of DAD. The following section uses this design methodology to determine the performance and safety of the DAD in a Speed-Compliance experiment.

6.1.3 DAD Evaluation: Speed Compliance Experiment

Experiment Details

We test four strategies of speed alerts; each driver is asked to drive the same route for each alert strategy. The alerts are presented in different orders for each driver, and the drivers are already familiar with the area. We measure the amount of time the driver spent above the speed limit, the ratio of time spent observing the alert or dash or the road in general, and the distribution of speeds measured for roads with various speed limits.

For each drive, we vary the display in one of the following four ways: (1) *No Disp* - No DAD alert is given. (2) *Warning* - A triangular exclamation point warning sign appears and bounces as soon as the driver exceeds speed limit. (3) *Numbers* - A textual alert constantly shows the driver's current speed and the road speed limit (e.g.: 43/45). The text representing the driver's speed zooms in and out if the driver is above the speed limit. (4) *Graphic* - A graphical alert constantly shows a vertical status bar with the driver's speed and the speed limit clearly marked. The entire graphic bounces if the driver is above the speed limit.

A graphical representation is shown in Figure 6.6.

On each of four iterations (each using a different display condition) of the experiment, the subject is told to drive on a given road course lasting approximately 20 minutes. This path is shown in Figure 6.7. The route is carefully chosen to include a variety of situations and environments. The speed limits vary from 15 to 65 miles per hour, and the roads range from small local roads through campus to major highways. The distances were calculated such that approximately 3-4 minutes was spent driving in each speed range.

During the drive, speed limits are acquired by determining the current global position in longitude and latitude via GPS and searching the list of road way-points for the closest match. Associated with each way-point is a speed limit that was manually annotated with the speed limit. The distance between each way-point is approximately 0.1 miles. When the current position deviates from all way-points by more than the width of the widest road, the speed limit is defaulted to a non-valid value.

Head pose is measured using a marker-based motion capture system, and eye gaze is measured using a camera-based eye tracking system. The vehicle speed data, as part of over 20 other vehicle parameters, is recorded via the vehicle's Controller Area Network (CAN) bus and passed as an input to the display module, to inform the subject of the speeds. A millisecond-accurate clock in the PC is used to time-stamp all entries of data recorded. The set up for the experimental test-bed LISA-P is shown in Fig. 6.1.

The subject is asked to drive as they would normally, but to pay particular attention to obey the speed limits. Each driver was familiar with the roads and path before beginning the drive. Data was collected from a total of eleven test subjects with varying experience levels ranging from age 22 and 50, several with glasses, totaling over 14 hours of driving data. All

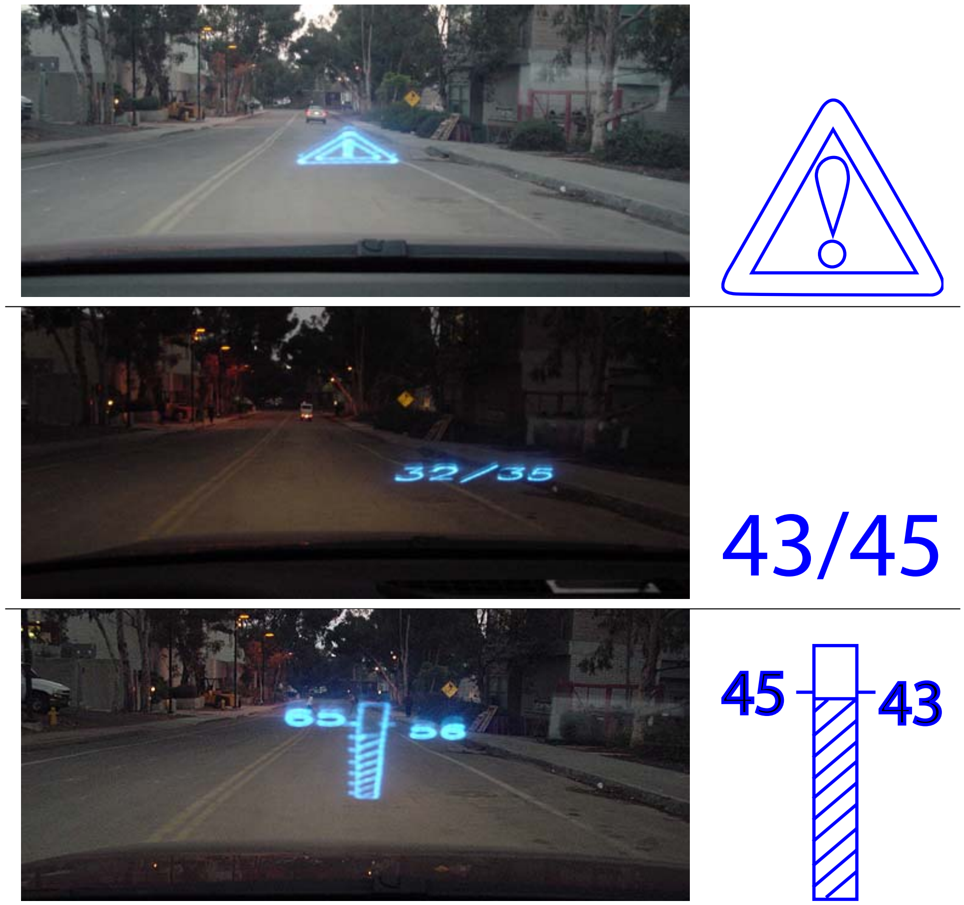


Figure 6.6: Illustration of the three alerts used.

drives were during the early evening hours, free from rush-hour traffic. Over the set of drivers, the order of the 4 display conditions was varied, hopefully mitigating the impact of any learning effects over the course of the study.

Results and Analysis

Plots of a sample drive showing speed vs. time, and the corresponding speed limits, for Display Conditions 1 and 2 are shown in Fig. 6.8.

For each display condition, the driver was asked to drive normally, paying attention to speed limits. To analyze the ability of the driver to stay under the limit, one statistic measured was the "Time-to-slow-back-down" or the average amount of time the driver spent over speed limit before returning back to under the limit. This measure was chosen to clearly represent how immediate were the effects of the different warnings. This measure also ignores route timing differences due to traffic lights, congestion, and environment changes, all of which would cause biases in other absolute measures such as "total amount of time spent over speed limit".

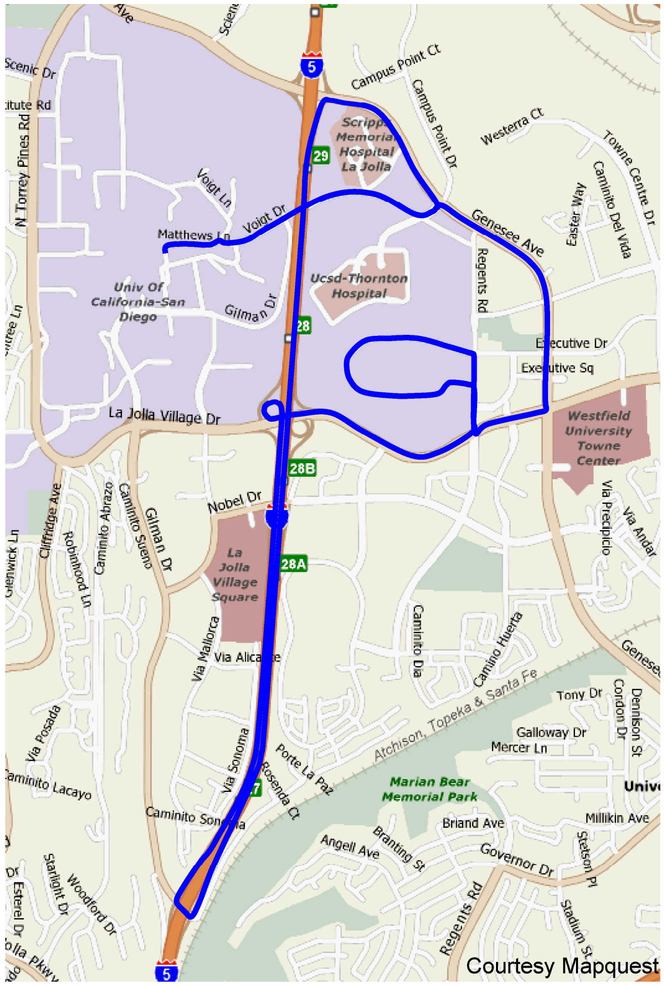


Figure 6.7: Driving Test Path, which includes local roads, main roads, and highways, with speed limits ranging from 15 to 65 mph.

The results are shown in Fig. 6.9, and statistics are listed in Table 6.2. With the second display condition, there is a clear drop in the amount of time it took each subject to return to driving below speed limit once the warning was shown. For all test subjects, the caution symbol from the second condition caused a drop of 2.24 seconds in the average time-to-slow-back-down. We then normalize these times relative to the times of Display Condition 1 in order to better compare relative effects over each driver, arriving at the values shown at the bottom of Fig. 6.9. Using this normalized metric, we can conclude that on average, the second display condition caused a drop of 38% in time-to-slow-back-down.

The other two warnings, involving the displays of numbers and graphics, were quite effective but not as much as the warning sign. As discussed below, this can be attributed to the two "active" signs being constantly displayed and thereby not catching as much of the driver's attention when the drivers were over the speed limit. Additionally, their information takes a bit

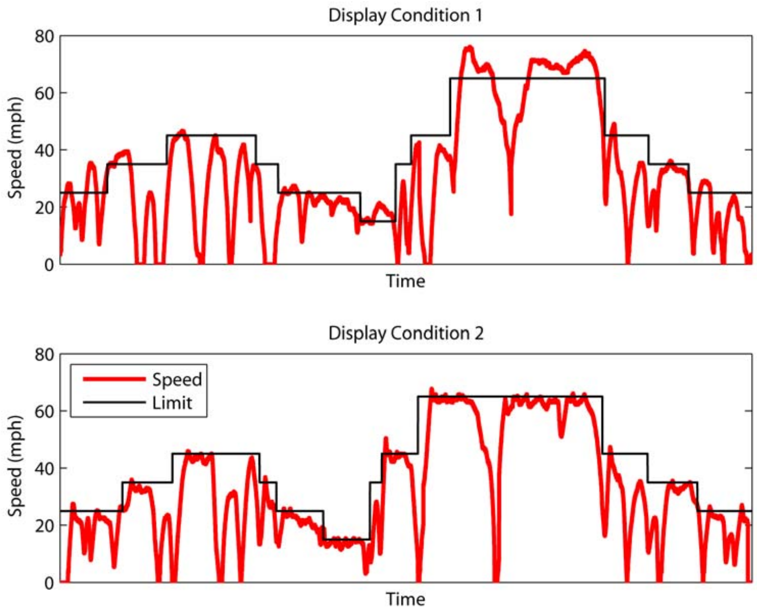


Figure 6.8: Results of sample Test Run for Condition 1 - No Display (top) and Condition 2 - Warning sign (bottom). The driver's ability to maintain speed is evidenced clearly by the reduced amount of time accidentally spent over the speed limit in Condition 2.

Table 6.2: Average time-to-slow-back-down with different alerts over all drivers.

Exp 1 - No Disp	5.09 sec
Exp 2 - Warning	2.85 sec
Exp 3 - Numbers	3.66 sec
Exp 4 - Graphic	3.28 sec

of time to process, compared with the static display which can be understood immediately.

To further understand this data, an analysis of variance (ANOVA) was performed on the normalized statistic, 'Time-to-Slow-Back-Down'. The test was conducted to ascertain whether the pattern seen in the bottom plot of Fig. 6.9 was a coincidence, or whether the reduction in time in experiments 2, 3, and 4 actually represent general patterns. Analysis was done by comparing two conditions at a time using ANOVA, which is essentially equivalent to a t-test. The analysis showed the second display condition implies a statistically significant reduction in time-to-slow-back-down ($p = .0039 < .01$), whereas the third and fourth conditions did not have statistically significant effects. Based on the calculated confidence intervals from the eleven test subjects, we may conclude that 99% of the population would experience between 4.94% and 71.76% reduction in time-to-slow-back-down using this second display condition.

Further, one could divide the population of test subjects into a "compliant" and "non-compliant" group. The "non-compliant" drivers (consisting of the first four in Fig. 6.9) exhibited

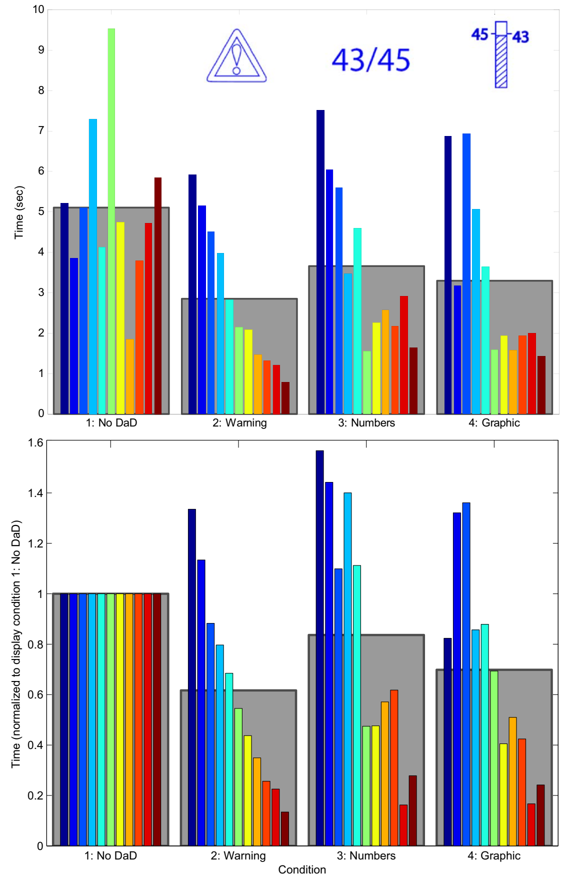


Figure 6.9: Time to slow back down, or the amount of time spent over the speed limit before slowing back down with different alerts. Each experiment consists of 4 trials by 11 different drivers. The overall averages are in gray, superimposed by the individual averages. The top plot represents the raw data, and the bottom plot shows the same data normalized by the values of the Condition 1: No DAD. See Table 6.2 for numerical figures.

no clear pattern in response to any of the alerts ($p = .10 > .01$). Interestingly, there was no clear common trait among these non-compliant drivers, as they were of varying age, gender, and backgrounds. The rest of the drivers were extremely responsive to all the alerts($p = .000082 < .0001$), and could be labeled as "compliant". Thus while we can conclude that the warning, Display Condition 2, causes a 38% overall average drop in time-to-slow-back-down($p < .01$), we can also identify that some users may be "non-compliant" in which case they are less likely to respond. As discussed below, the active nature of the DAD could allow the system to identify such non-compliant drivers by analyzing historical responses, and accordingly adjust or remove the alerts.

For each section of the route with a given speed limit, Fig. 6.10 displays the histogram of speeds from all test subjects. Subjects found it more difficult to maintain speed limits at 15mph, as evidenced by the top graph in Fig. 6.10a. However comparing this across all four conditions

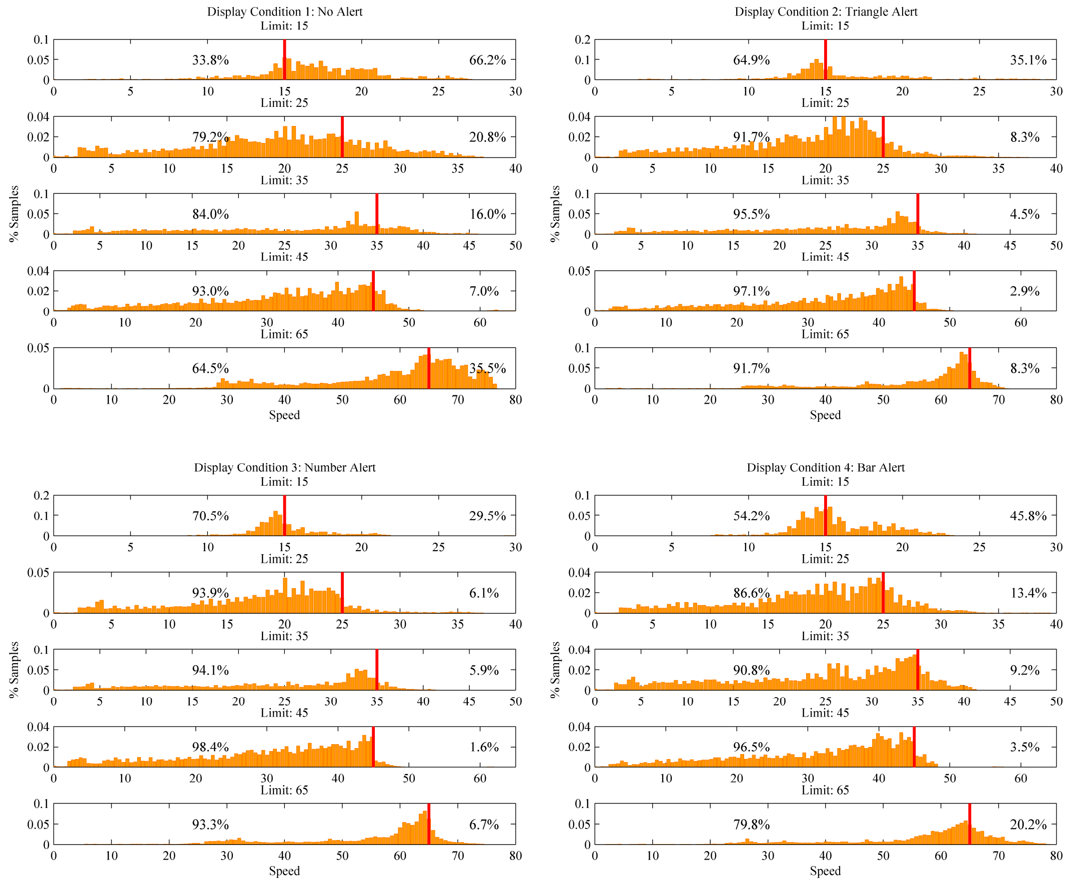


Figure 6.10: Histograms of speeds for each section of road with the given speed limit. (a) Condition 1: No DAD Display, (b Condition 2: Warning, (c) Condition 3: Numbers, and (d) Condition 4: Graphic. Significant differences can be seen by comparing over-speed cases across the 4 conditions, especially at 15 and 65mph limits.

shows the effectiveness of the HUD in reducing the amount of time spent over speed limit. The same result can also be clearly seen by comparing the histograms of the 65mph zones, where without the display there is a significant amount of time spent above the speed limit. These patterns demonstrate the effects of the DAD in assisting the subject to maintain their speed.

Driver Distraction: Pose Analysis

By analyzing the driver's behavior, it becomes possible to gauge the distraction level of the alerts and determine whether they were taking the driver's focus away from the road. Automatic analysis of pose and gaze can be done using the LISA-P testbed setup, with the motion capture system and the near-IR vision-based gaze tracking system. We can use these to determine that when the driver drifted above the speed limit, whether the HUD kept the driver from taking her attention off the road, ostensibly to look at the dashboard. This would determine drivers' attention and distraction level in response to various visual cues, which has implications on the safety of the alerts.

Eye Movement and Head Pose

Several of the studies mentioned in the related works section focus on systems monitoring the driver, especially eye gaze tracking systems ([164, 195, 6]). However while modern eye-gaze trackers have become extremely sophisticated, they still suffer under fast-changing in-vehicle conditions [112, 212, 219]. Indeed others who have measured eye gaze in vehicles have usually tracked eyes in more stable lighting environments. NHTSA [6] also conducted a thorough study into driver workload metrics, using eye gaze as a measurement tool. However in that study the eye gaze patterns were manually marked up by humans after the data was recorded, in order to achieve more accurate results.

We found that after collecting automatically marked eye gaze data, it was not accurate or robust enough to draw systematic conclusions. Specifically, many of the drivers in the current experiment wore glasses, which under strong illumination changes heavily affect the performance of the eye tracker, effectively serving as occlusions. Moreover, when the driver turns their head out of range of the eye tracking cameras, the gaze estimates are no longer valid. Finally, the eye gaze tracking system was cumbersome in that it required training for each individual subject, and slight errors in training would decrease the accuracy of the gaze estimates.

We determined that head pose estimates were reliable and thereby a better estimate of the attention and distraction levels of the driver. The marker-based head pose estimation system used in these experiments in comparison is extremely accurate and precise [217]. It does not suffer under lighting changes, as it is based on detecting IR reflections off of markers placed on the head. With the LISA-P instrumented with this powerful measurement tool, we were able to draw reliable data on driver reactions to the DAD alerts.

Distraction Results

The input head pose data uses labeled calibration data as a reference to cluster into three regions: Looking "Up," or forward at the windshield, "Down," or at the dashboard, and "At DAD," at the specific location of the DAD alert. These classes are indicated in Fig. 6.11. To calibrate the regions, several subjects were asked to look around each region, and the measurements were labeled and stored. Input data was then clustered using an $L-2$ norm-based nearest neighbor classifier.

Since each driver is unique in reaction speeds, the absolute amount of time spent looking in each direction is not a very reliable metric, especially since the times being considered are so quick. Therefore for each driver, we considered the relative amount of time spent looking in each direction with an alert (display conditions 2-4), compared to the time spent when there was no alert (display condition 1). In other words we measured the time in each direction *with alerts* as a percentage of the time in each direction *without alerts*. This normalized metric provides greater insight into the relative glance patterns of the drivers over each condition.

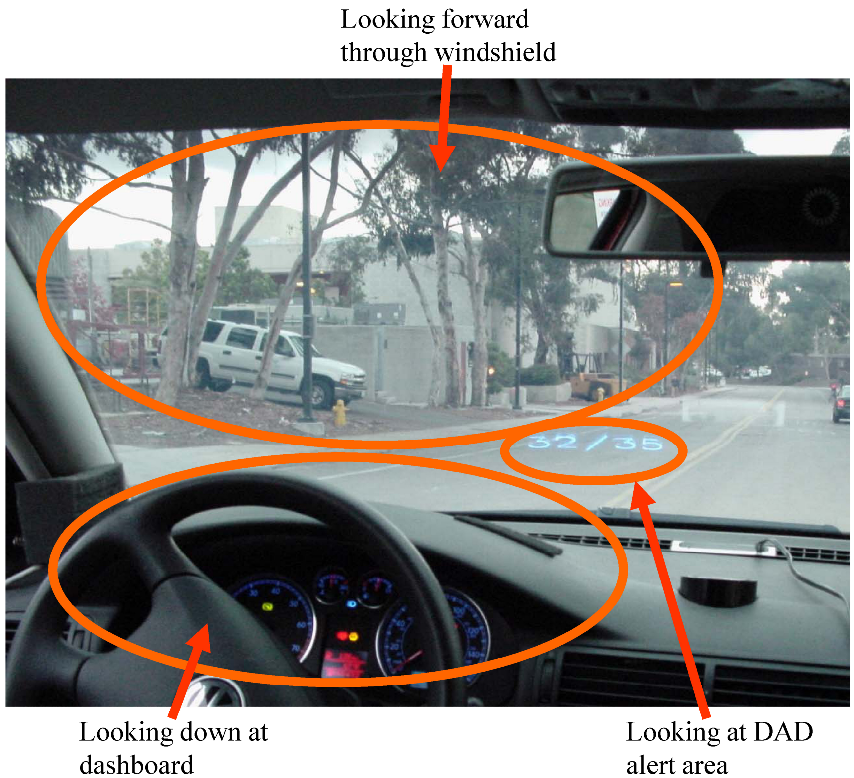


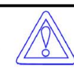
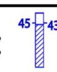
Figure 6.11: The head pose estimates are classified into three clusters, each corresponding to a certain region in front of the driver: "Up", "Down", or "At DAD"

Results of each display condition are shown in Table 6.3, again noting the behavior while the driver was traveling beyond speed limit.

The effects of the warning in Display Condition 2 can clearly be seen, in that the driver was not warned of his current speed, and so he had to look down to the dash to find out how much he needed to brake. This notion is verified by the head movements during Display Conditions 3 and 4, in which the speed was dynamically displayed to the driver, precluding the need to look down.

We can draw from these results that the alert type from Display Condition 3 caused the least distraction. The driver spent only 37% of the normal time looking down, and the time spent looking forward through the windshield increased by 10%. This amount of time is especially important when considering that every second is precious when it comes to avoiding accidents. An ANOVA analysis implies that the increase in forward-looking time might not be as statistically significant, whereas the time spent looking down is indeed a pattern with $p = .0034 < .01$.

Table 6.3: Normalized ratios of time spent "looking" in each direction while above the speed limit, with vs. without alerts, averaged over all drivers.

Alert Type:	1: [no alert]	2: 	3: 43/45	4: 
Ratio of time spent "looking **down**" with & without alert	1.00	3.40	0.37	0.31
Ratio of time spent "looking **at DAD**" with & without alert	1.00	6.80	1.37	5.50
Ratio of time spent "looking **forward**" with & without alert	1.00	0.95	1.10	1.03

6.1.4 Discussion and Concluding Remarks

A novel interface for communicating information to a driver was introduced and its motivations over other interfaces were presented. As part of an active driver assistance system, the DAD is a unique and demonstrably capable display.

Results were presented for a series of experiments conducted to discover the most effective and least distracting class of alerts using a HUD to assist a driver in maintaining speed. Head pose data was analyzed to determine the effects of the alerts on the driver's attention and focus.

The overall results in Table 6.2 show that the warning display that appears in the driver's peripheral vision while they are above the speed limit is most effective in assisting the drivers to maintain speed limit. With the warning display, the driver tended to speed only 62% as much as compared to a conventional speedometer. ANOVA implies that these results are indeed general patterns with $p = .0039$. Also, results from the head pose data imply that the warning display actually increased the time looking away from the road, however the numerical display decreased the time looking down by 63% overall ($p = .0034$).

These results were echoed by the test subjects themselves. Among the most prevalent comments were that the warning display was the most helpful because it caught their attention better than the active displays and was able to inform them that they were driving above the speed limit without causing them to move their focus away from the road. They did have to look down to gauge their speed more often, which could ultimately decrease safety. The numerical display allowed them to concentrate on the road more, as they did not have to look down to see their speed, however, it was not effective enough in grabbing their attention while drifting above the speed limit. Finally, the graphical display took a bit of time to register the information, and so it did not prove as useful, even though it was effective in slowing the drivers down.

One possibility to improve the utility of the alerts would be to combine the better aspects of each of them. The warning sign would prove more effective if the speed was also displayed in the driver's field of view. This kind of alert would still have the ability to 1) quickly grab the driver's attention, 2) include information about how much to slow down, and 3) allow the driver to maintain focus on the road. At the same time, the alert could recognize based on the current driver's responses to recent alerts, whether the driver is "non-compliant", which as shown above, decreases the likelihood of responding to the DAD. If this is the case, it may be wise to dynamically remove alerts to reduce distractions and annoyances.

The active capabilities of DAD are shown to be useful in several other situations, including a Backup Warning Aid and a Navigation Aid. Future experiments will evaluate the usefulness and critical safety improvements using these systems. It would also be interesting to consider the problem of overlaying objects or destinations with an alert on the windshield, which would require accurate calibration and registration mechanisms.

It is interesting to note that certain displays can increase distraction, while other displays decrease distraction. Design would thus become paramount in considering safe driver assistance systems. The ability of the DAD to test and implement many different designs and placements of alerts is thereby quite valuable. Because it is an active display, the DAD can even adjust its display types and information to suit the particular driver.

The Dynamic Active Display system has the potential to play a clear role in improving driver assistance systems. This alert modality can decrease distraction levels by alerting drivers without taking their gaze of the road. It also can actively alert the driver only if necessary, based on the state and pose of the driver, vehicle, and environment. The experiments conducted in this study quantitatively show the improvements in drivers' abilities to control speed, as well as decreased distraction levels, using the Dynamic Active Display - Speed Control system in real traffic conditions with the LISA-P testbed. Future research includes harnessing the capabilities of DAD in improving and analyzing intelligent Driver Assistance Systems.

6.2 Effects of Style

As with many recent developments in Advanced Driver Assistance System (ADAS) technology, we have so far neglected the notion that drivers may have independent styles which could influence their driving behavior in unique ways. These styles may make it difficult for a pattern recognition system to detect common patterns across entire populations of drivers. Intuitively, it may be more useful for a Driver Assistance System first to recognize the driver style, if any, and then to associate patterns to that particular style.

By distinguishing drivers into several different groups based on driving behavior, an ADAS could potentially improve performance for each group. Forrest Council, chair of the Strategic Highway Research Program (SHRP 2) Safety Technical Coordinating Committee, elucidates that "The more we know about driver behavior in particular circumstances, the more we can design particular treatments or countermeasures that can begin to reduce our collision rate and thus the number of deaths and injuries on our highways." Even a 1% improvement in safety

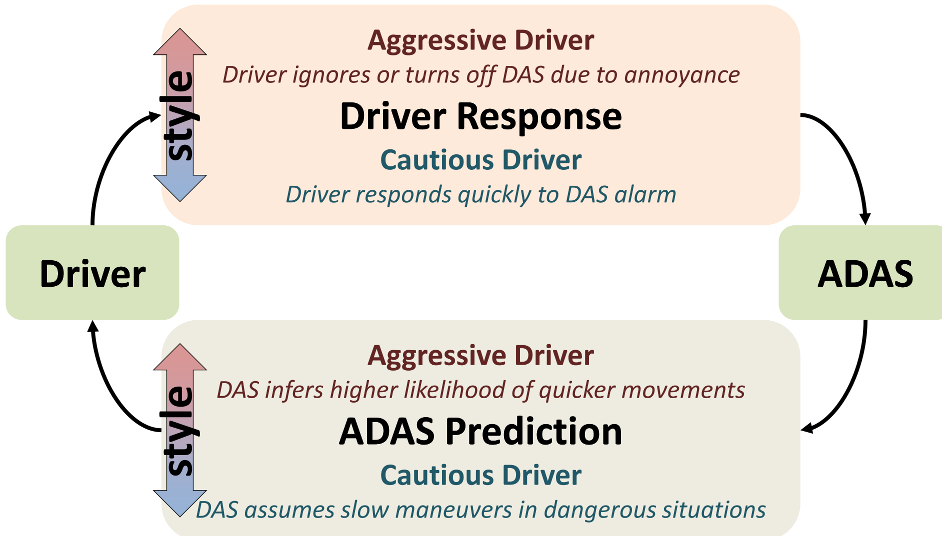


Figure 6.12: Overview of driver-vehicle interaction. The performance of the Advanced Driver Assistance System (ADAS) in interacting with the driver is colored by the different styles and personality types of individual drivers. In this article we examine the responsiveness and predictability of aggressive and non-aggressive, or cautious, drivers.

countermeasures could lead to saving 400 lives, 30,000 injuries, and $2.3 billion annually [206].

The aim of this research is to understand how to reduce false alarms and annoyances given the particular driver's style. Since every driver behaves differently behind the wheel, and reacts differently to potential warnings and events, this driver assistance system should adaptively update and provide more suitable recommendations based on the particular driver. We propose using data-driven approaches whereby we can discover various categories of drivers; in the following research we assume, without loss of generality, just two categories of drivers, "aggressive" and "non-aggressive" or cautious drivers. We analyze various metrics, including lateral and longitudinal acceleration and jerk, that may distinguish these groups of drivers; and how those measures correlate with various notions of driver "predictability" and "responsiveness."

The analysis in this work draws on experimental data from both simulator and naturalistic real-world drives. Results show that "aggressive" drivers tend to be significantly more *predictable*, most likely because their actions are more *consistent* than "non-aggressive" drivers' actions. Moreover, it is clear that the two groups of drivers tend to behave in different ways in similar situations. Finally, we find that "non-aggressive" drivers are quantifiably and significantly more *compliant* to feedback from an Intelligent ADAS.

6.2.1 Related Works and Driver Style Metrics

It has long been noted that drivers engage in vehicular maneuvers in various ways, which may differ from person to person [23]. Some have argued that drivers may change their style within the course of a single drive. This has been evidenced by the large amount of research into the causes and consequences of "road rage" [85]. Murphey et al. [136] have examined the longitudinal jerk statistics of individual drivers in order to distinguish aggressive drivers and optimize fuel economy.

Other studies have considered the notion that drivers may consistently behave differently than other drivers. Canale and Malan [24] analyzed longitudinal speed and acceleration, as well as distance to and speed of a leading vehicle, to cluster driver behaviors in "Stop-and-Go ACC" scenarios.

Several groups have patented driver style classifiers [78, 101]. These classifiers use parameters such as lateral and longitudinal acceleration in different maneuvers and environments, to classify driver reaction times and "sportiness." This information is then used to modify the dynamics of the vehicle, to respond better to the driver's own style.

A number of studies have attempted to distinguish driver styles to assist driver training and understand the psychology of driving. A couple of studies have found a relationship between the driving styles of parents and their children, possibly indicating some genetic component to driving style [197, 17]. These groups have relied upon responses to key questionnaires to distinguish different kinds of drivers.

In this study we examine a number of the style-distinguishing parameters listed above, in order to classify drivers as aggressive or non-aggressive. This information is then used to quantify the "predictability" and "responsiveness" of the driver. To the knowledge of the authors, this is the first study to analyze the relative abilities of DASs to predict future behavior of drivers of different styles. Additionally, we report the effectiveness of ADAS feedback mechanisms on the different classes of drivers.

Measures of Driving Style

In this study we consider several metrics of driving style. These include acceleration, as well as the derivative of acceleration, or jerk, in both lateral and longitudinal directions. Jerk measures how quickly the driver accelerates in a particular direction. It is presumed that more aggressive drivers tend to accelerate faster and thus have higher jerk profiles than non-aggressive drivers [136]. Where other studies have included both brake and throttle data [24], we just consider the acceleration, as that encodes information both about braking and throttling.

By examining lateral and longitudinal statistics, we are able to capture the relative behaviors of drivers in various situations. In particular, in the first two experiments in this study, we focus on the behavior of drivers in highway lane changes. This maneuver includes both lateral movement, and in many cases, to match speeds of the adjacent lane, a lane change would include longitudinal acceleration as well. In the third experiment, we focus on a speed-compliance task, which has much greater correlation to longitudinal motions.

In the case of the naturalistic driving experiments in Section 6.2.3, lateral motion is determined from a commercially available lane departure warning system. Longitudinal information is derived from the vehicle sensors via the vehicle's CAN Bus.

For each of these parameters, the most relevant information is encoded in the frequency and variability of accelerations. Thus we find it convenient to measure the standard deviation of the distributions of accelerations and jerks over time. As an example, drivers who are less prone to sudden maneuvers will tend to have tighter distributions of acceleration, and thus the variance of the distribution will decrease.

Additionally, real-world highway driving involves interaction with the environment, and in particular with vehicles around the ego-vehicle. Aggressive drivers may act differently when approaching or following slow lead vehicles. Thus we also record the average time gap between the ego-vehicle and the leading (preceding) vehicle. This information comes via an ACC radar system.

In summary, the following parameters are explored in this research to classify the driving style: (1) Std. Dev. of Longitudinal Acceleration, $\sigma_{lon-acc}$; (2) Std. Dev. of Longitudinal Jerk, $\sigma_{lon-jerk}$; (3) Std. Dev. of Lateral Acceleration, $\sigma_{lat-acc}$; (4) Std. Dev. of Lateral Jerk, $\sigma_{lat-jerk}$; and (5) Average Time-gap to Lead Vehicle, m_{TTLV}.

6.2.2 Experiment 1: Simulator-based Analysis

The first set of experiments utilizes a driving simulator, LISA-S, to examine indicators of driver style and test the predictability of each class of drivers. In the experiments the driver is cued to change lanes on occasion, and otherwise tasked with maintaining the current lane. The dynamics of the drive are extracted and used to classify the driver style, which is then examined in relation to the performance of the driver during the tasks.

Experimental Setup

The LISA-S simulator testbed is shown and described in Chapter 3. The main monitor is configured to show a PC-based interactive open source "racing" simulator, TORCS [204], which was modified to use a two-lane highway with a fixed maximum speed, to approach realistic driving. In this particular experiment the track contained no other vehicles, to limit the complexity of interactions and better analyze the dynamics of lane changes.

After taking time to acclimatize themselves to the simulator, the subject's comfort level was verified by subsequently being asked to keep the vehicle in a single lane for at least 60 seconds. For the remainder of the experiment, the subject was tasked with maintaining the current lane

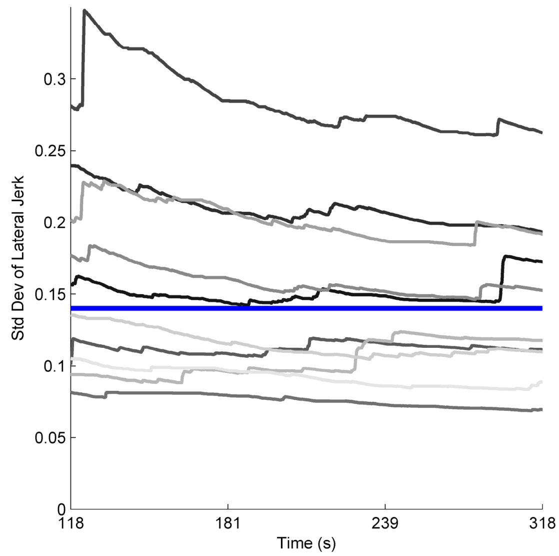


Figure 6.13: Evolution of $\sigma_{lat-jerk}$ over time, for each of the 10 drivers in the simulator-based experiment. The blue line shows a separation between "aggressive" (above) and "non-aggressive" (below) drivers. Notice that for each driver, the measured $\sigma_{lat-jerk}$ does not vary greatly over time, indicating that this is a consistent measure of the driver's style.

to the best of their ability. This allowed the subject to be actively engaged in the driving process throughout the experiment.

Occasionally then a cue to change lanes would come from the secondary monitor, whose entire screen would change color suddenly, at a set of predefined times unknown to the subject. Upon noticing the change, the driver was tasked with maneuvering to the appropriate lane as soon as was safe to do so.

A total of 10 subjects participated in the experiment, of varying age, driving experience, and simulator experience. Each data collection segment included about 10 minutes of data, with 10-15 lane changes per subject. This data is the same as was collected in Condition 2 of the Attention Shifts experiment of Chapter 3.

Data Analysis

For each of the drivers, the distributions of lateral and longitudinal jerk and acceleration were calculated over the whole drive. Over time, the distribution of $\sigma_{lat-jerk}$ did not change greatly, as indicated by Figure 6.13. This measure showed consistency over time, and thus was chosen as the best candidate to classify the style of the driver. As seen in Figure 6.13, a clear cluster of drivers emerges who accelerate with less force than others. These two groups are separated into "aggressive" and "non-aggressive" drivers.

The predictability of the driver is directly correlated with how consistently they behave in making maneuvers. In the case of lane changes, we can measure how long the driver takes

Figure 6.14: Differences in measured time from lane change initiation until lane crossing (Time-to-lane-crossing), for different driver styles in the simulator experiment. ANOVA showed a significant difference ($p < 0.05$) between the groups; additionally the distribution of times for "aggressive" drivers is much more compact. Each boxplot shows the median, 25^{th} and 75^{th} percentiles, and min and max extent of the data.

from the initiation of the maneuver until the point of lane crossing. Figure 6.14 represents the distribution of the "time-to-lane-crossing" as a function of the driver style. ANOVA analysis shows a statistically significant difference between these distributions ($p < 0.05$).

Similar statistical analysis can be done with each of the other style indicators; each of these p-values is shown in Table 6.4. Since Lateral Jerk and Acceleration classify the drivers in the same way, the distribution of Time-to-lane-crossing is just as distinguishable ($p = 0.0472 < 0.05$) in both cases. However the distributions of longitudinal acceleration and jerk are not as telling in this case.

Table 6.4: ANOVA: Time-to-lane-crossing in Non-aggressive vs. Aggressive Drivers, based on each style metric

Style Metric	*p-value*
$\sigma_{lon-acc}$	$p = 0.3938$
$\sigma_{lon-jerk}$	$p = 0.3938$
$\sigma_{lat-acc}$	$p = 0.0472$
$\sigma_{lat-jerk}$	$p = 0.0472$

Figure 6.14 shows that the "aggressive" drivers always tend to start their lane changes around 1.5 seconds prior to the lane crossing. This indicates that they are quite consistent in that maneuver, and may thus be more predictable. The non-aggressive drivers have a much wider distribution, starting the maneuver as much as 5 seconds prior to the lane crossing.

We might conclude that a behavior prediction scheme designed to detect the progression of lateral deviation would be much more precise and accurate in the case of the "aggressive" drivers. To further validate this proposition, in the next section we analyze data from naturalistic drives in real-world experiments, as described in previous chapters, using our intent analysis framework [48] to predict behaviors.

6.2.3 Experiment 2: Real-world Analysis - Predictability and Intent

In the following section we aim to validate the hypothesis that aggressive drivers are more predictable, in a real-world setting using results from a driver behavior-based lane change intent prediction system.

Data Collection

For this experiment, we re-analyze the data from Chapter 5.2. Data is collected in naturalistic, real-world driving using the LISA-X testbed shown in Figure 5.6. A total of 15 drivers of varying age, gender, and experience were asked to drive naturally on mostly highways, during periods of light traffic. Each driver completed approximately an hour of driving, resulting in a total of 500 naturalistic lane changes for the whole dataset.

Data was logged synchronously and post-processed to determine lane change timings based on lateral deviation. This dataset was then used to calculate each of the style parameters listed above, including lateral and longitudinal jerk and acceleration, as well as the time-to-lead-vehicle. We then use this data to classify the style drivers, as done for the simulator experiment above.

The predictability of the drivers, in this case, is measured from the output of a driver behavior inferencing system, described in more detail in Chapter 5.

As shown in Figure 5.7, the classifier outputs a class membership likelihood, which can then be thresholded to determine a true positive and false positive rate for the predicted lane change intent. We average the class membership probabilities for the positive examples, as well as for the negative examples. The difference between these metrics is defined as the Intent Prediction Confidence (IPC), as this is a reasonable measure of the separability of the positive and negative lane change examples.

Analysis

Training occurs on all the data, and then we test on each individual's dataset. The intent prediction confidence measure, is then calculated for each person. This represents the predictability of that person with respect to this particular lane change intent classifier.

Once again, as done in the simulator experiments, the drivers are classified based on their lateral jerk, $\sigma_{lat-jerk}$. Figure 6.15 shows the distribution of IPCs as a function of the driver style, derived from lateral jerk.

Figure 6.15: Distributions of Intent Prediction Confidences as a function of Driver Style (derived from $\sigma_{lat-jerk}$). ANOVA shows a significant increase in IPCs for "aggressive" drivers ($p < 0.05$).

It is apparent that "aggressive" drivers who exhibit a higher level of lateral jerk, tend to produce a higher confidence in the intent prediction. The distribution of "aggressive" IPCs is significantly greater than the distribution of "non-aggressive" IPCs ($p < 0.05$). The conclusion can be drawn that aggressive drivers, possibly due to their consistency in maneuvering (as demonstrated in Section 6.2.2), are easier to classify than "non-aggressive" drivers.

Similar significance results can be seen by classifying the drivers according to $\sigma_{lat-acc}$, as seen in Table 6.5. Aggressiveness in the longitudinal direction does not seem to be as related to the predictability of the driver in terms of lane changes, though $\sigma_{lon-acc}$ trends toward significance.

However when classifying the driver style based on the average "time-to-lead-vehicle," i.e. the time gap between the ego-vehicle and the vehicle directly preceding the ego-vehicle, a significant pattern arises. As seen in Figure 6.16, the drivers who tend to stay closer to the lead vehicle, and thus who are more "aggressive," are significantly easier to predict than the "non-aggressive" drivers.

As mentioned below, it may be the case that the population of "non-aggressive" drivers could be further sub-divided in order to make classification on that group more accurate. However, in the datasets used in these experiments, no clear or significant sub-grouping appeared which correlated with more accurate performance. It is important to note that the number of test examples in each of the classes of styles were chosen in approximately similar numbers, so as not to skew the significance testing results.

6.2.4 Experiment 3: Real-world - Responsiveness to feedback

We have noted in prior sections that "non-aggressive" drivers are more difficult to predict, due possibly to their variability, but it is also important to determine how receptive they may be to different forms of feedback from a DAS. In this section we expand upon results from the

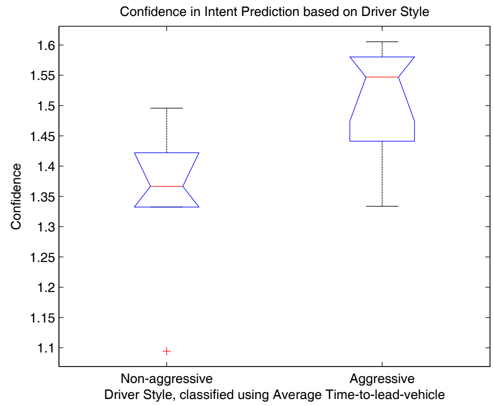


Figure 6.16: Distributions of Intent Prediction Confidences as a function of Driver Style (derived from m_{TTLV}). ANOVA also shows a significant increase in IPCs for "aggressive" drivers ($p < 0.05$). Thus drivers who tend to stay closer to the lead vehicles, are the ones who are easier to predict.

Table 6.5: ANOVA: IPC in Aggressive vs. Non-Aggressive Drivers, based on each style metric

Style Metric	*p-value*
$\sigma_{lon-acc}$	$p = 0.0990$
$\sigma_{lon-jerk}$	$p = 0.5557$
$\sigma_{lat-acc}$	$p = 0.0067$
$\sigma_{lat-jerk}$	$p = 0.0026$
m_{TTLV}	$p = 0.0184$

first part of this chapter, in which the effects of a novel heads-up display, the "Dynamic Active Display (DAD)," were analyzed during a speed-compliance experiment.

In those experiments, several different display conditions were used to assist drivers in maintaining the speed limit. In particular, one condition displayed the numerical speed and speed limit in the driver's para-foveal field of view, and another condition displayed the same information graphically. When going over the speed limit, the display started flashing or "bouncing," in order to catch the driver's attention and bring them back down to the speed limit.

For each of the conditions of the experiment, 10 subjects of varying age, gender, and experience were told to drive on a given road course lasting approximately 20 minutes. The route included all ranges of road types and speed limits, with approximately 3-4 minutes spent driving in each speed range, ultimately totaling over 13 hours of driving data.

To analyze the responsiveness of the driver to the DAD feedback, the main statistic measured was the "Time-to-slow-back-down" or the average amount of time the driver spent over speed limit before returning back to under the limit. The results are shown in Fig. 6.9. The top graph shows the raw times, for each individual driver. We then normalize these times relative to the driver's times in a control condition without displays, in order to better compare

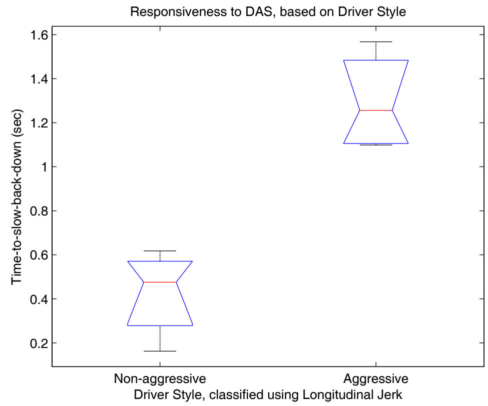


Figure 6.17: Distributions of "Time-to-slow-back-down" (a measure of responsiveness to the DAD feedback) as a function of Driver Style (derived from $\sigma_{lon-jerk}$). ANOVA shows a significant decrease time, and thus an increase in responsiveness, for the "non-aggressive" drivers ($p < 0.01$).

relative effects over each driver, arriving at the values shown at the bottom of Fig. 6.9.

The population of test subjects easily divides into a **compliant** and **non-compliant** group [44]. The "non-compliant" drivers (i.e., the first four in Fig. 6.9) exhibited no clear pattern in response to any of the alerts ($p = .10 > .01$). Interestingly, there were also no reported common traits among these non-compliant drivers, as they were of varying age, gender, and backgrounds [44]. The rest of the drivers were extremely responsive to all the alerts($p = .000082 < .0001$), and could be labeled as "compliant".

Style and Responsiveness

In order to classify the driver style for these experiments, both longitudinal acceleration and jerk were measured for each driver (lateral measures were unavailable in this experiment). Classification of drivers by $\sigma_{lon-acc}$ produced no significant results. The drivers were also classified using the longitudinal jerk metric, $\sigma_{lon-jerk}$, to approximately equal groups on either side of the mean. Upon comparing these groups using the data above, a significant pattern of non-responsiveness arose in the "aggressive" driver case.

Analysis can be seen in Figure 6.17. The "time-to-slow-back-down," which corresponds to responsiveness, as indicated above, is significantly different in the two cases. It can be seen that "aggressive" drivers tend to take much longer to respond to the numerical condition of the DAD display, as opposed to "non-aggressive" drivers.

By classifying drivers according to $\sigma_{lon-jerk}$, we can identify that those drivers who are "aggressive" also tend to be "non-compliant" to driver feedback, in which case they are less likely to respond to suggestion. By minimizing alerts to this class of drivers, a DAS stands to benefit from increased trust and thus it has a better ability to save lives. Further, though "non-aggressive" drivers are harder to predict, they may also be more receptive to feedback from

Driver Assistance Systems.

6.2.5 Discussion and Future Work

Given the range of style, experience, and personalities of drivers on the roads, intelligent assistance systems must be able to function either in spite of or in harmony with each individual driver. Measures of the driver's likely behaviors are also affected by the style, as the individuals act and respond in different manners and patterns. In this study we have presented several measures of driving style and show how they correlate with the *predictability* and *responsiveness* of the driver in several experimental conditions. Different measures for classifying "aggressive" drivers are useful depending on the maneuver.

Results show that more "aggressive" drivers tend to be significantly *more predictable*, yet *less responsive* to feedback. Their predictability increases most likely because their actions are more consistent than "non-aggressive" drivers' actions. Moreover, it is clear that the two groups of drivers tend to behave in different ways in similar situations. "Non-aggressive" drivers are quantifiably and significantly more compliant to feedback from an Intelligent DAS. These results could clearly affect the design, effectiveness, and feedback mechanisms of future Driver Assistance Systems.

It may be that the population of "non-aggressive" drivers need to be further split in order to detect more significant behavioral trends. However within the datasets used in these experiments, no significant sub-grouping of the non-aggressive data was found. Further work should include data collection of more subjects, in order to further break down and classify groups of driving styles.

Another avenue for future research includes Driver feedback. While it may be the case that certain styles of drivers are less predictable, we have also shown that "non-aggressive" drivers tend to be "compliant," or more susceptible to suggestions by a Driver Assistance System. By detecting on-line the differences among these classes of drivers, an intelligent ADAS could modify its own behavior to respond differently to an expert, rushed driver as opposed to a novice, cautious driver. Such customizability in the design of intelligent DAS will improve performance, comfort, and safety of the driving environment for a wide group of drivers.

6.3 Potential of Communicating Intent

With the anticipated increasing market penetration of modular V2V and V2I communications frameworks such as Intellidrive [209] and Car2Car [25], we now examine the potential of communicating intent. The scope of this research aims directly at the problem of understanding driver behavior, and the driver's interactions with the surrounding environment that influence their behaviors, to improve *cooperative active safety* systems. We propose a design for an inte-

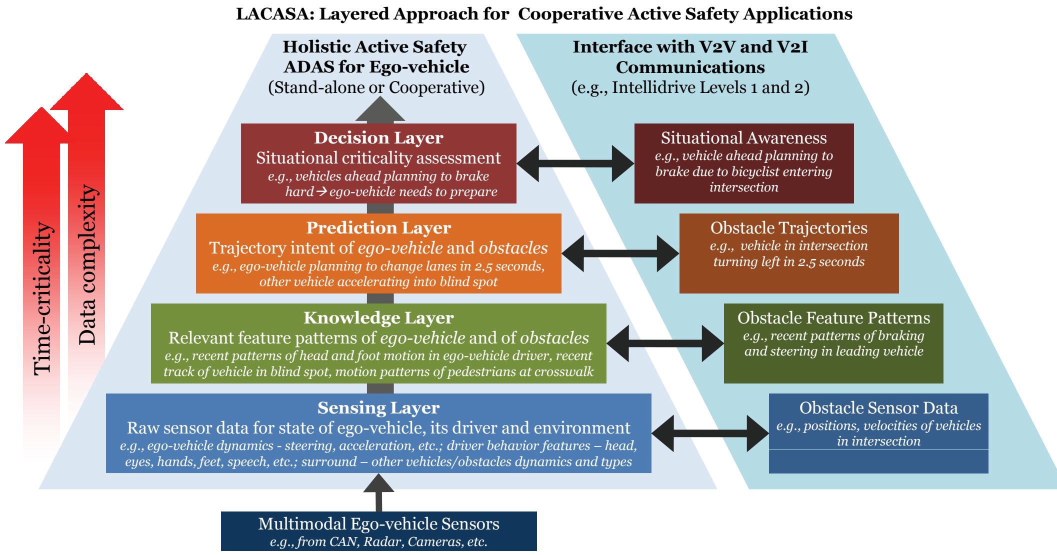


Figure 6.18: Layered Architecture for Cooperative Active Safety Applications (LACASA). A LACASA Driver Assistance System could operate independently without communication; and various layers could communicate over V2V and V2I depending on the implementation in each vehicle.

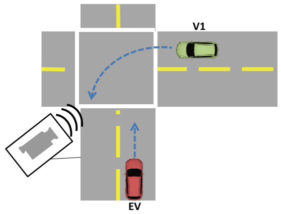
grated, intelligent Layered Architecture for Cooperative Active Safety Applications (LACASA), with a special focus on a human-centered advanced driver assistance systems implementation of LACASA. The key design components of the LACASA framework include the following features:

Holistic. The system should incorporate any available information about the driver, vehicle, and environment, all through sensors on the ego-vehicle itself.

Cooperative. The system should be able to operate in stand-alone mode, but should also be capable of improved performance through communications with other vehicles and infrastructure.

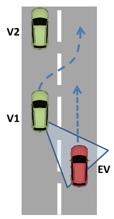
Modular. It should be able to cooperate at different "levels": new vehicles and infrastructure will enter the market with varying sensory and communications capabilities, and each LACASA node must be able to utilize the best available information from all the other systems.

In this section we demonstrate how the proposed layered framework could have an immediate effect in improving active safety in stand-alone vehicles, through the incorporation of Driver Intent detection [121, 48, 120, 30] as well as obstacle trajectory prediction [130]. As market penetration of the cooperative system increases, there will be a significant quantitative improvements in safety at various levels. The framework as a whole does not rely on a particular rate of market penetration to start improving active safety, thus overcoming a fundamental problem with many communication-based driver assistance systems. By utilizing all available production-level sensors existing in the ego-vehicle, as well as whatever level of information may be available from cooperative vehicles and infrastructure, the proposed framework demonstrates



Situation 1. Ego-vehicle (EV) with inattentive driver approaching (intelligent) intersection where Vehicle 1 (V1) is slowing in preparation to make a left turn.

Ego-Vehicle LACASA DAS Implementation	Vehicle 1 LACASA DAS Implementation	Intelligent Infrastructure with LACASA	Situational Criticality
None	None	None	EV could collide with V1 since both are unaware of the other's intentions
HC-ADAS	None	None	Radar on EV detects V1 and notifies driver of potential obstacle
HC-ADAS	None	Camera or Loop Detectors	Infrastructure detects V1's trajectory and notifies EV driver of likely obstacle
HC-ADAS	After-market-GPS	None	EV radar detection is complemented by communication from V1 indicating more accurate position and velocity information
HC-ADAS	After-market-GPS	Camera or Loop Detectors	EV radar detection is complemented by communication from infrastructure and from V1, indicating more accurate position and velocity information
HC-ADAS	HC-ADAS	Camera or Loop Detectors	V1 communicates trajectory intent based on V1's driver behaviors, to infrastructure and to EV, which issues a critical warning to EV driver about upcoming obstacle



Situation 2. Highway lane change. Vehicle 1 (V1) intends to change lanes, due to slow lead vehicle (V2), unaware of the ego-vehicle (EV) in its blind spot. EV driver is unaware of the potentially dangerous situation.

Ego-Vehicle LACASA DAS Implementation	Vehicle 1 LACASA DAS Implementation	Situational Criticality
None	None	EV could collide with V1 if EV driver is inattentive or unable to respond fast enough.
HC-ADAS	None	EV detects lane change of V1 soon after the maneuver starts, and notifies EV driver to slow down
HC-ADAS	After-market-GPS	EV's calculated Trajectory Intent of V1 is complemented by (time-critical) communication from V1 with more accurate lane-position information. EV is able to predict V1's lane change earlier, in order to notify EV driver.
HC-ADAS	HC-ADAS	V1 communicates lane change intent based on V1 driver's head motions and situational awareness (slow V2 ahead), to Ego-vehicle. EV-driver is then alerted to prepare to slow down, even before V1 starts changing lanes.

Figure 6.19: Motivational examples for modular, cooperative LACASA framework. The ego-vehicle, equipped with a proposed human-centered advance driver assistance system (HC-ADAS) based on LACASA, can interact with the obstacle vehicle (V1) in various manners depending on V1's implementation. Without any communication from V1, the advanced version of EV+LACASA is still able to improve performance through advanced sensing. With each additional bit of information from V1 (more details in Table 6.6), the ego-vehicle is able to make more accurate and timely assessments of the situational criticality. The same holds as increasing amounts of information come in from Intelligent Infrastructures, as seen in Example 1 (top).

an elegant approach to implementing future active safety systems.

The examples in Figure 6.19 embody the overall objectives of the proposed LACASA framework: By sensing and analyzing relevant information from both the interior and exterior of the vehicle, we hypothesize that active safety systems will be able to provide more accurate predictions and allow the driver earlier awareness of dangerous situations. Additionally, cooperation at different levels with surrounding vehicles could provide a more accurate and useful context to determine situational criticality, and provide alerts and assistance to drivers even earlier.

6.3.1 Cooperative Implementation of Intent-based Advance Driver Assistance Systems using LACASA

Market Penetration and "Layered" Framework

A major consideration in the design of cooperative active safety systems is the requirement for a significant number of vehicles to be equipped, in order for the system to work reasonably. However the market penetration of such systems is bound to advance slowly.

According to NHTSA [144], assuming every new vehicle on the road is equipped with an active safety system in each year since deployment, a best-case scenario, just 8% of vehicles on the roads would have the system after 3 years, and 27% of vehicles after 10 years. In order to overcome this penetration issue, as shown in Figure 6.18, each layer of the cooperative DAS is capable of communicating and integrating information at various levels. This is designed explicitly to allow for various types of V2V protocols and devices that will come online in near future. Some vehicles may have after-market communications devices with limited sensors and communications. Other vehicles will have top-of-the-line sensors with built-in time-critical communications protocols.

Information received from either of these vehicles, should be useful to an ego-vehicle equipped with the proposed LACASA framework. As more informative information about obstacle positions, trajectories, and intents become available from more advanced systems, the ego-vehicle's estimate of the situational criticality should tangibly improve. Thus increasing market penetration of V2V Intellidrive-style systems, while not critical to the performance of a LACASA ADAS, would systematically improve its performance in an elegant manner.

Additionally, V2I systems could be extremely useful for enhancing operational capabilities of vehicle-based active safety systems. Recently researchers have been successful in being able to predict vehicle trajectories and patterns from intelligent infrastructure [129, 131]; this information could feed directly into the "Prediction Layer" of a vehicle-based LACASA system.

A number of situations would benefit from intelligent cooperative DASs, including Forward Collision Warning, Lane Change Assist, and Intersection Assist. The improvements in these systems due to cooperation were somewhat dismissed by NHTSA [144], due to low estimated market penetration. However an implementation of the proposed framework does not suffer for lack of other such systems on the road. Indeed, as various systems come online, the ego-vehicle system would adapt and correspondingly update its performance, as demonstrated in the next section.

An example of a proposed protocol is the DOT-sponsored Intellidrive project [209]. The consortium has proposed several levels of communication protocols, where Level 1 includes stand-alone devices without access to the vehicle computer, and Level 2 includes built-in access to vehicle parameters. These levels are further subdivided into time-critical and non-time-critical applications. The proposed LACASA framework can incorporate information from either level of communication, with basic Level 1 position and velocity information sufficient to establish other vehicle's trajectories and baseline intentions. Level 2 information can provide more detailed information about other vehicle's trajectories, intents, and even their sense of the situational criticality.

In the simulated results reported below, as well as the sample situations in Figure 6.19, we consider several different sample implementations of the driver assistance system. These are based on the proposed framework of Figure 6.18, including the vision of a modular roll-out of Intellidrive-style communications. Table 6.6 compares some example features of each implementation of the LACASA system.

Table 6.6: Sample "Levels" of Implementations of LACASA. Future vehicles may have one of these two LACASA-based ADASs, or include some combination or subset of the sensor and communications equipment.

LACASA Implementation	*Sensors*	*Communications*
After-Market with GPS (*AM-GPS*)	GPS, Lane-Camera	*Intellidrive Level 1*: Pos, Vel; Basic Features and Simplistic Trajectory Intent
Human-Centric, Holistic ADAS (*HC-ADAS*)	+ Face-Camera, Radar, Vehicle Data	*Intellidrive Level 2*: Complex Features, Advanced Trajectory Intent and Situational Awareness

An important consideration in communication networks is the limited data rate of the channel. Protocols could be considered under the proposed framework, where messages are only broadcast in potentially critical situations, as has been done in prior research [172].

6.3.2 Quantitative Improvements of Cooperative LACASA-based Systems

To demonstrate the improvements in safety using the cooperative LACASA framework for driver assistance systems, the following section discusses some quantitative assessments of several example situations. These situations are built upon recent research into driver intent prediction [121, 120, 30, 48], with the aim to determine the relative improvement in safety gained through transmission of intents using the layered LACASA approach.

Forward Collision Warning - Risk Assessment Calculations

Recent research has shown how significant gains in safety can be achieved with relatively modest reductions in speed [95, 1]. These studies determined that at $60km/h$, every $5km/h$ reduction in speed results in between 33 and 50% reduction in crash fatalities. This is a clear case for such ADAS applications as Forward Collision and Early Brake Warning Systems. Advanced sensing through ACC radar could potentially alert the driver up to 2.6 seconds in advance of a crash [22].

An inattentive driver in a vehicle with a holistic LACASA HC-ADAS might be alerted to the vehicle ahead slowing down through the ACC radar. With such input, the driver may become alert and begin braking; potentially reducing risk of involvement in a fatal crash by 93.36%.

Assuming the preceding vehicle also had a holistic, cooperative HC-ADAS as proposed, the preceding driver's intent to brake could be detected 1 to 2 seconds in advance of the action [120]. With a one second lead time in the alert of the inattentive driver in the ego-vehicle, the driver in the lag vehicle could begin braking earlier, reducing their speed even further and reducing risk of involvement in a fatal crash by 98.47%.

Figure 6.20 shows the fatal crash risk reduction as a function of alert time, with a fixed response time ($0.75s$) and deceleration ($0.5g$). These sample calculations show the power of

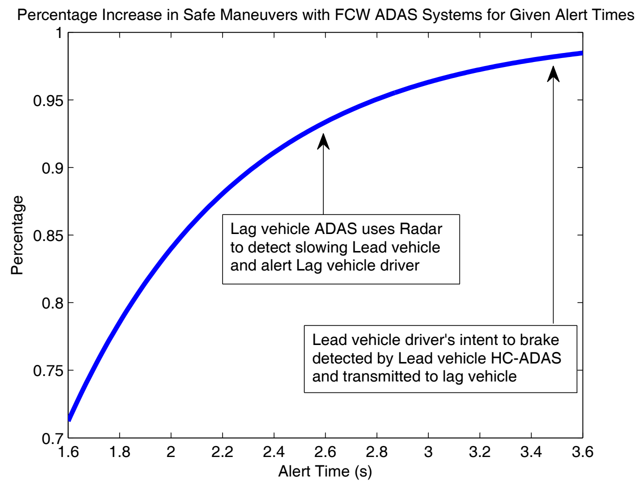


Figure 6.20: Forward Collision Warning - Reduction in risk with earlier alert times. Recent research showed ACC radar affords a 2.6 second alert time before a crash [22]. By incorporating a lead vehicle driver's intent to brake [120] and transmitting that information using LACASA to the lag vehicle, a significant reduction in fatal crashes could result.

adding the HC-ADAS system to cars, even if other cars are not equipped. As soon as other cars are able to transmit information, that becomes useful to the ego-vehicle to improve its situational awareness.

Lane Change Warning - Monte Carlo Simulation

Recent estimates find that 2% of traffic accidents every year occur due to poor lane changes, resulting in over 800 fatal collisions each year [143]. One such example is shown in situation 2 of Figure 6.19, where an unsafe lane change by the lead vehicle in front of the ego-vehicle, could lead to a potential collision.

A Monte Carlo Simulation was performed to quantify statistics about how different levels of LACASA Driver Assistance Systems would affect the collision rate in such circumstances. In the simulation, four conditions were considered, corresponding to the four conditions shown in Figure 6.19. In the case when both V1 (front car) and EV (rear car) have an advanced Human-Centric version of LACASA, termed the HC-ADAS, the front car is assumed to transmit its own intent to change lanes, 3 seconds before the actual lane crossing. This is in line with expected results from the lane change intent system developed in prior works [121, 48].

In the case when the front car V1 has only an after-market implementation of LACASA, it may still be able to accurately assess its lane position and transmit a confident lane change intent 2 seconds prior to the lane crossing, slightly after the maneuver has started [121, 48]. In the third case, V1 might have no DAS, in which case EV must rely on its own sensors, such as radar and camera systems, to detect the drifting V1; the system will at least be able to detect the maneuver as V1 begins to touch the lane boundary, 1 second before the center of V1 crosses the lane boundary. Finally, without any assistance from an ADAS, the rear (ego-vehicle) driver may not notice and be able to react to the lane change until the front vehicle crosses the lane boundary.

We define t_{alert} as the time of communication of the lane change alert from the front vehicle to the rear vehicle. t_{alert} is set to $[-3, -2, -1, 0]$ seconds, for the four conditions respectively. In each of these conditions, we uniformly vary the initial position $(-5m : +1m)$, velocity $(-2.25m/s : 2.25m/s)$, and acceleration $(-.1g : .1g)$ of the rear vehicle relative to the front vehicle, as well as the braking force $(-.7g : -1g)$ and reaction time $(abs[normal(m = .25sec, \sigma^2 = .5)])$ of the driver to the alert. We assume it takes an average of 6 seconds to complete the lane change [198], with the lane-crossing at $t_{LC} = 3$ seconds. The average braking force of the drivers is assumed to reduce to $0.4g$, 1.5 seconds after the initial brake. One million trials of each condition were conducted to obtain the results in Table 6.7 and Figure 6.21.

Table 6.7: Lane Change Simulation Outcomes (when front-vehicle crosses lane boundary)

Front-car LACASA	Rear-car LACASA	t_{alert}	Collisions	Close Calls	Safe Maneuvers
HC-ADAS	HC-ADAS	-3 sec	00.02%	28.90%	71.09%
AM-GPS	HC-ADAS	-2 sec	02.33%	96.11%	01.55%
none	HC-ADAS	-1 sec	24.62%	75.38%	00.00%
none	none	0 sec	35.48%	64.51%	00.00%

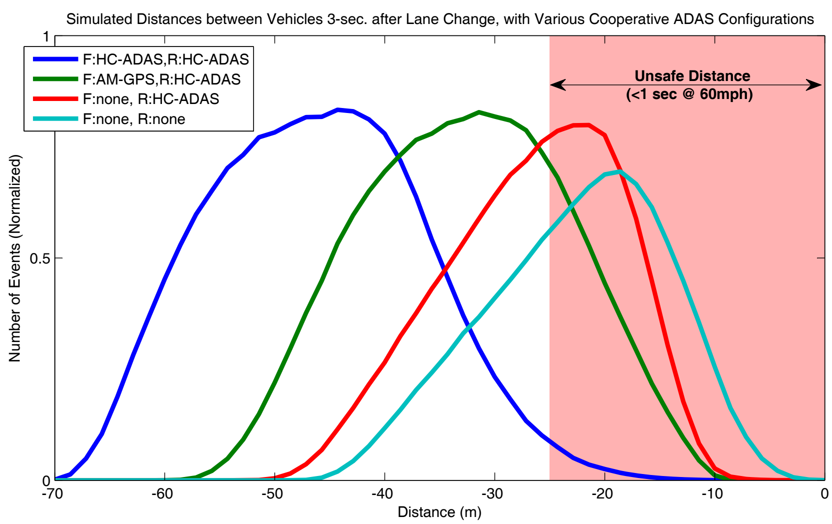


Figure 6.21: Lane Change Simulation Results: For those vehicles that did not get into collisions, the distribution of distances between the front (F) and rear (R) vehicles after the finish of the lane change maneuver, 3 seconds after the lane-crossing. Each of the curves represents a simulated scenario in which the lead and lag cars have different levels of cooperation.

We find that the percentage of collisions (which occur when the preceding vehicle crosses the lane boundary and overlaps with the lag vehicle) is much higher when there are no alerts or driver assistance systems. However when the rear vehicle has an alert from the LACASA-based HC-ADAS, the amount of collisions reduces by a third, from 35.48% to 24.62%. As soon as the cooperative framework is introduced in the lead vehicle, the number of collisions nearly disappears - down to 2.33% with an "Intellidrive Level-1" style system, and .02% with a more advanced human-centric system. The number of "close calls", defined as any situation with a lag time less than 1 second, also reduces significantly when the lead vehicle ADAS is upgraded from an after-market system to a more advanced, built-in system.

Intersection Turn Warning - Monte Carlo Simulation

Table 6.8: Intersection Turn Simulation Outcomes (when turning-vehicle leaves intersection)

Turning-car	Oncoming-car	t_{alert}	Collisions	Close Calls	Safe Maneuvers
HC-ADAS	HC-ADAS	0 sec	00.00%	03.01%	96.99%
AM-GPS	HC-ADAS	+1 sec	00.67%	24.57%	74.67%
none	HC-ADAS	+2 sec	10.46%	45.51%	44.03%
none	none	+3 sec	30.45%	52.69%	16.86%

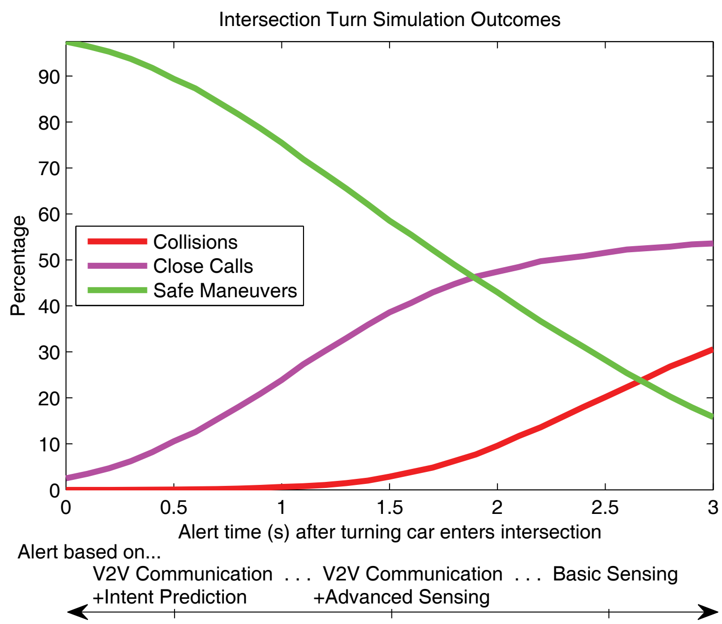


Figure 6.22: Intersection Turn Simulation Results: Percentage of outcomes resulting in Collisions, Close Calls, or Safe Maneuvers, as a function of Alert Time. Earlier alert times are more likely to be generated by advanced sensing with intent prediction in a framework of V2V communications.

A similar Monte Carlo simulation was performed on an intersection turn scenario. In this case, as seen in Figure 6.19, a driver is pulling out into an intersection with the intention of turning left. This driver is unaware of a driver in the oncoming lane who has right of way. The oncoming driver could be alerted in several ways, of a dangerous situation. In the most basic instance, neither vehicle has a LACASA system, so they may not notice the situation until too late. If the oncoming driver had an advanced LACASA implementation, with trajectory

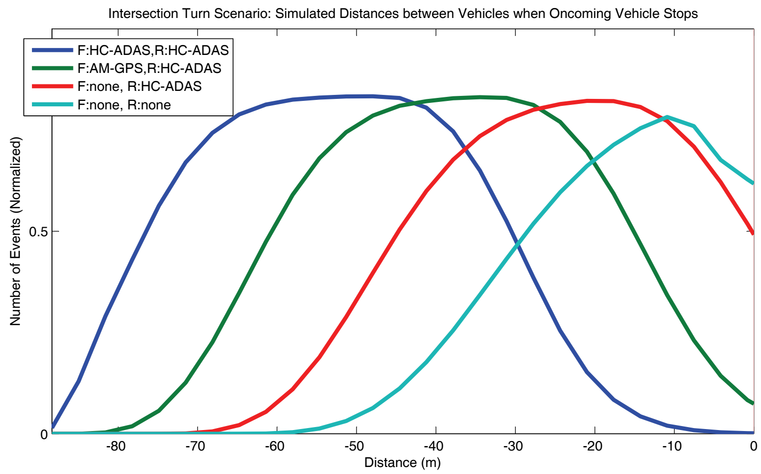


Figure 6.23: Intersection Turn Simulation Results: the distribution of distances at which the oncoming car stops prior to the intersection. A position greater than zero indicates that the oncoming car was in a collision with the turning car; the percentage of cars in accidents can be seen in Table 6.8. Each of the curves represents a simulated scenario in which the oncoming(R) and turning(F) cars have different levels of cooperation.

prediction capabilities, it may be able to detect the car using ACC radar and alert the oncoming driver.

In this case, t_{alert} corresponds to the time of alert, after the turning car has entered the intersection. For each of the four conditions of t_{alert}, we vary the parameters as in the previous simulation. In this case the oncoming car starts to brake with a force between $0.7g$ and $1g$, with a delay corresponding to the given alert time plus a variable reaction time as stated above. The turning car is assumed to take 4 seconds to fully exit the path of the oncoming vehicle, and the simulation ends when the oncoming car comes to a complete stop.

Recent research into turn-intent prediction [30] has shown the ability of an ADAS to predict a driver's intent to turn, 1 to 2 seconds before the turn. Given the turning vehicle in this case with a LACASA system with intent prediction, the system could predict the turning driver's intent, up to 2 seconds in advance of the turn. In the sample situations shown in Table 6.8, a situation where both cars are able to predict and communicate intents shows a complete elimination of collisions, and an 87% reduction in close calls (where the vehicles come within an unsafe distance).

Figure 6.22 shows the percentage of outcomes which correspond to collisions, given the alert time of the ADAS system after the turning car has entered the intersection. The results demonstrate that an earlier alert time, achieved by a combination of advanced sensing and intent prediction, could reduce the number and severity of collisions significantly. Figure 6.23 shows the results of the four different LACASA configurations, for those vehicles that did not get into accidents. A clear advantage can be gained even with limited market penetration of the layered cooperative architecture.

6.3.3 Concluding Remarks

The future of Intelligent Transportation Systems is intertwined with the development and incremental implementation of distributed sensing and communications networks [210]. We have proposed a general, cooperative, holistic, Layered Architecture for Active Safety Applications, LACASA, that can make immediate and significant impacts on safety as part of a stand-alone driver assistance systems in vehicles. A fundamental contribution in the model is a layered framework for active safety which also incorporates a Human-Centric model for Intent and Trajectory Prediction. The framework draws on recent results into the predictability and responsiveness of drivers in various situations, as well as recent improvements in machine vision and artificial intelligence.

The cooperative modularity of the LACASA framework is such that as the market penetration of V2V and V2I sensing and networking improves, the safety benefits of the system grow tangibly as well. In other words, the system does not rely on a specific level of market penetration to see immediate quantitative safety benefits. In one particular application, the lane change warning, a single vehicle with the proposed LACASA framework can reduce the likelihood of collision by 30%, without any V2V communication.

By adding V2V or V2I communication, there is an opportunity to eliminate a significantly larger chunk of collisions. Vehicles may have different sets of sensors, or they may even just be enabled with after-market implementations of driver assistance systems. The layered approach to the integration of information in the LACASA framework allows such diverse ADASs to interoperate seamlessly. “Here-I-am” signals from simple ADASs [210] can provide more accurate positional information to a more advanced LACASA systems in a neighboring vehicle, allowing it to improve its sense of obstacle trajectories and situational awareness. More complex informational signals can be drawn from an integrated, advanced LACASA implementation which incorporates *human-centric* information to get a more accurate prediction of the ego-vehicle’s *intented trajectory.* Transmitting such detailed information could lead, in the case of the lane change warning, to a 70% decrease in the number of “close-call” dangerous situations, over the case when transmitting simple positional information.

6.3.4 Acknowledgements

Chapter 6 is in part a reprint of material that is published in the IEEE Transactions on Systems, Man, and Cybernetics - Part B (2009), by Anup Doshi, Shinko Y Cheng, and Mohan M. Trivedi, and submitted to IEEE Intelligent Transportation Systems Magazine (2010), by Anup Doshi and Mohan M. Trivedi, and published in the IEEE Intelligent Transportation Systems Conference (2010), by Anup Doshi and Mohan M. Trivedi. The dissertation author was the primary investigator and author of these papers.

This research was supported by research grants from the University of California Transportation Center (US Department of Transportation Center of Excellence), the UC Discovery Program, Volkswagen, and the National Science Foundation. The author is grateful for the support and guidance of Dr. Shinko Y. Cheng, along with the advice and support of colleagues from the CVRR laboratory.

Chapter 7

Conclusions

In this dissertation we have tackled the challenges of using data from cameras and other multimodal focused on drivers, vehicles, and the environment, to infer the driver's cognitive states and intent. Several relevant research tasks are covered including behavioral attention analysis and cue selection, data fusion, and model development.

Our contributions range from basic results in how to detect attentive processes from body language, to include analysis of the most relevant internal (driver) and external (environment) cues for intent inference. We demonstrate for the first time significant sequential effects on driver response times and errors, a novel finding in such a complex environment. Based on this analysis of human behavior, we implement and analyze a cost-effective framework for a holistic (joint human-environment), real-time intent inference system on a vehicular testbed. Finally, we study the interactivity and appropriate design of such style-conscious assistance systems. Among these results, we find that aggressive drivers are more predictable, but less responsive to feedback, than cautious drivers.

We have detailed a history and review of the literature related to the prediction of human behaviors, specifically in vehicular environments. In particular, we find that the inference of human intent is a useful cue for vehicular trajectory prediction. One of the most useful indicators of intent is the visual search, and we find first that the interaction of head and eye gaze foretells whether a visual search is premeditated and intent-related. Further, we propose a system to tightly integrate environmental sensing with driver sensing to estimate the driver's attentive state and target. We then take a step deeper into the cognitive state of the driver, to understand how driver behaviors are affected by recent actions. We find in particular that certain sequences of cues or responses tend to prime the driver, significantly altering their response behavior. We are able to observe a consistent contextual effect on response times and errors, or pedal misapplications.

This led us to ask which cues are the most indicative of a driver's intentions, so as to utilize the most robust and effective features and sensors. We found that head pose is indeed more informative than eye-gaze in predicting driver intent. We then explored the real-world performance of such a real-time, cost-conscious intent inference system, with a particular focus on characterizing system performance.

Indeed, style and interactivity play a big role in these systems, leading us to an analysis of different manners of feedback. Using the novel DAD interactive display, we found that certain alerts were effective communicators, where other alerts decreased distraction levels by reducing the time spent looking away from the road. Further, we observe that "aggressive" drivers tend to be more predictable, but less responsive, than "cautious" drivers. We end with a view towards collaborative vehicles of the future, in which the potential for anticipatory communication between vehicles and infrastructure could lead to greatly improved safety.

Significant amount of work remains to be done in implementation and design of driver assistance systems, to cooperate constructively with human operators. However the opportunity to improve active safety and save lives through behavior and intent prediction is extremely promising. Through the use of advanced sensor-based intelligence and interactivity, the next generation of transportation systems will ultimately strive for the goal of accident-free roadways.

Bibliography

[1] Letty Aarts and Ingrid van Schagen. Driving speed and the risk of road crashes: A review. *Accident Analysis & Prevention*, 38(2):215 – 224, 2006.

[2] G Abe and M Itoh. How drivers respond to alarms adapted to their braking behaviour? *Journal of Mechanical Systems for Transportation and Logistics*, 1(3):331–342, 2008.

[3] G. Abe and J. Richardson. Alarm timing, trust and driver expectation for forward collision warning systems. *Applied Ergonomics*, 37(5), 2006.

[4] A. Amditis, H. Kubmann, A. Polychronopoulos, J. Engstrom, and L. Andreone. System architecture for integrated adaptive HMI solutions. *IEEE Intelligent Vehicles Symposium*, 2006.

[5] N. Anderson. Effect of first-order conditional probability in a two-choice learning situation. *Journal of Experimental Psychology*, 59(2):73–93, 1960.

[6] L. Angell, J. Auflick, P. A. Austria, D. Kochhar, L. Tijerina, W. Biever, T. Diptiman, J. Hogsett, and S. Kiger. Driver workload metrics task 2 final report. *Report DOT HS 810635, NHTSA, U.S. Department of Transportation*, Nov 2006.

[7] G. S. Aoude and Jonathan P. How. Using support vector machines and bayesian filtering for classifying agent intentions at road intersections. *Technical Report ACL09-02*, 2009.

[8] Stylianos Asteriadis, Paraskevi Tzouveli, Kostas Karpouzis, and Stefanos Kollias. Estimation of behavioral user state based on eye gaze and head poseapplication in an e-learning environment. *Multimedia Tools and Applications*, 41(3):469–493, February 2009.

[9] S. Ba and J. M. Odobez. Recognizing visual focus of attention from head pose in natural meetings. *IEEE Transactions on Systems, Man, and Cybernetics Part B*, 39(1), 2009.

[10] S. Baker, I. Matthews, J. Xiao, R. Gross, T. Ishikawa, and T. Kanade. Real-time non-rigid driver head tracking for driver mental state estimation. Technical Report 04-10, Carnegie Mellon University Robotics Institute, Feb 2004.

[11] D. H. Ballard and M. M. Hayhoe. Modelling the role of task in the control of gaze. *Visual Cognition*, 17(6):1185–1204, August 2009.

[12] A. E. Bartz. Eye and head movement in peripheral vision: nature of compensatory eye movements. *Science*, 152:1644–1645, 1966.

[13] Jorge P. Batista. A real-time driver visual attention monitoring system. *Iberian Conference on Pattern Recognition and Image Analysis*, 2005.

[14] L. M. Bergasa, J. Nuevo, M. A. Sotelo, R. Barea, and M. E. Lopez. Realtime system for monitoring driver vigilance. *IEEE Transactions on Intelligent Transportation Systems*, 7:63–77, March 2006.

[15] H Berndt, J Emmert, and K Dietmayer. Continuous driver intent recognition with hidden markov models. *in Proceedings of the IEEE Intelligent Transportation Systems Conference*, pages 1189–1194, 2008.

[16] V. D. Bhise, J. Meldrum, D. Jack, G. Troell, D. Hoffmeister, and L. Forbes. Driver head movements in left outside mirror viewing. *Society of Autmotive Engineers*, -(810761), 1981.

[17] Alessandra Bianchi and Heikki Summala. The "genetics" of driving behavior: parents' driving style predicts their children's driving style. *Accident Analysis & Prevention*, 36(4):655 – 659, 2004.

[18] E. Bizzi, R. E. Kalil, and P. Morasso. Two modes of active eye-head coordination in monkeys. *Brain Research*, 40:45–48, 1972.

[19] K. Bock and Z. M. Griffin. The persistence of structural priming: Transient activation or implicit learning? *Journal of Experimental Psychology: General*, 129:177–192, 2000.

[20] W Bouslimi, M Kassaagi, D Lourdeaux, and P Fuchs. Augmentd naive bayesian network for driver behavior modeling. *in Proceedings of the IEEE Intelligent Vehicles Symposium*, pages 236–242, 2005.

[21] P. Boyraz, M Acar, and D Kerr. Signal modelling and hidden markov models for driving manoeuvre recognition and driver fault diagnosis in an urban road scenario. *in Proceedings of the IEEE Intelligent Transportation Systems Conference*, pages 987–992, 2007.

[22] Joerg J Breuer, Andreas Faulhaber, Peter Frank, and Stefan Gleissner. Real world safety benefits of brake assistance systems. *NHTSA Paper Number 07-0103*, 2007.

[23] G. O. Burnham, J Seo, and G A Bekey. Identification of human drivers models in car following. *IEEE Transactions on Automatic Control*, 19(6):911–915, 1974.

[24] M Canale and S Malan. Analysis and classification of human driving behaviour in an urban environment. *Cognition, Technology & Work*, 4:197–206, 2002.

[25] Car2Car. Car to car communication consortium.

[26] Kim M. Cardosi. Human factors for air traffic control specialists: A user's manual for your brain. *Report DOT/FAA/AR-99/39, Federal Aviation Administration, U.S. Department of Transportation*, 1999.

[27] Ran Carmi and Laurent Itti. Visual causes versus correlates of attentional selection in dynamic scenes. *Vision Research*, 46(26):4333–4345, December 2006.

[28] M. Cerf, J. Harel, W. Einhauser, and C. Koch. Predicting human gaze using low-level saliency combined with face detection. *Neural Information Processing Systems Conference*, 2007.

[29] Anita Chang. China's massive traffic jam could last for weeks. Associated Press, August 24 2010.

[30] S. Y. Cheng and M. M. Trivedi. Turn intent analysis using body-pose for intelligent driver assistance. *IEEE Pervasive Computing, Special Issue on Intelligent Transportation Systems*, 5(4):28–37, October-December 2006.

[31] Shinko Y. Cheng, Sangho Park, and Mohan M. Trivedi. Multi-spectral and multi-perspective video arrays for driver body tracking and activity analysis. *Computer Vision and Image Understanding* in press, 2007. doi:10.1016/j.cviu.2006.08.010.

[32] Raymond Y Cho, Leigh E Nystrom, Eric T Brown, Andrew D Jones, Todd S Braver, Philip J Holmes, and Jonathan D Cohen. Mechanisms underlying dependencies of performance on stimulus history in a two-alternative forced-choice task. *Cognitive Affective Behavior Neuroscience*, 2(4):283–299, 2002.

[33] SangJo Choi, J H Kim, D G Kwak, P Angkititrakul, and J H L Hansen. Analysis and classification of driver behavior using in-vehicle can-bus information. *in the Biennial Workshop on DSP for In-Vehicle and Mobile Systems*, 2007.

[34] B. D. Corneil and D. P. Munoz. Human eye-head gaze shifts in a distractor task. ii. reduced threshold for initiation of early head movements. *Journal of Neurophysiology*, 86:1406–1421, 1999.

[35] I Dagli, M Borst, and G Breuel. Action recognition and prediction for driver assistance systems using dynamic belief networks. *in Proceedings of the Agent Technology Workshops*, pages 179–194, 2003.

[36] I Dagli and D Reichardt. Motivation-based approach to behavior prediction. *in Proceedings of the IEEE Intelligent Vehicles Symposium*, 2002.

[37] A Demcenko, M Tamosiunaite, A Vidugiriene, and L Jakevicius. Lane marker parameters for vehicles steering signal prediction. *WSEAS Transactions on Systems*, 2(8):251–260, 2009.

[38] G DiLorenzo, S Phithakkinukoon, and C Ratti. Context-aware navigation: improving urban living experience with predictive navigation systems. *in Proceedings of the UBI Challenge Workshop*, 2010.

[39] P. Dixon and S. Glover. Action and memory. In B.H. Ross, editor, *The psychology of learning and motivation*, pages 143–174. Elsevier, San Diego, CA, 2004.

[40] P. Dixon and S. Glover. Perseveration and contrast effects in grasping. *Neuropsychologia*, 47(6):1578–1584, May 2009.

[41] R. Dodge. The latent time of compensatory eye movements. *Journal of Experimental Psychology*, 4:247–269, 1921.

[42] U Dogan, H Edelbrunner, and I Iossifidis. Towards a driver model: Preliminary study of lane change behavior. *in Proceedings of the IEEE Intelligent Transportation Systems Conference*, pages 931–938, 2008.

[43] B. Donmez, L. Ng Boyle, J.D. Lee, and D.V. McGehee. Drivers' attitudes toward imperfect distraction mitigation strategies. *Transportation Research Part F: Psychology and Behaviour*, 9(6):387–398, 2006.

[44] A. Doshi, Shinko Y. Cheng, and Mohan M. Trivedi. A novel, active heads-up display for driver assistance. *IEEE Transactions on Systems, Man, and Cybernetics Part B*, 39(1), 2009.

[45] A. Doshi and M. M. Trivedi. A comparative exploration of eye gaze and head motion cues for lane change intent prediction. *IEEE Intelligent Vehicles Symposium*, June 2008.

[46] A. Doshi and M. M. Trivedi. Examining the impact of driving style on the predictability and responsiveness of the driver: Real-world and simulator analysis. *IEEE Intelligent Vehicles Symposium*, 2010.

[47] A. Doshi and Mohan M. Trivedi. Head and gaze dynamics in visual attention and context learning. *CVPR Workshop on Visual and Contextual Learning (VCL)*, 2009.

[48] A. Doshi and Mohan M. Trivedi. On the roles of eye gaze and head dynamics in predicting driver's intent to change lanes. *IEEE Transactions on Intelligent Transportations Systems*, 10(3), 2009.

[49] Anup Doshi, B T Morris, and M M Trivedi. Context-sensitive prediction of intentions for on-road driver assistance. *submitted to IEEE Pervasive Computing, Special Issue on Automotive Pervasive Computing*, 2010.

[50] Anup Doshi and Mohan M. Trivedi. Investigating the relationships between gaze patterns, dynamic vehicle surround analysis, and driver intentions. *IEEE Intelligent Vehicles Symposium*, June 2009.

[51] Anup Doshi and Mohan M. Trivedi. Attention estimation by simultaneous observation of viewer and view. *IEEE Workshop on Computer Vision and Pattern Recognition for Human Communicative Behavior Analysis (CVPR4HB)*, June 2010.

[52] Paula J. Durlach. Change blindness and its implications for complex monitoring and control systems design and operator training. *Human-Computer Interaction*, 19:423–451, 2004.

[53] Miguel P. Eckstein, Avi Caspi, Brent R. Beutter, and Binh T. Pham. The decoupling of attention and eye movements during multiple fixation search. *Journal of Vision*, 4(8):165–165, 8 2004.

[54] W. Estes. Toward a statistical theory of learning. *Psychological Review*, 57:94–107, 1950.

[55] L. Evans. *Traffic Safety and the Driver.* Van Nostrand Reinhold, New York, 1991.

[56] I Fallon and D O'Neill. The world's first automobile fatality. *Accident Analysis & Prevention*, 37:601–603, July 2005.

[57] Paula Finnegan and Paul Green. The time to change lanes: A literature review. Technical Report 90-13, University of Michigan Intelligent Vehicle-Highway Systems (IVHS), Sept 1990.

[58] P H Foo, G W Ng, K H Ng, and R Yang. Application of intent inference for surveillance and conformance monitoring to aid human cognition. *in Proceedings of the 10th International Conference on Information Fusion*, 2007.

[59] Edward G. Freedman. Coordination of the eyes and head during visual orienting. *Experimental Brain Research*, 190(4):369–387, October 2008.

[60] J Froelich and J Krumm. Route prediction from trip observations. *in Proceedings of the SAE World Congress & Exhibition*, 2008.

[61] J. H. Fuller. Comparison of head movement strategies among mammals. In A. Berthoz, W. Graf, and P. P. Vidal, editors, *Head-neck sensory-motor system*, pages 101–114, New York, 1992. Oxford Press.

[62] Tarak Gandhi and Mohan M. Trivedi. Vehicle surround capture: Survey of techniques and a novel omni video based approach for dynamic panoramic surround maps. *IEEE Transactions on Intelligent Transportation Systems*, Sept. 2006.

[63] Tarak Gandhi and Mohan M. Trivedi. Computer vision and machine learning for enhancing pedestrian safety. *Computational Intelligence in Automotive Applications of Studies in Computational Intelligence*, 132:59–77, May 2008.

[64] A Gerdes. Driving manoeuvre recognition. *in DLR Electronic Lirary*, 2006.

[65] H. H. L. M. Goossens and A. J. Van Opstal. Human eye-head coordination in two dimensions under different sensorimotor conditions. *Experimental Brain Research*, 114(3):542–560, May 1997.

[66] P. Griffiths and R.B. Gillespie. Sharing control between human and automation using haptic interface: Primary and secondary task performance benefits. *Human Factors*, 47(3):574–590, 2005.

[67] J Gunnarsson, L Svensson, F Bengtsson, and L Danielsson. Joint driver intention classification and tracking of vehicles. *in Proceedings of the IEEE Nonlinear Statistical Signal Processing Workshop*, pages 95–98, 2006.

[68] F Han, Y Tan, and J Eledath. Preceding vehicle trajectory prediction by multi-cue integration. *in Proceedings of the IAPR Conf on Machine Vision Applications*, pages 575–578, 2007.

[69] R. Hanowski, W. W. Wierwille, S. A. Garness, and T. A. Dingus. A field evaluation of safety issues in local/short haul trucking. *Proceedings of the IEA 2000/HFES 2000 Congress*, 2000.

[70] J. Harel, C. Koch, and P. Perona. Graph-based visual saliency. *Neural Information Processing Systems Conference*, 2006.

[71] K Hayashi, Y Kojima, K Abe, and K Oguri. Prediction of stopping maneuver considering driver's state. *in Proceedings of the IEEE Intelligent Transportation Systems Conference*, pages 1191–1196, 2006.

[72] James R. Healey and Sharon Silke Carty. Driver error found in some toyota acceleration cases. *USA Today*, July 14 2010.

[73] H. V. Helmholtz and J. P. C. Southall. Helmholtz's treatise on psychological optics, 3. *The Optical Society of America, Rochester, NY*, 1924.

[74] John M. Henderson, James R. Brockmole, Monica S. Castelhano, and Michael Mack. Visual saliency does not account for eye movements during visual search in real-world scenes. In Roger P.G. Van Gompel, Martin H. Fischer, Wayne S. Murray, and Robin L. Hill, editors, *Eye Movements*, pages 537 – 562. Elsevier, Oxford, 2007.

[75] M. J. Henning, O. Georgeon, and J. F. Krems. The quality of behavioral and environmental indicators used to infer the intention to change lanes. *Proceedings of the Fourth Int'l Driving Symposium on Human Factors in Driver Assesment, Training, and Vehicle Design*, July 2007.

[76] A. N. Herst, J. Epelboim, and R. M. Steinman. Temporal coordination of the human head and eye during a natural sequential tapping task. *Vision Research*, 41(25):3307–3319, 2001.

[77] James E. Hoffman. Visual attention and eye movements. In Harold Pashler, editor, *Attention*, chapter 3. Psychology Press, East Sussex, UK, 1998.

[78] Jihua Huang, William C. Lin, and Yuen-kwok Chin. Adaptive vehicle control system with driving style recognition based on vehicle launching, January 2010.

[79] Kohsia S. Huang, Mohan M. Trivedi, and Tarak Gandhi. Driver's view and vehicle surround estimation using omnidirectional video stream. *Proc. IEEE Intelligent Vehicles Symposium*, pages 444–449, 2003.

[80] T Hulnhagen, I Dengler, A Tamke, T Dang, and G Breuel. Maneuver recognition using probabilistic finite-state machines and fuzzy logic. *in Proceedings of the IEEE Intelligent Vehicles Symposium*, 2010.

[81] M. Itoh, K. Yoshimura, and T. Inagaki. Inference of large trucke driver's intent to change lanes to pass a lead vehicle via analyses of driver's eye glance behavior in the real world. *Proceedings of the SICE Annual Conference*, Sept. 2007.

[82] L. Itti and C. Koch. A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, 40:1489–1506, 2000.

[83] Laurent Itti. Quantifying the contribution of low-level saliency to human eye movements in dynamic scenes. *Visual Cognition*, 12(6):1093–1123, 2005.

[84] Laurent Itti and Pierre Baldi. Bayesian surprise attracts human attention. *Vision Research*, 49:1295–1306, 2009.

[85] L. James and D. Nahl. *Road Rage and Aggressive Driving: Steering Clear of Highway Warfare*. Prometheus Books, Amherst, NY, 2000.

[86] I Jentzsch and W Sommer. Functional localization and mechanisms of sequential effects in serial reaction time tasks. *Perception and Psychophysics*, 64(7):1169–1188, 2002.

[87] J Jerhot, M-M Meinecke, T Form, T-N Nguyen, G Stanek, and J Knaup. Integrated probabilistic approach to environmental perception with self-diagnosis capability for advanced driver assistance systems. *in Proceedings of the 12th International Conference on Information Fusion*, pages 1347–1350, 2009.

[88] W Jesteadt, R D Luce, and D M Green. Sequential effects in judgments of loudness. *Journal of Experimental Psychology Human Perception and Performance*, 3(1):92–104, 1977.

[89] Qiang Ji and Xiaojie Yang. Real-time eye, gaze, and face pose tracking for monitoring driver vigilance. *Real-Time Imaging*, 8(5):357–377, Oct 2002.

[90] M Jones, T Curran, M Mozer, and M H Wilder. Sequential effects reflect learning of temporal structure. *Presented at the 50th Annual Meeting of the Psychonomic Society*, November 2009.

[91] Jelena Jovancevic, Brian Sullivan, and Mary Hayhoe. Control of attention and gaze in complex environments. *Journal of Vision*, 6:1431–1450, 2006.

[92] A Karbassi and M Barth. Vehicle route prediction and time of arrival estimation techniques for improved transportation system management. *in Proceedings of the IEEE Intelligent Vehicles Symposium*, pages 511–516, 2003.

[93] W Khaisongkram, P Raksincharoensak, M Shimosaka, T Mori, T Sato, and M Nagai. Automobile driver behavior recognition using boosting sequential labeling method for adaptive driver assistance systems. *in Proceedings of the 31st Annual German Conference on AI*, pages 103–110, 2008.

[94] Aarlenne Z. Khan, Gunnar Blohm, Robert M. McPeek, and Philippe Lefvre. Differential influence of attention on gaze and head movements. *Journal of Neurophysiology*, 101:198–206, 2009.

[95] C. N. Kloeden, A. J. McLean, V. M. Moore, and G. Ponte. Travelling speed and the risk of crash involvement, volume 1 - findings. Technical report, NHMRC Road Accident Research Unit, The University of Adelaide, November 1997.

[96] Arni Kristjansson. Rapid learning in attention shifts: A review. *Visual Cognition*, 13(3):324–362, 2006.

[97] J Krumm. A markove model for driver turn prediction. *in Proceedings of the SAE 2008 World Congress*, 2008.

[98] J Krumm and E Horvitz. Predestination: Inferring destinations from partial trajectories. *in Proceedings of the Eighth International Conference on Ubiquitous Computing*, pages 243–260, 2006.

[99] N. Kuge, T. Yamamura, O. Shimoyama, and A. Liu. A driver behavior recognition method based on a driver model framework. *Proceedings of the Society of Automotive Engineers World Congress*, 2000.

[100] T Kumegai, Y Sakaguchi, M Okuwa, and M Akamatsu. Prediction of driving behavior through probabilitsic inference. *in Proceedings of the Eight International Conference on Engineering Applications of Neural Networks*, 2003.

[101] Gerhard Kurz, Armin Mller, Thomas Rhrig-Gericke, Reinhold Schb, Harry Trster, and Andy Yap. Method and device for classifying the driving style of a driver in a motor vehicle, September 2002.

[102] Michael F. Land. Predictable eye-head coordination during driving. *Nature*, 359:318–320, September 1992.

[103] Steven J. Landry, T. B. Sheridan, and Yan M. Yufik. A methodology for studying cognitive groupings in a target-tracking task. *IEEE Transactions on Intelligent Transportation Systems*, 2(2), June 2001.

[104] O. Le Meur, P. Le Callet, D. Barba, and D. Thoreau. A coherent computational approach to model bottom-up visual attention. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(5):802–817, May 2006.

[105] J. D. Lee, J. D. Li, L. C. Liu, and C. M. Chen. A novel driving pattern recognition and status monitoring system. *Pacific-Rim Symposium on Image and Video Technology (PSIVT)*, pages 504–512, 2006.

[106] S. E. Lee, C. B. Olsen, and W. W. Wierwille. A comprehensive examination of naturalistic lane changes. *Report DOT HS 809702, NHTSA, U.S. Department of Transportation*, March 2004.

[107] Y. C. Lee, J. D. Lee, and L. N. Boyle. Visual attention in driving: the effects of cognitive load and visual disruption. *Human Factors*, 49(4):721–733, 2007.

[108] J. Levy and H. Pashler. Task prioritization in multitasking during driving: Opportunity to abort a concurrent task does not insulate braking responses from dual-task slowing. *Applied Cognitive Psychology*, 22:507–525, 2008.

[109] J. Levy, H. Pashler, and E. Boer. Central interference in driving: Is there any stopping the psychological refractory period? *Psychological Science*, 17(3):228–235, 2006.

[110] Yulan Liang, Michelle L. Reyes, and John D. Lee. Real-time detection of driver cognitive distraction using svm. *IEEE Transactions on Intelligent Transportation Systems*, 8(2):340–350, June 2007.

[111] X. Liu, F. Xu, and K. Fujimura. Real-time eye detection and tracking for driver observation under various light conditions. In *IEEE Intelligent Vehicles Symposium*, pages 18–20, June 2002.

[112] X. Liu, F. Xu, and K. Fujimura. Real-time eye detection and tracking for driver observation under various light conditions. In *Proceedings of IEEE Intelligent Vehicle Symposium*, pages 18–20, June 2002.

[113] Yung-Ching Liu and Min-Hui Wen. Comparison of head-up display (HUD) vs. head-down display (HDD): driving performance of commercial vehicle operators in Taiwan. *Int. J. of Human-Computer Studies*, 61(5):679–697, 2004.

[114] C C Macadam. Understanding and modeling the human driver. *Vehicle System Dynamics*, 40(1):101–134, 2003.

[115] Vijay Mahadevan and Nuno Vasconcelos. Spatiotemporal saliency in dynamic scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(1):171–177, Jan 2010.

[116] L. T. Maloney, M. F. Dal Martello, C. Sahm, and L. Spillmann. Past trials influence perception of ambiguous motion quartets through pattern completion. *Proceedings of the National Academy of Sciences*, 102:3164–3169, 2005.

[117] L Malta, C Miyajima, and K Takeda. A study of driver behavior under potential theats in vehicle traffic. *IEEE Transactions on Intelligent Transportation Systems*, 10(2):201–210, 2009.

[118] Y. Matsumoto, T. Ogasawara, and A. Zelinsky. Behavior recognition based on head pose and gaze direction measurement. *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, 2000*, 3:2127–2132 vol.3, 2000.

[119] Joel C. McCall and Mohan M. Trivedi. Video based lane estimation and tracking for driver assistance: Survey, system, and evaluation. *IEEE Transactions on Intelligent Transportation Systems*, 7(1), March 2006.

[120] Joel C. McCall and Mohan M. Trivedi. Driver behavior and situation aware brake assistance for intelligent vehicles. *Proceedings of the IEEE, Special Issue on Advanced Automobile Technoology*, 95(2):374–387, Feb 2007.

[121] Joel C. McCall, David Wipf, Mohan M. Trivedi, and Bhaskar Rao. Lane change intent analysis using robust operators and sparse bayesian learning. *IEEE Transactions on Intelligent Transportation Systems*, Sept. 2007.

[122] N. R. Mennie, M. M. Hayhoe, and B. T. Sullivan. Look-ahead fixations: anticipatory eye movements in natural tasks. *Experimental Brain Research*, 179(3):427–442, May 2007.

[123] N. R. Mennie, M. M. Hayhoe, B. T. Sullivan, and C. Walthew. Look ahead fixations and visuo-motor planning. *Journal of Vision*, 3(9), October 2003.

[124] H Mima, K Ikeda, T Shibata, N Fukaya, K Hitomi, and T Bando. Estimation of driving phase by modeling brake pressure signals. *in Proceedings of the 16th International Conference on Neural Information Processing: Part 1*, pages 468–475, 2009.

[125] D Mitrovic. Experiments inj subsymbolic driving pattern prediction. *in Proceedings of the 6th International Conference on Neural Information Processing*, pages 673–678, 1999.

[126] D Mitrovic. Machine learning for car navigation. *in Proceedings of the 14th International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems*, pages 670–675, 2001.

[127] D Mitrovic. Reliable method for driving events recognition. *IEEE Transactions on Intelligent Transportation Systems*, 6(2):198–205, 2005.

[128] Pietro Morasso, Giulio Sandini, Vincenzo Tagliasco, and Renato Zaccaria. Control strategies in the eye-head coordination system. *IEEE Transactions on Systems, Man and Cybernetics*, 7(9):639–651, Sept. 1977.

[129] Brendan Morris and Mohan M. Trivedi. Learning, modeling, and classification of vehicle track patterns from live video. *IEEE Transactions on Intelligent Transportation Systems*, 9(3), September 2008.

[130] Brendan Morris and Mohan M. Trivedi. Unsupervised learning of motion patterns of rear surrounding vehicles. *IEEE International Conference on Vehicular Electronics and Safety*, November 2009.

[131] Brendan Morris and Mohan M. Trivedi. VECTOR: Trajectory analysis for advanced highway monitoring. *Presented at ITS America's Annual Meeting*, June 2009.

[132] Brendan Morris and Mohan M. Trivedi. Vehicle iconic surround observer: Visualization platform for intelligent driver support applications. *IEEE Intelligent Vehicles Symposium*, June 2010.

[133] R. R. Mourant and R. J. Donahue. Mirror sampling characteristics of drivers. *Society of Automotive Engineers*, -(740964), 1974.

[134] M. C. Mozer, S. Kinoshita, and M. Shettel. Sequential dependencies offer insight into cognitive control. In W. Gray, editor, *Integrated Models of Cognitive Systems*, pages 180–193. Oxford University Press, Oxford, 2007.

[135] M. C. Mozer and M. Sitton. Computational modeling of spatial attention. In Harold Pashler, editor, *Attention*, chapter 3. Psychology Press, East Sussex, UK, 1998.

[136] Y L Murphey, R Milton, and L Kiliaris. Driver's style classification using jerk analysis. *IEEE Workshop on Computational Intelligence in Vehicles and Vehicular Systems*, pages 23 –28, April 2009.

[137] Erik Murphy-Chutorian, Anup Doshi, and Mohan M. Trivedi. Head pose estimation for driver assistance systems: A robust algorithm and experimental evaluation. *IEEE International Transportation Systems Conference*, Sept 2007.

[138] Erik Murphy-Chutorian and Mohan M. Trivedi. 3d tracking and dynamic analysis of human head movements and attentional targets. *IEEE/ACM International Conference on Distributed Smart Cameras*, June 2008.

[139] Erik Murphy-Chutorian and Mohan M. Trivedi. Head pose estimation in computer vision: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(4):607–626, April 2009.

[140] Erik Murphy-Chutorian and Mohan M. Trivedi. Head pose estimation and augmented reality tracking: An integrated system and evaluation for monitoring driver awareness. *IEEE Transactions on Intelligent Transportation Systems*, 11(2):300–311, June 2010.

[141] S. Nagirir, Y. Amano, K. Fukui, and S. Doi. A study of a personally adaptive driving support system using a driving simulator. *R&D Review of Toyota CRDL*, 39(2), 2004.

[142] A Nakano, H Okuda, T Suzuki, S Inagaki, and S Hayakawa. Symbolic modeling of driving behavior based on hierarchical segmentation and formal grammar. *in Proceedings of the International Conference on Intelligent Robots and Systems*, pages 5516–5521, 2009.

[143] National Highway Traffic Safety Administration. Traffic safety facts 2008.

[144] National Highway Traffic Safety Administration. Vehicle safety communications project task 3 final report: Identify intelligent vehicle safety applications enabled by DSRC, March 2005.

[145] National Highway Traffic Safety Administration. Fatal Analysis Reporting System (FARS). *available online - www-fars.nhtsa.dot.gov*, 2010.

[146] Vidhya Navalpakkam and Laurent Itti. Modeling the influence of task on attention. *Vision Research*, 45:205–231, 2005.

[147] NHTSA. Vehicle backover avoidance technology study: Report to congress. Technical Report NHTSA-2006-25579, U.S. Department of Transportation, National Highway Traffic Safety Administration, November 2006.

[148] NHTSA National Center for Statistics and Analysis. Traffic safety facts: 2005 data - speeding. Technical Report DOT HS 810 629, U.S. Department of Transportation, National Highway Traffic Safety Administration, 2006.

[149] I Nizetic, K Fertalj, and D Kalpic. A prototype for the short-term prediction of moving object's movement using markov chains. *in Proceedings of the International Conference on Information Technology Interfaces*, pages 559–564, 2009.

[150] K Ohashi, T Yamaguchi, and I Tamai. Humane automotive system using driver intention recognition. *in Proceedings of the SICE Annual Conference*, 2004.

[151] B.R Okombi-Diba, M. Okuwa, Y. Uchiyama, A. Kozato, and T. Hongo. Evaluation of a night driver support system: driver eye movement behavior. *SICE 2004 Annual Conference*, 1(4), August 2004.

[152] N Oliver and A P Pentland. Driver behavior recognition and prediction in a smartcar. *in SPIE Enhanced and Synthetic Vision*, 4023:280–290, 2000.

[153] N. Oliver and A. P. Pentland. Graphical models for driver behavior recognition in a smartcar. *IEEE Intelligent Vehicles Symposium*, 2000.

[154] N.M. Oliver, B. Rosario, and A.P. Pentland. A bayesian computer vision system for modeling human interactions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):831–843, Aug 2000.

[155] Erik C. B. Olsen, Suzanne E. Lee, and Walter W. Wierwille. Eye glance behavior during lane changes and straight-ahead driving. *Transportation Research Record*, 1937:44–50, 2005.

[156] M. Pantic, A. Pentland, A. Nijholt, and T.S. Huang. Human computing and machine understanding of human behavior: A survey. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 4451 LNAI:47–71, 2007.

[157] Harold Pashler, James C. Johnston, and Eric Ruthruff. Attention and performance. *Annual Review of Psychology*, 52(1):629–651, 2001.

[158] Margie Peden, Richard Scurfield, David Sleet, Dinesh Mohan, Adnan A. Hyder, Eva Jarawan, and Colin Mathers, editors. *World report on road traffic injury prevention*. World Health Organization, Geneva, 2004.

[159] A Pellacchia, C Igel, J Edelbrunner, and G Schoner. Making driver modeling attractive. *IEEE Intelligent Systems*, 20(2):8–12, 2005.

[160] J. Pelz, M. Hayhoe, and R. Loeber. The coordination of eye, head, and hand movements in a natural task. *Exp Brain Res.*, 139(3):266–277, August 2001.

[161] A. Pentland. Social signal processing [Exploratory DSP]. *IEEE Signal Processing*, 24(4):108 –111, July 2007.

[162] Alex Pentland and Andrew Liu. Modeling and prediction of human behavior. *Neural Computation*, 11:229–242, 1999.

[163] Robert J. Peters and Laurent Itti. Beyond bottom-up: Incorporating task-dependent influences into a computational model of spatial attention. *IEEE Conference on Computer Vision and Pattern Recognition*, 2006.

[164] L. Petersson, L. Fletcher, and A. Zelinsky. A framework for driver-in-the-loop driver assistance systems. *IEEE Transactions on Intelligent Transportation Systems*, September 2005.

[165] M Plochl and J Edelmann. Driver models in automobile dynamics application. *Vehicle System Dynamics*, 45(7):699–741, 2007.

[166] M. I. Posner and Y. Cohen. Components of visual orienting. *Chapter in Attention and Performance X, Bouma H. and Bouwhuis D., eds*, pages 531–556, 1984.

[167] P Raksincharoensak, T Mizushima, and M Nagai. Direct yaw moment control system based on driver behaviour recognition. *Vehicle System Dynamics*, 46:911–921, 2008.

[168] P Rammelt. Planning in driver models using probabilstic networks. *in Proceedings of the IEEE International Workshop on Robot and Human Interactive Communication*, 2002.

[169] T Ranney. Models of driving behavior: A review of their evolution. *Accident Analysis and Prevention*, 26(6):733–750, 1994.

[170] M. A. Recarte and L. M. Nunes. Effects of verbal and spatial-imagery tasks on eye fixations while driving. *Journal of Experimental Psychology: Applied*, 6(1):31–43, 2000.

[171] R. Remington. Analysis of sequential effects in choice reaction times. *Journal of Experimental Psychology*, 82(2):250–257, 1969.

[172] Shahram Rezaei, Raja Sengupta, Hariharan Krishnan, Xu Guan, and Raman Bhatia. Tracking the position of neighboring vehicles using wireless communications. *Transportation Research Part C: Emerging Technologies*, In Press:–, 2009.

[173] Matthew Rizzo and Ida L. Kellison. Eyes, Brains, and Autos. *Arch Ophthalmol*, 122:641–645, Apr. 2004.

[174] Matthew Rizzo and Ida L. Kellison. Eyes, Brains, and Autos. *Arch Ophthalmol*, 122:641–645, Apr. 2004.

[175] G. H. Robinson, D. Erickson, G. Thurston, and R. Clark. Visual search by automobile drivers. *Human Factors*, 14(4):315–323, 1972.

[176] S. Ron, A. Berthoz, and S. Gur. Saccade vestibuloocular reflex cooperation and eye head uncoupling during orientation to flashed target. *Journal of Physiology*, 464:595–611, 1993.

[177] A. L. Rothenstein and J. K. Tsotsos. Attention links sensing to recognition. *Image and Vision Computing*, 26:114–126, 2008.

[178] Constantin A. Rothkopf, Dana Ballard, and Mary Hayhoe. Task and context determine where you look. *Journal of Vision*, 7(14):1–20, 2007.

[179] M. S. Ryoo and J. K. Aggarwal. Hierarchical recognition of human activities interacting with objects. *IEEE Conference on Computer Vision and Pattern Recognition*, 0:1–8, 2007.

[180] D. D. Salvucci. Inferring driver intent: A case study in lane-change detection. In *Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting*, pages 2228–2231, Santa Monica, CA, 2004. Human Factors and Ergonomics Society.

[181] D D Salvucci. Modeling driver behavior in a cognitive architecture. *Human Factors*, 48(2):362–380, 2006.

[182] Dario D. Salvucci and Andrew Liu. The time course of a lane change: Driver control and eye-movement behavior. *Transportation Research Part F: Traffic Psychology and Behavior*, 5(2):123–132, June 2002.

[183] Dario D. Salvucci, Hiren M. Mandalia, Noboyuki Kuge, and Tomohiro Yamamura. Lane-change detection using a computational driver model. *Human Factors*, 49(3), June 2007.

[184] A Sathyanarayana, P Boyraz, Z Purohit, R Lubag, and J H L Hansen. Driver adaptive and context aware active safety systems using can-bus signals. *in Proceedings of the IEEE Intelligent Vehicles Symposium*, 2010.

[185] T Sato and M Akamatsu. Modeling and prediction of driver preparations for making a right turn based on vehicle velocity and traffic conditions while approaching an intersection. *Transportation Research Part F*, 11:242–258, 2008.

[186] T Sato and M Akamatsu. Analysis of drivers' preparatory behaviour before turning at intersections. *IET Intelligent Transportation Systems*, 3(4):379–389, 2009.

[187] T W Schaap, A R A van der Horst, and B van Arem. Influence of unexpected events on driving behavior at different hierarchical levels: a driving simulator experiment. *in Proceedings of the European Conference on Human Design for Intelligent Transport Systems*, 2008.

[188] Seeing Machines. facelab head and eye tracker.

[189] T. Sharon, T. Selker, L. Wagner, and A. Frank. Carcoach: A generalized layered architecture for educational car systems. *Proceedings of the IEEE international Conference on Software - Science, Technology & Engineering*, February 2005.

[190] R Simmons, B Browning, Y Zhang, and V Sadekar. Learning to predict driver route and destination intent. *in Proceedings of the IEEE Intelligent Transportation Systems Conference*, 2006.

[191] Matthew Smith and Harry Zhang. Safety vehicles using adaptive interface technology (task 8): A literature review of intent inference. *SAVE-IT Phase 1 Final Report, http://www.volpe.dot.gov/hf/roadway/saveit/*, 2004.

[192] R. Smith, M. Shah, and N. da Vitoria Lobo. Determining driver visual attention with one camera. *IEEE Transactions on Intelligent Transportation Systems*, 4:205–218, December 2003.

[193] David L. Strayer and Frank A. Davis. Cell-phone-induced driver distraction. *Current Directions in Psychological Science*, 16(3):128–131, 2007.

[194] J Sun and K He. Statistic-based context recognition in smart car. *in Proceedings of the 4th European conference on Smart sensing and context*, 2009.

[195] K. Takemura, J. Ido, Y. Matsumoto, and T. Ogasawara. Drive monitoring system based on non-contact measurement system of driver's focus of visual attention. *Proceedings of the IEEE Intelligent Vehicles Symposium*, June 2003.

[196] K Tanaka, Y Kishino, T Terada, and S Nishio. A destination prediction method using driving contexts and trajectory for car navigation systems. *in Proceedings of the ACM Symposium on Applied Computing*, 2009.

[197] Orit Taubman-Ben-Ari, Mario Mikulincer, and Omri Gillath. From parents to children– similarity in parents and offspring driving styles. *Transportation Research Part F: Traffic Psychology and Behaviour*, 8(1):19 – 29, 2005.

[198] L. Tijerina and S. Hetrick. Analytical evaluation of the effectiveness of minimum separation distance and turn-signal onset rules for lane change crash avoidance system warning onset. *Human Factors and Ergonomics Society Annual Meeting Proceedings, Surface Transportation*, 5:949–953, 1997.

[199] Louis Tijerina, W. Riley Garrott, Duane Stoltzfus, and Edwin Parmer. Eye glance behavior of van and passenger car drivers during lane change decision phase. *Transportation Research Record*, 1937:37–43, 2005.

[200] M. E. Tipping. Sparse bayesian learning and the relevance vector machine. *Journal of Machine Learning Research*, 1:211–244, 2001.

[201] M. E. Tipping and A. C. Faul. Fast marginal likelihood maximisation for sparse bayesian models. *Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics*, Jan 3-6 2003.

[202] T Toledo, H N Koutsopoulos, and M Ben-Akiva. Estimation of an integrated driving behavior model. *Transportation Research Part C*, 17, 2009.

[203] R Toledo-Moreo, M Pinzolas-Prado, and J M Cano-Izquierdo. Maneuver prediction for road vehicles based on a neuro-fuzzy architecture with a low-cost navigation unit. *IEEE Transactions on Intelligent Transportation Systems*, 11(2):498–504, 2010.

[204] TORCS. The open racing car simulator, 2010.

[205] K Torkkola, K Zhang, H Li, H Zhang, C Schreiner, and M Gardner. Traffic advisories based on route prediction. *in Workshop on Mobile Interaction with the Real World*, 2007.

[206] Transportation Research Board. Strategic highway research program (SHRP 2): Overview and safety research plan. *Available online: www.TRB.org/SHRP2*, 2009.

[207] Mohan M. Trivedi, Tarak Gandhi, and Joel C. McCall. Looking in and looking out of a vehicle: Computer vision-based enhanced vehicle safety. *IEEE Transactions on Intelligent Transportation Systems*, January 2007.

[208] Mohan Manubhai Trivedi and Shinko Yuanhsien Cheng. Holistic sensing and active displays for intelligent driver support systems. *IEEE Computer, Special Issue on Human-Centered Computing*, May 2007.

[209] US Department of Transportation. Intellidrive.

[210] US DOT Research and Innovative Technology Administration. ITS strategic research plan, 2010-2014: Executive summary, 2010.

[211] Paul Viola and Michael J. Jones. Rapid object detection using a boosted cascade of simple features. *IEEE Conference on Computer Vision and Pattern Recognition*, 2001.

[212] P. Wang, B. G. Matthew, J. Qiang, and W. James Wayman. Automatic eye detection and its validation. In *Proceedings of the IEEE Workshop on Face Recognition Grand Challenge Experiments (in Conjunction With CVPR)*, June 2005.

[213] Q Wang, Jingyu Y, M Ren, and Y Zheng. Driver fatigue detection: A survey. *in Proceedings of the 6th World Congress on Intelligent Control and Automation*, 2006.

[214] H. Watanabe, H. Yoo, O. Tsimhoni, and P. Green. The effect of HUD warning location on driver responses. In *International Transportation Systems World Congress*, pages 1–10, 1999.

[215] M Wilder, A Ahmed, M Mozer, and M Jones. Sequential dependencies in motor tasks. *(in preparation)*, 2010.

[216] M. Wilder, M. Jones, and M. C. Mozer. Sequential effects reflect parallel learning of multiple environmental regularities. In Y. Bengio, D. Schuurmans, J. Lafferty, C.K.I. Williams, and A. Culotta, editors, *Advances in Neural Information Processing Systems*, volume 22, pages 2053–2061. NIPS Foundation, La Jolla, CA, 2010.

[217] Andrew D. Wiles, David G. Thompson, and Donald D. Frantz. Accuracy assessment and interpretation for optical tracking systems. *SPIE Proceedings on Medical Imaging, Visualization, Image-Guided Procedures, and Display*, -(5367), 2004.

[218] M. Wittman, M. Kiss, P. Gugg, A. Steffen, M. Fink, E. Poppel, and H. Kamiya. Effects of display position of a visual in-vehicle task on simulated driving. *Applied Ergonomics*, 37(2), 2006.

[219] J. Wu and M. M. Trivedi. A binary tree for probability learning in eye detection. In *Proceedings of the IEEE Workshop on Face Recognition Grand Challenge Experiments (FRGC) in conjunction with CVPR 2005.*, volume 3, pages 170– 170, 2005.

[220] Junwen Wu and Mohan M. Trivedi. Simultaneous eye tracking and blink detection with interactive particle filters. *EURASIP Journal on Advances in Signal Processing*, Oct 2007.

[221] S. Yantis. Control of visual attention. *in Attention, ed. H. Pashler. 13-74 Psychology Press, Hove, UK.*, 1998.

[222] A. J. Yu and J. D. Cohen. Sequential effects: Superstition or rational behavior? *Advances in Neural Information Processing Systems*, 21:1873–1880, 2009.

[223] W. H. Zangemeister and L. Stark. Types of gaze movement: Variable interactions of eye and head movements. *Experimental Neurology*, -(77):563–577, 1982.

[224] Gregory J. Zelinsky, Wei Zhang, Bing Yu, Xin Chen, and Dimitris Samaras. The role of top-down and bottom-up processes in guiding eye movements during visual search. *Neural Information Processing Systems Conference*, 2005.

[225] L Zhang, M H Tong, T K Marks, H Shan, and G W Cottrell. Sun: A bayesian framework for saliency using natural statistics. *Journal of Vision*, 8(7):1–20, 2008.

[226] H. Zhou, M. Itoh, and T. Inagaki. Detection of lane change intention via analyses of eye and head movement behaviors. *Proc. SICE Symposium on Systems and Information*, pages 231–236, 2006.

[227] H. Zhou, M. Itoh, and T. Inagaki. Influence of cognitively distracting activity on driver;s eye movement during preparation of changing lanes. *Proceedings of the SICE Annual Conference*, Aug. 2008.

[228] Y Zhou, W Xu H Ning, Y Gong, and T S Huang. Detecting unsafe driving patterns using discriminative learning. *in Proceedings of the IEEE International Conference on Multimedia and Expo*, 2007.

[229] Y. Zhu and K. Fujimara. Headpose estimation for driver monitoring. *IEEE Intelligent Vehicles Symposium*, 2004.

[230] B D Ziebart, A Maas, J A Bagnell, and A K Dey. Human behavior modeling with maximum entropy inverse optimal control. *in Proceedings of the AAAI Spring Symposium on Human Behavior Modeling*, 2009.

[231] B D Ziebart, A Maas, A K Dey, and J A Bagnell. Navigate like a cabbie: probabilistic reasoning from observed context-aware behavior. *in Proceedings of the 10th International Conference on Ubiquitous Computing*, 2008.