1-1-2017

Human-Machine Interface Development For Modifying Driver Lane Change Behavior In Manual, Automated, And Shared Control Automated Driving

Walter Joseph Talamonti
Wayne State University,

HUMAN-MACHINE INTERFACE DEVELOPMENT FOR MODIFYING DRIVER LANE CHANGE BEHAVIOR IN MANUAL, AUTOMATED, AND SHARED CONTROL AUTOMATED DRIVING

by

WALTER JOSEPH TALAMONTI

DISSERTATION

Submitted to the Graduate School

of Wayne State University,

Detroit, Michigan

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

2017

MAJOR: INDUSTRIAL ENGINEERING

Approved By:

____

Advisor Date

____

____

____

DEDICATION

To my parents

ACKNOWLEDGMENTS

Conducting the research for this dissertation has been a truly humbling experience. It is without a doubt that the accomplishments described herein depended on the support of a much broader team that deserves mention.

First, I would like to thank my adviser, Dr. Ellis, for his willingness to take me on as your PhD student without knowing my character or work prior. Your continued guidance throughout my dissertation, along with additional research not originally planned, created unique experiences and taught many unexpected lessons. Your academic advice and guidance through this journey will carry on into my future endeavors. Our next discussion at CBW can carry a lighter note, or should I say a 'hoppy' one; the next Saison is on me.

To my WSU Committee members - thank you for taking the time to be on my committee, shaping my educational course, and making me a better scholar.

To my external committee member, Louis, you take on many names - dissertation committee member, supervisor, mentor, and friend. Your unwavering support and technical guidance challenged me to think differently, overcome fears, and grow as a Research Engineer. Thank you for the countless hours of help and mentorship. I accomplished a personal dream, something that would not have been possible without you.

The tremendous effort and the time dedicated by the Ford Motor Company VIRTTEX driving simulator team at the Ford Research and Innovation Center (RIC) made these experimental results possible. Many thanks to Mike Blommer, Reates Curry, and Radhakrishnan Swaminathan. Without this team's assistance in designing and executing each study as well as recruiting over 277 participants, this study and research would not be viable.

Thanks to my family for their years of support and countless edits.

A special thanks to my wife, Sandy. Thank you for your unwavering support, patience, and love and for all the sacrifices you have made on my behalf. Words cannot express how grateful I am.

Without all of the aforementioned individuals' contributions, this research would not have been possible.

TABLE OF CONTENTS

DEDICATION .... **ii**

ACKNOWLEDGMENTS .... **iii**

LIST OF TABLES .... **viii**

LIST OF FIGURES .... **ix**

LIST OF ACRONYMS .... **xi**

CHAPTER 1 INTRODUCTION **1**

1.1 Motivation 1

1.1.1 Rear-End Crashes 1

1.1.2 Steering versus Braking Response 4

1.1.3 Steering Behavior 5

1.2 Evasive Steer Assist (ESA) 6

1.3 Dissertation Contributions 6

1.4 Dissertation Outline 7

1.4.1 Study 1: Initial Problem Characterization: Driver Response to Sudden Forward Collision Events in Manual and Fully Autonomous Driving (see Chapter 2) 7

1.4.2 Study 2: Further Problem Characterization Study (see Chapter 3) 7

1.4.3 HMI Concepts (see Chapter 4) 8

1.4.4 Study 3, Part 1: Driver Haptic Steering Alerts and Urgency Scaling (see Chapter 5) ... 8

1.4.5 Study 3, Part 2: Driver Haptic Steering Alerts, “Mirage Events”, and Multi-Attribute Utility Theory: A Novel Method for Selecting an Optimal HMI (see Chapter 6) 9

1.4.6 Study 4: Screening Study to Evaluate Factors Influencing HMI Effectiveness (see Chapter 7) 9

1.4.7 Study 5: HMI Effects in Fully Automated versus Shared Control Automated Driving (see Chapter 8) 10

1.4.8 Conclusions and Recommendations for Future Research 11

CHAPTER 2 STUDY 1: DRIVER RESPONSE TO SUDDEN FORWARD COLLISION EVENTS IN MANUAL AND FULLY AUTONOMOUS DRIVING—INITIAL PROBLEM CHARACTERIZATION STUDY **12**

2.1 Abstract 12

2.2 Introduction 12

2.3 Method 12

2.4 Results 14

2.5 Discussion 16

CHAPTER 3 STUDY 2: FURTHER PROBLEM CHARACTERIZATION STUDY 17

3.1 Abstract 17

3.2 Introduction 17

3.3 Method 18

3.3.1 Participants 18

3.3.2 Apparatus 18

3.3.3 Secondary Tasks 19

3.3.4 Experimental Design 19

3.3.5 Response Variables 19

3.3.6 Simulated Driving Scenario 20

3.3.7 Procedure 21

3.4 Results 22

3.4.1 Response Time Results 22

3.4.2 Response Type Results 23

3.5 Discussion 24

CHAPTER 4 HMI CONCEPTS 27

4.1 Abstract 27

4.2 Introduction 27

4.3 Theoretical Motivation for a Field of Safe Travel Display 27

4.3.1 Gibson’s Ecological Optics and Field of Safe Travel 27

4.4 Field of Safe Travel (FOST) Display Elements 29

4.4.1 Elements of the FOST Display 29

4.4.2 Auditory-Visual Display Design Guidance 30

4.5 Theoretical Motivation for Haptic Displays 31

4.6 Haptic Display Elements 33

4.6.1 Haptic Displays and the Active Control Element 33

4.6.2 Haptic Steering Patterns 34

4.6.3 Prototyping Apparatus 36

4.7 Lane Change Adviser 36

CHAPTER 5 STUDY 3, PART 1: DRIVER HAPTIC STEERING ALERTS AND URGENCY SCALING 37

5.1 Abstract 37

5.2 Introduction 37

5.3 Method 48

5.4 Results 51

5.5 Discussion 58

CHAPTER 6 STUDY 3, PART 2: DRIVER HAPTIC STEERING ALERTS, MIRAGE EVENTS AND MAUT: A NOVEL METHOD FOR SELECTING AN OPTIMAL HMI 64

6.1 Abstract 64

6.2 Introduction 64

6.3 Method 68

6.3.1 Participants 68

6.3.2 Apparatus 68

6.3.3 Experimental Design 68

6.3.4 Response Measures 68

6.3.5 Simulator Driving Scenario 69

6.3.6 Procedure 72

6.4 Results 73

6.5 Discussion 83

CHAPTER 7 STUDY 4: SCREENING STUDY TO EVALUATE FACTORS INFLUENCING HMI EFFECTIVENESS 87

7.1 Abstract 87

7.2 Introduction 87

7.3 Method 90

7.3.1 Participants 90

7.3.2 Apparatus 90

7.3.3 Secondary Tasks 91

7.3.4 Experimental Design 91

7.3.5 Response Variables 91

7.3.6 Simulated Driving Scenario 91

7.3.7 Procedure 94

7.4 Results ... 97
7.4.1 Response Time Results ... 98
7.4.2 Response Type Results ... 99
7.5 Discussion ... 100
CHAPTER 8 STUDY 5: HMI EFFECTS IN FULLY AUTOMATED VERSUS SHARED CONTROL AUTOMATED DRIVING ... 102
8.1 Abstract ... 102
8.2 Introduction ... 102
8.3 Method ... 104
8.3.1 Participants ... 104
8.3.2 Apparatus ... 105
8.3.3 Secondary Tasks ... 105
8.3.4 Experimental Design ... 105
8.3.5 Response Variables ... 106
8.3.6 Simulated Driving Scenario ... 106
8.3.7 Procedure ... 107
8.4 Results ... 109
8.4.1 Response Time Results ... 109
8.4.2 Response Type Results ... 111
8.5 Discussion ... 111
CHAPTER 9 CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH ... 114
9.1 Summary and Conclusions ... 114
9.2 Future Work ... 119
9.2.1 Lane Change Adviser ... 119
9.2.2 Further HMI Development ... 119
9.2.3 Automated Driving and Shared Control ... 119
APPENDIX A: VIRTTEX DRIVING SIMULATOR ... 120
APPENDIX B: STUDY 4: 2^{6-1} EXPERIMENTAL DESIGN DETAILS ... 122
BIBLIOGRAPHY ... 124
ABSTRACT ... 130
AUTOBIOGRAPHICAL STATEMENT ... 132

LIST OF TABLES

Table 2.1 Distribution of Brake-Only vs. Brake-&-Steer Responses by Drive Mode 15

Table 3.1 Response Type by Drive Mode per TTC 24

Table 4.1 Haptic Profile Properties 35

Table 5.1 Relative Object Circularity: Counts of Circular Triads (#CT) and Z-values Per Object (Haptic Steering Pattern) for First (Left) and Second (Right) Set of Pair Comparisons 55

Table 5.2 Ordinal Scaling of Haptic Steering Patterns for First and Second Blocks of Pair Comparisons 56

Table 6.1 Proportions of Positive Appropriateness and Acceptability Ratings by TTC per Haptic Pattern 74

Table 6.2 Proportions of Ratings Jointly Positive on Both Appropriateness and Acceptability by TCC per Haptic Pattern 77

Table 6.3 'Generality' Scores or Counts of the Number of TTC Conditions each Haptic Pattern is Typcially Rated Positively on Both Appropriateness and Acceptability 80

Table 7.1 Independent Variables and Levels 92

Table 7.2 Description of Experimental Conditions 93

Table 7.3 Fisher's Exact Test on Main Effects 100

Table 8.1 Count of Brake-Only Responses Observed by Condition 111

Table A.1 VIRTTEX Motion Performance Specifications 121

LIST OF FIGURES

Figure 1.1 Brake vs. Steer Curve 5

Figure 1.2 The Lane Change Steering Maneuver 6

Figure 2.1 Uncovering Maneuver at FC Event 14

Figure 2.2 Response Time Results 15

Figure 2.3 Driver response type, distance, and speed at response onset in automated driving mode with forward collision alarm and autobraking, in automated driving mode but without alarm or autobraking, and in manual driving mode 16

Figure 3.1 Uncovering Maneuver at FC Event 21

Figure 3.2 Response Time as a function of TTC and Drive Mode 23

Figure 3.3 Driver Response Types by TTC and Drive Mode Across Several Studies 25

Figure 4.1 Gibson's Field of Safe Travel. Source (Gibson and Crooks, 1938) 28

Figure 4.2 Auditory-Visual Elements of the Lane Change Adviser 29

Figure 4.3 Hierarchical Control Model of Automobile Steering 32

Figure 4.4 Proprioceptive Haptic Wheel Profiles 34

Figure 5.1 Scale Additivity Example Source: (Maxwell, 1974) 46

Figure 5.2 Timeline of Haptic Steering Pattern Presentation in Pairs 50

Figure 5.3 Chart of Proportion Matched (median proportion of 0.78 reference line added) 52

Figure 5.4 Chart of Consistence (Participant Circular Triads). (median consistence of 0.9 reference line added) 53

Figure 5.5 Chart of Consistence for 2nd Set of 45 Pairs 54

Figure 5.6 Visualization of Significantly Different Patterns using the Studentized Range Test 57

Figure 5.7 Bradley-Terry Scale Values for First Set and Second Set of Pair Comparisons 58

Figure 5.8 Actual vs. Estimated Vote Counts, First Set and Second Set of 45 Pair Comparisons 58

Figure 5.9 Matrix Plot of Haptic Steering Patterns Rank-Ordered Urgency and Various Physical Attributes 62

Figure 5.10 A Scatterplot of Urgency (rank) vs. Peak Torque with Angular Impulse Labels per Haptic Steering Pattern 62

Figure 6.1 Balanced Bipolar Scale: For Appropriateness and Acceptability Ratings 69

Figure 6.2 Timing Sequence for Mirage Events 71

Figure 6.3 Mirage Event 72

Figure 6.4 Star plots, per haptic steering pattern, of the probability of a positive rating on appropriateness (blue solid line) and acceptability (red broken line) ratings at various TTCs 76

Figure 6.5 Star plots of the proportion of participants who rated a haptic pattern positively on both appropriateness and acceptability at various TTCs 78

Figure 6.6 Mean Appropriateness vs. Acceptability for haptic steering patterns by TTC condition 80

Figure 6.7 MAUT results based on joint Appropriateness and Acceptability ratings (sum of products) and Generality 82

Figure 6.8 MAUT results based on joint Appropriateness and Acceptability ratings (participant counts) and Generality 83

Figure 6.9 Rank Order of Urgency Scaling of Haptic Steering Patterns and Significant Differences by Studentized Range Test 85

Figure 7.1 Description of the LCA 96

Figure 7.2 Auditory-Visual Elements of the LCA 96

Figure 7.3 Main Effects Plot for Response Time 99

Figure 8.1 Main Effects 110

Figure 8.2 Boxplot of Response Times by Drive Mode by LCA State 111

Figure 8.3 Boxplot of Response Times by LCA On or Off 113

Figure A.1 Ford’s VIRtual Test Track EXperiment (VIRTTEX) simulator 120

Figure B.1 Resolution VI Design 122

Figure B.2 Resolution VI Treatment Combinations 123

LIST OF ACRONYMS

ABS Anti-lock Braking System 1, 2, 15

ACC Adaptive Cruise Control 131

ADAS Advanced Driver ASistance 118

AEB Automatic Emergency Braking 2, 6, 131

AR Augmented Reality 4

CWIM Collision Warning Interface Metrics 65

dGPS differential Global Positioning System 36

ESA Evasive Steer Assist .... 2, 3, 4, 6, 7, *et seq.*

FC Forward Collision .... 10, 12, 13, 14, 18, *et seq.*

FCW Forward Collision Warning .... 2, 6, 7, 10, 12, *et seq.*

FOST Field of Safe Travel .... 3, 7, 9, 11, 23, *et seq.*

HHDD High Head Down Display .... 7, 17, 18, 19, 20, *et seq.*

HMI Human-Machine Interface .... 3, 4, 6, 7, 8, *et seq.*

HUD Head-Up Display .... 7, 12, 13, 17, 29, *et seq.*

ISI InterStimulus Interval 39, 50, 73

LCA Lane Change Adviser .... 10, 11, 23, 36, 86, *et seq.*

LDW Lane Departure Warning 35, 66, 92, 108

MAUT Multi-Attribute Utility Theory .... 64, 73, 81, 82, 84, *et seq.*

NVH Noise Vibration and Harshness 34

PBA Panic Brake Assist 2, 15, 16

SSQ Simulator Sickness Questionnaire 22, 51, 73, 97, 109

TH Time-Headway 94, 107

TOR Take Over Request 4, 7, 65

TTC Time To Collision .... 4, 6, 7, 8, 9, *et seq.*

VIRTTEX VIRtual Test Track EXperiment .... 18, 21, 22, 48, 49, *et seq.*

CHAPTER 1 INTRODUCTION

1.1 Motivation

1.1.1 Rear-End Crashes

Rear-end crashes are common on U.S. roads and accounted for one-third of police reported crashes in 2010 (NHTSA, 2010). Among rear-end crashes, the most common subtypes are 'lead-vehicle stopped', 'lead-vehicle decelerating', and 'lead-vehicle traveling at a constant slower speed' scenarios (Victor et al., 2014). Kusano and Gabler (2011) analyzed rear-end-crash data collected in the National Motor Vehicle Crash Causation Survey (NMVCCS) and found that drivers generally do not steer in forward collision situations. They reported that, among drivers who were involved in rear-end collisions and for whom pre- crash maneuver data were available, 28% did no pre-crash avoidance maneuver at all; about half (49%) braked-only; 19% steered and braked; and 4% steered only. These data are for drivers who crashed. Earlier research (e.g., test track research by Rice and Dell'Amico (1974); simulator testing by Lechner and Malaterre (1991)) reviewed in Adams (1994) reported that drivers in field tests and simulator tests generally avoided collisions more successfully when they steered, or braked and steered, around obstacles rather than braking only. Exceptions have been reported. For example, Adams et al. (1995) found in fixed-base simulator testing that un-alerted drivers generally steered to avoid a rock in the road. They noted that the earlier studies involved larger obstacles (e.g., a car, a large barrel) rather than a rock and this difference might prompt different strategies. More recently, Mazzae et al. (2003) reported that nearly all participants in test track testing of the Anti-lock Braking System (ABS) both braked and steered during their crash avoidance maneuvers in response to an intersection hazard. Future research is needed to determine how an intersection hazard from the left or right affects steering versus braking behavior when compared to a rear-end collision hazard in car following. Nonetheless, the bulk of real-world data indicate that drivers in rear-end crash scenarios do not steer when they could or should.

Automated driving also faces a similar challenge depending on the level of automation. SAE J3016 (SAE, 2014) defines several levels of automated driving. Level 2 (partial automation) assigns the driver the role of monitoring the driving environment so as to intervene if necessary. Level 3 (conditional automation) assigns the automation the role of monitoring the driving environment, and the driver is expected to respond appropriately when prompted by the automation. For Level 2 and Level 3 automation, ongoing research focuses on how to keep or get the driver back "in the loop" and how to best support the transition to manual driving. Gold et al. (2013) recently reported that in simulated highway driving with an obstruction (e.g., an accident) ahead, drivers manually controlling vehicles typically brake-only rather than steer if they are closer to the obstacle, but steer rather than brake the further they are from the obstacle. They also found that braking rather than steering (or braking and steering) is even more likely to be observed in automation-to-manual control transitions. Blommer et al. (2017) conducted a moving-base simulator study comparing driver responses to a sudden forward collision hazard and found that drivers in manual mode tended to use evasive steering more than braking only, compared to drivers in autonomous driving who suddenly needed to manually intervene. In Level 2 or Level 3 automated driving, then, if there is a collision hazard that requires a manual takeover, the problem is how to get people to steer when they need to do so.

Advanced Driver Assistance Systems for Forward Collision Avoidance

There are many forms of Advanced Driver Assistance Systems (ADAS) relevant to the rear-end crash problem. Forward Collision Warning (FCW) alerts the driver to a forward collision hazard. FCW is sometimes paired with autonomous emergency braking (AEB). Cicchino (2016) analyzed crash rates from 27 states from 2010 through 2014 and compared passenger vehicles with FCW only, versus FCW and AEB, against the same vehicle models without those systems. The results indicated that FCW alone and FCW with AEB both reduced rear-end striking crashes, but that only FCW with AEB was statistically significant beyond chance in reducing rear-end striking injury crashes. Panic-Brake Assist (PBA), and Collision Mitigation by Braking (CMbB) are examples of other ADAS technologies intended to reduce rear-end crash occurrence and/or severity by braking. PBA systems detect attempted emergency braking by indicators such as a rapid or forceful brake pedal press (combined with forward collision sensing) and provide additional pressure to the brakes or activate ABS. CMbB systems provide automatic braking in crash-imminent situations at relatively short ranges (to increase confidence in correct detection) in order to reduce the impact velocity.

Evasive Steer Assist (ESA) is a recent rear-end collision avoidance feature designed to support the driver in evasive steering or emergency "swerve" maneuvers. The kinematics of the vehicle generally support later steering than braking at higher speeds for crash avoidance (Dang et al., 2012). ESA dynamically changes the vehicle's steering gains and applies a form of differential braking during a swerve or emergency lane change so that the driver can steer around the obstacle ahead in a stable manner, preventing overshooting or undershooting the adjacent lane. ESA can be an effective crash countermeasure in certain circumstances where there is not enough room to brake without collision, but there is a steerable path around the obstacle. Ideally, it would keep the driver from slowing or stopping in the lane and becoming a new obstacle. However, when manually controlling the vehicle (or manual takeover from automated driving) the driver must initiate steering for ESA to activate. The human factors problem, as suggested earlier, is to get drivers to steer when necessary.

Another steering-related ADAS concept under development is a lane change adviser. The lane change adviser is an algorithm that alerts a driver when it predicts that he/she will want to change lanes to maintain a consistent speed. In principle, this feature would operate in normal traffic and would reduce travel time, increase road capacity, and improve fuel efficiency by minimizing speed variations. It might also familiarize drivers with the Human-Machine Interface (HMI), which would be useful in ESA scenarios. An adequate sensor set (radar, vision system, ultrasonic sensors, etc.) would provide data to suppress the HMI if no steerable path was detected.

In summary, rear-end crashes are among the most common crash types in the United States. Drivers attempting to avoid rear-end crashes generally brake rather than steer, even when there is a steerable path to avoid the collision. Research suggests that drivers tend to steer rather than brake when an obstacle is farther away and brake rather than steer when the obstacle is closer, i.e., crash-imminent. This tendency conflicts with the physics of the vehicle, which may sometimes support later steering than braking to avoid collisions. Drivers can also be more reluctant to steer if they are closer to vehicles ahead of them during emergency manual takeovers in simulated Level 2 or Level 3 automated driving. ESA is an automated driver assistance feature that supports manual driving to avoid collisions, but requires the driver to initiate steering. This dissertation describes an approach to develop an auditory-visual field-of-safe travel (FOST) display and haptic steering display. This HMI will advise the driver to steer around an obstacle ahead when feasible, thereby enabling the ESA feature to ensure the stability of that maneuver. It is intended for use when the driver has his or her hands on the wheel in manual driving. It is also suitable for a shared-control framework for automated driving (Petermeijer et al., 2015). In shared haptic control, the driver and the automated driving system simultaneously share control over steering, and the driver can feel the automated steering and interactively modulate the steering torques. In the present case, drivers in either manual or automated driving modes might receive a discrete haptic steering profile to signal the possibility of a lane change in manual driving or a lane change in response to a Takeover Request (TOR) in automated driving. The lane change adviser recommends the driver make a lane change when feasible in benign driving conditions to keep a consistent speed. It may also familiarize drivers with a consistent HMI that indicates a steerable path in an emergency.

Lorenz et al. (2014) investigated the influence of Augmented Reality (AR) information presented as a roadway overlay and whether it positively influences the takeover process. They investigated two AR concepts: an AR Red and AR Green display. Each display projected a corridor onto the road: the AR Red display projected one that the driver should avoid, and the AR Green display projected one that he/she could safely steer through. Of the manual drivers who made up the control condition, 80% steered-only, 10% braked-only, and the remaining 10% steered-and-braked. In the AR Red takeover scenario, 18.8% steered only, 25% braked only, and the remaining 56.2% steered-and-braked. In the AR Green takeover scenario, 47.1% steered-only, and the remaining 52.9% steered-and-braked. The findings indicate that supporting the driver in his/her information processing by providing AR pathways does not significantly affect takeover times, but it did affect the response type by category (AR Red vs AR Green).

1.1.2 Steering versus Braking Response

Sometimes steering is better for collision avoidance (Beckman, 1998). The kinematics of braking and steering are such that a driver may be able to steer around an obstacle later and closer than braking only.

Figure 1.1 depicts a “brake versus steer” curve showing the unrealized benefit of ESA for a crash- imminent situation, beginning at about 60 kph and 12.5 m to a stopped object. The green shaded area depicts optimal steer versus brake, showing where ESA may be of greatest benefit. However, that is not the only benefit of ESA. At even greater Time to Collision (TTC) values, the driver who steers rather than brakes avoids becoming a new obstacle for vehicles approaching from behind.

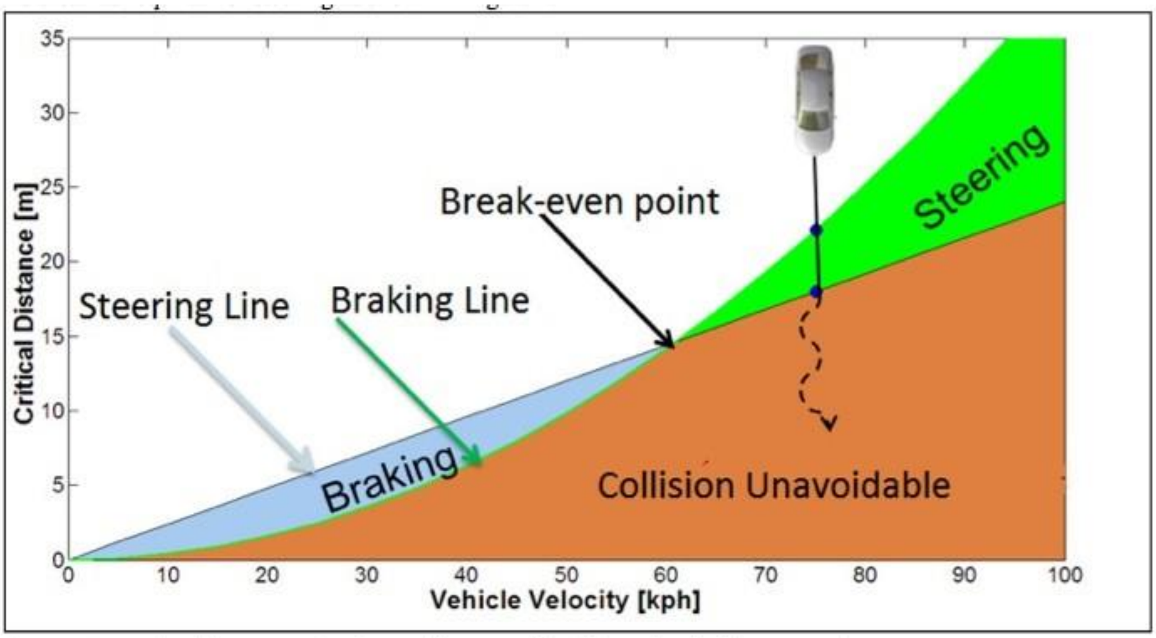

Figure 1.1: Brake vs. Steer Curve

1.1.3 Steering Behavior

There are many reasons why a driver may not steer when it would be effective. One reason is that steering through an emergency lane change can be hard to do. Steering is a challenging manual control task when done in emergencies or other less-than-ideal circumstances such as a reduced coefficient of friction. To show this, we will use Figure 1.2, below. The top plot shows that one begins a lane change by generating a sinusoidal steering profile. This is integrated over time to generate appropriate heading changes. Those are integrated in turn to make the desired move to an adjacent lane. The vehicle steering system constitutes a second-order, highly cross-coupled system. Second-order control systems can be difficult to control because a driver must anticipate and predict the vehicle's response (Wickens et al., 1997), a difficult task in aggressive maneuvers. This is why we are not all race car drivers. For this reason, driver assistance can help with steering and lane changes.

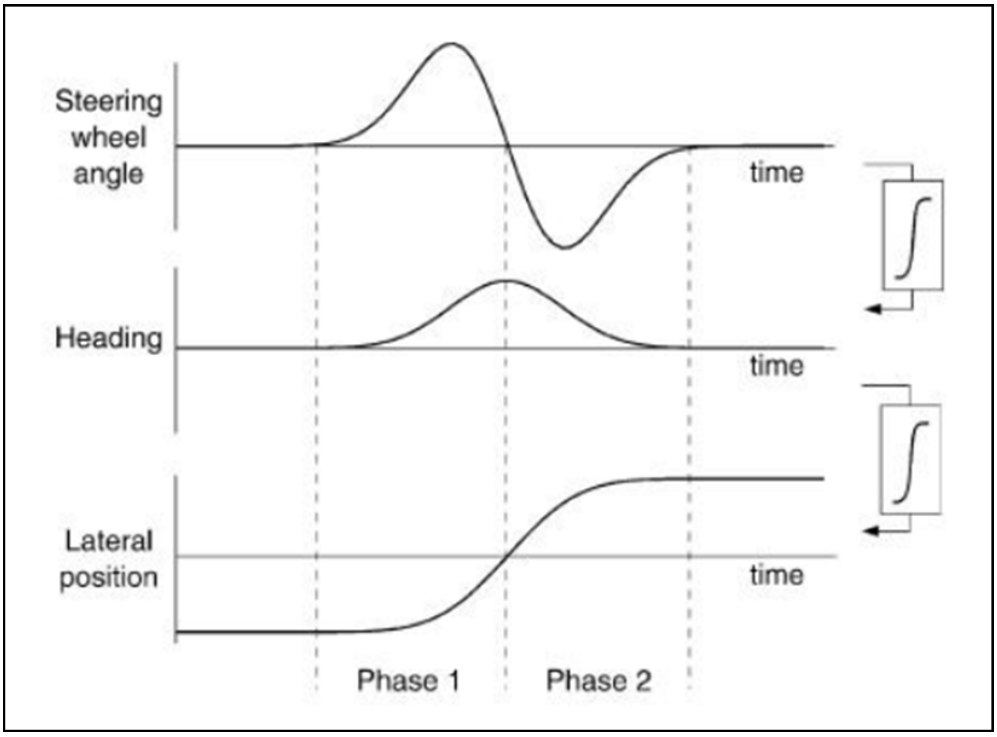

Figure 1.2: The Lane Change Steering Maneuver

1.2 Evasive Steer Assist (ESA)

The purpose of ESA is to help drivers avoid collisions by changing lanes. The technology is intended to help in a crash-imminent scenario where the TTC is such that braking alone will not prevent collision. The ESA controller limits the driver's steering input, effectively restricting how quickly the vehicle tires can turn in order to prevent the driver from steering the vehicle beyond its dynamic limits and to let the vehicle follow a predetermined steerable path. However, the driver must initiate steering for ESA to engage. This dissertation is directed toward the development of an HMI that can facilitate the initiation of steering for lane changes when appropriate.

1.3 Dissertation Contributions

Driver assistance and automated driving technologies can address the rear-end crash problem in various ways. Braking is commonly assumed for driver assistance technologies such as FCW and AEB. Driver assist technologies like ESA and Level 2 or Level 3 automated driving systems (NHTSA 2013; 2015) might facilitate manual emergency lane changes, but may require the driver to manually initiate the maneuver, something drivers are often reluctant to do. As explained in SAE J3016 (SAE, 2014), in Level 2 automation the human driver must monitor the driving environment and be ready to manually take over driving when necessary. Level 3 automated driving involves an automated system that monitors the environment with the expectation that the human driver will respond appropriately to manual TORs to intervene. An HMI might advise the driver of a steerable path around a forward collision hazard when feasible. Steering around an obstacle may be advisable if the adjacent lane is unoccupied and braking is unlikely to avoid a collision. Changing lanes to pass an obstacle could also promote smooth traffic flow and keep one's own vehicle from becoming a new forward collision hazard.

This dissertation's contributions include the following: a) audio-visual FOST display concepts; b) discrete haptic steering display concepts; c) a paired-comparison approach to scaling urgency for haptic displays applied while driving; d) a new mirage scenario methodology for eliciting subjective assessments in the context of a forward collision hazard, briefly presented then removed, without risk of simulator sickness; and e) evidence of the role of mental set in collision avoidance and a means to suggest a driver may steer around an obstacle in certain circumstances.

1.4 Dissertation Outline

1.4.1 Study 1: Initial Problem Characterization: Driver Response to Sudden Forward Collision Events in Manual and Fully Autonomous Driving (see Chapter 2)

This study uncovered differences in the type of response at manual takeover (brake-only vs. steering behavior) as a function of manual driving, fully autonomous (i.e., hands-off) driving (warning presented but driver intervention required), and fully autonomous driving with automation failure (i.e., without any system warning and driver intervention required). A phase plane analysis indicated that at greater TTCs, drivers were more likely to steer, with or without braking, as opposed to braking alone. The analysis also indicated that an ESA feature, operating while drivers attempted an emergency lane change, would be beneficial. This study was published as Blommer et al. (2017).

1.4.2 Study 2: Further Problem Characterization Study (see Chapter 3)

This study extended the problem characterization to longer TTCs so as to complement other published research (e.g., Gold et al. (2013)). The first study used a Head-Up Display (HUD) for auto- mated driving status and FCW display. This study used a High Head-Down Display (HHDD) mounted on top of the instrument panel, with similar iconography and auditory alerts. Results indicated that response time increased as TTC increased across 1.7, 3.0, 5.0, and 7.0 s; that response time also varied more with increased TTC; and that although brake-only responses were more common in autonomous to manual takeovers than in manual driving, all drivers in fully autonomous mode steered without braking at the 1.7 s TTC condition. This discrepancy in hands-off fully autonomous driving between the first and second studies was unexpected. All the other conditions were generally consistent with prior research. One methodological difference was that the second study increased the number of lane changes executed during the simulator driving session because there were very few drivers steering during pilot testing, regardless of TTC. The increased number of lane changes was introduced to familiarize the driver with “driving the simulator and its handling characteristics.” This increase had unexpected consequences that led to a methodological investigation carried out in later studies as part of this dissertation. Since the absence of brake-only responses might be a methodological artifact associated with the increased number of lane changes, the research effort continued on the development of an HMI to facilitate both normal and emergency lane changes.

1.4.3 HMI Concepts (see Chapter 4)

Auditory-visual HMI concepts were identified that might be suitable for manual driving or fully autonomous (i.e., hands-off) automated driving. In addition to fully automated hands-off driving, an alternative exists that may promote greater driver engagement in Level 2 and Level 3 automation. That alternative is referred to here as hands-on shared control automated driving. In this alternative, the driver has his/her hands on the wheel continuously while in autonomous driving mode. The virtual driver generates steering torques for lanekeeping that the human driver can perceive and override if necessary. These auditory-visual displays were augmented for hands-on shared control for automated driving by means of discrete haptic steering displays intended to indicate opportunities to steer around an obstacle ahead in either normal or emergency driving.

1.4.4 Study 3, Part 1: Driver Haptic Steering Alerts and Urgency Scaling (see Chapter 5)

A haptic display or alert is perceived by kinesthetic-tactile sensations rather than by sight or hearing. This is a novel concept that required methodical development because it influences the driver’s primary lateral control element, i.e., the steering wheel. Drivers’ assessed the urgency conveyed by each of several candidate haptic steering wheel torque profiles by pairwise comparisons obtained while participants drove in the moving-base simulator (but without a hazard present, as might be done for on-road acceptance testing). This dissertation develops a method of pairwise comparisons for haptic steering displays (suitable for simulator testing and possibly on-road testing); analyzed the data for internal consistency and reliability; and applied both interval-level scaling and ordinal-level scaling to the participants' judgments. The results place the various haptic profiles along a continuum of urgency (with scaling intervals assumed to be meaningful) as well as in rank order (without making strong scaling assumptions about interval distances). This study is in review for *Applied Ergonomics*.

1.4.5 Study 3, Part 2: Driver Haptic Steering Alerts, "Mirage Events", and Multi-Attribute Utility Theory: A Novel Method for Selecting an Optimal HMI (see Chapter 6)

This dissertation developed a novel method for in-situ subjective assessments of haptic steering profile appropriateness and acceptability in the face of different TTC scenarios. While driving in the moving-base simulator, drivers experienced each haptic profile in the context of a forward collision hazard, briefly presented and then removed. This method did not necessitate braking or steering responses and thus eliminated the risk of simulator sickness. A range of TTCs from crash-probable (5 s or 7 s) to crash-imminent (3 s or 1.7 s) were used to mimic both normal and emergency conditions. This unique approach provided data that were analyzed in several ways, including Multi-Attribute Utility Theory (MAUT), to arrive at a single "best" haptic steering display. This "best" pattern was then evaluated in subsequent testing. A key finding was that the "best" profile in terms of appropriateness and acceptability was not the most "urgent" as identified in the pair-comparisons urgency scaling effort. Thus, it was discovered that a psychophysical scaling effort without additional in-situ testing might result in a design selection that is not optimal in terms of consumer acceptance. This study has been accepted for publication in *Applied Ergonomics*.

1.4.6 Study 4: Screening Study to Evaluate Factors Influencing HMI Effectiveness (see Chapter 7)

An alert that has never been seen, heard, or felt before may be incomprehensible to a driver (Lerner et al., 2011). Exposure to a novel display in everyday driving may enhance the effectiveness of that display in an emergency situation. The novel haptic steering display operated on a primary driver control element (the steering wheel), so it might be perceived as a fault unless the driver was familiar with it. Therefore, as part of the overall HMI, a system concept was developed that is here referred to as a lane change advisor. The lane change advisor is the same in normal driving (to suggest lane changes when approaching slower vehicles) and in emergency situations. An initial screening study therefore used the FOST with or without the haptic steering display for both lane change advising and emergency steering alerts. This study was conducted for crash-imminent conditions (1.7 s or 3.0 s TTCs) in both manual driving and hands-on shared control automated driving modes. Results of the screening study indicated the following: a) manual drivers responded reliably faster than drivers in shared-control autonomous conditions; b) drivers in the 1.7 s TTC scenario responded faster than drivers in the 3.0s TTC scenario regardless of driving mode; c) drivers who were distracted when the forward collision event arose responded later than those who were not distracted (as expected for Level 2 and Level 3 automated driving); and d) all participants in all conditions did some form of steering - no driver out of 63 participants braked-only. These results suggest that the HMI system was highly successful in facilitating lane changes for emergency conditions. However, concerns remained regarding the possibility of a methodological artifact. This potential artifact might be associated with the high emphasis on lane changes provided in training. This concern led to a series of pilot tests to identify conditions that would bring back brake-only behavior.

1.4.7 Study 5: HMI Effects in Fully Automated versus Shared Control Automated Driving (see Chapter 8)

Most haptic steering displays and driver warnings are oriented toward keeping the driver in the lane. In contrast, this dissertation investigated a method to suggest that the driver may steer out of lane when feasible for forward collision avoidance. Results from the screening study (Study 4) indicated a significant response time difference between participants in manual driving and those in hands-on, shared-control automated driving. Furthermore, not a single participant in the screening study braked-only in response to the FC hazard. Pilot testing revealed that hands-off fully automated driving with low levels of lane change practice and no exposure to the LCA brought back brake-only response behavior like that found in Study 1. In this final study, the research effort compared two different styles of automated driving (hands-off, fully automated driving versus hands-on shared-control automated driving). Within each of these two automated driving modes, this study compared the effects of a lane change adviser On versus Off. In this study, this factor is referred to as LCA State where LCA On means the participant received the LCA in both benign driving and at the FC hazard event, and LCA Off means that the participant did not receive the LCA, but only FCW at the FC hazard. This manipulation was done in order to investigate the potential effects of mental set that might increase the driver's propensity to steer rather than brake-only.

Results of this study indicated that participants in hands-off automated driving responded slower than those in hands-on shared-control automated driving. Furthermore, drivers with LCA off responded more slowly than drivers with LCA on regardless of automated drive mode. Results also indicated a statistically significant difference in response type distribution. Specifically, the incidence of brake-only responses were significantly reduced with LCA on as compared to with LCA off, regardless of type of automation. These results collectively suggest that the LCA used to administer FOST or FOST plus haptic displays in benign conditions increase the propensity of lane change behavior in crash-imminent situations.

1.4.8 Conclusions and Recommendations for Future Research

In this final chapter, all study results are summarized and interpreted. The LCA and FOST, with or without haptics, had a demonstrable effect on reducing brake-only behavior in both hands-off fully automated driving, hands-on shared control automated driving, and manual driving. The source of this effect involves familiarization with the displays and developing a mental set in normal driving that a driver may sometimes steer rather than just brake to get around an obstacle ahead. Recommendations for further development of the HMI concepts are suggested.

CHAPTER 2 STUDY 1: DRIVER RESPONSE TO SUDDEN FORWARD COLLISION EVENTS IN MANUAL AND FULLY AUTONOMOUS DRIVING - INITIAL PROBLEM CHARACTERIZATION STUDY

2.1 Abstract

The study reported here measured a non-distracted driver's response to a sudden Forward Collision (FC) event, in which the driver would assume manual control of an autonomous vehicle. Forty-eight volunteers participated in three driving scenarios conducted in a high-fidelity motion-base simulator: fully (i.e., hands-off) autonomous driving with collision avoidance support; fully autonomous driving without collision avoidance support; and manual driving without Forward Collision Warning (FCW). Results indicated that drivers in manual mode were more likely to use evasive steering rather than braking as compared to drivers in the autonomous modes. Between-subjects variations in speed were higher for automation with collision support than for the other two scenarios. For both autonomous driving scenarios, drivers' reaction times were longer than for manual driving. The results of this study suggest that an ESA feature, operating while drivers attempted an evasive lane change maneuver, would be of benefit.

2.2 Introduction

In autonomous vehicle operation, drivers may sometimes need to regain manual control of the vehicle. Control handoff from vehicle to human may arise in a structured (i.e., preplanned) or unstructured (i.e., sudden or unexpected) manner. As expected for Level 2 or Level 3 automated driving systems (NHTSA, 2013: SAE, 2014), the study reported here was designed to measure a non-distracted driver's response to a sudden FC event. Three driving scenarios were investigated in Ford Motor Company's high-fidelity motion-base driving simulator: autonomous vehicle driven with full collision avoidance support, autonomous vehicle driven without collision avoidance support, and vehicle driven in manual mode.

2.3 Method

A Head-Up Display (HUD) was installed in the simulator and used for both the manual and automated driving modes tested. In the automated driving mode, the HUD presented system information and alerts such as whether automation was available, whether it had been engaged, and whether the system was in manual control mode. The alerts included notifications that the system was approaching a construction zone or had detected an emergency that required the driver to take over immediately (only for the final critical event with full collision avoidance support). Voice prompts and chimes to deliver the information alerts, all except for the forward collision threat where no auditory or voice alerts were presented, augmented the HUD displays.

Near the end of the simulator drive, an “uncovering” event suddenly occurred that required the participant to intervene and brake or steer to avoid a stopped passenger vehicle ahead of the vehicle they had been following throughout the drive. Figure 2.1 shows the uncovering maneuver; the host vehicle is light blue, the cut-out vehicle is black, and the stopped vehicle is dark blue. The lead vehicle started an aggressive swerve maneuver to reveal a vehicle stopped in the travel lane at a TTC of 2.7 s from the host vehicle. For automated driving with support, the automated driving system detected the stopped vehicle when the lead vehicle was approximately halfway across the lane line. Thus, the Time to Collision (TTC) with the stopped vehicle was approximately 1.7 s when the FC alert in the HUD and warning sound commenced, concurrently with the autobraking. For automated driving with no support and for manual driving, there were no alerts or autobraking. Drivers could, of course, see the event whether they were driving manually or with automation. In all instances, the driver needed to respond manually to the FC event. Driver responses from 48 participants were recorded.

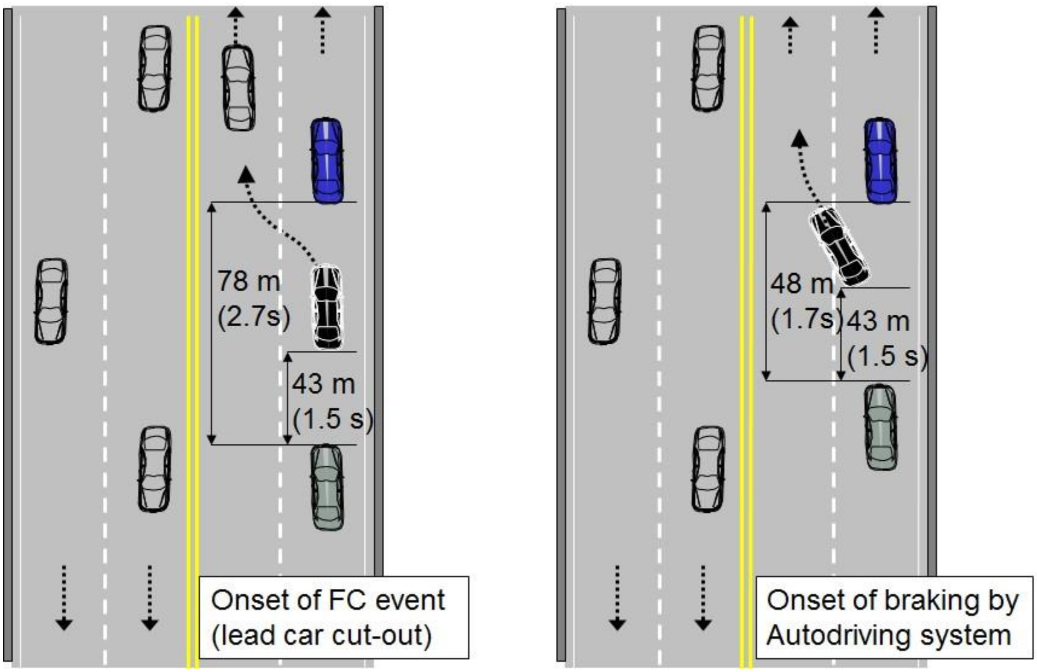

Figure 2.1: Uncovering Maneuver at FC Event

2.4 Results

Video of driver eye-glances at hazard onset was manually reviewed to verify that test participants were looking forward as appropriate for Level 2 or Level 3 automated driving. Test participants who happened to be looking away from the road ahead at hazard onset were replaced until the planned sample size was filled. Video of each driver's response was also used to manually code the response type (i.e., brake-only or brake-and-steer). Table 2.1 presents the type of response to the forward collision hazard as a function of driving condition. Fisher's exact test indicated no significant difference between the two automated driving conditions: there was roughly a 50-50 split between brake-only and brake-and-steer maneuvers. On the other hand, about 85% of drivers in manual driving steered rather than braked-only. Because there were no differences between the two automated driving conditions, those data were combined into one automated driving condition. The difference in response type between automated and manual driving was significant at $p = 0.054$ when tested by a one-tailed Fisher's exact test. One interesting observation was that all drivers in this study did some form of braking. Consider the response time results. Figure 2.2 shows jittered box plots for each of the three driving conditions and the median response times per group. Manual driving was associated with reliably faster response times than automated driving with full support (Mann-Whitney test of medians, $W = 340.0$, $p = 0.0044$).

Table 2.1: Distribution of Brake-Only vs. Brake-&-Steer Responses by Drive Mode

Drive Mode	**Response Type**		**Totals**
	Brake-Only	**Brake-&-Steer**	
Automated	8	8	16
Automated w/o Support	7	9	16
Manual	3	13	16
All	**18**	**30**	**48**

The same result was found in a comparison of manual driving and automated driving without collision avoidance support (W = 325, p = 0.0226). The median response times between the two automated driving groups were not statistically different (W = 286.5, p = 0.407).

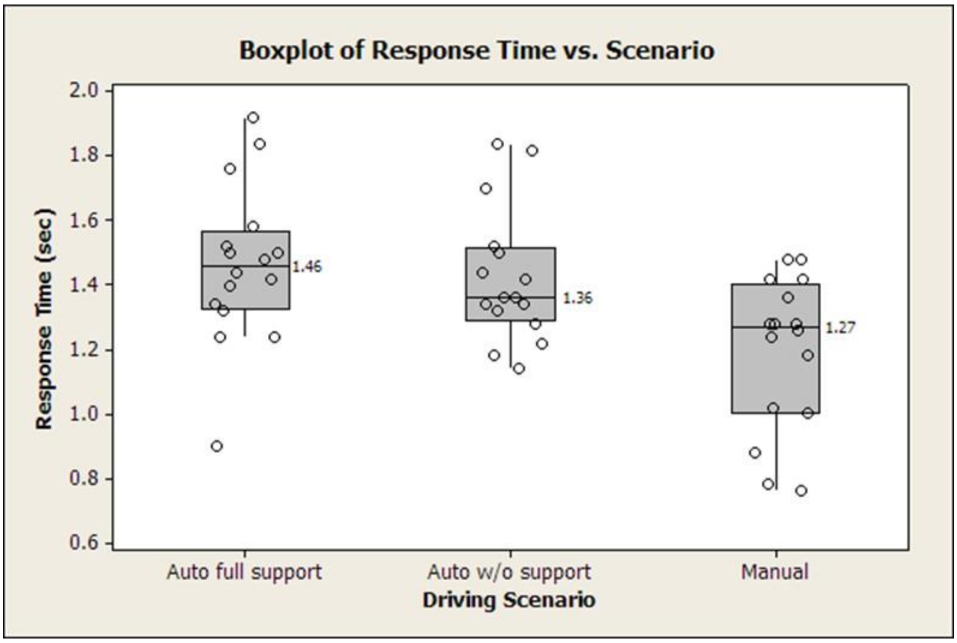

Figure 2.2: Response Time Results

Even a high-fidelity motion-base simulator like the one used in this study cannot fully reproduce the motion response of emergency braking and steering. Therefore, response measures such as peak lateral acceleration, mean braking levels, or number of crashes in manual versus automated driving are inappropriate. Instead, driver response type, response time, and initial range and range-rate at response onset can be analyzed in terms of driver assistance technology. Panic Brake Assist (PBA) applies full ABS braking at 0.8 g. ESA provides sufficient lateral acceleration to help a driver steer around an obstacle, e.g., at $a_y = 0.4g$, and then allows the driver to resume lane keeping. An automated vehicle may have PBA and ESA available during the automation-to-manual handoff. Therefore, an analysis of kinematic outcomes of the experiment given was carried out assuming these driver assistance technologies were in place.

Figure 2.3 shows the speed and distance to the forward obstacle at response onset for all participants. Each plotting point is coded by the scenario or condition run and annotated for response type. In addition, the upper line in the figure indicates the required distance to brake in line to a complete stop; and the lower line in the figure indicates the required distance to steer only, just to avoid contact with the stopped vehicle. The region between the two lines indicates cases where there is not enough distance to brake only, even with PBA, but there is enough distance remaining to steer around the obstacle with ESA support. Cases beneath the lower line involve responses so late that there is not enough distance remaining to steer or brake successfully. Cases above the upper line indicate responses so quick that a driver can avoid the obstacle by either braking or steering. For the high-intensity hazard conflict used in this study, Figure 2.3 suggests that evasive steering, if feasible, may be preferable to braking only for collision avoidance.

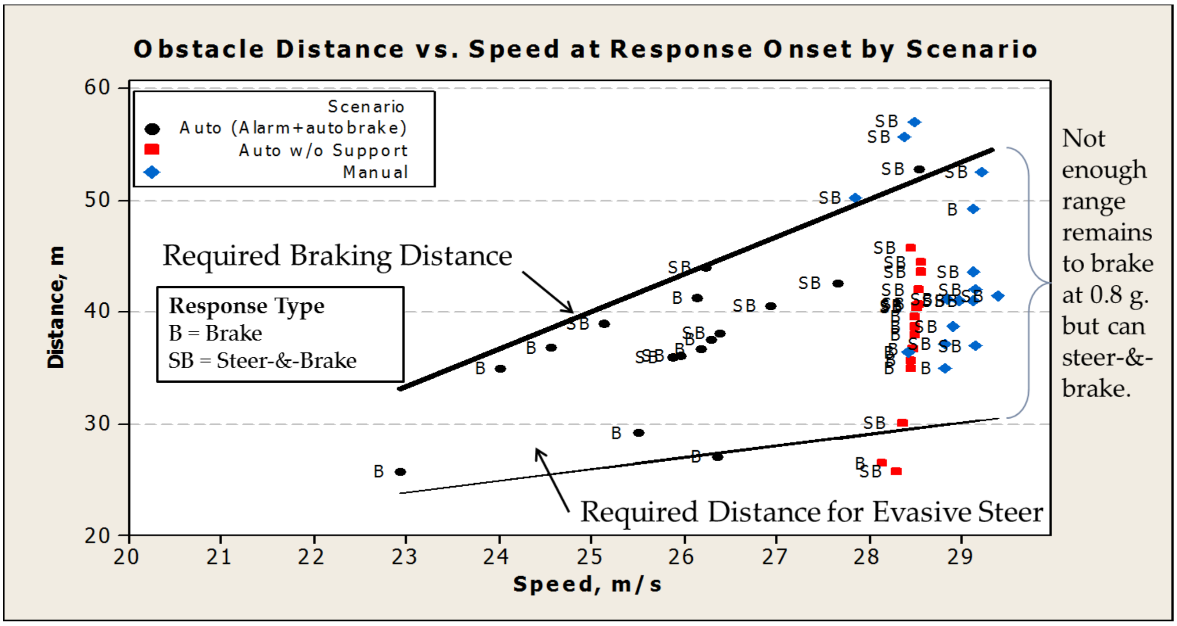

Figure 2.3: Driver response type, distance, and speed at response onset in automated driving mode with forward collision alarm and autobraking, in automated driving mode but without alarm or autobraking, and in manual driving mode

2.5 Discussion

This study indicated that drivers may be better able to avoid collisions by steering around the obstacle rather than braking-only, particularly in high-conflict forward collision situations. The results also showed that drivers may not steer when they should. Fully automated driving was associated with more brake-only behavior than manual driving. Thus, there appears to be a challenge to facilitate lane changing in such scenarios when the adjacent lane is clear. If such driver response can be facilitated, technologies like ESA can help ensure successful maneuvering around the obstacle. Facilitating lane changing in automated and manual driving was the primary focus of this dissertation research.

CHAPTER 3 STUDY 2: FURTHER PROBLEM CHARACTERIZATION STUDY

3.1 Abstract

This study extended the problem characterization to longer TTCs (1.7 s, 3.0 s, 5.0 s, and 7.0 s) so as to complement other published research (e.g., Gold et al. (2013)). The first study used a HUD for automated driving status and Forward Collision Warning (FCW) display. However, this study used a High Head-Down Display (HHDD) mounted on top of the instrument panel, with similar iconography and auditory alerts. Results indicated that, across TTCs, a) response time increased as TTC increased; b) response time variability increased with increasing TTC; and c) brake-only responses were more common in fully autonomous manual takeovers than in manual driving, but all drivers in fully autonomous mode steered without braking at the 1.7 s TTC condition. This discrepancy in response type in fully autonomous driving between the first and second studies was the subject of further inquiry. All the other conditions were generally consistent with prior published simulator research. One methodological difference between the first study and this study was an increased in the number of lane changes executed during the simulator driving session. This procedural change was introduced because, in contract to other published research (e.g., Gold et al. (2013)), there were very few drivers steering during pilot testing, particularly at longer TTCs. The increased number of lane changes was introduced to familiarized drivers with driving the simulator and its handling characteristics. This increased practice had unexpected consequences that led to a methodological investigation to be described later. Since the absence of brake-only responses might be a methodological artifact associated with the increased lane changes (and contrary to prior published research, including our own), the research effort continued on the development of an HMI to facilitate both normal and emergency lane change maneuvers.

3.2 Introduction

Blommer et al. (2017) (see previous chapter) found that in a crash-imminent forward collision hazard situation, manual drivers are more likely to steer (rather than brake-only) than drivers in fully automated driving mode who must manually take over the driving task. Gold et al. (2013) reported related results in an automated driving study also conducted in a simulator. They too used an uncovering maneuver to reveal a stopped vehicle. However, they used longer TTCs at hazard onset (either 5 s or 7 s). Their baseline study involved manual driving, which they compared against fully automated (hands-off) driving. In manual driving, all participants steered-only at the 7 s TTC, and some drivers exhibited a combined steering and braking response at the 5 s TTC. For drivers in fully automated driving who had to take over manually at a 7 s TTC, there was a roughly 50 − 50 split between steer-only and brake- and-steer responses; at a 5 s TTC, a few brake-only responses were added to those response types. This suggests that the more time a driver has to assess his/her options, the more likely he/she is to steer.

This chapter seeks to characterize the problem further by building upon the earlier results and extending the TTC from 1.7 s to 3 s, 5 s, and 7 s to a stopped obstacle that is suddenly revealed. This study, over a wide range of TTCs, was designed to measure a non-distracted driver's response while driving in one of two modes: manual driving and fully automated (i.e., hands off) driving with full collision avoidance support. It was expected that the likelihood of braking only would increase with decreasing TTC and would be even more prominent in automated driving than in manual driving.

3.3 Method

3.3.1 Participants

The sample consisted of 64 volunteer Ford Motor Company employees between 18 and 62 years of age (M = 38.73 years, SD = 12.82 years), evenly balance by gender and age category. Age was subdivided into two categories: "younger" (18-45 years) and "older" (46-62 years). Most of the participants had non-engineering backgrounds (e.g., finance, administration, information technology), and none had prior "surprise event" experience in the Ford VIRTTEX simulator. Participants also had to have valid drivers licenses.

All participants in all conditions analyzed herein had their eyes directed toward the road ahead at the onset of the FC event. Any test participant who was looking away from the road ahead at hazard onset was replaced, until the fully balanced sample of participants was met.

3.3.2 Apparatus

Driving Simulator

See Appendix A for details on the VIRTTEX driving simulator.

High Head-Down Display (HHDD)

A High Head Down Display (HHDD) was installed in the simulator, above the center stack. This 3.5-inch HHDD displayed information and alerts in both the manual and automated modes of driving. In the automated mode of driving, the HHDD presented information about whether automation was available, whether it was engaged, or whether the system had returned to manual driving mode. The HHDD also presented alerts to turn automation off in response to an upcoming construction zone and, if an emergency was encountered, a high-intensity alert to take over manual control immediately. The visual information and alerts were augmented by voice prompts and chimes. In manual mode, the HHDD simply displayed the text "Manual driving" throughout the drive.

3.3.3 Secondary Tasks

To counteract driver fatigue and/or boredom during the study drive, three secondary tasks of varied difficulty were requested by means of voice prompts. While the task requests to each participant appeared to be at random time intervals, all secondary tasks followed a schedule that was consistent across the ensemble. Each task was requested on multiple occasions throughout the drive. One task was to adjust the cabin temperature setting by 5 degrees (either up or down) from its current setting. A second task involved reading aloud 6 random digits, presented sequentially at a rate of 2 digits per second, on a video display mounted in the center stack area. The third task was a Drawsome task that required the participant to "connect the dots" of a figure presented on an iPad mounted near the center console. For this task, the participant used his/her finger to trace a path between points to enclose the figure without taking any direct-line path more than once. Drivers in the manual or baseline condition were required to complete the same secondary tasks as those in the automated-to-manual handoff condition. Note that before and during the critical event, there was no secondary task underway.

3.3.4 Experimental Design

The study was designed as a one-trial-per-participant, completely randomized, between-subjects experimental design with two independent factors:

1. Drive mode: Manual versus Automated
2. TTC to stopped obstacle: 1.7 s, 3 s, 5 s, or 7 s

This set of independent variables creates $2x4 = 8$ experimental conditions. Within each condition, $n = 8$ participants were assigned according to a plan that provided for equal gender and age distribution. The total number of participants for a balanced design was $8x8 = 64$ persons.

3.3.5 Response Variables

The dependent variables were a) driver response time from event onset until first response onset and b) the nature of the response (steer only, steer-and-brake, brake only, do nothing). Driver eye-glance video was also captured but used only to verify whether participants were looking toward the road scene at hazard onset.

3.3.6 Simulated Driving Scenario

The simulated driving scenario was essentially the same as that described in detail in Blommer et al. (2017) with the following exceptions. The lead vehicle was changed from a black van to a white box delivery truck. Also, the stopped vehicle was changed from the blue Mercedes sedan to a yellow Ford Fusion. These changes were made to increase the vehicles' visibility at greater following distances and TTCs. The timing of the hazard onset for the 1.7 s TTC condition was identical for both studies (see Figure 3.1). However, for the 3.0, 5.0, and 7.0 s TTC conditions, the inter-vehicle separation between the subject vehicle and the lead vehicle was increased so that at the FC event, the distance to the stopped object was appropriate for the assigned TTC treatment. In both studies, the participant had experienced a variety of inter-vehicle distances so that he/she did not encounter a unique situation at hazard onset. In both studies, the lead vehicle changed lanes leftward.

Near the end of the simulator drive, an "uncovering" event suddenly occurred, revealing a stopped vehicle ahead that the driver had to avoid (if driving manually) or take over from automation and then avoid. Figure 3.1 shows the uncovering event. The lead vehicle started an aggressive swerve maneuver to reveal a vehicle stopped in the travel lane at a time-headway of 2.7 s from the host vehicle. For automated driving, the automated driving system detected the stopped vehicle when the lead vehicle was approximately halfway across the lane line. Thus, the TTC with the stopped vehicle was predetermined to be either 7, 5, 3, or 1.7 s. If the FC alert in the HHDD and warning sound commenced, concurrently with the autobraking (Figure 3.1 right panel), it meant that the vehicle was at or below a 2.1 second TTC, which is standard FCW warning time in production Ford Motor Company vehicles (if equipped). (Note that in manual driving, there were no alerts or autobraking unless the TTC was 1.7 s) Drivers could, of course, see the event whether they were driving manually or with automation. In all instances, the driver needed to respond manually to the emergency (FC) event.

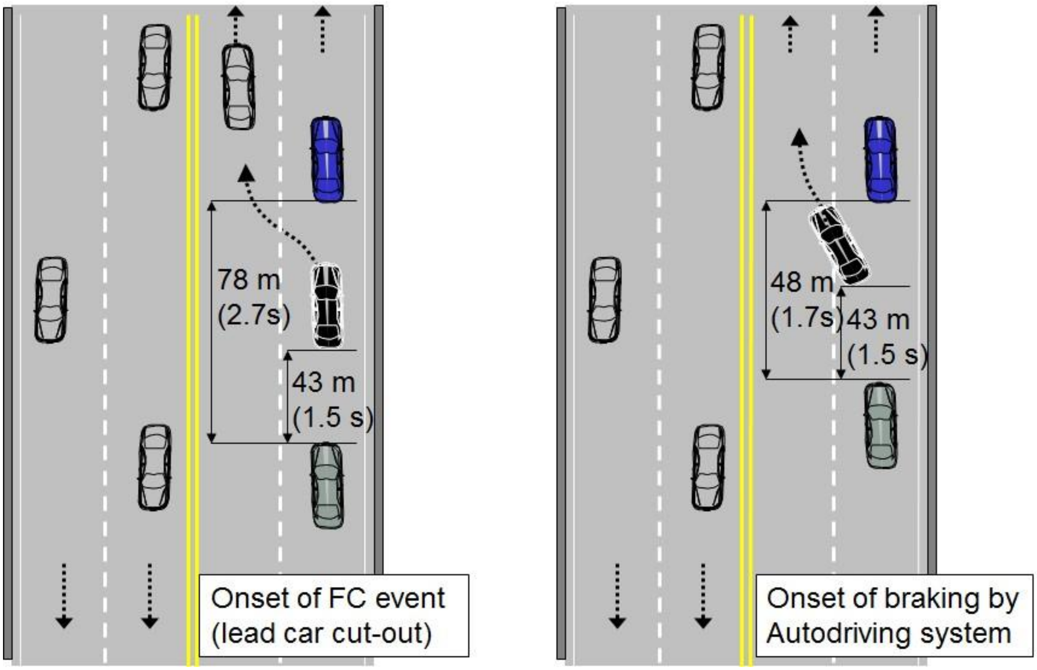

Figure 3.1: Uncovering Maneuver at FC Event

3.3.7 Procedure

Test participants arrived at the VIRTTEX simulator facility, signed the informed-consent form, reviewed a safety video, and were shown a presentation that introduced the automated driving concept. The introduction explained in general terms how an automated vehicle might operate to manage longitudinal control, lateral control, and car following separation. The training included slides that indicated that the system could not necessarily handle all driving situations. Drivers were told their first responsibility was to safely control the vehicle at all times even when automation was engaged. The slides also introduced the secondary tasks that the drivers would need to perform. No mention was made of a forward collision warning or event. After this, each participant was accompanied to the VIRTTEX simulator cab.

After the driver was comfortably seated in the simulator, training began, while the driver was stationary, with review of the requested secondary tasks. The ride-along observer, seated in the back passenger seat, then remained silent until the end of the simulator session while the simulator operator continued the driving portion of the training. The driving portion of the training involved practicing the secondary tasks while driving, getting a feel for making lane changes in the simulator, and learning how to activate and deactivate automation. There were 4 lane changes for acclimation to the simulator handling characteristics and an additional 4 lane changes associated with deactivating automation by steering.

Half of all participants were assigned to manual driving for baseline comparison purposes. Drivers assigned to manual driving were told of this assignment at the end of the training period in the simulator just before the study drive began. Thus all participants received the same training (i.e., experiencing automation and learning to drive VIRTTEX) regardless of their assigned drive condition for the study portion of the drive.

The VIRTTEX operator in the control room and the ride-along observer in the back seat did not interact with the test participant until the study drive was completed, i.e., after the critical event was over. At that point, test participants were asked if they recalled receiving a warning and, if so, whether they could recall any details about the warning. Each participant also completed a Simulator Sickness Questionnaire (SSQ). Once the simulator was settled, both the participant and observer departed the VIRTTEX dome, and the participant was thanked for his/her participation. The entire experimental session took approximately one hour to complete.

3.4 Results

The data were reduced and manually verified for accuracy and quality. Compliance with the requirement that test participants have their heads up and eyes looking toward the road ahead was confirmed by review of the eye-glance video for the data reported here. The response time after the FC event onset and the nature of each participant's response were analyzed.

3.4.1 Response Time Results

Figure 3.2 presents the mean response times for each TTC by drive mode. Analysis of variance for a two-factor, completely randomized design revealed a significant main effect for mode ($F(1,56) = 3.99$, $p = 0.05$). Manual driving had a typically faster response time than fully automated driving ($M = 1.93$ vs. 2.26 s). There was also a main effect for TTC ($F(3, 56) = 42.93$, $p < 0.0001$). The mean response times for 1.7, 3.0, 5.0, and 7.0 s TTCs were 1.16, 1.47, 2.18, 3.56 s, respectively. Tukey post-hoc tests revealed that there was no significant difference in response times to the *TTC* = 1.7 s vs *TTC* = 3.0 s hazard conditions. However, *TTC* = 1.7 s was significantly different from the 5 and 7 s conditions. The *TTC* = 3 s condition was significantly different from both the 5 and 7 s conditions. Jointly, these results support the classifications of 3 s or shorter TTCs as crash-imminent and significantly different from 5 s or longer TTCs, which may be termed as crash-probable or crash-possible. With regard to the latter, Tukey's test indicated a significantly longer response time for the *TTC* = 7 s condition as compared to the *TTC* = 5 s condition. The mode by TTC interaction was not statistically significant ($F(3,56) = 0.33$, $p = 0.802$).

In this final chapter, all study results are summarized and interpreted. The LCA and FOST, with or without haptics, had a demonstrable effect on reducing brake-only behavior in both hands-off fully automated driving, hands-on shared control automated driving, and manual driving. The source of this effect involved familiarization with the displays and developing a mental set in normal driving that a driver may sometimes steer rather than just brake to get around an obstacle ahead. Recommendations for further development of the HMI concepts are suggested.

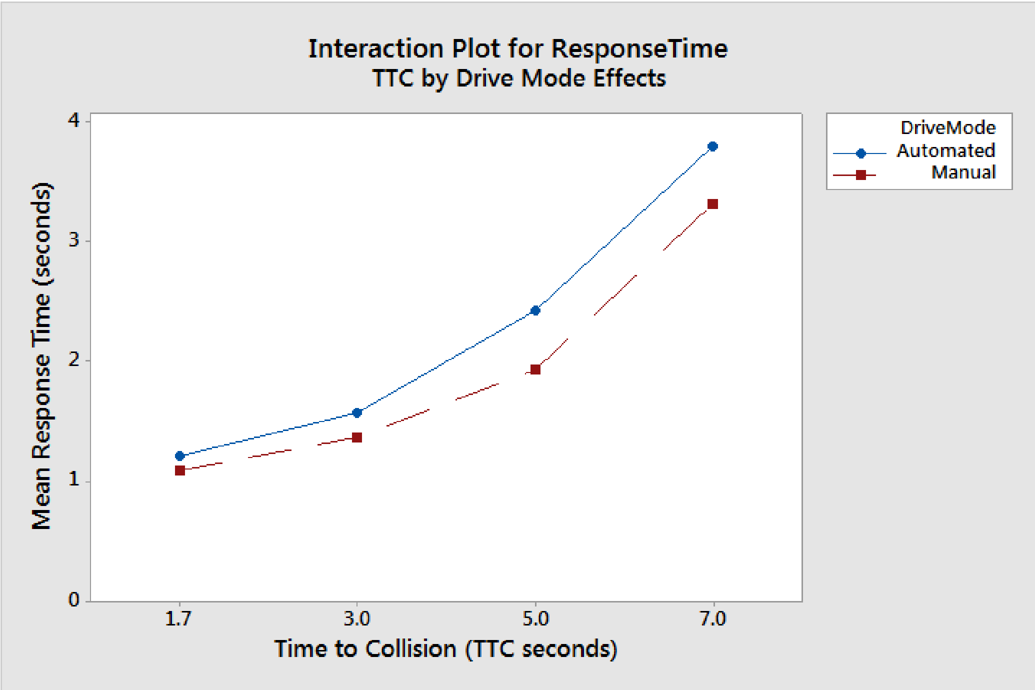

Figure 3.2: Response Time as a function of TTC and Drive Mode

3.4.2 Response Type Results

The Freeman-Halton extension of the Fisher exact probability test was used to evaluate response type differences between manual and fully automated driving at each TTC. Table 3.1 presents the results of the analysis in terms of the two-tailed probability that a difference exists between the two drive modes by response type. There was no difference in the distribution of response types by drive mode for the 1.7 s TTC condition. In contrast to the Blommer et al. (2017) findings, there were no brake-only responses in fully automated driving at 1.7 s TTC. At 3 and 5 s TTCs, there were significant differences in the distribution of response types, with more smooth steering (i.e., steer-only responses) exhibited in manual driving. Finally, there was a non-significant trend in the same direction for the responses at 7 s TTC.

Table 3.1: Response Type by Drive Mode per TTC

		Response Type			Freeman-Halton Test
TTC	Drive Mode	Brake Only	Brake & Steer	Steer Only	p-value
1.7 s	Manual	1	7	0	0.4999
	Automated	0	8	0	
3.0 s	Manual	0	3	5	0.0256
	Automated	1	7	0	
5.0 s	Manual	0	3	5	0.0401
	Automated	2	5	1	
7.0 s	Manual	0	2	6	0.1319
	Automated	1	5	2	

3.5 Discussion

The results of this study indicate that hands-off, fully autonomous driving is associated with slower response times at manual takeover than with manual driving. This difference in mean response times is approximately 330 ms and may be explained by the movement time required to move the hands from the lap to the steering wheel. For example, Niebel (1976) provides MTM-1 models that, for a 'reach' movement, predicts approximately 390 ms for a 14-inch precision hand motion (i.e., the distance from the lap to the 3 and 9 o'clock positions on the steering wheel). Given that there were also braking responses, the movement time is within a plausible range of the response time difference.

The TTC effects on response time were consistent with expectations. For alert drivers, it is sensible to expect quicker responses when there is less time remaining before contact. As indicated in Figure 3.2, there was no reliable interaction between TTC and drive mode. This suggests there are limits to the time required to manually take over in fully automated driving, as compared to manual driving, regardless of the driver's effort to respond quickly.

The TTC results also support a distinction between crash-imminent and crash-probable hazard conditions. There was no reliable difference between the first two conditions in terms of mean response time. However, each was significantly different from both the 5 s and the 7 s TTC conditions. This pattern of results in post-hoc paired comparisons suggests that a psychological difference exists between the shorter (1.7 and 3 s) TTCs and longer (5 and 7 s) TTCs. The two shorter TTCs may be considered crash-imminent conditions; the two longer TTCs may be termed crash-probable. Finally, a significant difference in mean response times between the 5 and 7 s TTC conditions may be indication of yet another distinction, between crash-probable and crash-possible (Allen, 1995).

In reference to driver response type, Figure 3.3 indicates overall agreement between this study and that of Gold et al. (2013) for the 7 and 5 s TTC conditions. The results for manual driving are similar to each other, and those for automation-to-manual handoff are also similar to each other. These TTC conditions might be termed crash-probable rather than crash-imminent.

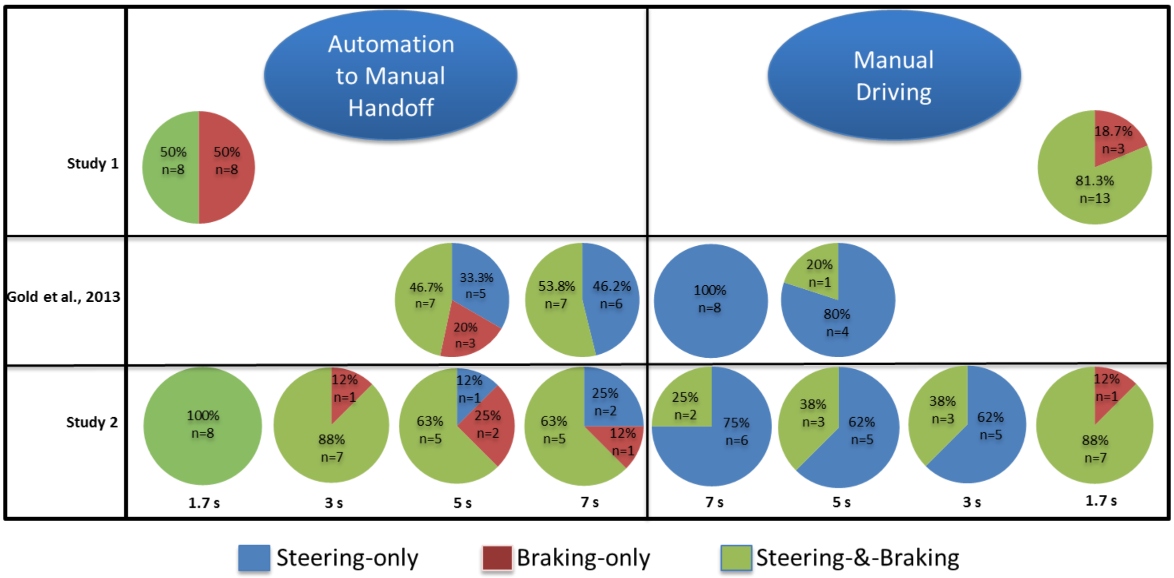

Figure 3.3: Driver Response Types by TTC and Drive Mode Across Several Studies

When compared to Blommer et al. (2017), this study's results for manual driving are more or less consistent. However, the results diverged for fully automated driving with manual takeover at 1.7 s TTC. As indicated in Figure 3.3, the earlier study resulted in a roughly 50−50 split between brake-only and brake-and-steer responses. In contrast, the present study resulted in no brake-only responses. In addition to the simulation differences noted in the Methods section, the procedure for this study introduced many more lane changes during training. In pilot testing, 3 participants, one each at 3, 5, and 7 s TTC conditions, all braked only. This unexpected result prompted a change in the simulator training to increase the number of lane changes requested. The increase was intended to provide better familiarization with the handling characteristics of the simulator and to increase steering and steering-and-braking responses. It had the unintended effect of completely eliminating brake-only responses in the 1.7 s TTC condition. This unexpected outcome is the focus of a later chapter.

In summary, drivers in fully automated driving responded later, and more often with either braking only or braking and steering, than manual drivers. An opportunity exists to facilitate faster responses and smoother steering in autonomous driving and manual driving as well. In autonomous driving, one way to reduce response time delays is to move from hands-off fully automated driving to some form of hands-on automated driving. In manual driving, there may be an opportunity to provide cues that might facilitate steering. The next chapter describes a HMI to accomplish this.

CHAPTER 4 HMI CONCEPTS

4.1 Abstract

Both auditory-visual and haptic steering displays were developed to facilitate steering around obstacles when appropriate. The theoretical foundations, design guidance, and specific design elements for auditory-visual as well as haptic displays are discussed.

4.2 Introduction

The development of a Human-Machine Interface (HMI) to facilitate driver lane change behavior in manual and automated driving is described. The HMI took the form of an auditory-visual Field of Safe Travel (FOST) display administered under both emergency conditions and benign conditions. The latter was accomplished through a feature referred to as the lane change adviser. An auditory-visual HMI was chosen because it could be compatible with both manual driving and hands-off, fully automated driving systems. Separately, a discrete haptic steering display was developed to augment the FOST display. This concept requires drivers to have their hands on the steering wheel and so is compatible with manual driving. In addition to manual driving, a form of automated driving exists that involves hands on the wheel while the automation is engaged; this is referred to as shared control. A discrete haptic steering display could also be compatible with a hands-on shared-control automated driving system.

4.3 Theoretical Motivation for a Field of Safe Travel Display

4.3.1 Gibson's Ecological Optics and Field of Safe Travel

Gibson (1960) developed a theory of direct perception based on the idea that the light in an optic array reaching the eye allows humans and other animals to move through the world. Humans are sensitive to higher-order invariants in the optic array information. For example, "τ" is defined as the visual angle subtended at the eye by an object one is approaching divided by rate of change in visual angle (expansion rate). TTC is defined as range divided by range-rate to an obstacle ahead and is a physical analogue to τ. Both τ and TTC are higher-order invariants that arise from specific relationships between lower-order observables (respectively, visual angle and expansion rate; range and relative speed). The study of ecological optics (Gibson's term) is to specify and measure the structure of the information available in optic arrays (both the fine structure and overall pattern).

Gibson demonstrates that laws exist by which some variables in the optic array specify environmental facts. For example, this theory holds that the invariants in an optic array specify relative motions of objects and the observer.

One key environmental fact identified by Gibson (1966) is an affordance. An affordance is an opportunity for action. In the case of driving, the principal affordance is in the form of sufficient space and time to move from one location to the next without collision. The perceptual process is dynamic and perceived affordances change from moment to moment. Gibson and Crooks (1938) provide a depiction (shown in Figure 4.1 below) of the affordances for safe passage while driving, termed a Field of Safe Travel (FOST). The elongation from the front of the rightmost vehicle through the field of other vehicles is a perceived one. Among other things, it indicates pinch points and potential hazards that a driver typically navigates smoothly through direct perception of the optic array elements.

A display of a steerable path might be considered a tailored form of Gibson's FOST. It is tailored in the sense that the path depends on the driver's intent as well as the current physical environment (i.e., other vehicles, pedestrians, roadside objects, etc.). In certain situations, a driver might focus on one apparent field of safe travel and ignore other options. The FOST display makes these options more apparent. For example, a driver in a FC situation will use visual looming to manage braking behavior (Pepping and Grealy, 2012). Because of the stress of the hazard, he/she may not notice that evasive steering is also feasible if the adjacent lane is clear.

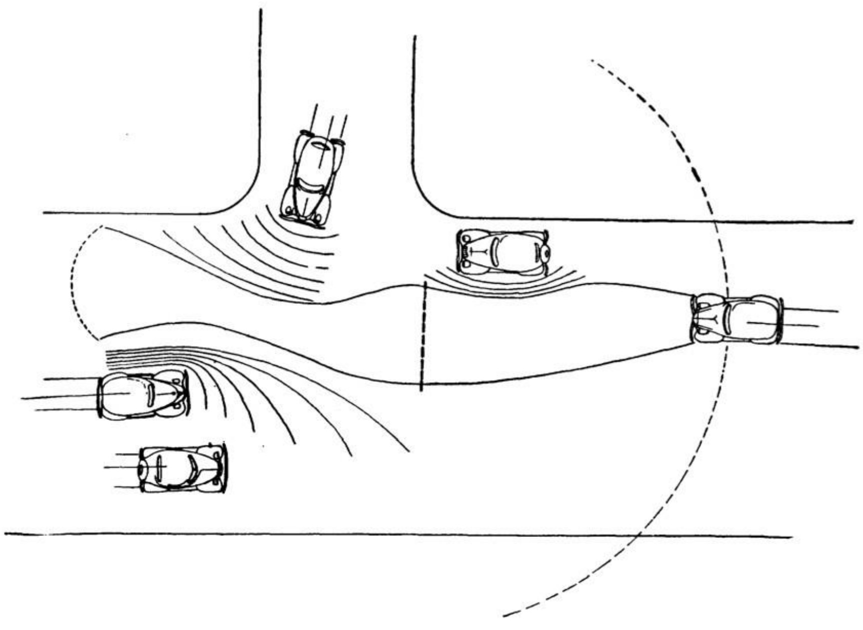

Figure 4.1: Gibson's Field of Safe Travel. Source: (Gibson and Crooks, 1938)

4.4 Field of Safe Travel (FOST) Display Elements

The FOST display developed as part of this dissertation is given in Figure 4.2. This display was delivered through the instrument cluster (located behind the steering wheel where the speedometer would be located). This location was used rather than a head-up display (HUD) to be compatible with design directions within Ford Motor Company that arose over the course of this project. In reviewing the design elements, it is important to keep in mind that the instrument cluster is directly in front of the seated driver. Therefore, display elements to the left or right are bilaterally spatially accurate with regard to the driver's spatial frame of reference.

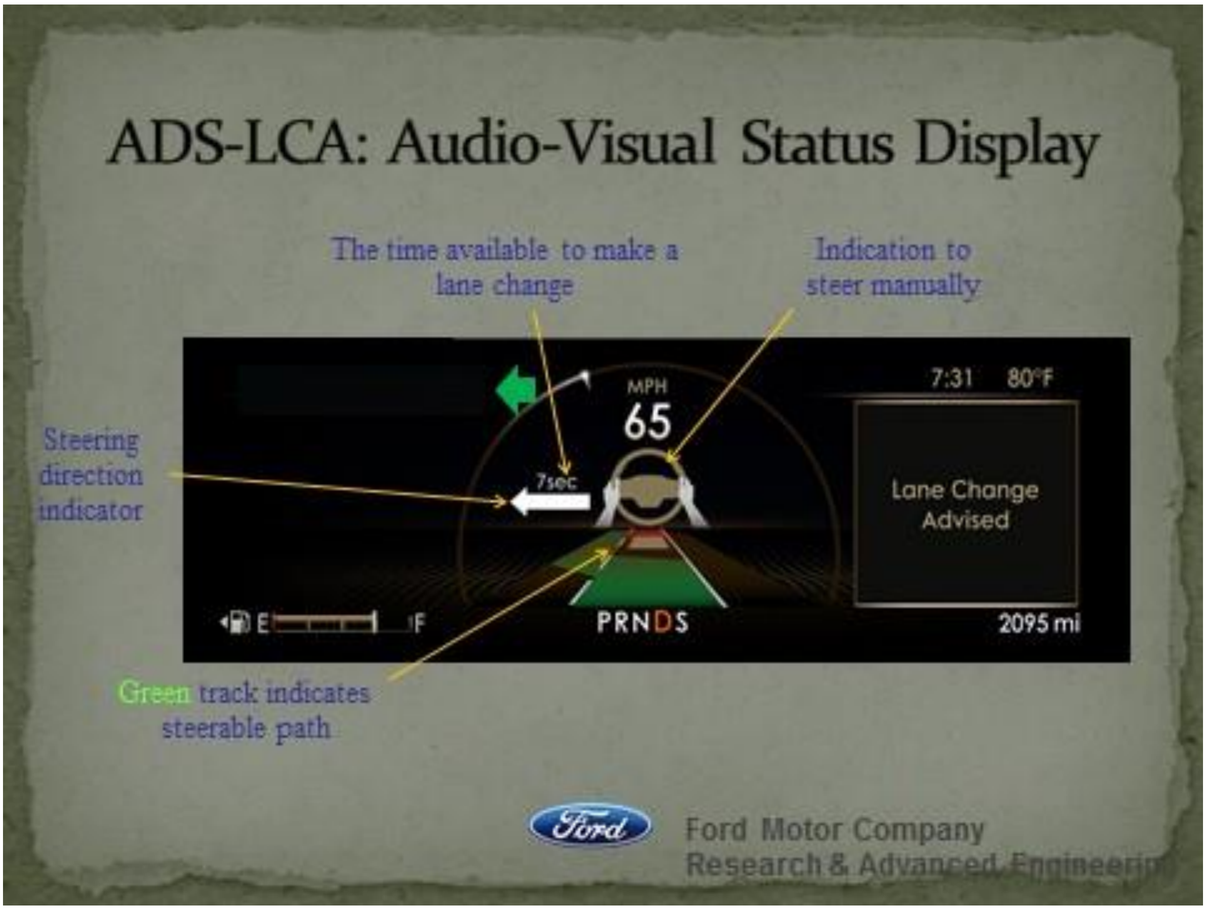

Figure 4.2: Auditory-Visual Elements of the Lane Change Adviser

4.4.1 Elements of the FOST Display

The display in Figure 4.2 contains a number of design elements presented as a discrete display. Starting from the top, a green turn signal arrow is shown; it is activated at the onset of the display. An auditory component in the form of the "tick-tock" sound of the turn signal accompanies the onset of the directional steering display. In addition to current speed, an icon of hands on the wheel is intended to convey the notion to steer manually. To the left of this is a white arrow, indicating the direction to steer, with the time available to make a lane change displayed above it. Beneath this is a roadway mimic. It depicts the lead vehicle ahead plus any additional vehicles that come into sensor range. Overlaid is green shading for the adjacent lane that indicates the steerable path, i.e., the field of safe travel. On the right, in the message center, is a concurrent verbal notification of the FOST display's meaning. The message center was originally intended for post-event notification. However, in pilot testing, an opportunity was identified to provide event notification, both during and shortly after the display concluded.

The timing of the display was meant to be brief, meaning that at the detection threshold for display onset, a recommendation would display for 1.5 s and then time out. The message center, on the other hand, would display an appropriate verbal message both during and after the display concluded for another 2.5 s. Other temporal elements of the display include the time available to make a lane change, time headway, and a depiction of TTC to the vehicle ahead.

4.4.2 Auditory-Visual Display Design Guidance

Wickens (2004) proposed several principals of display design used to guide the development of the FOST display. Of those, the categories considered for design of the FOST display are perpetual principles and mental model principles. The perceptual principle of redundancy gain states that a message is more likely to be interpreted correctly when the same message is expressed more than once, particularly when presented in alternate physical forms (Wickens, Lee, Liu, & Gordon Becker, 2004, p. 188). This is manifested in the display three times over: the white arrow, the turn signal, and the field of safe travel. It is also augmented by the haptic profiles discussed below. As for mental model principles, 2 in particular apply: 1) the principle of pictorial realism and 2) the principle of the moving part. The principle of pictorial realism states that the display should look like (i.e., be a picture of) the variable that it represents. If a display contains multiple elements, it is critical to arrange them to correspond with how they are configured in the environment or how the operator conceptualizes that environment. This relates to any dynamic elements such as the field of safe travel itself, the bilateral configuration of the display, and the representation of the obstacles (in this case vehicles) in the environment. The second relevant mental model principle, the principle of the moving part, states that the moving element(s) of any display of dynamic information should move in a spatial pattern and direction that is compatible with the users model of how the represented element actually moves in the physical system (Wickens et al., 2004, p. 189). The high stimulus-response compatibility between the directional turn signals and the steerable path in the world is an example of this. Relating to the FOST display, if the mental model of the driver is such that the vehicle moves left when the steering wheel is turned off center left, then the corresponding indicator in the FOST display should also move left or right with increasing steering wheel angle input. An example of the principle of ecological display might be redundant use of turn signal blinkers and a “tick-tock” sound, commonly associated with changing lanes.

4.5 Theoretical Motivation for Haptic Displays

“Haptic” comes from a Greek word meaning “to lay hold of” and refers to the perception of objects by touch and proprioception (Gibson, 1966). A haptic display, therefore, is one that is felt through the kinesthetic and tactile senses rather than being seen or heard.

A theoretical motivation for a haptic display is that it can contain high stimulus-response compatibility, also known as “idiomotor compatibility.” Greenwald (1972) hypothesized that the stimulus of highly idiomotor-compatible combinations should effectively select the response without burdening the limited-capacity decision processes of the central nervous system. Idiomotor theory proposes that responses are centrally coded by representations of their sensory feedback. The dimension denoting the extent to which a stimulus corresponds to sensory feedback from its required response will be referred to as “idiomotor compatibility.” Low compatibility is associated with greater response time or more response error; high compatibility has the reverse associations.
Greenwald’s work suggests an advantage to a stimulus that matches the kinesthetic-tactile feedback of the appropriate response. For instance, steering behavior might be prompted by a haptic steering display, i.e., a steering torque in the direction needed to drive around an obstacle (Tijerina et al., 2000). The perceived FOST and the direction in which the vehicle directs the driver to steer must never differ. Kelley (1968) presented a theory of hierarchical control that serves as the central theoretical foundation for the proposed research. Figure 4.3 below depicts Kelley’s hierarchical control model in relation to steering an automobile. The control system hierarchy comprises a series of events that are used throughout the control process, with smaller, more immediate events that employ less energy bringing about larger, more distant events involving more energy (Kelley, 1964, p. 24). The fundamental characteristic of the hierarchical control system is that the outer loop output (*X*) directly results from lower loop output (*Y*), that is directly controlled by the inner loop (*Z*) output. The progression is from inner to outer loop as the control hierarchy is ascended. Kelley refers to an inner loop display as a command display. The discrete haptic profiles described below provide onset cues for steering and thus constitute an example of a

command display.

The outer loop is what people monitor in terms of a goal state. For lane changes this is lateral position in the adjacent lane. It is here, in the outer loop, that visual display concepts based on the FOST may help drivers choose a steering goal by helping them choose an escape path.

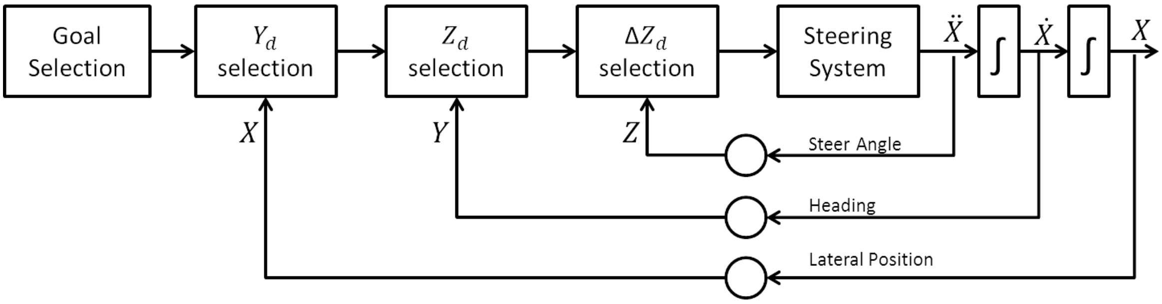

Figure 4.3: Hierarchical Control Model of Automobile Steering

The inner loop is what people control: in this case the steering wheel. As explained earlier, automobile steering is a second-order, highly cross-coupled system that can be difficult for humans to control in emergencies. This can be described from a hierarchical point of view as pertaining to the drivers strategy. First, the driver must set goals for the outer loop variable (e.g. lateral position). The driver's input to the steering wheel controls the lateral position through changes in heading. The vehicle's lateral position cannot be changed independently of the vehicle's heading. Then the driver must translate this outer loop goal into control of lower-order variables of the next-lowest level (i.e., heading). Finally, this places constraints on the inner loop (steering wheel angle) managed by an innermost controller element (steering wheel angle). This inner loop variable is controlled by only one element on an automobile: the steering wheel. Thus the inner loop was chosen to be affected by the haptic steering display. The control system, then, will have multiple feedback loops, both visual and haptic. Kelley (1968) refers to the inner loop display as a command display. A command display via the control element can be continuous or discrete. The command display that was developed was a discrete haptic display delivered through steering wheel torque profiles.

Humans control the innermost loop variable but monitor the result of the vehicle motions (i.e., the heading error or lateral displacement). Irwin Boers ideas (cite here) cover the inner loop, but not the outer loop, which can be observed in the form of lateral position. The outer loop is what the driver is trying to achieve. The inner loop is largely below consciousness. This is noted by the fact that drivers do not think about the angle at which they are turning the steering wheel at every instant as they drive. The goal of this dissertation research is to affect the monitoring aspect. Boer's research suggests that a haptic display affects the inner loop in an almost subconscious fashion. Thus a haptic display is hypothesized to more effectively influence driver strategy than a visual display alone. The innermost loop (steering manipulation) is the state that is most rapidly changing, reflexive in nature, and suitable for automation support like ESA. ESA both supports and automates innermost loop control. If activated, ESA allows drivers to focus on goal selection and consciously monitor the outer loop error in relation to their goal state (lateral position).

4.6 Haptic Display Elements

Described herein is the theoretical motivation for shifting from a hands-off, feet-off driving paradigm to one of hands-on shared control automated driving.

4.6.1 Haptic Displays and the Active Control Element

Haptic feedback and driving have been investigated on numerous occasions, including haptic driver seats (Fitch, 2008; Fitch, Hankey, Kleiner, & Dingus, 2011; Fitch, Kiefer, Hankey, & Kleiner, 2007; Kochhar & Tijerina, 2006), pulse braking displays for forward collision warning (IVBSS project, Sayer, et al. 2012), haptic steering wheels for forward collision avoidance (Schumann, 1985; Tijerina, 1999), and continuous ‘shared’ control in automated driving (Mulder, Abbink, & Boer, 2012). The intent of this dissertation is to build upon this research and develop an ecologically valid discrete haptic display that suggests a lane change when appropriate.

The literature is divided into two types of haptic feedback that can be provided through a steering wheel: a haptic vibration and a directional haptic torque (Beruscha et al., 2011). Vibrations are haptic signals induced by motors with eccentric weights such as those used in mobile phones. The other set of signals is based on a torque supplied by an electric motor, applied to, and perceivable, at the steering wheel. The direction of torque oscillation corresponds to the direction of steering wheel rotation.

Currently, Forward Collision Warning system warnings are presented to the driver via the auditory and visual channels and are intended to get the driver to look up and react. The reaction of the driver is assumed to be braking. Recent studies indicate that FCW is effective and FCW with automatic braking is even more effective (Cicchino, 2016). However, the focus of this dissertation is a warning system to prompt steering rather than braking only. Studies have shown that haptic alerts can bypass auditory and visual distractions to deliver more effective warnings (Ulusoy and Sipahi, 2014). To advise steering, the planned steering wheel torque should be unambiguous and intuitive. That driver’s action should be to steer in the direction of the presented torque.

Regarding the haptic vibrations and torques, the shape, amplitude, and frequency of the signal have to be distinguishable from road noise, vibrations, and feedback through the steering wheel, commonly referred to as Noise Vibration and Harshness (NVH). This needs to be accounted for in the development and tuned accordingly, as NVH is mostly engineering, but often objective measurements fail to predict or correlate well with the subjective impression on human observers.

The intensity of a proprioceptive-tactile (haptic) warning signal via the active steering wheel should be determined together with the signal's shape. Of course there must be a minimum intensity; since random jolting at the steering wheel is always possible, the proprioceptive-tactile warning signal has to have a unique shape that creates a clearly perceptible signal-to-noise ratio between the warning signal and possible random noise at the steering wheel (Joseph Schumann p 124).

4.6.2 Haptic Steering Patterns

Per recommendations for haptic steering alerts provided by Campbell et al. (2007), the fundamental building block for the haptic steering patterns was a symmetric triangle wave of torque over time. Ten different proprioceptive haptic wheel profiles were developed to be tested. These were divided into six "crash-cautionary" and 4 "crash-imminent" subsets. The haptic steering pattern profiles are depicted in Figure 4.4, and their properties are given in Table 4.1.

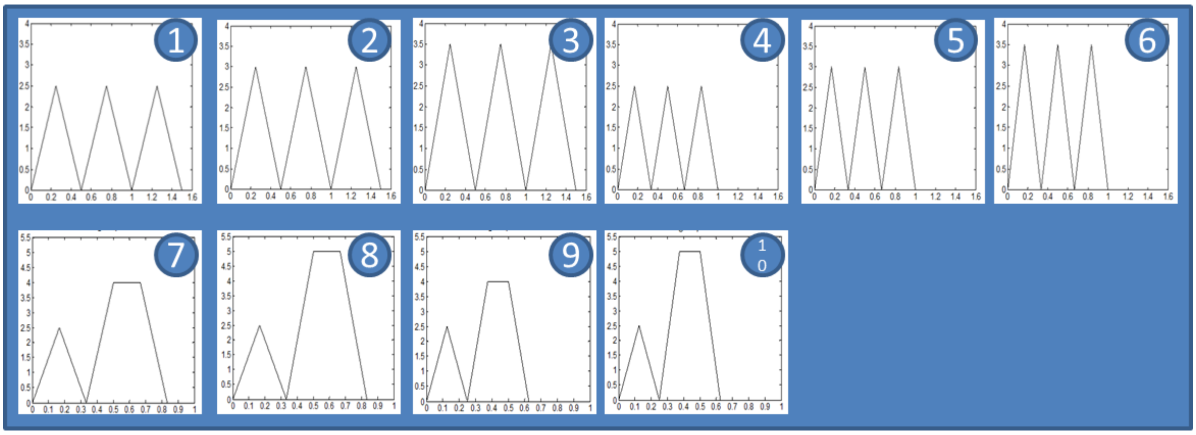

Figure 4.4: Proprioceptive Haptic Wheel Profiles

Table 4.1: Haptic Profile Properties

Object #	Label	Total Duration (sec)	Number of Pulses	Peak Amplitude (Nm)	Angular Impulse (Nm*sec)
1	Crash-Cautionary	1.50	3	2.5	1.875
2	Crash-Cautionary	1.50	3	3.0	2.250
3	Crash-Cautionary	1.50	3	3.5	2.625
4	Crash-Cautionary	1.00	3	2.5	1.250
5	Crash-Cautionary	1.00	3	3.0	1.500
6	Crash-Cautionary	1.00	3	3.5	1..750
7	Crash-Imminent	0.84	2	4.0	1.752
8	Crash-Imminent	0.84	2	5.0	2.085
9	Crash-Imminent	0.63	2	4.0	1.314
10	Crash-Imminent	0.63	2	5.0	1.564

The six crash-cautionary patterns were developed to be suitable for a wide range of Time to Collision (TTC) conditions. That is, patterns were developed that might work for a lane change advisor operating during non-emergency conditions as well as for emergency steering assistance. Crash-cautionary patterns were always 3-pulse triangle wave patterns (thought to be distinct from road noise such as running over a pothole). They did, however, vary in duration and peak torque. A 1.5 s presentation (the duration of Ford Motor Company's FCW warning signal on current production vehicles) was contrasted with a 1.0 s presentation because, within limits, more rapidly repeating stimuli are thought to convey greater urgency (Campbell et al., 2007). The peak torques for the triangle wave patterns varied from 2.5 to 3.5 Nm. This is within the range of directional torque magnitudes used in prior research for Lane Departure Warning (LDW) systems (Beruscha et al., 2011). The crash-cautionary patterns were proto- typed in a real vehicle and reviewed by a jury of human-factor experts while driving on public highways. Note that "Angular Impulse" in Table 4.1 refers to the total area under the torque wave curve for each torque profile. Since each profile was presented as a combination of torques delivered over periods of time, this is another way to examine the physical nature of each profile.

Four additional crash-imminent patterns were developed to convey an even greater sense of urgency (see Table 2 and Table 4.1). These patterns were intended particularly for high-intensity conflicts with relatively short TTCs (i.e., 3 s or less), so the total durations were shortened to either 0.63 or 0.84 s and consisted of 2 pulses each. In each of the four crash-imminent patterns, a short-duration directional triangle wave was presented as a pre-cue to the driver. This pre-cue was followed by a second directional torque, with the peak torque (either 4.5 Nm or 5 Nm) held briefly to increase the salience of the directional cue. The shorter durations also limited the amount of vehicle yaw induced by the torque hold.

While research led to the selection of a triangle pattern for the final profile development, it wasn't without trying a multitude of patterns first to see what "felt" right. Although the author trusted that the literature was sound, he felt it was worth experiencing the gamut firsthand to see why they might not be as good as the triangle pattern. This was purely as part of his learning experience and not as part of any controlled experiment or formal research study.

Literature shows that an oscillatory torque can be confusing to drivers unless the frequency is high enough to be perceived as a vibration and hence doesn't fit the intended design intent of this dissertation. In addition, an oscillatory torque is non-directional as it crosses the zero point. This was another requirement of the discrete torque: that it provide to the driver the intended direction of the recommended lane change.

4.6.3 Prototyping Apparatus

The apparatus used for prototyping was a production vehicle, a 2013 Lincoln MKZ, modified with a prototype steering rack, differential Global Positioning System (dGPS) receiver, and dSpace microAutobox. The development environment used was Matlab 2013a for the Simulink model development and the MicroAutobox for the implementation of the haptic pattern prototypes on a vehicle.

4.7 Lane Change Adviser

The third element of the HMI is referred to as a Lane Change Adviser (LCA). The LCA uses algorithms to advise the driver of a lane change opportunity when encountering slower traffic ahead and a clear adjacent lane. Its purpose was to familiarize the driver with the HMI in benign circumstances. This was thought to be especially important for the haptic steering display to distinguish it from some type of malfunction. It is a system concept that integrates both the FOST and the haptic steering profiles into a unified concept that might work for both normal driving and FC hazard scenarios.

CHAPTER 5 STUDY 3, PART 1: DRIVER HAPTIC STEERING ALERTS AND URGENCY SCALING

5.1 Abstract

Pair comparisons are commonly used in the behavioral sciences to gather data with which to build a psychological scale for a dimension of interest. This paper reviews statistical methods to check the following key attributes of 2-alternative forced choice (2AFC) pair-comparison data: a) reliability of a judges repeated assessments of the same stimuli; b) the consistence of a participant's data in terms of transitivity of judgments; c) concordance among participants; and d) stimulus assessment for outliers before application of a scaling routine. The chapter includes an application of the 2AFC pair-comparison method to alternative haptic steering wheel displays where the dimension of interest is the urgency or imperativeness of changing lanes.

5.2 Introduction

This chapter investigates pair comparisons, consistency checks, and scaling exemplified in developing a scale of urgency for driver haptic steering alerts. In subjective assessment experiments, the individual participant is the "instrument" used to measure a stimulus on a psychophysical or psychological dimension of interest. Subjective assessments are preferable when participants are better able to perceive certain attributes of a stimulus (e.g., a HMI); cases where suitable objective measures do not exist; cases where the subjective assessments are easier to obtain within time, budget, or other constraints yet are reliable and valid).

Subjective assessment data are often used to construct a unidimensional scale on which stimuli may be arrayed. For such scaling, the simplest data collection technique is the 2AFC pair comparison. A participant is presented with a pair of stimuli and indicates which stimulus in that pair dominates on the dimension of interest. The 2AFC pair-comparison procedure is attractive because people often make a better binary, relative judgment on a dimension of interest compared to the numerical estimates that would be required with rating scales, magnitude estimations, and so forth (Stillwell and Seaver, 1983). A major drawback to the pair-comparison method is that the number of pairs grows as the square of the number of stimuli to be examined. Therefore, the method is generally only used when the number of stimuli is small (e.g., 10 or fewer). On the other hand, in a classic paper on pair comparisons and scaling of social values, Thurstone (1927) had students compare 19 crimes in a total of 171 pairs and indicate which was more serious.

The 2AFC pair-comparison method of scaling has been used in a wide variety of applications in human factors and applied ergonomics. Applications include vehicle design (Bhise, 2011), workload assessment (Hart and Staveland, 1988), warning design (Wogalter et al., 2002), four-burner stove design (Hoffmann and Chan, 2011), preferences for drowsy-driver alarms (Fairbanks et al., 1995), human reliability estimates (Stillwill and Seaver, 1983), video image quality (Satgunam, Woods, Bronstad, and Peli, 2013), visual discomfort with 3D television (Li, Barkowsky, and Le Callet, 2013), realism of alternative moving-base simulator motion drive algorithms (Grant et al., 2009), and cell-phone features preferred by senior citizens (Chen et al., 2014). The International Organization for Standardization (ISO) has also developed two methods for prioritizing in-vehicle messages, one of which is for expert evaluators to make pairwise comparisons among all possible pairs of messages, determining which of each pair should receive priority, to create a priority scale with messages arrayed along it (ISO, 2004).

Despite its popularity, published applications of 2AFC pair-comparison procedures and scaling seldom make use of several available statistical tools to test the assumptions of the methods involved. Key references for such tests are Dunn-Rankin et al. (2004), Cliff and Keats (2003), Maxwell (1974), and Kendall (1975). The purpose of this chapter is to review key assumptions for scaling based on 2AFC pair-comparison data, illustrate some of the available tools to test those assumptions, ordinally rank the stimuli and statistically evaluate the pairwise ordinal differences among them, and apply a common interval-level scaling solution known as the Bradley-Terry model along with assessments of the goodness of fit of the solution to the data. These methods were used to develop a psychometric scale of “urgency” for haptic steering displays intended to promote a lane change when appropriate. The goal of this chapter is to improve the quality of the research using the 2AFC pair-comparison procedure and identify gaps in current knowledge regarding certain scaling assumptions.

The 2AFC Pair-Comparison Method

In a pair-comparison study, a set of n stimuli are presented two at a time to a participant (David, 1963). A participant is typically presented a total of $\binom{n}{2} = \frac{n(n-1)}{2}$ unique pairs in an evaluation of n stimuli or stimuli. The participant decides which stimulus in a pair “dominates” the other with respect to some dimension of interest. Dominance is observed when two stimuli are compared with respect to some attribute and one is judged to be greater in that attribute than the other (Cliff and Keats, 2003). For analytical tractability and ease of application, 2AFC pair-comparison studies use a forced-choice format. Either *A* dominates *B* or *B* dominates *A*; ties and “can’t say” answers are not allowed (the participant is instructed to arbitrarily pick a stimulus in such cases). A pair of stimuli is generally presented only once. Only in the occasional study is participant reliability checked by presenting the participant with the same pair of stimuli again but with the presentation sequence reversed. Finally, a stimulus is never compared against itself, as this violates the forced-choice dominance concept. As mentioned earlier, pair-comparison data often form the basis of a psychometric scale of the dimension of interest. The dimension of interest can be quite varied; they include product preference, apparent loudness, subjective expected utility, alarm urgency, and social desirability, to name a few.

The number of stimuli, manner of presentation, and sequencing of stimuli must be carefully considered in a pair-comparison study. The number of pair comparisons grow as roughly the square of the number of stimuli to be judged and thus must be limited if all pairs are to be presented, for reasons explained hereafter. How many stimuli might be evaluated in pair comparisons depends on what the stimuli are (sounds, sights, words, tastes, etc.), how they might be administered without undue fatigue in the participants, whether a pair will be presented more than once, and whether incomplete but orderly subsets of stimuli can be presented to subsets of participants. Regarding manner of presentation for perceptual stimuli (e.g., haptic displays), it is important to present pairs of stimuli with an Intertimulus Interval (ISI) long enough to keep the stimuli distinct but short enough that the first stimulus is not for- gotten. Pilot testing is the best route for determining the appropriate acrshortisi. For proper sequencing, stimuli within a pair should be counterbalanced (e.g., *AB* and *BA*, *AC* and *CA*) for all participants, if a pair is presented more than once. Furthermore, the sequence of pairs should be randomized so as to minimize systematic bias or carryover effects. A Matlab function developed by Tillemans (2011) was used to generate what are known as Ross sequences (Ross, 1934). Ross sequences provide pair orders such that a given stimulus is presented to a participant first and second an equal number of times over a single complete pair comparison if the number of stimuli being compared is odd. It will also generate pairs when the number of stimuli being compared is even, but equal number-presentation orders for all objects cannot be achieved in a single full presentation because $\frac{n(n-1)}{2}$ is odd whenever n itself is even. In such a case, reversing the pair orders for half the test participants works to counteract any order effects (e.g., people's tendency to pick the last object presented).

When there are too many stimuli to present to a single participant, alternatives to complete pair comparisons have been proposed, each based on assumptions. Dunn-Rankin et al. (2004) discuss balanced incomplete blocks (BIB) designs that define small subsets of the stimuli (e.g., 3 stimuli per subset) grouped so that all possible paired comparisons are inferred from the ranking of objects in each small subset. However, if the researcher seeks to maintain pair comparisons and avoid direct rankings of 3 or more stimuli for logistical or other reasons, see Chapter 8 of Cliff and Keats (2003) for details on procedures based on assuming transitivity among pair judgments.

After $(k(k-1))/2$ pair comparisons for $k = 1, n$ stimuli are collected, they are organized in a $(k * k)$ matrix called an adjacency matrix. For a participants data from a single set of comparisons, the adjacency matrix has a 1 in each (i, j) off-diagonal cell to indicate that a given row stimulus dominated a given column stimulus, i.e., $i > j$, or a 0 otherwise; the diagonal generally is ignored or may be filled with zeros. By symmetry, if (i, j) has a 1 in it, then (j, i) has a 0 in it and vice versa. For a group or ensemble of participants, the adjacency matrix contains the total number of votes from N participants. The total number of votes in a composite adjacency matrix is N times the total number of pair comparisons presented to each participant for evaluation. The adjacency matrix forms the basis of the data analysis for pair-comparison scaling.

Consistency Checks for 2AFC Pair-Comparison Data

Before constructing a psychometric scale for the dimension of interest, it is useful to assess the quality of the pair-comparison data by answering a number of questions (Dunn-Rankin et al., 2004); (Cliff and Keats, 2003). The answers to the first two of the following four questions identify participants who should be removed from further analysis. The answer to the third question identifies stimuli that should be removed from further analysis. The answer to the fourth question indicates the degree to which there is a common underlying dimension for scaling.

1. How reliable is a given participant when asked to compare the same pair of stimuli twice?
2. How consistent is a given participants pair-comparison data in terms of transitivity of judgments?
3. Are there stimuli that are over-involved in inconsistent judgments?
4. To what extent is there a unitary scale, i.e., to what extent do participants concur in their

rankings of the stimuli?

Participant reliability can be assessed by asking a participant to compare each pair of stimuli a second time. Generally a “complete” pair-comparison study has only $\frac{n(n-1)}{2}$ pair comparisons, i.e., only one unique comparison per pair. Thus, judgment reliability is rarely tested. However, by presenting each pair of stimuli twice, one can evaluate reliability of judgments. The same stimulus in a pair is expected to be selected both times in half of the pairs by chance, but may differ for the other half of the pairs (all within sampling variation). The binomial sign test allows for an exact test of the null hypothesis of random responding against a directional alternative hypothesis that the reliability is better than $50 - 50$. High reliability across pair comparisons presented twice (with stimulus order reversed) would support the notion of orderly responding from a competent participant. Otherwise, the null hypothesis of random responding would be retained and would be grounds to drop a participant’s data from further analysis. Any of a number of commonly available statistical analysis programs can be used to carry out a sign test. As a simple example, consider a participant who has rated 4 items in $\frac{4(4-1)}{2}$ or 6 pair comparisons twice each and given matching responses for 5 out of 6 of the pairs. The null hypothesis is that the participant’s responses are random (i.e., p = 0.5 for match). The one-tailed directional alternative hypothesis is that the reliability is better than chance and this is supported by a sign test with a p = 0.1094.

Consistence is the term given by Kendall (1975) for consistency measured in terms of circular triads or lack of transitivity. Intransitivity in judgments creates tied ranks or scale values in scaling solutions. Consider three stimuli, *A*, *B*, and *C*, and the symbol ‘>’, which means a judge reports that the stimulus to the left of the symbol dominates the one on its right on the dimension of interest. If $A > B$ and $B > C$, transitivity logically implies that $A > C$. Transitivity in pair comparisons will result in a linear order $A > B > C$. However, it is sometimes the case that a participants complete pair comparison produces one or more intransitivities such as $A > B > C > A$. An intransitivity is also called a circular triad because the geometrical representation is a triangle in which all the dominance arrows go around in the same direction. Kendall and Smith (1940) derived g, the maximum number of circular triads among n stimuli presented for comparison in $\frac{n(n-1)}{2}$ pairs. As stated above, for a single participant, the pair comparisons for n stimuli can be organized into an $(n * n)$ matrix known as an adjacency matrix. Each cell in the adjacency matrix is filled with a 1 if the row stimulus dominates the column stimulus for that cell and a 0 otherwise. If the (*i*, *j*) cell has a 1, this implies that the (*j*, *i*) cell has a 0, and vice versa. (The adjacency matrix diagonal is left blank or filled with 0s.) The sum across rows for a given column stimulus, a_j, is an ordinal indicator of the column stimulus on the dimension of interest. This value reflects the number of comparisons in which that stimulus dominated the others in the set being evaluated. The value *g* is evaluated as follows:

$$g = \begin{cases} \dfrac{(n^3 - n)}{24}, for\ n\ is\ odd \\ \dfrac{(n^3 - 4n)}{24}, for\ n\ is\ even \end{cases}$$

Kendall (1975) also showed that *d*, the number of circular triads actually present in a participant's data, can be calculated from a participant's adjacency matrix of results as follows:

$$d = \frac{n(n-1)(2n-1)}{12} - \frac{1}{2}\sum a_j^2$$

where a_j is the sum of 1s in a row of an $n * n$ adjacency matrix for stimulus j.

Kendall (1975) then defined a coefficient of consistence, here termed zeta, with a range from 0 to 1.0 as an index of the degree of transitivity in pair comparisons achieved by a participant:

$$zeta = \frac{1 - d}{g}$$

For example, for *n* = 6 stimuli compared on some dimension, $g = \frac{(6^3 - 4(6))}{24} = 8$. If a participant's data show 5 circular triads, then $1 - \left(\frac{5}{8}\right) = 0.375$. Only if there are no circular triads will zeta equal 1.0, and if every possible circular triad is present, then zeta will equal 0.0. A statistical assessment is possible for the presence of circular triads under the null hypothesis that a participant is responding randomly, compared to the directional alternative hypothesis that he/she is responding more consistently (i.e., with fewer circular triads). Tables of exact probabilities for up to *n* = 10 stimuli have been published by Kendall (1975) and others (e.g., Dunn-Rankin et al. (2004); Iida (2009)). For larger sets of stimuli (i.e., more than 10 objects), Chi-Square approximations are available (Dunn-Rankin et al., 2004). The relevant one-tailed or directional alternative hypothesis is that the participant is generating fewer circular triads than might be expected by chance or with random responding. If the null hypothesis of random responding is rejected in favor of the directional alternative there are fewer circular triads than expected with random responding (e.g., 0.05 alpha level), then the data should be retained for further analysis. Otherwise, that participants data might be omitted. The TRICIR program included in Dunn-Rankin et al. (2004) calculates the number of circular triads present in a participant's data.

Kendall (1975) presents a table of the exact probability (rounded to 4 decimal places) of observing *d* or fewer circular triads in an ajaceny matrix for *n* stimuli under the null hypothesis that the judgments are random. The table indicates *f*, the frequency of observing *d* circular triads in a single complete set of *n*(*n* – 1)/2 pair comparisons. For example, for *n* = 3, there are 3(3-1)/2 or 3 pair comparisons (*A* – *B*, *A* – *C*, and *B* – *C*). Therefore, any given participant may only generate at most one circular triad from his or her three pair comparisons. But there are *f* = 2 ways in which a circular triad might arise: judgments of *A* > *B*, *B* > *C*, but *C* > *A* (i.e., *A* > *B* > *C* > *A*) or judgments of *C* > *B, B* > *A,* but *A* > *C* (i.e., *C* > *B* > *A* > *C*). There are also *f* = 6 ways in which judges will be consistent and generate no circular triads: *C* > *B, B* > *A, C* > *A; B* > *C, C* > *A, B* > *A*; *B* > *A, A* > *C, B* > *C*; *C* > *A, A* > *B, C* > *B*; *A* > *C, C* > *B, A* > *B*; and *A* > *B, B* > *C, A* > *C*. Thus, the probability of generating no circular triads is 6/8 = 0.75 and probability of generating 1 circular triad is 2/8 = 0.25. These exact probabilities are derived from several sources. Iida (2009) presents the results for *n* up to 9, whereas Kendall (1975) presents exact probabilities for up to 10 stimuli. Dunn-Rankin et al. (2004) present cumulate probability tables for 5 to 15 objects but omit the extreme values of *d* that may nonetheless be of interest to some researchers; Kendall (1975) shows all values of *d* for each number to be compared. Kendall (1975) provides the original and most complete source of probabilities for up to 10 stimuli. However, his probabilities indicate the cumulative probability that a given value of *d* will be exceeded, the reverse direction to what an applied researcher would normally require. For larger stimulus sets, a Chi-Square approximation implemented in the TRICIR program (Dunn-Rankin et al., 2004) works well.

In addition to assessing the consistence of a participant, the stimuli being assessed can also be evaluated for over-involvement in inconsistent judgments. That is, stimuli may be analyzed for consistency of effects in pair comparisons on an ensemble of participants. Knezek (1978) considered the case of a stimulus involved in a relatively high number of circular triads where this pattern is found across multiple participants. In such circumstances, he argued that it is reasonable to infer that the stimulus has one or more characteristics that cause participants to make inconsistent judgments. For example, the stimulus may be sufficiently close to another stimulus in some attribute that it is beyond the resolution of participants to distinguish between them. The subsequently developed scale might be improved by dropping the troublesome stimulus from further analysis. The TRICIR program (Dunn-Rankin et al., 2004) calculates a Z statistic for each stimulus j, Z_j, based on the number of circular triads (#CTs) across all participants that a given stimulus was involved in and identifies stimuli with an unusually large number of circular triads at a 0.05 alpha level ($Z > 1.64$):

$$Z_j = \frac{(\#CT_j - mean(CT)}{stdev(CT)}$$

Agreement across participants may be assessed with Kendall's coefficient of concordance Kendall (1975). This coefficient, *W*, measures the degree of agreement among raters in terms of the rank-ordering of stimuli along the dimension of interest. It ranges from 1.0 (complete agreement) to 0.0 (no agreement). In a table if *N* participants with *K* stimulus rankings per participant, the sum of ranks are computed per stimulus. If there were complete agreement, the sum of ranks would be some multiple of the number of participants and would produce the largest variability. Kendall's *W* forms a ratio between the rank sum variability (the numerator) and the maximum variability (the denominator). If there are ties in the rankings, the formula used to correct for ties is as follows:

$$W = \frac{SS_{ranks}}{SS_{max} - (N * \frac{\sum t^3 - t}{12})}$$

where SS_{ranks} is the sum of square ranks; SS_{max} is the maximum sum of squares for the data; *N* is the number of participants; and *t* is the number of ties per participant. Kendall's *W* has a sampling distribution that can be used to test the hypothesis that all participants are independent in their judgments versus the alternative that they have some degree of commonality in their rankings of stimuli.

Psychophysical Scaling with Pair-Comparison Data

To scale things is to put them in some order, comparatively speaking, with respect to an attribute of interest (e.g., perceived urgency). Once the quality of the pair-comparison data is evaluated for unreliable or inconsistent participants, problematic stimuli, and the possibility that there is no single dimension to scale, scaling solutions may be sought. The scaling approach with the fewest assumptions ranks or orders the stimuli on the dimension of interest based on the pair comparison data. Dunn-Rankin et al. (2004) describe the variance stable rank method of ordinal scaling based on the assumption that scale values are proportional to the sum of ranks (i.e., number of votes) that participants assign to each stimulus. The ranks are "variance stable" because "a specific difference in rank sums has the same probability of occurrence wherever the rank totals (and the scales scores) may be located" (Dunn-Rankin et al., 2004, p. 57). This method of scaling allows the researcher to specify a nonparametric statistical test that involves a critical range in terms of votes. A difference in total votes between any two stimuli equal to or greater than the critical range value is reliably different beyond chance variation. In pair comparison studies, the sum of ranks is the number of votes or dominance judgments a stimulus gets across all participants and all other stimuli presented in pairs. For n stimuli and N participants, the total number of votes that will be cast is calculated as $N * \frac{n(n-1)}{2}$. The maximum number of votes possible for any single stimulus is$(n-1) * N$. For example, if there were K = 4 stimuli (A, B, C, D) and N = 10 participants, the adjacency matrix for all participants combined would be $10 * \frac{4(4-1)}{2}$ or 60 votes. The most votes that any given stimulus could receive across the group of participants is $(4-1) * 10 = 30$ votes. This might happen if, say, all 10 participants agree that $A > B$, all 10 agree that $A > C$, and all 10 agree that $A > D$. This is how Stimulus A gets $(4-1) * 10$ or 30 votes. This is an ordinal method, meaning the rankings are meaningful, not the magnitudes or "step sizes" of differences. However, variance stable ranks ranging from 0 to 100 may be calculated for each stimulus by dividing the number of votes by the total number of votes and multiplying that quotient by 100.

Researchers often desire a scale that not only rank-orders stimuli but also indicates how big the separation is between them (Grant et al., 2009). For this, scale additivity is required; Maxwell (1974) presents the example in Figure 5.1 to illustrate scale additivity. Three objects, A_1, A_2, and A_3, are arranged on a subjective preference scale at scale distances from arbitrary zero points of x_1, x_2, and x_3, respectively. With such a scale, "differences" between pairs of preferences should be additive, i.e., $A_1 - A_2 = x_1 - x_2$; $A_2 - A_3 = x_2 - x_3$; and $A_1 - A_3 = x_1 - x_3$. Additivity is demonstrated if $(A_1 - A_2) + (A_2 - A_3) = (A_1 - A_3)$. Maxwell (1974) points out that additivity is by no means assured. For instance, consider 100 judges making pair comparisons, with 90% preferring A_1 to A_2 (meaning 10% prefer A_2 to A_1) and 80% preferring A_2 to A_3 (so 20% prefer A_3 to A_2). By taking differences as before, $(A_1 - A_2) + (A_2 - A_3)$ = (90 − 10) + (80 − 20) = 80 + 60 = 140% = $(A_1 - A_3)$. Clearly, percentage preferences as scale values will not support additivity, since there are only 100 percent of participants to work with.

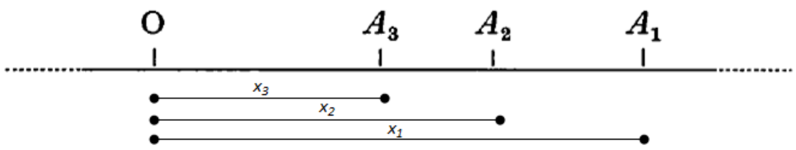

Figure 5.1: Scale Additivity Exampe. Source (Maxwell, 1974)

Empirical addition is the real-world concatenation operation that is the bedrock of fundamental measurement (Newman, 1956). The concatenation of measurement rods end to end to measure an objects length on a ratio scale, or the concatenation of graduated brass weights on one side of a pan balance to measure an object's weight (or mass) on the other side of the pan balance, is also justified on a ratio scale of measurement with a true zero. On the other hand, subjective assessments generally do not readily lend themselves to direct concatenation. However, the theory of conjoint measurement was developed to restore the operation of empirical addition to behavioral sciences by considering experimental design as a concatenation of "differences" rather than direct quantities (Coombs, 1983). As a simple example, from the Analysis of Variance (ANOVA), a main effect for a factor at two levels can yield a "significant difference"; an interaction between two such factors can yield a "significant difference in differences"; and so on (Cohen, 1988). In the framework of conjoint measurement, it is intervals, changes, differences among stimuli that are being counted. Without a real, absolute zero point, the derived psychometric scales would be at the interval level of measurement. Whether or not an interval scale of measurement is achieved depends on whether it meets the conditions of conjoint measurement theory (Michell, 1999). Unfortunately, the theory of conjoint measurement is currently without a corresponding error theory to distinguish random measurement or response error from systematic discrepancy with the additivity requirements, and the issue is beyond the scope of this paper. However, two popular approaches to psychometric scaling purport to provide interval scale of measurement.

Two of the most common scaling approaches for 2AFC data are Thurstone scaling and Bradley- Terry modeling. The older of the two, Thurstone scaling, is based on Thurstone's law of comparative judgment (Thurstone, 1927); (Dunn-Rankin et al., 2004). Thurstone postulated that a participant's reactions to any stimulus are subjective and vary randomly from moment to moment. However, there is a most frequent reaction, called the modal reaction. This reaction can be estimated from repeated judgments from a single participant or from the frequency of single judgments from many participants (assuming there is a common underlying subjective dimension of interest).

Thurstone further assumed that the subjective reactions were normally distributed. Because the mean and mode of a normal distribution are the same, the mean can serve as the scale value for a stimulus on the psychological scale or dimension of interest. The proportion of pair-comparison dominance judgments, e.g., stimulus $A > B$, are tabulated in an evaluation and, because of the normality assumption, are then expressed as standard normal deviates obtained from a standard normal table. The differences between pairs of stimuli are also obtained by use of the normality assumption. Specifically, the average z-score for each stimulus is computed Guilford (1954) (pp. 161-163) and this provides the interval scale of the subjective continuum of interest. Because interval scales have arbitrary origins, the z-score averages for the stimuli may be rescaled for convenience. For instance, one can assign the smallest (or most negative) average z-score to zero and shift all other scale values accordingly to obtain an interval scale of positive values.

As a scaling procedure, the Bradley-Terry (BT) model (Bradley and Terry, 1952) can be viewed as equivalent to the Thurstone model (Abelson, 1958). The principal difference is that the BT model assumes that the random quality difference $A > B$ (i.e., object A dominates object B) has a logistic distribution, but the Thurstone model assumes that the random quality difference $A > B$ is normally distributed (Tsukida and Gupta, 2011, p. 6). Logistic regression can be used to fit a BT model (SAS, 2005; Agresti, 2002). For any pair comparison, if n_{ij} denotes the number of votes out of N judges that object i dominates object j, if n_{ji} denotes the number of votes that object j dominates object i, and if $n_{ij} + n_{ji} = N$, then the logistic transform or "logit" of the probability that object i dominates j is given as follows:

$$\pi_{ij} = \ln(\frac{n_{ij}}{N - n_{ij}})$$

If β_1, β_2, β_3 and so on are regression coefficients from a logistic regression associated with each of the objects in the stimulus set being modeled, then the probability that object i dominates j is defined in So (1995) as follows:

$$\pi_{ij} = \frac{\exp(\beta_i)}{\exp(\beta_i) + \exp(\beta_j)}$$

The likelihood function for the Bradley-Terry model for pair-comparison data for n objects in set A is defined in So (1995):

$$\mathcal{L}(\beta_1 \ldots \beta_n) = \prod_{(i,j) \ni Set\ A} \pi_{ij}$$

The likelihood for the Bradley-Terry model is identical to the binary logistic model assuming that the multiple evaluations of a given pair are independent, with a fixed probability of preferring one item, and that the evaluations of different pairs are independent of one another (SAS, 2005). As the name implies, the maximum likelihood estimation (MLE) method estimates beta weights that maximize the likelihood of obtaining the observed data. The BT model is more popular than the Thurstone model because it can take advantage of more statistical tools, i.e., those associated with logistic regression. In practice, the scaling solutions for the Thurstone model and the BT model will generally be quite close, despite substantial differences in the psychological theory of judgment attached to each (Coombs, 1983). The primary objective of the study reported here is to obtain perceived urgency judgments for 10 different haptic steering patterns that might be used to alert a driver to change lanes or steer around an obstacle in crash-imminent situations and also in less-severe crash-cautionary situations. Scaling for urgency will allow for more informed matching of patterns to collision hazard conditions in subsequent studies. Procedurally, the study applied the method of pair comparisons and various data quality and consistency evaluations before deriving both an interval-level scaling solution (BT scaling) and a rank- order scaling solution.

5.3 Method

Participants

Sixteen male and 16 female volunteers from 18 to 63 years of age participated in the study. Participants were predominantly Ford Motor Company employees with non-engineering backgrounds (finance, administration, etc.), and none were part of the research team for this study.

Apparatus

VIRTTEX - See description in Appendix A

Simulator Scenario

The basic simulator scenario was a simulated daytime drive that started at a traffic signal on the edge of a suburban environment with a straight 4-lane divided road (two lanes in each direction). The drive transitioned to a rural environment with a straight 4-lane divided highway and a posted speed limit of 65 mph (105 kph). To impose a steering task, road crown was simulated to cause the host vehicle to drift toward the near shoulder after approximately 7 s without steering input by the driver. Drivers were instructed to stay in the right lane and return to the center of the right lane as needed after a haptic steering pattern was presented. Cruise control was set to 65 mph and was active throughout the pair-comparison evaluations. Additional details of the simulator scenarios are provided in the Procedures section.

Procedure, Design of Experiments, and Response

After arriving at the VIRTTEX facility, participants signed an informed consent form, watched a safety video about the simulator, and completed initial orientation by going through a presentation with the experimenter that described the objectives and procedures of the study. Then they were escorted into the VIRTTEX simulator dome, made adjustments to seat and steering wheel for comfort, and were given additional training on the use of the simulated vehicle. Once driving, drivers received no driving support other than conventional cruise control. During initial training, drivers were asked to begin accelerating up to 65 mph and then engage cruise control. Participants were told that the vehicle would drift out of its lane without the drivers steering input; and this was demonstrated by asking participants to remove their hands from the steering wheel and then correct the lane drift. Drivers were then instructed to stay in their lane and not make any lane changes. There were no secondary tasks during driving.

The pair comparisons were part of a larger study, and each participant drove for approximately one hour in the right lane of the 4-lane divided highway. The data collection session was in two parts. The first part of the session used a novel method, referred to as mirage events, to provide situational context for driver judgments. A mirage event is a simulated hazard event briefly presented and then removed. In the present case, a 0.3 s 500-Hz beep was presented, and shortly thereafter a lead vehicle appeared and suddenly swerved out of lane to reveal a stopped vehicle ahead at a predetermined TTC. Each mirage event was accompanied by a haptic steering pattern activated at the appropriate time and then removed from the driving scene before the subject vehicle closed the gap, after which the participant continued to drive. Four counterbalanced blocks of mirages were presented, each with a fixed TTC (7, 5, 3, or 1.7 s). In each block, the mirages were repeated once for each of the 10 haptic steering patterns, randomly sequenced, with approximately 17.5 s between beeps. No evasive maneuvers were required, drivers were instructed not to change lanes or brake in response to the mirages, and drivers continued driving throughout the session. Ratings of appropriateness and acceptability for each haptic steering pattern were collected after each mirage event.

The second part of the simulator data collection session involved the pair comparisons and followed the first portion with continued driving and active cruise control. Six practice paired comparisons were presented in order to familiarize participants with the study protocol and timing. None of the patterns used for practice was one of the 10 patterns used for the study proper. No lead vehicle was ever present during the pair-comparison presentations, and no lane changes were ever required.

Figure 5.2 shows an overview of the timing for a paired comparison. With an initial lead-in time of 1.2 s prior to each pair of haptic steering patterns, a 0.3 s 500-Hz tone was presented to alert the driver, followed by a 0.9 s silent pause. Then the first haptic steering pattern was presented, an Intertimulus Interval (ISI) of 2.2 s passed to allow the driver to resettle the vehicle in the lane as needed, and then the second haptic steering pattern was presented. When the second pattern of a pair was completed, the participant had 5.0 s to indicate which pattern (1 or 2) was perceived as more urgent, where urgency was defined as follows: "the level of timeliness and imperativeness to change lanes implied by the profile. In short, this is asking you which profile calls for more prompt action."

Pilot testing was used to determine the ISI that would keep the patterns distinct but not tax sensory memory too much. Because of the repetitive nature and duration of the pair-comparison portion of the study, each participant's progress was displayed in the center stack as a counter (e.g., 29 of 90 for 90 pairs).

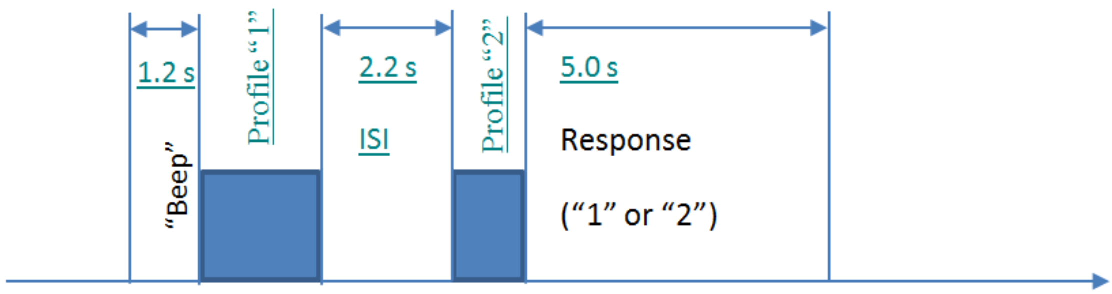

Figure 5.2: Timeline of Haptic Steering Pattern Presentation in Pairs

The paired-comparison portion of the study consisted of two blocks, each consisting of a complete pair comparison set of $\binom{10}{2} = \frac{10(10-1)}{2} = 45$ unique pairs of the 10 haptic steering patterns. Each participant received a randomly ordered sequence of all 45 paired comparisons in the first block. The second block was a replicate of the first block and started immediately after the first block was completed. The order of the 45 paired comparisons in Block 2 was the same as in Block 1 for each test participant, but the order of patterns within each pair had been reversed. This part of the drive took approximately 20 minutes from start to finish. After the participant concluded his/her evaluation of the final pair, he/she was asked to slowly bring the vehicle to a stop in the lane, shift the transmission into park, and not unbuckle until instructed to do so. As the simulator docked, the ride-along observer verbally conducted a SSQ (Kennedy et al., 1993), and the session concluded.

Data Analysis Approach

Several different software tools were used for the data analysis. Reliability of judgments across the two blocks of pair comparisons was assessed for each participant by use of the sign test in Minitab. The TRICIR program of Dunn-Rankin et al. (2004) was used to analyze the circular triads of the paired comparison data for each participant for comparison against the exact probabilities given by (Kendall, 1975). TRICIR also provided a rank-order scale of the objects, using the simplified rank method; calculated Kendalls coefficients of consistence (zeta) and Kendalls coefficients of concordance (*W*); and provided probability values for each as well as the Z-statistics for stimuli (haptic steering patterns) to identify outliers. The Bradley-Terry model was fitted by a custom program in Matlab. As a verification, Bradley-Terry interval scaling of the haptic steering patterns was also fitted to the data using PROC LOGISTIC and PROC GENMOD in SAS according to the method described in SAS (2005).

5.4 Results

The cumulative adjacency matrix was calculated for the first 45 pair comparisons among 31 participants (Participant 12 was omitted, as will be explained below) and the second 45 pair comparisons among 31 participants (Participant 6 was omitted, as will be explained below). These two tables, along with the individual participant's adjacency matrices, form the core data set for the analysis presented below.

Per-Judge Test-Retest Consistency

Figure 5.3 shows the proportion of 45 paired comparisons that were matched by each judge across the two blocks of pair comparisons. Multiplying by 100, the percentage match ranged from a low of 58% to a high of 96%, with a median value of 78%. To test the null hypothesis that judgments were not random, the closest exact binomial probability for an α level of 0.05 is exactly 0.0364 with a critical value of 29 out of 45 (a proportion match of 0.64) or better. As shown in the figure, only one participant (Participant 22) failed to meet this test. This participant might be dropped from the final scaling. However, if this participant’s consistence in terms of circular triads is acceptable for the first full paired-comparison data set, that persons data may be retained.

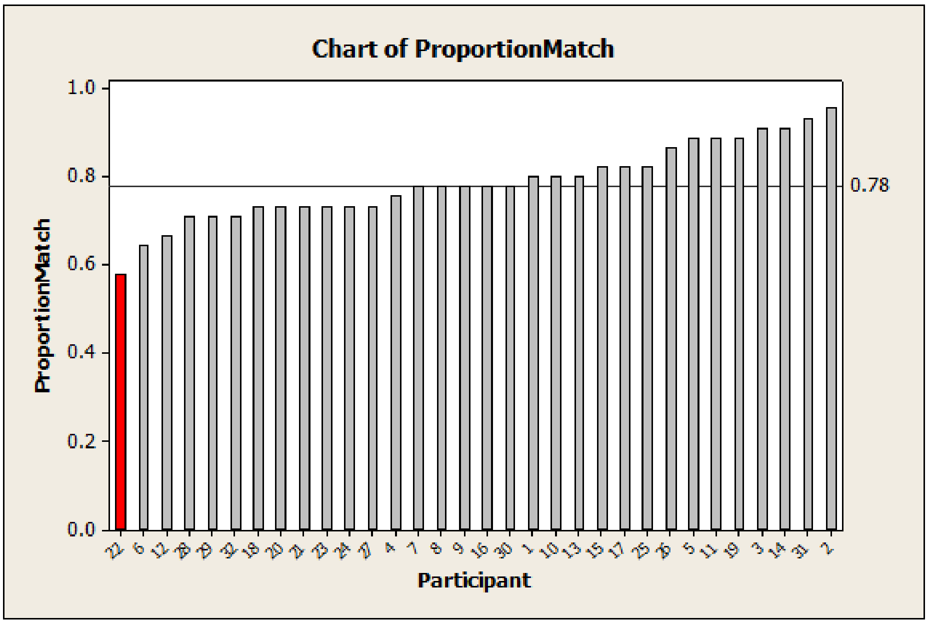

Figure 5.3: Chart of Proportion Matched (median proportion of 0.78 reference line added)

Per-Judge Consistence: Circular Triad Analysis

Figure 5.4 shows the consistence values for each judge from the first full pair-comparison data set (i.e., the first block of 45 paired comparisons). Consistence scores ranged from a minimum of 0.4 to a maximum of 1.0, with a median value of 0.9. Kendall’s test for participant circular triads reports a critical value of 21 out of 45 (*consistence* = 121/45 or 0.53) or fewer to reject the null hypothesis of random responding. As indicated by the red bar, Participant 12 failed to meet this requirement; Participant 12 was therefore dropped from further analysis. Note that Participant 22, though among the lowest in consistence for Block 1 paired comparisons, nonetheless did not generate so many circular triads as to warrant dismissal from the BT modeling and rank order scaling.

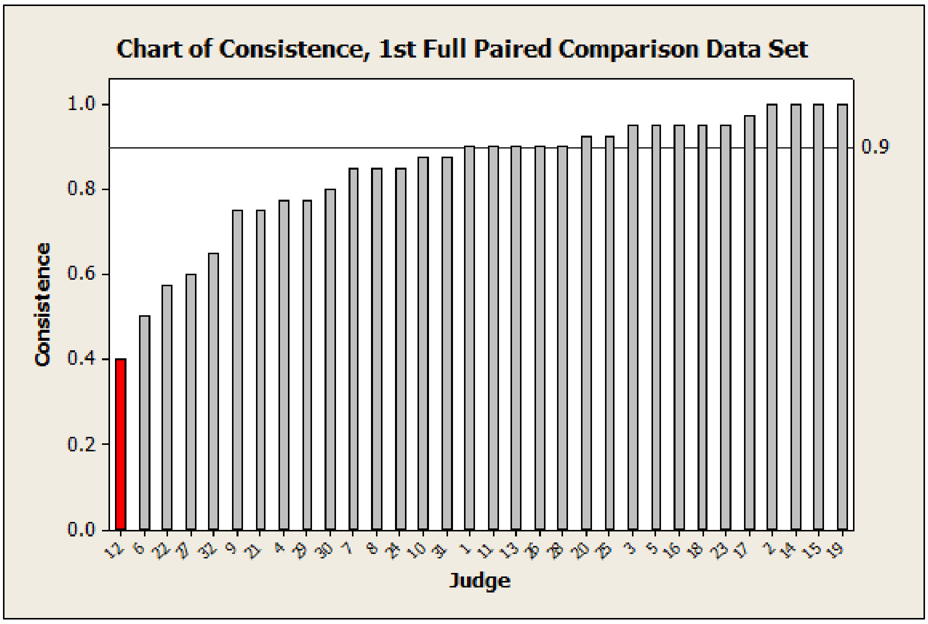

Figure 5.4: Chart of Consistence (Participant Circular Triads). (median consistence of 0.9 reference line added)

In the second set of 45 pair comparisons (i.e., the complete replicate), Participant 6 failed to exceed the critical value of 21/45 needed to reject the null hypothesis of random responding with his/her reported 24 circular triads (see Figure 5.5). Thus he/she would be dropped from further analysis. No other participant had a reported circular triad count in excess of the critical value. Interestingly, Participant 6 was the second-most inconsistent from the first 45 pair comparisons. The median consistence for the second set of pair comparisons, 0.875, was slightly lower than the median for the first 45 pair comparisons, but this difference is not statistically significant (Wilcoxon signed-ranks test, $W = 52$, two-tailed $p = 0.49$).

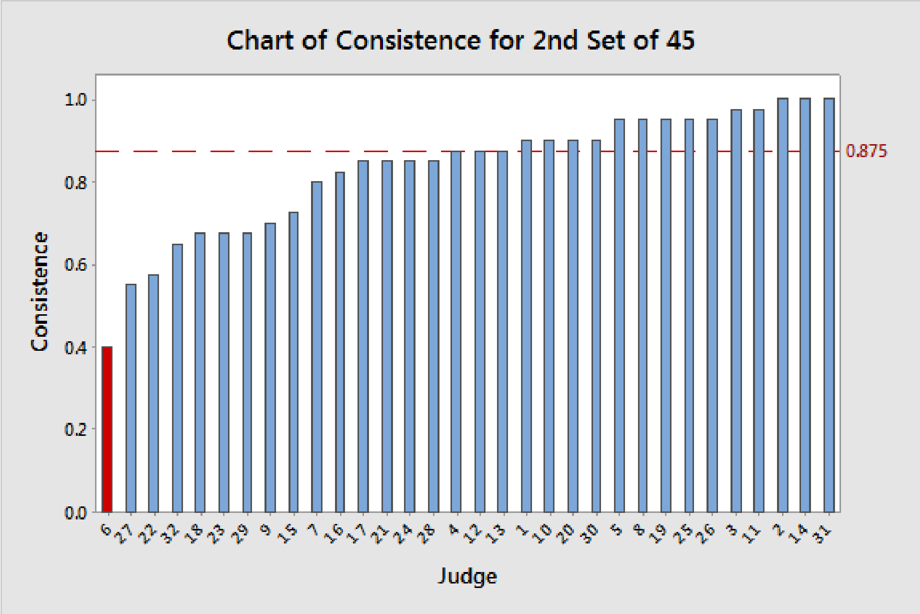

Figure 5.5: Chart of Consistence for 2nd Set of 45 Pairs

Relative Object Circularity Analysis

The TRICIR program provides a relative test of whether a specific object (haptic steering pattern) is involved in more circular triads than other objects. The program calculates the Z-statistic described earlier as a way to determine extreme values. Table 5.1 presents the 10 haptic steering pattern codes along with their associated Z-values for the first set of pair comparisons (left side) and separately for the second set of pair comparisons (right side). As can be seen, none of the objects has a Z-score greater than 1.64, the one-tailed 0.05 z-value. This was taken to indicate that no objects should be removed from the set because they were over-involved in circular triads. Note that $Z < -1.64$ indicates that an object was involved in fewer circular triads than others. For example, Object 1 was involved in significantly fewer circular triads in both the first 45 and second 45 pairs than the rest of the objects.

Table 5.1: Relative Object Circularity: Counts of Circular Triads (#CT) and Z-values Per Object (Haptic Steering Pattern) for First (Left) and Second (Right) Set of Pair Comparisons

First Set					Second Set			
Object	# CTs	# Votes	Grp Z		Object	# CTs	# Votes	Grp Z
1	16	11	-1.78		1	31	21	-1.69
2	52	69	-0.07		2	52	61	-0.37
3	80	118	1.26		3	79	121	1.33
4	26	59	-1.30		4	52	52	-0.37
5	74	125	0.98		5	76	128	1.14
6	68	191	0.69		6	72	186	0.89
7	71	161	0.84		7	56	162	-0.12
8	40	234	-0.64		8	45	234	-0.81
9	55	189	0.08		9	71	184	0.82
10	52	238	-0.07		10	45	246	-0.81

Note: Participant 12 was omitted from the first set of pair-comparison data and Participant 6 from the second set.

Inter-Participant Concordance

Kendall's coefficient of concordance (W) for the 31 participants (after omitting Participant 12) using the first block of pair comparisons was calculated as W = 0.76, highly significant (p < 0.0001). This indicates that not all participants were independent in their judgments. Instead, they had a relatively high degree of agreement, though not perfect agreement, in the judged urgency of the haptic steering patterns. Analyzing the second 45 pair comparisons results in findings similar to those reported above for the first 45 pairs. Kendall's coefficient of concordance (W) for the 31 participants (after omitting Participant 6) was found to be W = 0.79, also highly significant (p < 0.0001).

Ordinal Scaling Solutions

The ordinal scaling of the 10 haptic steering patterns in terms of perceived urgency was obtained by summing the dominance counts per pattern using the TRICIR program of Dunn-Rankin et al. (2004) and ranking patterns from least urgent (fewest votes) to most urgent (most votes). The first and second sets of pair comparisons were ordered separately; the results are given in Table 5.2. Below each pattern code is the number of vote counts it received. Note that the order of the patterns is identical in the first and second sets of pair comparisons. For each haptic steering pattern, there are $N * (n - 1)$ or $(31) * (10 - 1) = 279$ possible votes in a complete pair comparison data set that might be cast indicating how many participants said that pattern dominated the other 9 patterns.

Table 5.2: Ordinal Scaling of Haptic Steering Patterns for First and Second Blocks of Pair Comparisons

		1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th
First 45 Pair Comparisons	Pattern	1	4	2	3	5	7	9	6	8	10
	# Votes	11	59	69	118	125	161	189	191	234	238
Second 45 Pair Comparisons	Pattern	1	4	2	3	5	7	9	6	8	10
	# Votes	21	52	61	121	128	162	184	186	234	246

The rank-order scaling solution allowed for statistical testing of differences between pairs of objects. Using the nonparametric methods described in Dunn-Rankin et al. (2004), the expected standard deviation in the sum of votes, $E(S)$, received by any pattern is given as follows:

$$E(S) = \sqrt{\frac{N * n * (n+1)}{12} = \sqrt{\frac{31 * 10 * (10+1)}{12}} = \sqrt{284.17} = 16.85}$$

The Studentized Range from Dunn-Rankin et al. (2004)(p. 217) for 10 objects and $\alpha = 0.05$ is $Q_a = 4.474$. The critical range or critical difference in number of votes between any two objects in the in the study, then, is the following product:

$$Critical\ Range = E(S) * Q_a = (16.85) * (4.474) = 75.39$$

Any pair of objects that differ in votes by more than 75 are reliably different from each other in perceived urgency beyond chance.

Despite some numerical differences in vote counts across the first and second sets of comparisons (refer to Table 5.1), the results of testing for statistically significant differences in total votes between pairs of objects were the same. Specifically, the following perceived urgency differences among pairs of patterns emerged (here “>” implies more urgency):

- 10 > 1, 4, 2, 3, 5, and 7; not different from 8, 6, or 9
- 8 > 1, 4, 2, 3, and 5; not different from 6, 9, or 7
- 6 > 1, 4, and 2; not different from 9, 7, or 5
- 9 > 1, 4, and 2; not different from 7, or 5
- 7 > 1, 4, and 2; not different from 5
- 5 > 1; not different from 3, 2, or 4
- 3 > 1; not different from 2, 4, or 1
- 2 not different from 4 or 1

- 4 not different from 1

Figure 5.6 shows the rank-order of haptic steering patterns and these significant and non-significant differences. Any two pattern codes not underlined by the same line are reliably different at the 5% level.

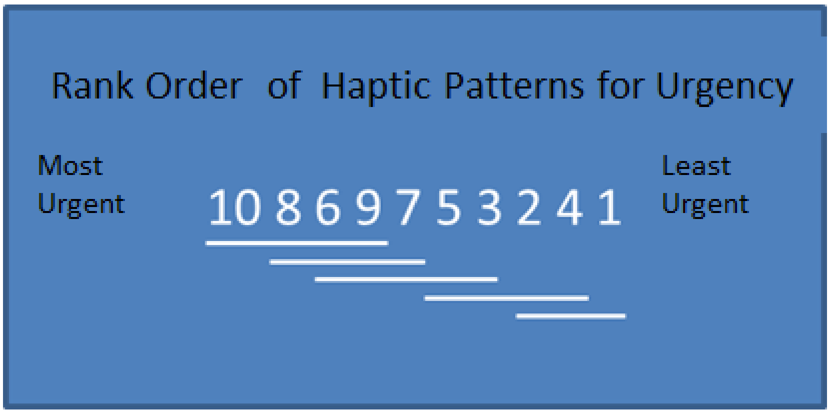

Figure 5.6: Visualization of Significantly Different Patterns using the Studentized Range Test

Bradley-Terry (BT) Scaling Solutions

Figure 5.7 shows the BT scaling solution for the first complete set of pair comparisons (minus Participant 12) and the scaling solution for the second set (minus Participant 6). The values on the x-axis are the regression weights for each object. The logistic regression routine assumed Pattern 10 was the default (coded 0), and since it had the maximum number of votes, all other patterns' beta weights are negative. Consistent with the ordinal results, the order of haptic steering patterns is the same in all cases. However, the intervals between objects change somewhat between the two sets of pair comparisons obtained under virtually identical circumstances. This calls into question the stability of the assumed interval level of measurement purportedly achieved by the BT (and Thurstone) models.

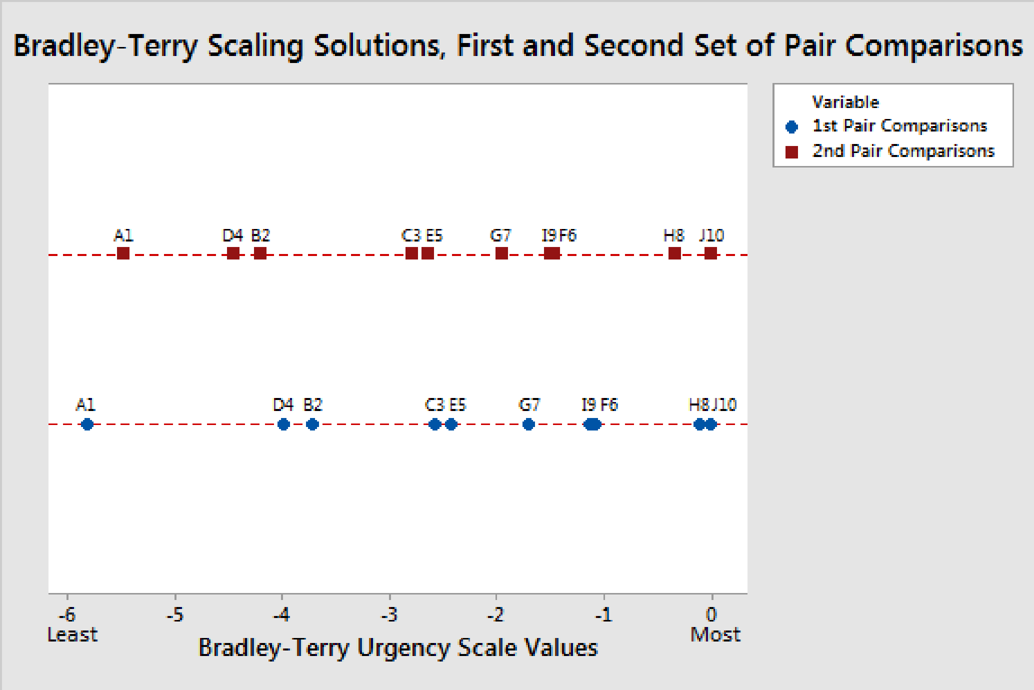

Figure 5.7: Bradley-Terry Scale Values for First Set and Second Set of Pair Comparisons

Figure 5.8 presents plots of the actual vote counts versus the vote counts estimated or fitted by the BT model for each of the of $\binom{10}{2} = 45$ unique pairs of haptic steering patterns in the first and second sets of comparisons. The estimated vote counts, per pair comparison, were obtained by multiplying the BT model probability of dominance by the number of judges (N = 31). Overall, there is a high correlation in both sets of data with slightly better fit in the second set of pair comparison data.

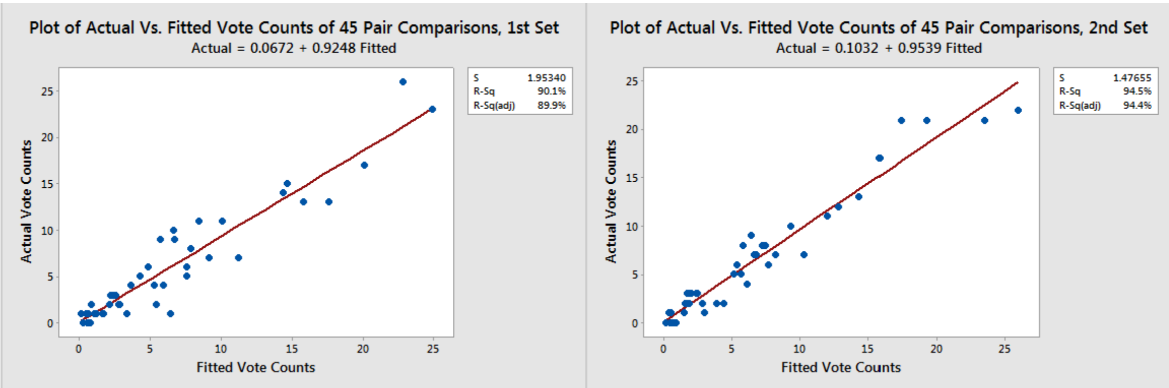

Figure 5.8: Actual vs. Estimated Vote Counts, First Set and Second Set of 45 Pair Comparisons

5.5 Discussion

Assessing Data Consistency

A 78% (median) participant test-retest percent match was observed, and only one participant (Participant 22) failed the reliability check for random responding. The level of reliability in judgments may have been enhanced by the exposure to the haptic steering patterns in the mirage scenarios prior to the pair comparisons. The mirage scenarios did not ask for participant judgments of urgency; ratings of appropriateness and acceptability were obtained instead. Nonetheless, these scenarios provided situational context that may be beneficial for psychophysical scaling judgments and should be considered when feasible.

In terms of circular triads, one participant (Participant 12) exceeded the critical value of 21/40 circular triads in the first block of pair comparisons, and this was used as a basis to exclude that individual's data in the analysis that followed. The second block of pair comparisons also revealed one participant (Participant 6) who appeared to be responding at random. Overall, the median consistence was high in the first and second sets of pair comparison data (0.875 and 0.9, respectively). The analysis of relative object circularity resulted in no haptic patterns being discarded. The relatively high transitivity in judgments may also have benefited from extensive exposure to the haptic steering patterns prior to the pair comparison data collection.

Participants may reliably make the same pair-comparison judgments in a test-retest sense and exhibit transitivity but disagree on what "urgency" means. If so, the answer to the question "Is there a common underlying scale?" would probably be answered in the negative. In the present case, the concordance coefficients suggested that it was reasonable to combine the judgments of multiple participants to create an urgency scale for the haptic steering patterns. The inter-participant concordance was found to be 0.76 in the first set of pair comparisons and 0.79 in the second, rejecting the null hypothesis that participants were independent in their judgments in both instances.

Across the two sets of pair-comparison data, there is no clear pattern or trend in the test-retest data to support the hypothesis that participants will learn and thus reduce the number of circular triads on the second round. While the median consistence did increase slightly in the second set of pair comparisons, this difference was not significant. Another concern is that participants will experience fatigue and thus become more inconsistent in their judgments in the second set of pair comparisons. If anything, as a group, the participants remained as consistent as they were in the first pass through, if not slightly better, in the study reported here. However, it is recommended that the propensity for fatigue be assessed on a case-by-case or study-by-study basis. In general, deviations from transitivity, whether due to fatigue or for other reasons, will reduce the quality of the fit of the data to a linear psychometric or psychophysical scale. Repeating a complete pair-comparison process is a question of logistics and potential benefits and drawbacks. In the present case, the two sets of data revealed the same ordinal results and similar, though not identical, interval-scale results. Given that repetition costs time and money, the belief is that the first pass of data, i.e., a full matrix of $\binom{n}{2} = \frac{n(n-1)}{2}$ unique pairs, is likely to be good enough to evaluate inconsistencies in the judgments and to achieve a scaling solution. Repeating a large matrix of paired comparisons is probably unnecessary and often unfeasible for applied research.

The consistency of pair comparisons may be degraded by one or more of several factors [Kendall and Smith (1940); Knezek (1978); Dunn-Rankin et al. (2004); Cliff and Keats (2003)]. In the broadest terms, the participant may be incompetent (e.g., a novice compared to an expert jeweler judging which of two gems is superior in quality); the objects may be too similar for the participant to tell them apart (e.g., which is higher-pitched out of two tones close in frequency); the participant may have had a momentary lapse in attention (e.g., lapses due to monotony, boredom, or time on task); the participant may have misunderstood the task (e.g., he/she forgot or misunderstood the precise definition of the dimension of comparison); there may have been a procedural error (e.g., the definition of the dimension for judgment was too vague); the participant may have made a response error (e.g., pressed the wrong button or uttered the wrong code for his/her preference); and/or the participant's decision threshold may have drifted over time (e.g., by recalibrating judgments with practice). In rare instances, a diabolical participant may systematically respond in a contrarian fashion. Although the reasons for any observed inconsistency may not be known in detail, the quality of the scaling solution and the suitability of that scaling for application may nonetheless be jeopardized. Therefore, it is useful to consider consistency checks like those presented in this paper before deriving a scaling solution.

Cliff (1996) and Cliff and Keats (2003) argue in favor of ordinal scaling and ordinal analysis methods in light of the quality of data typically obtained from human factors and psychology experiments. Cliff (1996) has attempted justify this stance. He argues that behavioral-science data often have only ordinal justification. Even an "objective" response like reaction times (RTs) to haptic steering warnings might at first glance seem to possess ratio scale properties associated with time. But RTs are generally surrogates for theoretical constructs like "distraction", "display conspicuity", "fatigue", and so forth. Even if one supposes that there is a monotonic relationship between the RT measure and the construct, it is hard to justify linearity. Thus, as Cliff (1996) put it, "as behavioral variables, the comforting interval or ratio status of physical variables is hard to justify" (p. 7). This is even more applicable in the case of subjective assessments like "urgency." The quality of the data obtained in the haptic steering study was relatively high considering the reliability, consistence, and concordance results. However, it was not perfectly in accord with the assumptions needed for interval-level scaling. As mentioned earlier, the interval-scaling that the Bradley-Terry model purports to provide required additivity in scale separations that currently cannot be adequately assessed without an error model of conjoint measurement. Thus, Cliff and Keats (2003) argue in favor of ordinal scaling because the data often do not possess the more elaborate data and extended set of properties (e.g., additivity) required to take that step to intervals.

Relating Profiles to Physical Attributes

One of the reasons to conduct a scaling study is to obtain data that are not easily derived from physical measurements alone. The authors nonetheless used exploratory data analysis techniques (Chambers, 1983) to look for patterns regarding what physical attributes of the haptic steering patterns were associated with greater perceived urgency. A matrix plot (a variation of a draughtmans plot) was prepared to relate physical attributes of the 10 different haptic steering patterns from Table 4.1 to their rank-ordered urgency. In Figure 5.9, red diamond plotting points represent what were originally thought to be crash-imminent patterns, and blue dots represent what were originally considered to be crash-cautionary patterns. Scanning across the physical attributes, greater urgency is generally related to two-pulse patterns over three-pulse patterns; shorter-duration patterns over longer ones; patterns with a hold over those with no hold; and patterns that delivered higher peak torques. The effects of angular impulse (the total area under the $torque * time$ curves) are less clear; this was explored further in a scatter plot of rank urgency as a function of peak newton-meters of torque (PeakNm) (see Figure 5.10). Each of the 10 haptic steering patterns is a plotting point, and each plotting point is labeled with the angular impulse or total area under the $torque * time$ curve for that pattern. Two trends emerge from examining this plot. The first is that, consistent with the matrix plot, the greater the peak torques, the greater the perceived urgency among the haptic steering patterns. Note, though, that each peak torque has two plotting points associated with it. This merely reflects Table 4.1 and the way the patterns were developed. The second trend is that for any given peak torque, the pattern with the lower angular impulse is always (in five pairs of instances) rated higher in perceived urgency. Why this is happening is unclear. One possibility is that the lower angular impulse creates a more “crisp” or punctate sensation at the wheel that the driver translates (generally) as a more urgent warning.

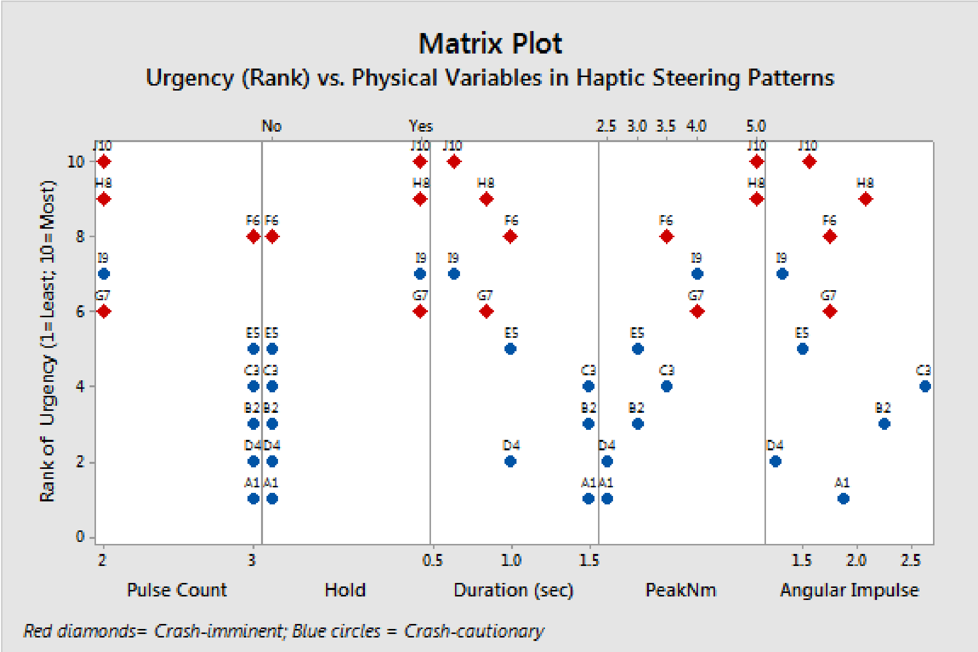

Figure 5.9: Matrix Plot of Haptic Steering Patterns Rank-Ordered Urgency and Various Physical Attributes

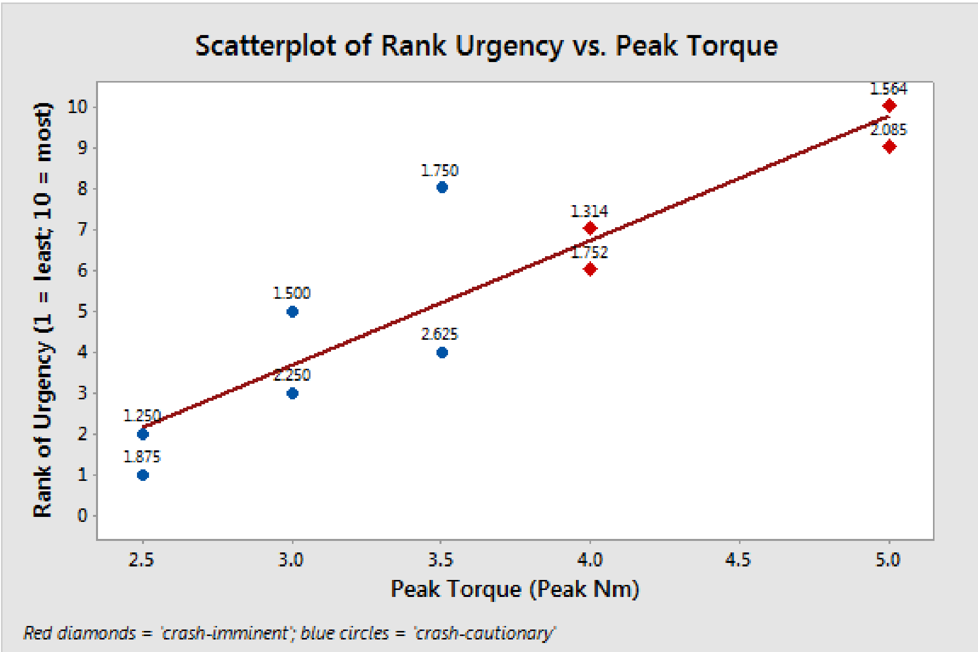

Figure 5.10: A Scatterplot of Urgency (rank) vs. Peak Torque with Angular Impulse Labels per Haptic Steering Pattern

Implications for Discrete Directional Haptic Steering Displays

One application for these findings is to relate perceived urgency to application in a lane change advisory system for crash-cautionary conditions (e.g., TTC values of 5-7 seconds or longer) and also as a warning display system for crash-imminent conditions (e.g., TTC values of 3 seconds or less). The test session in a simulated drive would consist of experiencing one or more haptic steering profiles as part of a multimodal (haptic, visual, and auditory) display in relatively benign crash-cautionary time frames and then, in the end, one surprise event in a crash-imminent time frame. The measurable responses would include the response type (e.g., braking, steering, or a combination), the driver's response time (RT) to the haptic warnings versus that of an auditory-visual warning, and the minimum TTC, among others. This type of paradigm will help evaluate the efficacy of the haptic steering system concept and guide subsequent test-track and on-road evaluations. This series of studies would contribute human-factor data to feature design.

The next chapter will discuss the results of the first part of the study (i.e., the mirage scenarios), including the rated appropriateness and acceptability of each profile in the selected scenario to the physics of the profile. It will be there that the psychophysics (i.e., paired comparison) results will be related meaningfully to the driving situation. The conclusion might be that more urgent profiles should be reserved for crash-imminent situations and that milder profiles should be used for crash-cautionary conditions. Alternatively, Coombs (1983) argues that human preference behavior is often describable by a single-peaked function, a function that increases monotonically to a maximum and then decreases monotonically. As he put it, it is often the case that "good things satiate...bad things escalate" (p. 21). If so, this suggests that perhaps a medium level of perceived urgency in the haptic steering patterns will be most preferred over a wide range of driving conditions.

CHAPTER 6 STUDY 3, PART 2: DRIVER HAPTIC STEERING ALERTS, MIRAGE EVENTS, AND MAUT: A NOVEL METHOD FOR SELECTING AN OPTIMAL HMI

6.1 Abstract

In-situ subjective assessments of 10 haptic steering profile appropriateness and acceptability measures in the face of different Time to Collision (TTC) scenarios. While driving in the moving-base simulator, drivers experienced each haptic profile in the context of a forward collision hazard, briefly presented and then removed. This novel method is called mirage scenario. It did not necessitate braking or steering responses; this eliminated the risk of simulator sickness. A range of TTCs from crash-probable (5 or 7 s) to crash-imminent (3 or 1.7 s) were used to both situations in both normal and emergency conditions. This novel approach provided data that were analyzed in several ways, including Multi- Attribute Utility Theory (MAUT), to arrive at a single 'best' haptic steering display. A key finding was that the 'best' profile in terms of appropriateness and acceptability was not the most 'urgent' as identified in the pair-comparison urgency scaling effort. Thus, it was discovered that a psychophysical scaling effort without additional in-situ testing might result in a design selection that is not optimal in terms of consumer acceptance.

6.2 Introduction

This chapter describes a new method, a mirage scenario, to support formative evaluation of driver alerting or warning displays for manual and automated driving. This method provides driving contexts (e.g., various TTCs to a lead vehicle) briefly presented and then removed. In the present study, during each mirage event, a haptic steering display was evaluated. This haptic display indicated that a steering response could be initiated to drive around an obstacle ahead. A motion-base simulator was used in a 32-participant study to present vehicle motion cues similar to the actual application. Surprise was neither present nor of concern, as it would be for a summative evaluation of a forward collision warning system. Furthermore, to reduce the risk of simulator sickness, participants never performed collision avoidance maneuvers. This chapter illustrates the mirage scenario procedures, the rating methods and definitions used with the mirage scenario, and analysis of the ratings obtained, together with a MAUT approach to evaluate and select among alternative designs for future summative evaluation.

Simulation for Driver Alert Development

Driving simulators play a role in the testing of driver assistance technologies, including driver warning and alerting systems used for collision avoidance during manual driving and in Take Over Requests (TOR) for automated driving. Simulators allow for various scenarios to be presented without physical risk and in a controlled fashion. FC hazards, for example, can be created in a virtual environment by having the vehicle ahead brake hard, a vehicle in the adjacent lane cut in suddenly, or a lead vehicle swerve sharply to reveal an obstacle in the travel lane. Duarte and Rebelo (2007) point out that a credible simulated situation should look realistically dangerous but be safe for the participants. They assert that the warnings should always be inserted in a context like that in which they will be presented in the real world. Otherwise, the absence of contextual cues may invalidate an assessment of a warning. The importance of providing context in warning design has long been acknowledged in auditory warning development (Edworthy and Adams, 1996). For example, Zwolinski and Sagot (1998) used a train simulator to design and refine auditory displays for emergency stop warnings. They pointed out that the simulator not only reproduces the operational environment and context of the train cab, but also provides key environmental factors that might affect driving, such as the forward movement of the rails, ambient sounds, etc. The variable configuration of this simulator also integrated all the new control and informational devices in relation to the new set of auditory signals chosen. More recently, Singer et al. (2015) reported on a series of laboratory studies that evaluated various auditory displays for detectability and perceived urgency under various ambient noise conditions. They reported that the presence of different ambient noise sources (e.g., heavy rain, driving with the window down) had more significant effects on some auditory warnings than others. They also reported that, overall, perceived urgency, noticeability, and annoyance were highly intercorrelated: the louder the alarm, the more noticeable, urgent, and annoying it is.

There is also growing evidence that the nature of the motion environment in a simulator (full motion, partial motion, no motion) can have an impact on driver responses to warnings. Lerner et al. (2015) summarized a series of experiments conducted under a program of research called the Collision Warning Interface Metrics (CWIM) program. In one study, they examined the effects of a simulator with full motion, partial motion (e.g., road vibration), and no motion (fixed-base simulation) for Forward Collision Warning (FCW) and Lane Departure Warning (LDW) display evaluations. They stated,

> "Full motion fidelity provides accurate vestibular motion feedback to drivers, whereas partial motion fidelity provides limited vestibular feedback. For LDW events, the no-motion condition resulted in shorter secondary task engagements, but larger lane exceedances (than full-motion conditions). The partial-motion condition also resulted in shorter secondary task engagements, but lane exceedances were smaller than in the no-motion condition. For FCW events, a number of interactive effects were observed, which makes definitive determinations difficult. Based upon the results of this study, partial or no motion is likely acceptable for evaluations of relative DVI performance, but the complex interactions in FCW events point to the need to match the real world as closely as possible" (p. x).

In addition to driving context and vehicle motions, driver factors are also important. One key driver factor in traffic accident causation is expectancy violation ((Rumar, 1990); Tijerina, 2015 (Smiley, 2015)). Real-world crashes often arise because the driver did not expect the hazard and responded late, inappropriately, or not at all. A common practice involves testing alternative driver warning designs using the same participant (Ljung-Aust, Engstrm, Vistrm, Nbo, Bolling, Hjort, and Kallgren, 2011). Repeated-measures evaluations can offer logistical convenience and statistical efficiency that make them among the most common experimental designs in applied human factors and ergonomics. This is problematic, since drivers are highly context-sensitive and adapt to the situation (Engstrom and Aust, 2011). Ample evidence exists that a driver may be surprised once and thereafter may not behave realistically. The altered expectancy can manifest itself in differential carryover effects that make interpreting reaction time data difficult (Tijerina et al., 2015). The repeated exposure to a hazard also alters the nature of the drivers response, e.g., from a closed-loop response initially to a reflexive, open- loop response thereafter (Ljung-Aust et al., 2011). The only real alternative to deal with differential carryover or driver expectancy effects associated with repeated-measures designs is to avoid them and use a one-trial-per-subject paradigm, i.e., a between-subjects experimental design Poulton (1981). This can be an expensive method in terms of time and resources. In general, a $k - treatment$ within-subjects design will require k times the number of persons for a between-subjects design to assess the same k treatments at the same level of statistical power (Cohen, 1988).

System developers face a dilemma: On one hand, it is useful to have persons compare multiple alternative driver warnings or alerts for subjective assessments of appropriateness, acceptability, urgency, annoyance, and the like. On the other hand, drivers are not easily surprised more than once, and this can compromise the usefulness of driver performance data if repeated exposures are used. One potential way out of this dilemma is to consider breaking up the evaluation process into phases of formative evaluation and summative evaluation (Wogalter et al., 2002). A formative evaluation is done while the system is still being developed. It may use objective performance measures to identify the usability of a design alternative. It may also involve subjective assessments collected in a repeated-measures format to develop a profile or subjective scale of one or more dimensions considered important for the application. For example, Tijerina et al. (2000) repeatedly exposed drivers to pulse braking events intended as haptic displays for forward collision warning applications. Drivers were repeatedly exposed to various levels of the braking, sometimes while distracted and sometimes while not distracted. Each time, drivers needed to indicate whether or when they noticed a braking event. This is an example of a formative evaluation. The purpose of the study was to characterize the conspicuity of the monopulse braking display concept under various configurations of vehicle jerk rate and braking event duration, not to assess its actual collision avoidance potential. Therefore, only some of the contextual factors were present: those needed to assess warning conspicuity (actual vehicle motions, driver attentional state). In contrast, a summative evaluation focuses on the final design, and as Wogalter et al. (2002) put it, ". . . the final product must be 'released' into the context of use, and then criterion measures can be gathered from participants" (p. 225). In the summative evaluation of driver warning systems for infrequent hazards, the presence of a hazard and the element of surprise are both required. Here the use of a one-trial-per-participant testing protocol may be necessary, but it is deployed in the product development process where it is most appropriate. The study reported here describes a new method, called a mirage scenario, for formative subjective evaluations of driver alerts or warning systems. The specific application is for a haptic steering display that warns of an obstacle ahead and indicates that a steering response could be initiated to drive around the obstacle.

Study Objectives

The goal of this study (as part of a larger evaluation) was to capture subjective evaluations of alternative haptic steering displays in terms of appropriateness and acceptability in situ. The new method, here called a mirage scenario, provided driving contexts (various TTCs to a stopped lead vehicle suddenly ahead) similar to those of the target application. Additionally, the use of motion-base simulator allowed for vehicle motion cues similar to what the actual application would provide.

However, to reduce the risk of simulator sickness, drivers executed no collision avoidance braking or steering maneuvers. Surprise on the part of the driver was not present or of concern as it would be for a summative evaluation.

6.3 Method

6.3.1 Participants

The participant sample consisted of 32 volunteer participants, 16 male and 16 female, who ranged from 18 years to 63 years of age. All participants held a valid driver’s license. Most participants were Ford Motor Company employees who had non-engineering backgrounds (i.e., finance, administration, IT, etc.), and none were part of the research team designing this study. Some participants may have had prior exposure to VIRTTEX with surprise hazard events.

6.3.2 Apparatus Driving Simulator

See Appendix A for details on the VIRTTEX driving simulator.

6.3.3 Experimental Design

The mirage events were the first part of a larger two-part simulator session. In the first part, each of the 10 haptic patterns was paired with each of the four TTC mirage events in order to collect subjective assessments (see Response Measures).The mirage portion of the experimental design was a digram-balanced Latin Square design (Lewis, 1989) to counterbalance sequential effects for the four TTC conditions (7, 5, 3, and 1.7 s TTCs). Within a TTC block, each of the 10 haptic steering patterns was serially presented in random order. Thus, not only were the TTC blocks balanced, the 10 haptic steering patterns were also randomized so the participant did not experience the same order across TTC blocks. The second block started immediately after the first block was completed, and the third and fourth blocks followed without pause.

The second part of the simulator session, always completed after the four blocks of mirage events, consisted of pair comparisons of the haptic steering profiles in terms of perceived urgency as the test participant continued to drive manually with cruise control engaged. The pair-comparison analysis and urgency scaling were described in detail in Chapter 5.

6.3.4 Response Measures

Immediately after each mirage event, the participant orally rated the haptic steering pattern in terms of appropriateness and acceptability. Appropriateness was operationally defined in training as follows: “How useful is the provided torque level to suggest a steering maneuver given the situation at hand?” Acceptability was operationally defined in training as follows: “How tolerable is the provided torque level given the situation at hand?” Each response was given by reference to a 6-point balanced bipolar scale. Each participant was also told at the outset that in their ratings, they should “assume that Ford would not present a haptic steering pattern unless you could actually steer around the obstacle.” The 6-point scale used for each dimension is presented in Figure 6.1. Note that the sign of the scale number called out by the participant provided a binary indication (*No* or *Yes*) of whether it was appropriate or acceptable. The magnitude of the scale value chosen indicated the degree of appropriateness or acceptability.

NO			YES		
-3	-2	-1	1	2	3
Completely Inappropriate	Inappropriate	Barely Inappropriate	Barely Appropriate	Appropriate	Completely Appropriate
Completely Unacceptable	Unacceptable	Barely Unacceptable	Barely Acceptable	Acceptable	Completely Acceptable

Example prompting: **Appropriateness?** *Your answer: e.g.,* “-1”.

Example prompting: **Acceptability?** *Your answer: e.g.,* “2”.

Figure 6.1: Balanced Bipolar Scale for Appropriateness and Acceptability Ratings

6.3.5 Simulator Driving Scenario

The basic simulator scenario was a simulated daytime drive that started at a traffic signal on the edge of a suburban environment with a straight 4-lane divided road (two lanes in each direction). The drive transitioned to a rural environment with a straight 4-lane divided highway and a posted speed limit of 65 mph (105 kph). To impose a steering task, road crown was simulated to cause the host vehicle to drift toward the near shoulder after approximately 7 s without steering input by the driver. Drivers were instructed to stay in the right lane and return to the center of the right lane as needed after a haptic steering pattern was presented. Cruise control was set to 65 mph and was active throughout the drive. Additional details of the simulator scenarios are provided in the Procedure section.

Mirage Events

Development of the Mirage Scenarios

The development of the mirage scenario was paramount to this research. Not only is the mirage scenario itself novel, the purpose to which it was applied is novel. The reason for the mirage scenario was to provide context for the haptic patterns in a variety of TTC scenarios. The scenarios of specific interest were the TTCs investigated as part of the experimental problem characterization described in Chapter 3: 7, 5, 3, and 1.7 s. The hypotheses tested and reported here is to see if select profiles are perceived to be more appropriate and more acceptable specific TTCs and whether crash-imminent situations cause a desire for an even more powerful profile.

A mirage event is defined here as a simulated hazard event briefly presented and then removed. Figure 6.2 shows a mirage scenario: a) the forward scene is initially an empty road; b) a lead vehicle appears; c) shortly after appearing, the lead vehicle begins to swerve out of its lane; d) this reveals a stopped vehicle in the travel lane at a given TTC; e) the mirage continues to a point of closest approach; and finally f) the mirage disappears.

Figure 6.3 shows the timing sequence for mirage events used in the present study. During driving on an empty stretch of rural highway, a 0.3 s 500-Hz tone was presented; shortly thereafter a lead vehicle suddenly appeared and then swerved to reveal a stopped vehicle in the lane ahead at a predetermined TTC. Each mirage event was accompanied by a haptic steering pattern that activated after the lead vehicle swerved and was halfway over the adjacent lane line, revealing the stopped vehicle to the simulated forward collision sensor, at around 1 s after the start of the lead vehicle swerve (cutout maneuver). The mirage then disappeared from the driving scene before the subject vehicle closed the gap, after which the participant continued to drive on an empty stretch of highway until the next mirage. Immediately after each mirage disappeared, the participant orally rated the haptic steering pattern for appropriateness and acceptability in that context using a six-point scale (see Response Measures section) while he/she continued to drive. After approximately 17.5 s, the next mirage started. This sequence of events continued through all blocks of TTC conditions.

(a) Empty road

(b) A vehicle suddenly appears ahead

(c) Vehicle swerves, uncovers a stopped vehicle (pattern presented)

(d) Point of closest approach

(e) Mirage disappears, empty road again

Figure 6.2: Mirage Event

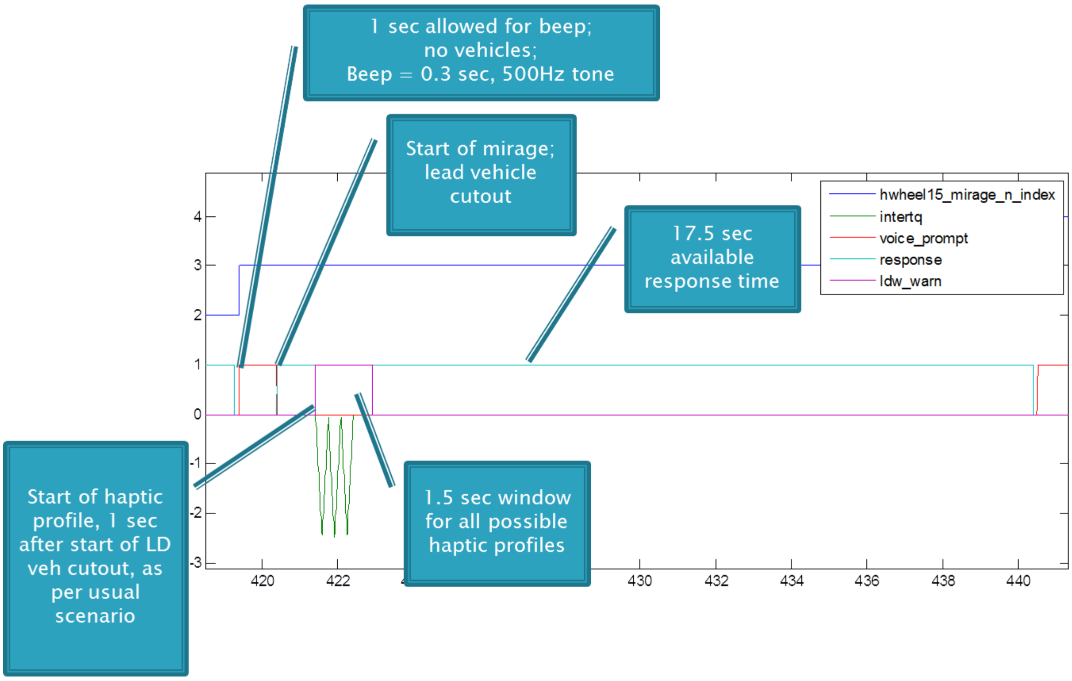

Figure 6.3: Timing Sequence for Mirage Events

6.3.6 Procedure

After arriving at the VIRTTEX facility, participants signed an informed consent form, watched a safety video about the simulator, and completed initial orientation by going through a PowerPoint presentation with the experimenter that described the objectives of the study and its procedures. Then they were escorted into the VIRTTEX simulator dome, made adjustments to seat and steering wheel for comfort, and were given additional training on the use of the simulated vehicle. Once driving, they received no driving support other than conventional cruise control. During initial training, drivers were asked to begin accelerating up to 65 mph and then engage cruise control. They were told that the vehicle would drift out of its lane without the driver's steering input; this was demonstrated by asking the participant to remove hands from the steering wheel and then correct the lane drift. Drivers were instructed to stay in their lane and not make any lane changes. There were no secondary tasks while driving.

The mirage portion of the study drive began with 6 practice mirage events presented while participants drove with cruise control engaged according to the timing sequence described earlier. None of the haptic steering patterns used for practice was used for the study proper. After the practice was complete, formal data collection began as the participant continued to drive. Completion of the TTC blocks of mirage trials took approximately 15 minutes from start to finish. Once the participant provided an appropriateness rating and an acceptability rating for the last mirage event in the fourth TTC block, that portion of the simulator session was complete. The participant was given a brief period just to drive before the second part of the larger simulator session. This second portion was a pair-comparison study to scale the 10 haptic patterns for perceived urgency. Urgency was operationally defined for participants as "the level of timeliness and imperativeness to change lanes implied by the profile. In short . . . which profile calls for more prompt action." For each of $\binom{10}{2}$ or 45 pairs of patterns presented with a short ISI, the participant indicated which of the two he/she perceived as conveying greater urgency. After all 45 pair comparisons had been collected while the driver continued driving, the pair-comparison block was repeated again without pause. The order of the 45 paired comparisons in the second block was the same as in the first block, but the order of patterns within each pair had been reversed. The order of pairs was randomized across the sample of test participants. This part of the drive took approximately 20 minutes from start to finish. Once the participant provided a choice for the 90th pair, the drive was deemed complete. The participant was asked to slowly bring the vehicle to a stop in the travel lane, shift the transmission into 'P' (park), and not unbuckle until instructed to do so. As the simulator docked, a ride-along observer verbally administered a SSQ, and the session concluded. Note that the second part of the simulator session is the subject of a separate report (Talamonti, Tijerina, Blommer, Swaminathan, and Curry, in review).

6.4 Results

The responses for the mirage events are categorical ratings of appropriateness and acceptability for the haptic steering patterns. As ordinal data reflecting driver attitudes toward the haptic patterns in a given TTC condition, these data were analyzed first in terms of the proportion of positive responses along each dimension and then with analysis of the magnitudes as well as signs of the ratings assigned to each haptic pattern, and finally with a MAUT approach to identify the best pattern both overall and for each TTC condition.

Analyses of the Proportion of Positive Ratings

The most basic judgment is whether a haptic pattern is appropriate or acceptable for a given TTC condition. Table 6.1 presents the proportion out of N = 32 participants who gave a haptic pattern a positive rating (i.e., +1, +2, or +3) on the appropriateness dimension and on the acceptability dimension separately. A proportion of 0.5 can be considered to reflect ambivalence among the participants as a group. The two-tailed sign test of the null hypothesis that the observed proportion is 0.5 is rejected at the 95% confidence level if the observed proportion is equal to or greater than 0.69 or if it is equal to or less than 0.31; these constitute a medium-to-large effect size (Cohen, 1988). Each of the values in large bold font is statistically significant with 95% confidence.

Inspection of the table reveals that participants were generally ambivalent about patterns 6 through 10 for the least severe TTC of 7.0 s and patterns 1 through 4 for the most severe TTC of 1.7 s. Additionally, pattern 10 (arguably the harshest torque pattern) was significantly unacceptable for the least severe TTC scenario (*TTC* = 7.0 s). Only pattern 5 was judged both appropriate and acceptable by a significant margin across all TTC conditions.

Table 6.1: Proportions of positive appropriateness and acceptability ratings by TTC per haptic pattern

		Time-To-Collision (TTC)			
Pattern	**Proportion(Dimension)**	**TTC = 1.7 s**	**TTC = 3.0 s**	**TTC = 5.0 s**	**TTC = 7.0 s**
1	Prop(Appropriate)	**0.31**	0.63	**0.81**	**0.72**
	Prop(Acceptable)	0.50	0.66	**0.75**	**0.78**
2	Prop(Appropriate)	0.44	**0.78**	**0.81**	**0.72**
	Prop(Acceptable)	0.66	**0.75**	**0.84**	**0.69**
3	Prop(Appropriate)	0.59	**0.75**	**0.69**	0.56
	Prop(Acceptable)	**0.69**	0.66	**0.75**	0.53
4	Prop(Appropriate)	0.53	**0.69**	**0.84**	**0.78**
	Prop(Acceptable)	**0.78**	**0.75**	**0.88**	**0.97**
5	Prop(Appropriate)	**0.75**	**0.81**	**0.88**	**0.84**
	Prop(Acceptable)	**0.88**	**0.84**	**0.88**	**0.75**
6	Prop(Appropriate)	**0.78**	**0.78**	**0.69**	0.56
	Prop(Acceptable)	**0.78**	**0.75**	0.59	0.56
7	Prop(Appropriate)	**0.72**	**0.88**	**0.81**	0.47
	Prop(Acceptable)	**0.78**	**0.78**	**0.78**	0.63
8	Prop(Appropriate)	**0.81**	**0.88**	0.47	0.44
	Prop(Acceptable)	**0.69**	**0.72**	0.41	0.22
9	Prop(Appropriate)	0.59	**0.81**	**0.69**	**0.69**
	Prop(Acceptable)	**0.69**	**0.78**	**0.72**	0.63
10	Prop(Appropriate)	**0.78**	**0.91**	0.63	0.44
	Prop(Acceptable)	**0.69**	**0.69**	0.50	**0.22**

Note: Each value in larger bold font is statistically significantly different from 0.50 with 95% confidence.

Another way to look at the proportion of positive appropriateness ratings and positive acceptability ratings is with star plots (Chambers, 1983). Star plots (also called radar plots) are a graphical method for data analysis used to present multivariate data concisely for visual inspection and comparison. In Figure 6.4, the arms of the star plot for a given haptic pattern are made up of the proportion of positive responses (between 0.0 and 1.0) at each of the four TTCs (1.7, 3.0, 5.0 and 7.0 s). Separate lines for appropriateness and acceptability judgments are drawn to connect the arms and provide a profile. Thus the star plots are a graphical rendering of the data in Table 1. The figures show that some haptic patterns were judged unfavorably for some TTCs. Pattern 5 stands out from the rest in that it was generally rated as an appropriate and acceptable HMI for all the TTCs tested with the mirage method. A consistent stimulus-response mapping promotes learning and skill acquisition (Schmidt and Wrisberg, 2008). Therefore a single haptic steering pattern that is judged appropriate and acceptable for the most TTCs is preferable to promote a consistent driver experience.

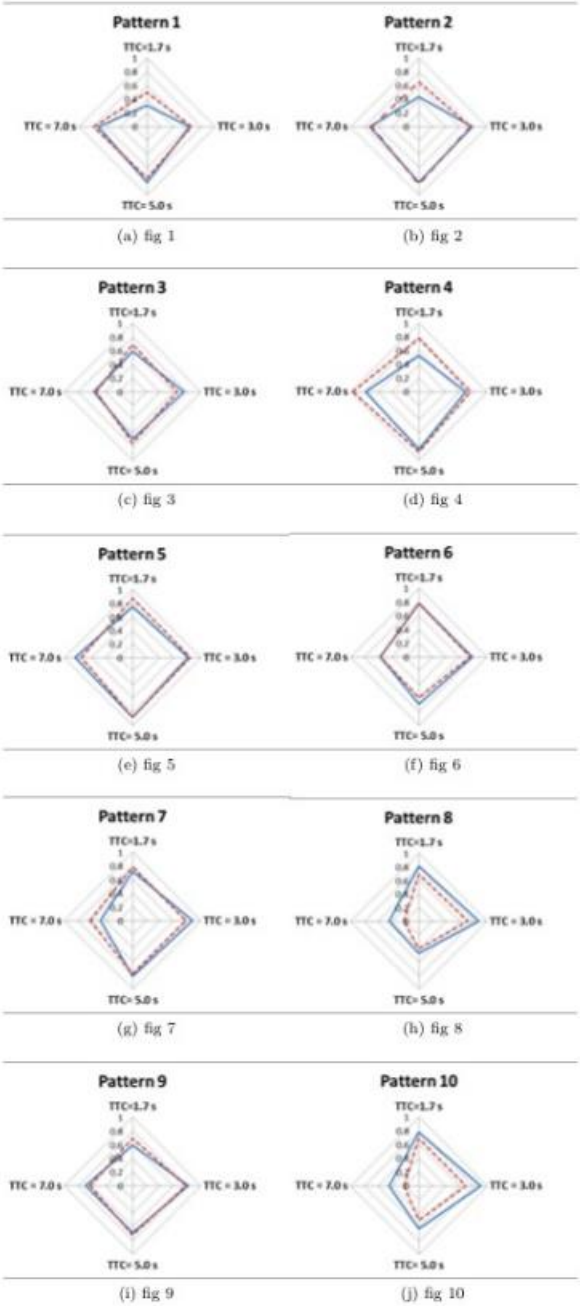

Figure 6.4: Star plots, per haptic steering pattern, of the probability of a positive rating on appropriateness (blue solid line) and acceptability (red broken line) ratings at various TTCs

The star plots also reveal something about the overall relationship between appropriateness ratings and acceptability ratings. For example, a higher percentage of participants found pattern 1 acceptable for the TTC = 1.7 s condition than found it appropriate. A similar result appears for pattern 2 and pattern 4. In contrast, pattern 10 was judged appropriate to a higher degree than it was judged acceptable in the TTC = 3.0 s condition. Patterns 5 and 6 more closely align both appropriateness and acceptability. These variations suggest that appropriateness and acceptability do vary together, though not necessarily perfectly. It is possible that a pattern might be judged acceptable but not necessarily appropriate or vice versa.

To investigate which patterns were judged both appropriate and acceptable, the proportion of participants who responded positively on both rating scales was calculated for each haptic pattern in a given TTC. These are shown in Table 6.2 below. Proportions in larger bold font are statistically significantly different from 0.50 with 95% confidence. A proportion of 0.5 can again be considered to reflect ambivalence among the participants as a group. The two-tailed sign test with 95% confidence is significant for table entries in large bold font. Patterns 1, 2, 8, and 10 were judged both appropriate and acceptable significantly less than 50-50 for at least one TTC condition. Pattern 5 was again favored as highly as, or more highly than, any other pattern in most TTC conditions. The exception is pattern 4, which was most favored for the *TTC* = 7 s condition. Figure 6.5 presents star plots of the proportion of 'joint' positive responses for both appropriateness and acceptability at each of the four TTCs and is a graphical rendering of Table 6.2. These plots resemble those in Figure 6.4, with pattern 5 again standing out favorably across TTCs.

Table 6.2: Proportions of ratings jointly positive on both appropriateness and acceptability by TTC per haptic pattern

		Time-To-Collision (TTC)			
Pattern	**Prop(Dimension)**	**TTC = 1.7 s**	**TTC = 3.0 s**	**TTC = 5.0 s**	**TTC = 7.0 s**
1	Prop(Appropriate & Acceptable)	**0.22**	0.50	0.63	0.66
2	Prop(Appropriate & Acceptable)	**0.31**	0.63	**0.75**	0.56
3	Prop(Appropriate & Acceptable)	0.50	0.53	0.66	0.44
4	Prop(Appropriate & Acceptable)	0.47	0.56	**0.81**	**0.78**
5	Prop(Appropriate & Acceptable)	**0.69**	**0.72**	**0.81**	0.66
6	Prop(Appropriate & Acceptable)	**0.69**	0.66	0.53	0.44
7	Prop(Appropriate & Acceptable)	0.56	**0.72**	**0.72**	0.41
8	Prop(Appropriate & Acceptable)	0.63	0.66	**0.31**	**0.16**
9	Prop(Appropriate & Acceptable)	0.53	**0.69**	0.59	0.53
10	Prop(Appropriate & Acceptable)	0.66	**0.69**	0.44	**0.16**

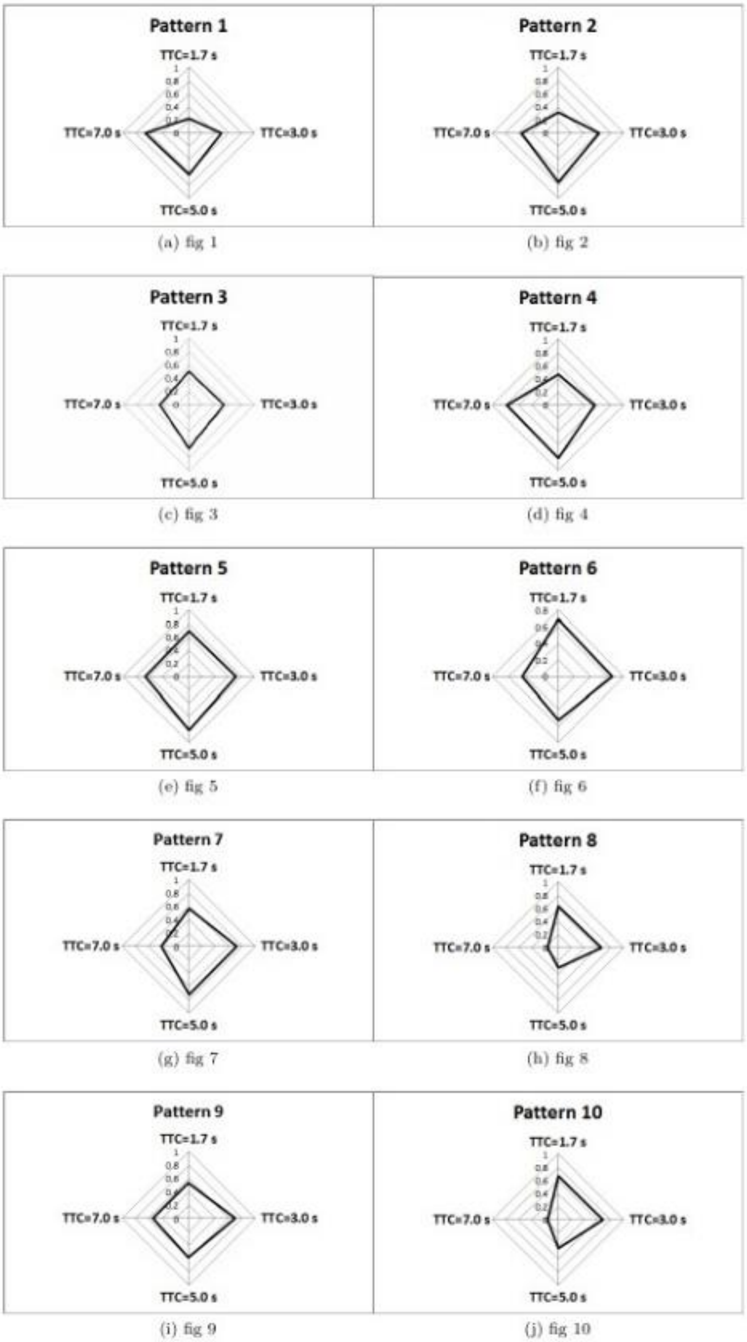

Figure 6.5: Star plots of the proportion of participants who rated a haptic pattern positively on both appropriateness and acceptability at various TTCs.

Analysis of the Rating Magnitudes

The analysis of rating magnitudes and signs (i.e., - 3, -2, -1, 1, 2, 3) to identify the “best” pattern or patterns was performed by first converting all ratings for appropriateness and separately for acceptability, into z-scores across all TTC conditions and all haptic steering patterns using the following formula:

$$Z = \frac{X - \bar{\bar{X}}}{sd(X)}$$

This transformation preserves the distribution of the original data. A negative z-score can be interpreted as indicative of a pattern being inappropriate or unacceptable, and the reverse for a positive z-score. Once the array of z-scores was calculated separately for both appropriateness and acceptability, the arithmetic mean z-score was calculated for each of the ten haptic patterns (i.e., profiles) in each of the four TTC conditions.

Figure 6.6 shows the mean z-scores for each haptic pattern, plotted per TTC condition. Reference lines at the zero point for each dimension have been added to define quadrants. The upper right quadrant in each plot indicates the patterns that were rated, on average, both acceptable and appropriate for that TTC condition. The remaining three quadrants indicate average ratings that were unacceptable in either or both dimensions.

Close scrutiny of the plots in Figure 6.6 reveals several important points. For example, patterns 8 and 10 (both relatively harsh haptic patterns) are scaled high in appropriateness and acceptability for the $TTC = 1.7$ s and $TTC = 3$ s conditions. On the other hand, they are scaled low in appropriateness as well as acceptability for TTC = 5 s and TTC = 7 s conditions. In contrast, pattern 1 (the mildest haptic pattern) is scaled very low in both appropriateness and acceptability for the $TTC = 1.7$ s and $TTC = 3$ s conditions but more favorably in the 5 s and 7 s TTC conditions. These results suggest a qualitative difference between TTCs of 3.0 s or less and TTCs of 5 seconds or more. Cacciabue (2013) defined a driver warning delivered at a TTC of between 5 seconds and 10 to 15 seconds as ‘crash-cautionary’ and driver warnings delivered at TTCs below 5 s as ‘crash-imminent.’ These values, in turn, were proposed earlier by COMSIS (1996) and Lerner (1996) based on considerations of driver perception-response times. The data presented here support the distinction between ‘crash-imminent’ and ‘crash-cautionary’, this time based on driver ratings of the haptic steering patterns presented in situ through the mirage events.

Another observation from Figure 6.6 is that strong positive linear correlations exist between the dimensions for three of the four TTC conditions. Significant Pearson correlation coefficients were found for the 1.7 s TTC condition ($r = 0.84$, $p = 0.002$), the 5.0 s TTC condition ($r = 0.94$, $p < 0.001$), and the 7.0 s TTC condition ($r = 0.96$, $p < 0.001$). However, in the 3.0 s TTC condition, there was less of a correlation between average z-scores for appropriateness and acceptability ($r = 0.50$, p = 0.15). The reason for this is unclear. It is possible that the 3 s TTC condition is perceived as more ambiguous than the others and this was manifested in more variation in driver judgments.

By counting the number of TTC conditions for which each haptic pattern appears in the upper right quadrant in Figure 6.6, we can obtain an index of “generality” for each pattern across the different TTC conditions (see Table 6.3). Only pattern 5 (the 1.0 s, 3-pulse, 3 Nm pattern) is rated both acceptable and appropriate, on average, across all TTC conditions tested. All of the other patterns are, on average, rated unacceptable, inappropriate, or both in one or more TTC conditions.

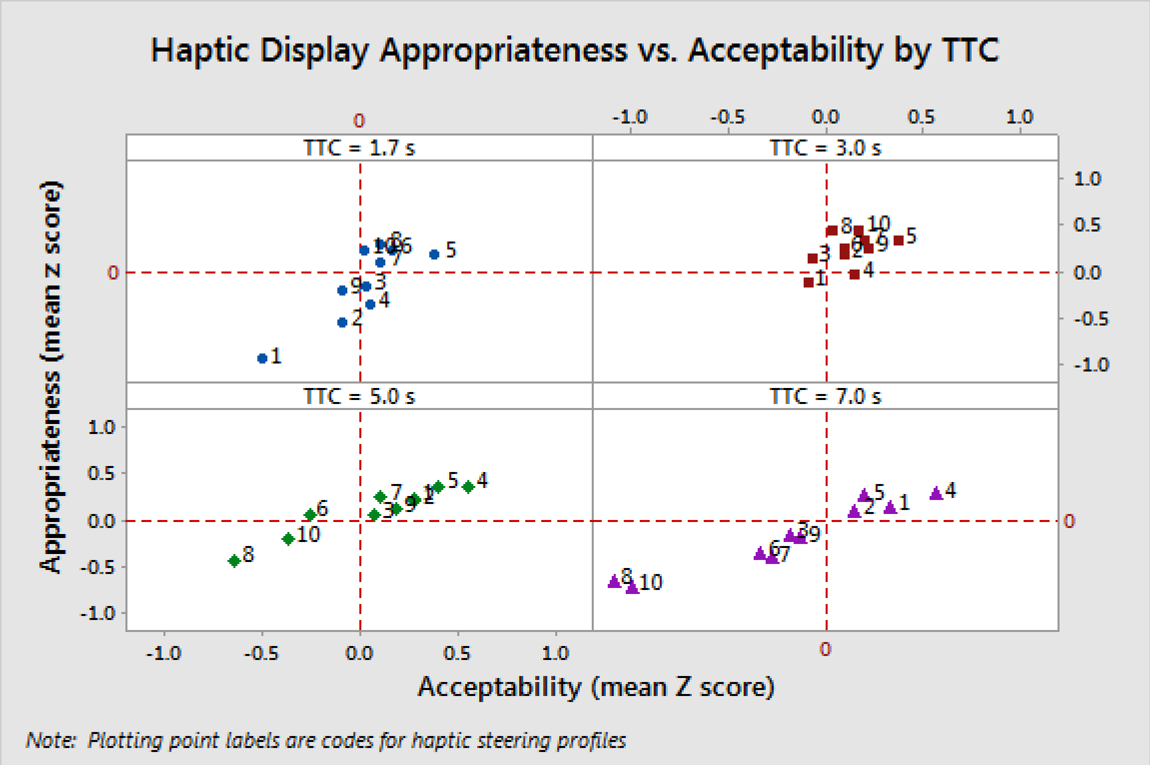

Figure 6.6: Mean Appropriateness vs. Acceptability for haptic steering patterns by TTC condition

Table 6.3: 'Generality' scores or counts of the number of TTC conditions each Haptic pattern is typically rated positively on both Appropriateness and Acceptability

Haptic Steering Pattern Number	TTC Condition with positive mean Appropriateness and mean Acceptability Z-scores (1 = 'Yes'; 0 = 'No')				Generality Score (TTC conditions out of 4 tested)
	1.7 s	3.0 s	5.0 s	7.0 s	Total
1	0	0	1	1	2
2	1	1	1	0	3
3	0	0	1	0	1
4	1	1	0	0	2
5	1	1	1	1	4
6	0	0	1	1	2
7	0	1	1	1	3
8	0	0	1	1	2
9	0	1	1	0	2
10	1	1	0	0	2

Additional analysis was undertaken to examine the values of appropriateness and acceptability ratings jointly. Figure 6.6 and Table 6.3 reflect the average standardized score ratings of each pattern on each rating dimension taken separately per TTC condition. For the analysis of joint ratings, participant responses were characterized as falling into one of three classes. The “joint positive” class encompasses pattern rated favorably on both dimensions by the same participant. The “joint negative” class encompasses patterns rated unfavorably on both dimensions by a participant. The “joint divergent” class encompasses patterns rated favorably on one dimension but unfavorably on the other dimension by a participant. For each TTC condition, participants ratings of each haptic pattern were segregated into each of these three classes. Then the product of the appropriateness and acceptability ratings was calculated, per participant, and these products were summed for each class per pattern within each TTC. Finally, the evaluation of the alternative haptic patterns was carried out in a Multi-Attribute Utility Theory or MAUT framework (Chelst and Canbolat, 2011).

To conduct the MAUT analysis, a simple additive model was developed with four attributes weighted equally: joint favorable ratings, joint unfavorable ratings, joint divergent ratings, and generality. For each attribute, a linear utility function was applied for simplicity and in the absence of any empirical or theoretical justification for a more complex form. An overall objective function to be maximized was defined for which joint favorable ratings’ product sums and generality values added to the objective function and for which joint unfavorable and joint divergent ratings product sums were subtracted from the value of the objective function.

Maximize for a given TTC condition j (i.e., 1.7, 3.0, 5.0, or 7.0 s):

$$Objective\ (Pattern_i | TTC_j) = JointPositive - JointNegative - JointDivergent + Generality$$

Where each of the following terms is normalized to fall within the dimensionless range of 0 to 100:

- JointPositive = sum of products of jointly favorable ratings
- JointNegative = sum of products of jointly unfavorable ratings
- JointDivergent = sum of absolute value of products of favorable ratings on one dimension, unfavorable on the other dimension
- Generality = the number of TTC conditions from Table 6.3

The MAUT model was applied to evaluate all haptic patterns in each TTC condition; the results are provided in Figure 6.7. The figure indicates that pattern 5 is the only pattern that is uniformly at the top of the utility scale across all TTCs. This is consistent with other analyses presented earlier. Pattern 1 (the mildest stimulus) is evaluated lowest for the 1.7 and 3.0 s TTC conditions, but more favorably for the 5.0 and 7.0 s TTC conditions. This is in keeping with earlier analyses that indicated there is a psychological reality to the difference between crash-imminent conditions (TTC of 3 s or less) and crash-cautionary conditions (TTCs of more than 3 s). Note also that patterns 10 and 8 (more aggressive torque profiles) are rated lowest for the 5.0 and 7.0s TTC conditions but higher for the 3.0 and 1.7 s conditions. This is also in keeping with the crash-imminent and crash-cautionary distinction.

The MAUT analysis was repeated with counts of joint positive, joint negative, and joint divergent ratings rather than the rating products themselves (see Figure 6.8). Although there is some shifting of the profiles, the results are essentially the same as those obtained with product sums. One noticeable difference is that pattern 4 swaps positions with pattern 5 as the top choice for the *TTC* = 7.0 s condition, though it remains a poor choice for crash-imminent TTC conditions.

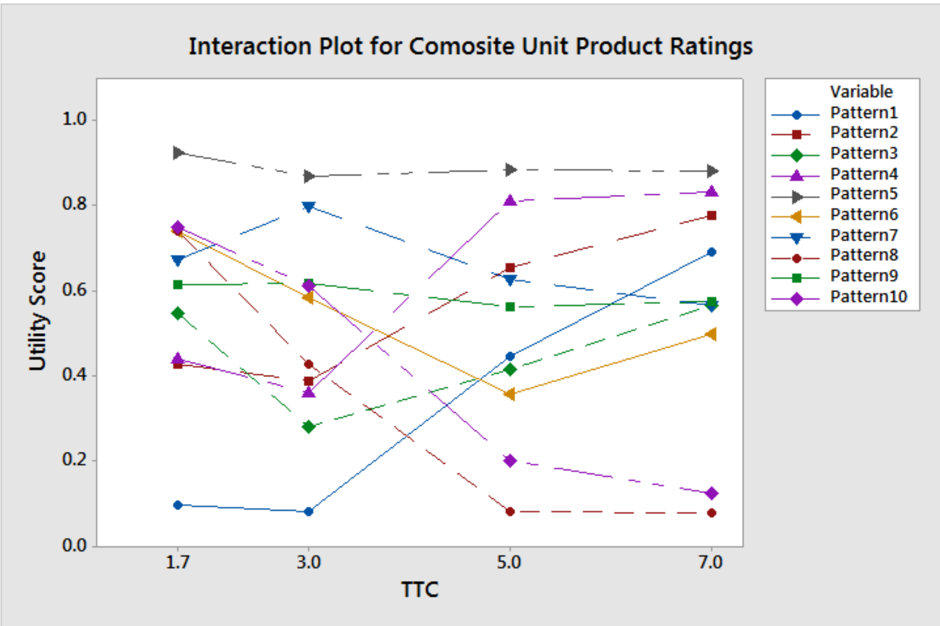

Figure 6.7: MAUT results based on joint Appropriateness and Acceptability ratings (sum of products) and Generality

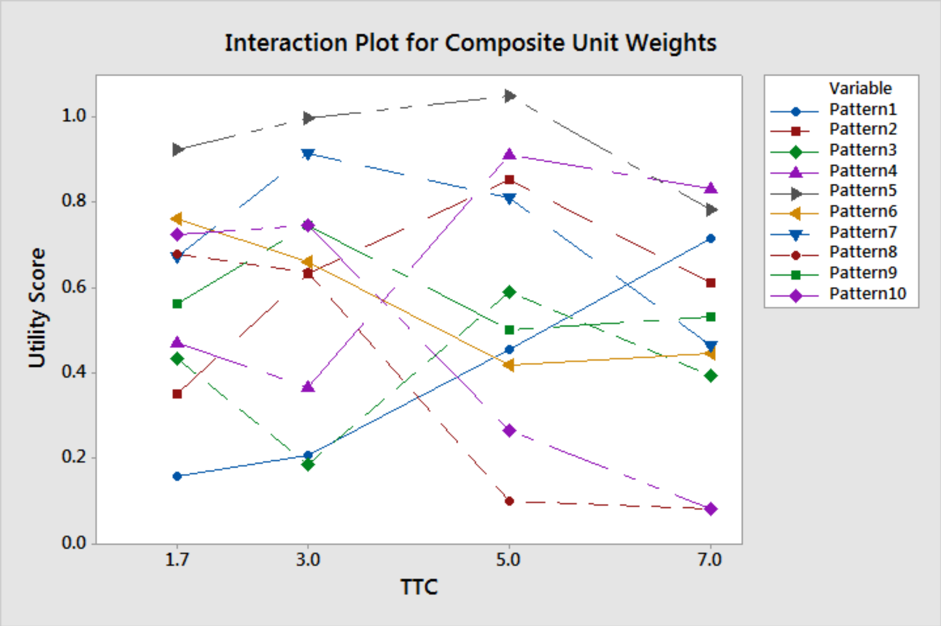

Figure 6.8: MAUT results based on joint Appropriateness and Acceptability ratings (participant counts) and Generality

6.5 Discussion

To guide system design, it can be useful to have persons compare multiple alternative driver warnings or alerts using subjective assessments of appropriateness, acceptability, conspicuity, or other dimensions. Alternatively, some evaluations have a driver react to a critical event repeatedly but with different warning displays. Unfortunately, drivers are not easily surprised more than once, and this can compromise the usefulness of driver performance data if repeated exposures are used. Furthermore, repeated braking or steering in a simulator (even a high-fidelity moving-base simulator) can lead to simulator sickness. One solution requires breaking up the evaluation process into phases of formative evaluation (done while the system is still being developed) and summative evaluation (done after the system has been designed and is ready for final testing prior to release). To support formative evaluation, this paper describes a new method, here called a 'mirage scenario'. This method provides driving contexts (various TTCs to a lead vehicle) similar to those of the target application, briefly presented and then removed. The specific application was for a haptic steering display concept that warns of an obstacle ahead and indicates that a steering response could be initiated to drive around the obstacle. A motion- base simulator allowed for vehicle motion cues similar to what the actual application would provide. Surprise was neither present nor of concern, as it would be for a summative evaluation. Furthermore, no collision avoidance maneuvers were performed, so the risk of simulator sickness was reduced. A one-trial-per-participant testing protocol might then be reserved for a subsequent summative evaluation of driver warning systems for infrequent hazards where the presence of a hazard and the element of surprise are both required.

The evaluations were appropriateness and acceptability ratings that used 6-point balanced bipolar rating scales at each of several TTCs. This subjective rating format allowed for two levels of analysis with which to select suitable haptic steering profiles for future studies. Analyses of the proportion of positive responses on each rating dimension, as well as joint positive ratings, were first pursued through graphical exploratory data analysis methods and inferential techniques. These results were then augmented with additional analyses of the magnitudes of the various positive and negative participant ratings, both for each rating dimension separately, and then for the ratings jointly combined in a MAUT analysis using a simple additive model. This model maximized an objective function that increased with joint favorable ratings on both dimensions by a participant; decreased with joint unfavorable ratings on both dimensions; decreased with divergent (favorable on one dimension, but unfavorable on the other) ratings; and increased with generality (the number of TTC conditions for which the pattern received both positive average appropriateness ratings and positive acceptability ratings). Joint favorable ratings, while never unanimous, are desirable just as joint negative ratings and divergent ratings are undesirable. Generality is desirable to provide display consistency and familiarization over a range of crash- cautionary and crash-imminent conditions. Based on these analyses, a single pattern (pattern 5) was identified that will be used in future evaluations.

As mentioned in the Method section, the second part of the larger simulator session involved pair comparisons of the haptic patterns in terms of urgency. Urgency was defined for the participants as “the level of timeliness and imperativeness to change lanes implied by the profile. In short…which profile calls for more prompt action.” Figure 6.9 shows the results of ordinal scaling of the patterns, from most urgent to least urgent, using pattern codes). Patterns that are not significantly different from one another by a studentized range test are underscored. (See Talamonti, et al., in review, for details of the analysis.) Note that pattern 5 landed in the middle of the ordinal urgency scale, patterns 10 and 8 were at the top in terms of urgency, and pattern 1 was ordinally scaled least urgent. Note also that patterns 10 and 8, though not significantly different from each other, are significantly different from patterns 5 and 1 and that pattern 5 is significantly different from pattern 1. Referring back to Table 4.1, the authors originally thought that patterns 1 through 6 were intended to be milder, or crash-cautionary, and patterns 7 through 10 were intended to be more aggressive, or crash-imminent. This appears to be largely borne out in the urgency ratings. With the exception of a reversal of patterns 6 and 7, the rank orders of all the crash-cautionary patterns are indeed lower on the urgency scale than all the crash- imminent patterns. These scaling results suggest that the urgency definition and scaling procedures used did produce reasonable experimental outcomes.

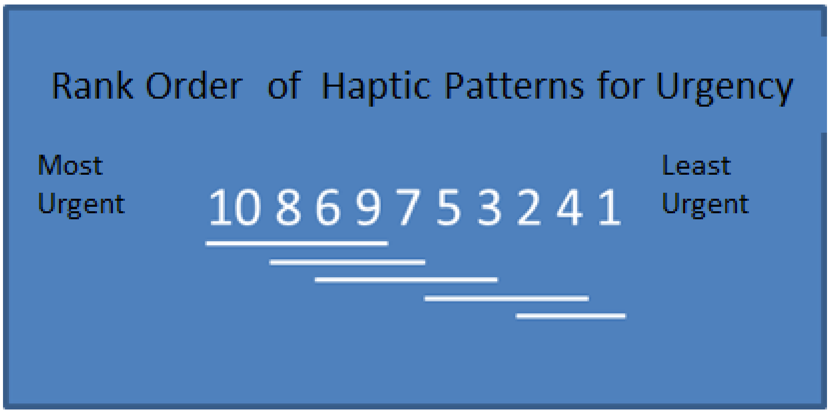

Figure 6.9: Rank Order of Urgency Scaling of Haptic Steering Patterns and Significant Differences by Studentized Range Test

An interesting question that arises from these results is why pattern 5, which sits in the middle of the ordinal urgency scale, was favored over all the remaining patterns when considered in terms of appropriateness and acceptability ratings across all TTCs. The authors originally thought that more urgent profiles should be reserved for crash-imminent situations and that milder profiles should be used for crash-cautionary conditions. Consistent with this expectation, the mirage data indicated low utility of the relatively mild pattern 1 for the crash-imminent (1.7 and 3.0 s) TTC conditions but higher utility in the crash-cautionary (5.0 and 7.0 s) TTC conditions. The higher utility of relatively harsh patterns 10 and 8 in the crash-imminent TTC conditions and their lower utility in crash-cautionary (5.0 and 7.0 s) TTC conditions are also consistent with this expectation. Why pattern 5 nonetheless appears most robust across all TTCs cannot be answered from the data in this study. However, one possible explanation comes from Coombs (1983) theory of preference behavior. He argues that human preference behavior is often describable by a single-peaked function, a function that increases monotonically to a maximum and then decreases monotonically. As Coombs puts it, it is often the case that “good things satiate...bad things escalate” (p. 21). If so, this suggests that perhaps a medium level of perceived urgency in the haptic steering patterns would be most preferred over a wide range of driving conditions, as appears to be the case in the present study. This would offer the added benefit of allowing a consistent stimulus-response mapping over multiple exposures via a LCA so that the driver receiving the haptic cues will more readily interpret it. This could be particularly useful with a distracted driver or a driver under stress, e.g., a driver in a crash-imminent situation with insufficient time or distance to brake. Whether this type of haptic cue can affect driver response in a surprise situation remains the subject of a future study.

If so, this would have the added benefit of allow a consistent stimulus-response mapping over multiple exposures so that the driver receiving the haptic cues will more readily interpret it. This could be particularly useful when the driver is distracted or under stress such as in a crash-imminent situation with not enough time or distance to brake.

CHAPTER 7 STUDY 4: SCREENING STUDY TO EVALUATE FACTORS INFLUENCING HMI EFFECTIVENESS

7.1 Abstract

If one is exposed to a novel display in normal driving, this exposure may enhance the effectiveness of that display in an emergency. This enhancement might arise from increased display familiarity, the introduction of a mental set for a wider range of response or maneuver options, or both. The haptic steering display was novel, and because it operated on a primary driver control element (the steering wheel), it could be perceived as a vehicle fault unless the driver was familiar with it. Therefore, a system concept was developed, as part of the overall HMI, that is here referred to as a lane change advisor. The lane change advisor HMI is the same in normal driving (to suggest lane changes when encountering slower vehicles ahead in the travel lane) and in emergencies. An initial screening study reported here used a combined FOST/haptic steering display system for both lane change advising and emergency steering alerts. This study was conducted for crash-imminent conditions (1.7 or 3.0 s TTCs) in both manual driving and shared control (hands-on) autonomous driving modes. Results of the screening study indicated the following: a) drivers in manual mode responded reliably faster than drivers in shared-control autonomous conditions; b) drivers in the 1.7 s TTC scenario responded faster than drivers in the 3.0 s TTC scenario regardless of driving mode; c) drivers who were distracted when the forward collision event arose responded later than those who were alert (as expected for Level 2 and Level 3 automated driving); and d) all participants in all conditions did some form of steering; no driver out of 63 participants braked only. These results suggest that the HMI system was highly successful in facilitating lane changes for emergency conditions. However, only three participants were observed even glancing to the FOST display during the critical event. Thus concerns remained regarding the possibility of a methodological artifact associated with high emphasis on lane changes given in training. This concern led to a series of additional conditions being investigated to identify conditions that would bring back the Study 1 results for brake-only behavior in a final study.

7.2 Introduction

For this study, the proprietary automated driving cluster HMI developed for another program was augmented with the new cluster-based auditory-visual FOST display. This auditory-visual FOST display concept is usable for manual driving or fully automated (i.e., hands-off) driving. The discrete haptic steering display identified in Study 3 is usable with the auditory-visual FOST display for manual driving. However, a haptic steering display is not usable for a fully automated, hands-off driving system because the driver is not necessarily touching the wheel when the alert is given. In contrast, the haptic steering alert might be used as part of a hands-on shared-control automated driving system. In the concept of shared control, the driver must keep at least one hand on the steering wheel during automated driving. The automations virtual driver provides torques to the wheel that the driver may feel and augment or override depending on circumstances.

A hands-on, shared-control automated driving system offers several potential advantages over a hands-off automated driving system for Level 2 or Level 3 automation. First, hands-on shared control may increase driver engagement with the driving task. Second, the steering wheel can be used as a haptic display system during automated driving. Since the driver has at least one hand on the wheel during automated driving, the system may be used to indicate system faults, convey automation intent, or present haptic alerts such as those evaluated in Study 3. Third, a manual takeover from automation in an emergency may be faster. Faster response might be expected if only because shared control eliminates the additional movement time needed to get hands back on the wheel. Given the significant differences between manual driving and hands-off fully automated driving found in Study 1 and Study 2, the decision was made to investigate what benefits might accrue from a hands-on shared control HMI in Study 4. Study 4, described here, therefore included the auditory-visual FOST display with and without haptics to assess the effects of the haptics on driver response in both manual and shared-control automated driving.

Test methodology questions arose because of a difference in results for the 1.7 s TTC crash- imminent hazard conditions tested in Study 1 and in Study 2. The initial study (see Chapter 2) that motivated this dissertation (Blommer et al., 2017) compared manual drivers with drivers in fully automated (i.e., hands-off) driving. All drivers had to manually respond to a high-intensity FC hazard. Results indicated that drivers who manually took over from fully automated driving responded more slowly than manual drivers. They also responded with a roughly 50 −50 chance of braking only versus steering, as compared to an 80 −20 chance of steering and braking versus braking only for manual drivers. A second study (see Chapter 3) investigated driver response across a wider range of TTCs. Results at 5 and 7 s TTCs were similar to those reported in other published research. However, this second study did not replicate the 50 − 50 chance of braking only at the shortest TTC for fully automated driving. This discrepancy in results was unexpected. The second study differed from the first in greater simulator lane change practice. The present study sought to understand the influence of high versus low levels of lane change practice on driver response to an imminent forward collision hazard.

The LCA concept presented the HMI components (FOST display with or without haptics) in normal driving contexts to familiarize the driver with the HMI. It was of interest to determine whether there might be a significant difference in driver responses as a function of low versus high exposure to the lane change advisor during a simulated highway commute of approximately 30 minutes.

The last factor included in the present study had to do with the drivers state of attention to the road. Level 2 and Level 3 automation assume an alert driver, and this was incorporated in Study 1 and Study 2. However, the driver is sometimes not alert in manual driving and therefore may not be alert at hazard onset in automated driving. Thus, the effects of a lapse in driver attention were investigated in the screening study as well.

Study 2 found no appreciable difference in driver response times between 1.7 s and 3 s TTC conditions. However, both were different from the 5 and 7 s TTC conditions. This suggests that 3 seconds or shorter TTCs constitute crash-imminent hazard situations whereas 5 s or longer TTCs may be considered crash-probable or crash-possible (Allen, 1995). Emergency lane changes may be most appropriate for crash-imminent FC situations. Therefore, the present study focused on 1.7 and 3 s TTC conditions and any differences in driver response that might arise from them with the new HMI concepts.

In summary, this study applied a fractional factorial experimental design to evaluate the effects of six factors thought to influence driver response to crash-imminent FC situations. These factors were as follows: manual versus hands-on shared-control automated driving; level of crash-imminent hazard intensity (1.7 vs 3 s TTC); lane change practice (2 lane changes versus 10); exposure to the Lane Change Adviser (low versus high); HMI (FOST display with and without haptics); and driver attentional state (eyes forward versus eyes not forward at FC event onset).

The following research questions motivated this study. In general, the questions seek to identify factors that “help” a driver’s steer response to an unexpected crash-imminent FC hazard. Note that the word “help” means faster response, more steering, or some combination of these. Key research questions are as follows:

1. Does manual driving “help” more than automated driving?
2. Does a haptic display “help” more than no haptics?
3. Does more exposure to the HMI “help” compared to lower exposure?
4. Does more lane change practice in VIRTTEX “help” compared to less?
5. Does shorter or longer crash-imminent TTC influence driver response?
6. Does attentive versus distracted driver state influence driver response?
7. Do these factors interact substantially in an understandable way?

7.3 Method

7.3.1 Participants

The sample consisted of 63 volunteer Ford Motor Company employees between 22 and 65 years of age (M = 43.36 years, SD = 11.18 years), approximately evenly balance by gender and age category. Age was subdivided into two categories: “younger” (22 to 45 years of age) and “older” (46 to 65 years of age). Most of the participants had non-engineering backgrounds (e.g., finance, administration, information technology) and none had prior “surprise event” experience in the Ford VIRtual Test Track EXperiment (VIRTTEX) simulator. Each participant also had to have a valid driver’s license.

7.3.2 Apparatus

Driving Simulator

See Appendix A for details on the VIRTTEX driving simulator. The vehicle dynamics and scenario details for this study were identical to those of Study 3 for the 1.7 and 3 s TTC conditions. Additional details for the HMI cluster are provided below.

HMI Cluster Display

The HHDD for automation state from Study 3 was replaced for this study by an instrument cluster HMI concept that was developed by Ford Motor Company as part of another vehicle program. This HMI concept, though not developed as part of this dissertation, defined design constraints on the auditory-visual FOST display that was created for this dissertation. The new HMI is described in the Procedure section.

7.3.3 Secondary Tasks

A set of secondary tasks (adjust climate control; VIRTTEX numbers task; iPad trails task) was given, identical to that described in Chapter 3. Drivers in all test conditions were required to complete the same secondary tasks. Note that before and during the critical event, there either was or was not a secondary task underway, as determined by the experimental design described below. If the experimental design required a distracted driver, the VIRTTEX Numbers task was administered to ensure that the driver's eyes were off the road when the FC event occurred.

7.3.4 Experimental Design

The study was designed as a one-trial-per-participant, between-subjects, 2^{6-1} fractional factorial experimental design (Box et al., 2005). The six independent factors and their levels are described in Table 7.1. Six factors at two levels yields 64 combinations. The 2^{6-1} design, as a half-fraction design, involves 32 treatment combinations selected with design generator $D = MTHEP$, using the independent variable codes from Table 7.1. This results in a Resolution VI design with the following properties:

1. Main effects are aliased with 5-way interactions or higher;
2. 2-way interactions are aliased with 4-way interactions;
3. 3-way interactions are aliased with other 3-way interactions.

Assuming higher-order interactions are non-significant or trivial, this experimental design efficiently screens for main effects and 2-way interactions. For n = 64 test participants, this design results in a $Power$ = 0.75 with 2 participants per cell assuming an alpha of 0.05 and a "large" effect size (Cohen, 1988). For a complete power analysis and a depiction of the experimental design, see Appendix B.

Table 7.1: Independent Variables and Levels

Factor	Code	Level	Description
Drive Mode	M	-1	Manual Driving
		+1	Automated Driving (Shared Control)
TTC	T	-1	3.0 second
		+1	1.7 second
HMI	H	-1	FOST
		+1	FOST + Haptics
HMI Exposure	E	-1	Low: 2 Exposures
		+1	High: 8 Exposures
Lane Change Practice	P	-1	Low: 1 pair
		+1	High: 4 pair
Distraction	D	-1	No distraction at FC event onset
		+1	Distracted at FC event onset

7.3.5 Response Variables

The dependent variables were a) driver response time from event onset until first response onset and b) the nature of the response (steer only, steer and brake, brake only). Driver eye-glance video was also captured and used to determine whether participants were looking toward the scene at hazard onset and also to assess whether the driver glanced at the FOST display at FC event onset.

7.3.6 Simulated Driving Scenario

The simulated driving scenario was the same as that described in detail in Study 3 with the following exceptions. After receiving static secondary task training while stopped, drivers were asked to put the vehicle into gear, accelerate to 65 mph, and then drive for a short period of time to become acclimated to the driving simulator. Participants were told that the vehicle was equipped with Lane Departure Warning and that they could experience it by intentionally steering out of the lane to either the left or right to simulate a drift. Once they received the LDW, they were told to re-center their vehicle in lane. All participants were then instructed on how to activate and deactivate automation. Additionally, each participant in the hands-on, shared control, automated drive group was given additional information on the 'maintain control' instance (described below).

After the initial training was the conditioned training. The conditioned training is where the manipulations of the assigned experimental conditions took place based off of the exposure level of each main effect assigned to the participant. Training described by condition is independent of whether the participant was in any of the following three main effects conditions: 1) automated or manual, 2) distracted or non-distracted, and 3) with or without haptics. The procedural difference between being in the manual drive mode and the hands-off shared-control automated drive mode was a few sentences changed or removed in the operator script. Those in automation received additional instruction regarding hands on/off warning and additional instruction on the input force required to break out of lateral control. Whether or not the participant was to be distracted at the final event had no effect on the training or operator script. The addition of haptics as a factor required only one additional line from the operator, reiterating to the participant that he/she should keep both hands on the wheel for the initial (haptic profile) exposure. In order to not only expose the participant to either high lane change practice or low lane change practice, but also contain the condition of either high or low FOST exposure, two simulation databases were created. The first database contained 8 shifts in the roadway, and the second database contained none. The database selection was then inherent to whether the participant was to receive low or high lane-change practice. Specific experimental conditions and their description are given in the Table 7.2, below.

Table 7.2: Description of Experimental Conditions

Experimental Condition	Description
High Lane Change Practice & High Exposure to FOST Display	Each participant assigned to this condition experienced a series of true positive conditions of the Lane Change Adviser which they then used to make a total of 8 lane changes.
High Lane Change Practice & Low Exposure to FOST Display	Participant in this condition made 8 lane changes without any exposure to the FOST display. This can be considered normal lane changing and was done so via (a construction zone or road geometry) rather than through a slow moving lead vehicle scenario as required by the Lane Change Adviser system
Low Lane Change Practice & High Exposure to FOST Display	Each participant in this condition received could be construed as a false activation of the Lane Change Adviser feature. Here participants are told that while they will experience the FOST display, they are not to make a lane change in response to it. The participant is also told that the activation of the display is purely for demonstration purposes. This was done using the construction zone database.
Low Lane Change Practice & Low Exposure to FOST Display	Each participant in this condition received no lane change practice and no exposure to the FOST display.,These participants received the introductory material mentioned above and had to continue the same length of drive as all other participants only with secondary tasks assigned to keep them from becoming bored or tired.

After the training portion of the drive was complete, the study portion began. This consisted of primarily straight highway driving. Approximately halfway through the drive, each participant experienced a true positive exposure to the LCA (i.e., FOST display with or without haptics) upon encountering a slow lead vehicle. Specifically, the participant was traveling in the right lane; then a slow lead vehicle became visible at approximately 10 seconds of car following Time-Headway (TH). At 7 s TH, the LCA activated, using the FOST display with or without haptics to recommend a leftward lane change to the participant. Once the participant made the lane change, he/she spent a short period in the left lane before another slower lead vehicle became visible. As in the first instance, at a TH of 7 s, the lane change adviser activated, this time signaling a lane change back to the right lane.

The hazard conditions tested were reduced to 1.7 and 3 s FC events, i.e., crash-imminent events only. The timing of the hazard events were kept identical to those in Study 3. Depending on their assigned condition, participants either were not distracted at FC event onset or were distracted at FC onset via the VIRTTEX numbers task. At the FC hazard event, each participant received the FOST display, either with or without haptics, in combination with FCW. The FOST display, again with or without haptics, and FCW alert warning sound commenced concurrently with each other in manual driving and concurrently with autobraking in the hands-on shared-control autonomous control mode.

7.3.7 Procedure

Test participants arrived at the VIRTTEX simulator facility, signed the informed consent form, reviewed a safety video, and were shown a presentation that introduced the automated driving concept and the lane change adviser concept. The introduction explained in general terms how an automated vehicle might operate to manage longitudinal control, lateral control, and car following separation. The training included slides that indicated that the system could not necessarily handle all driving situations. Drivers were told their first responsibility was to safely control the vehicle at all times even when automation was engaged.

In the training materials, the automated driving system was framed as a prototype system that was unable to differentiate between intentional and unintentional steering input. For that reason, participants were required to keep at least one hand lightly on the steering wheel at all times. Participants were told that if they took their hands off for 10 s, they would receive an audible alert instructing them to “keep hands on wheel”. This was the ruse used to implement a hands-on, shared-control, automated driving regime.

Automation (i.e., the automated driving system) was described through a series of temporally correct screen flows depicting exactly what the participant would see in the instrument cluster as a state of automation in time. The first illustration described to each participant as what their instrument cluster would look like in manual driving. Once conditions became proper for automation to be engaged, participants would see a pop-up message in the instrument cluster and hear an audible voice prompt telling them so. This was shown in a second illustration. Participants were told that the automation on/off button was a toggle button on the steering wheel that would be pointed out to them in the simulator. Once they pressed the toggle button to activate automation, the instrument cluster would transition to one that depicted what automation saw. This was the third illustration given. It was stressed to each participant that if he/she intentionally or unintentionally turned the steering wheel past a certain point, he/she would be preemptively given the lane departure warning and would see a "hands on wheel" icon in the center of the instrument cluster. This was the fourth illustration. Participants were told that at this point they should be steering but that automation would continue to maintain longitudinal control. Once they re-centered the vehicle in the lane they chose, automation would resume steering control. This was the fifth illustration, nearly identical to the third except for auditory confirmation and visual confirmation via a pop-up saying "lanekeeping resumed". If at any point they experienced an alert asking them to disengage automation, they could use the brake or the off button, at which point they received a positive indication that manual driving was reengaged.

The LCA was explained as a sub-function of automation that used the sensors and brains of automation to recommend when to make a lane change. Participants were told that they could choose to follow or disregard the recommendation. Most importantly, participants were told that the system would not recommend a lane change to them if one was not feasible.

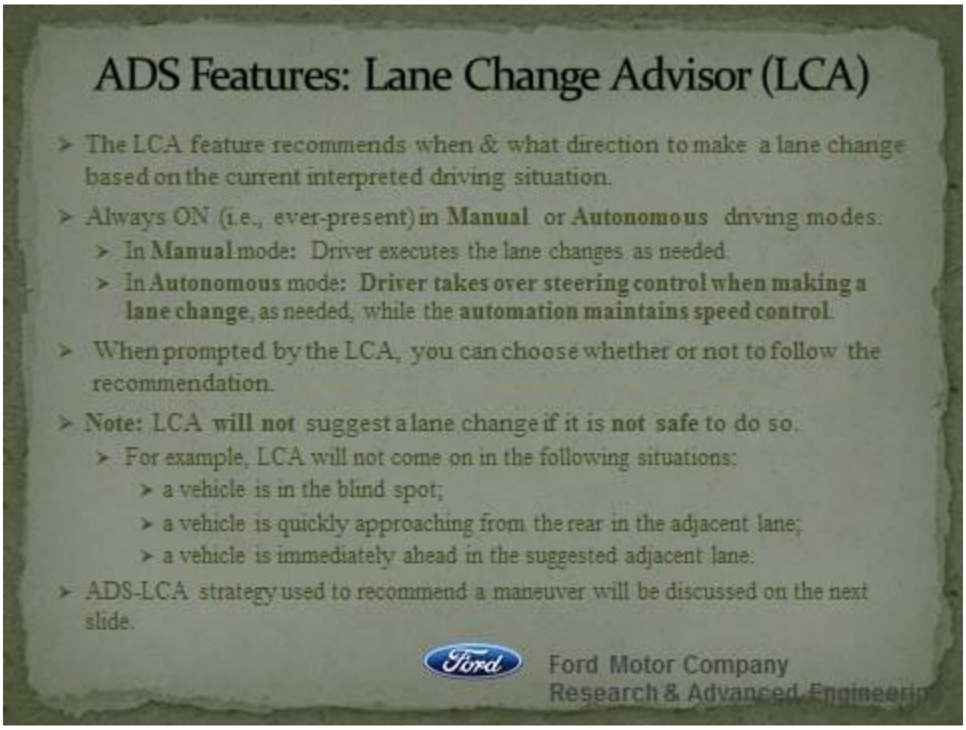

Figure 7.1: Description of the LCA

After the description of the LCA, each participant was shown a picture of what it looks like when activated and given a description of each element of the display.

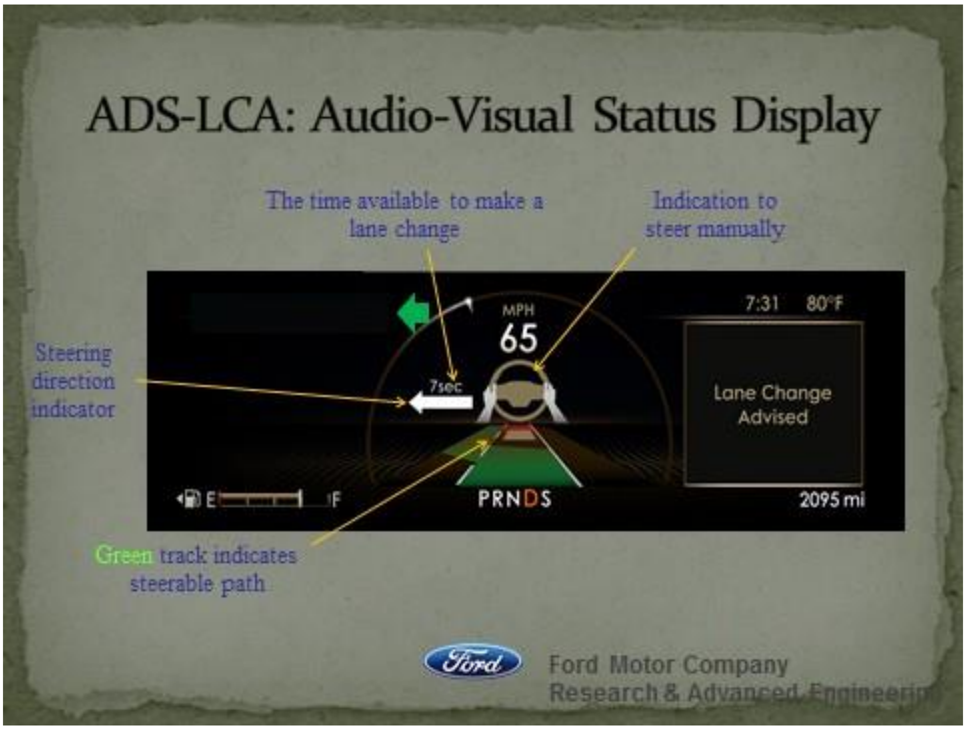

Figure 7.2: Auditory-Visual Elements of the LCA

Depending on the level of HMI that each participant was selected for, he/she was or was not given one more slide explaining the haptics. Participants were told that to enhance the lane change adviser auditory-visual display, an explicit haptic steering profile was used to reinforce the suggested steering direction. It was also clarified that participants would experience same haptic profile each time a lane change was recommended. The slides also introduced the secondary tasks that the drivers would need to perform. No mention was made of a forward collision warning or event. After this, the participant was accompanied to the VIRTTEX simulator cab.

After the driver was comfortably seated in the simulator, he/she began training, while stationary, with review of the requested secondary tasks. The ride-along observer, seated in the back passenger seat, then remained silent until the end of the simulator session while the simulator operator continued the driving portion of the training. This portion involved practicing the secondary tasks while driving, getting a feel for making lane changes in the simulator, and learning how to activate and deactivate automation. There were 4 lane changes for acclimation to the simulator handling characteristics and an additional 4 lane changes associated with deactivating automation by steering.

Half of all participants were assigned to manual driving for baseline comparison purposes. Drivers assigned to manual driving were told of this assignment at the end of the training period in the simulator just before the start of the study drive. Thus all participants received the same training (i.e., experiencing automation and learning to drive VIRTTEX) regardless of their assigned drive condition for the study portion of the drive.

The VIRTTEX operator in the control room and the ride-along observer in the back seat did not interact with the test participant until the study drive was completed, i.e., after the critical event was over. At that point, test participants were asked if they recalled receiving a warning and, if so, whether they could recall any details about it. They also completed an SSQ. Once the simulator was settled, both the participant and the observer departed the VIRTTEX dome, and the participant was thanked for his/her participation. The entire experimental session took approximately one hour to complete.

7.4 Results

The data were reduced and manually verified for accuracy and quality. The distraction factor required that test participants either be head up and eyes looking toward the forward road scene or eyes-down on task was confirmed by review of the eye-glance video for the data reported here. The response time after the FC event onset and the nature of each participants’ response were analyzed. In the final two days of participant scheduling, some trouble began with the VIRTTEX simulator hydraulic system causing unexpected system shutdowns. On the final day of participant scheduling, the motion system failed completely. Therefore, the final participant sample was short one participant.

7.4.1 Response Time Results

The end result was an *n* of 63 out of the planned 64 participants from the original study. A cell means model as explained by Kirk in (Schinka et al., 2003) was carried out using effects coding in Minitab. Also reported is a simple measure of effect size, eta-squared (η^2), which is an estimate of the proportion of total squared variation associated with an effect. Significant main effects were found for driving mode, TTC, and driver distraction state. The main effect of mode was significant (F (1, 41) = 8.46, p = 0.006, η^2 = 0.05). Manual drivers responded with a mean of 1.40 s; drivers in shared control responded more slowly (mean of 1.65 s). There was a significant main effect for TTC (F (1, 41) = 43.34, p < 0.0001, η^2 = 0.24), with drivers responding more quickly in 1.7 s TTC condition than in the 3.0 s condition (means of 1.25 and 1.79 s, respectively). A distraction main effect was significant (F(1, 41) = 60.25, p < 0.0001, η^2= 0.33), with distracted drivers responding more slowly than alert drivers (means of 1.20 and 1.85 s, respectively). Thus, mode accounted for 5% of the squared variability in response times, while TTC and distraction factors accounted for 24% and 33%, respectively.

There were also 2 statistically significant two-way interactions. There was an interaction be- tween exposure and distraction (F (1, 41) = 4.14, p = 0.05, η^2 = 0.02) and between TTC and attention (F (1, 41) = 4.79, p = 0.03, η^2 = 0.03). The first interaction accounted for only 2% of the variation in response times. The second interaction is roughly an order of magnitude smaller in effect size than either of the component main effects. For these reasons, these two-way interactions are not discussed further.

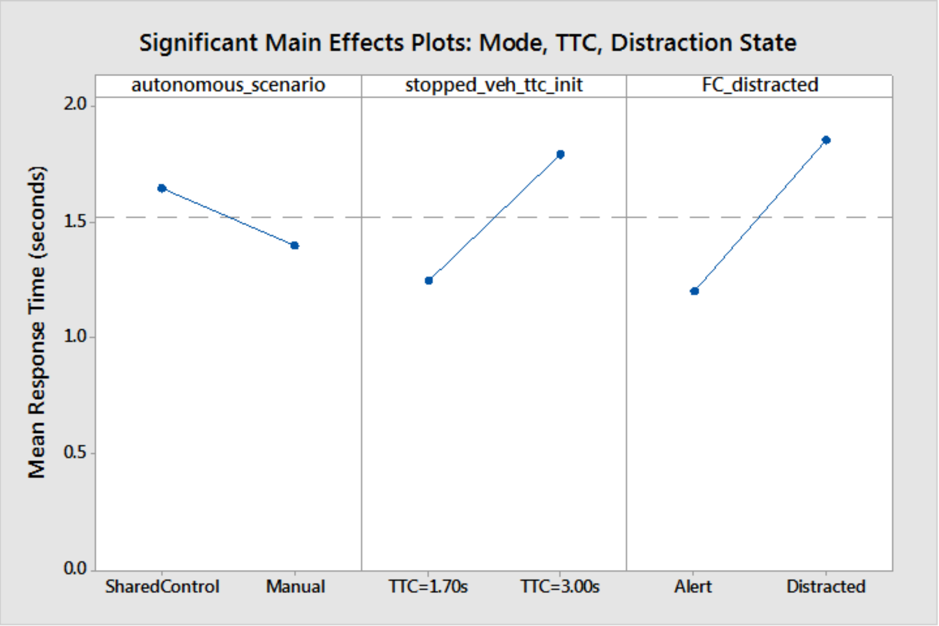

Figure 7.3: Main Effects Plot for Response Time

7.4.2 Response Type Results

Of the 63 participants in the study, no one braked only. All responses were either brake-and-steer or steer-only (smooth steering) responses. Fisher's exact test was applied to each of the main effects to see if any differences in response type existed (see Table 7.3). From the results, one can see that only TTC had a significant effect on response type. At $TTC = 1.7$ s there is a roughly $50 - 50$ split of steer-only and brake-and-steer responses; at $TTC = 3$ s there is a $70 - 30$ split in favor of steer-only responses. No other main effects were significant.

Table 7.3: Fisher's Exact Test on Main Effects

	Steer Only	Brake & Steer	N	Fisher's Exact Test (p-value)
Manual	20	12	32	0.613
Automated	17	14	31	
1.7sec TTC	14	17	31	0.042
3.0sec TTC	23	9	32	
Not Distracted	21	11	32	0.311
Distracted	16	15	31	
FOST	18	14	32	0.799
FOST+Haptics	19	12	31	
Low Exposure	19	13	32	1
High Exposure	18	13	31	
Low Lane Change Practice	18	13	31	1
High Lane Change Practice	19	13	32	

7.5 Discussion

The response time data indicated substantial effects of driver state of distraction and TTC. Averaged over all other conditions, distracted drivers responded about 0.65 s later than alert drivers. Drivers in the 1.7 s TTC condition responded about 0.55 s faster than those in the 3.0 s TTC condition. There was a smaller main effect of drive mode, such that manual drivers responded about 0.25 s faster than drivers in hands-on shared-control automated driving, averaging over all other factors. These results are consistent with expectations. Two two-way interactions were also significant, but accounted for a trivial proportion of the variability in response times and were not considered further.

A surprising result arose in the analysis of response type. Not a single participant braked only, in any condition. Thus, the goal of eliminating brake-only responses in favor of steering was achieved beyond expectations. This might be interpreted to indicate that the LCA and FOST (with or without haptics) promoted steering regardless of the levels of practice and exposure, at both TTCs and regardless of driver attentional state. However, it is also possible that the results arise from procedural influences, e.g., a subtle unintended emphasis on lane changing during orientation and training. As evidence for this, manual review of eye-glance video identified only 3 participants glanced at the FOST instrument cluster display during the FC hazard event. To investigate the possibility of a procedural artifact, another experiment was designed that avoided mention of the lane change adviser in training and used only a low number of lane changes. It also brought back the hands-off, fully autonomous driving mode and compared it to hands-on, shared-control autonomous driving with alert drivers as expected for Level 2 and Level 3 automated driving. This is the subject of Study 5.

CHAPTER 8 STUDY 5: HMI EFFECTS IN FULLY AUTOMATED VERSUS SHARED CONTROL AUTOMATED DRIVING

8.1 Abstract

Most haptics steering displays and driver warnings are oriented toward keeping the driver in the lane. In contrast this dissertation investigated a method to suggest that the driver may steer out of lane when feasible for forward collision avoidance. Results from the screening study (Study 4) indicated a significant response time difference between participants in manual driving and those in hands-on, shared control automated driving. Furthermore, not a single participant in the screening study braked-only in response to the Forward Collision (FC) hazard. Pilot testing revealed that hands-off fully automated driving with low levels of lane change practice and no exposure to a lane change adviser brought back brake-only response behavior like that found in Study 1. In this final study, the research effort compared two different styles of automated driving (hands-off fully automated driving versus hands-on shared control automated driving). Within each of these two automated driving modes, this study compared the effects of a Lane Change Adviser (LCA) on versus off. In this study, this factor is referred to as LCA State where LCA 'on' means the participant received the LCA in both benign driving and at the FC hazard event, and LCA 'off' means that the participant did not receive the LCA, but only Forward Collision Warning (FCW) at the FC hazard. This manipulation was done in order to investigate the potential effects of mental set that might increase the driver's propensity to steer rather than brake-only.

Results of this study indicated that participants in hands-off automated driving responded slower than those in hands-on shared control automated driving. Furthermore, drivers with LCA off responded more slowly than drivers with LCA on regardless of automated drive mode. Results also indicated a statistically significant difference in response type distribution. Specifically, the incidence of brake-only responses were significantly reduced with LCA on as compared to with LCA off, regardless of type of automation. These results collectively suggest that the LCA used to administer the Field of Safe Travel (FOST) display or FOST plus haptic displays in benign conditions increase the propensity of lane change behavior in crash-imminent situations.

8.2 Introduction

Study 4 examined a variety of factors and their influence on driver response including: manual versus hands-on shared control; alert versus distracted driver state; 1.7 versus 3.0 s TTC; high versus low lane change practice; and high versus low exposure to the LCA. Significant main effects for Drive Mode, TTC, and driver Attentional state all had expected effects on response time. On the other hand, results from study 4 revealed that all participants exhibited some form of steering. Not a single driver braked- only to the FC hazard event. These results suggest the LCA delivery of FOST with or without haptics was highly successful in eliminating brake-only behavior over a wide range of driver state, practice, exposure, and hazard intensity conditions. However, there are reasons to question this. Manual review of video revealed that only 3 participants looked at the instrument cluster display at FC event onset. Thus questions arose about the level of emphasis given to the LCA and lane change behavior given in simulator orientation. This prompted a series of pilot studies to see if the brake-only behavior identified in Study 1 could be recovered.

The enumerated items below describe a series of focused pilot tests manipulating a variety of procedural factors in order to identify conditions where brake-only behavior was again exhibited. For each of these pilot tests the shortest crash-imminent TTC of 1.7 s was used because Evasive Steer Assist (ESA) would be most appropriate in such a scenario. Also, only the alert driver state expected in Level 2 and Level 3 automated driving systems was employed. The pilot studies carried out were as follows:

1. Pilot Study 1; Manual driving was employed without FOST or Haptics in a drive scenario that involved high lane change practice (10 lane changes) achieved through changes in road geometry (i.e., lane drops through construction zones). Of 7 participants run, 5 participants were looking up at critical event onset. Results indicated 4 brake-&-steer and 1 steer only response. This method was not successful in bringing back brake-only responses.
2. Pilot Study 2; This was the same as Pilot Study 1 (manual driving without FOST/Haptics) but low lane change practice (i.e., 2 lane changes). For 6 participants, all of whom looked ahead at forward collision hazard onset, there were 5 brake-&-steer, 1 steer only response. This change in method was not successful in bringing back brake-only responses.
3. Pilot Study 3; Manual driving without FOST or haptics, but this method used high lane change practice with construction zone lane drops (no lead vehicle) for the 8 lane changes during training and encountering a slower lead vehicle in the main study (for each of 2 lane changes). Of 7

participants, 6 were looking forward at time of hazard onset. All 6 exhibited brake-&-steer behavior. This was not successful in bringing back brake-only responses.

4. Pilot Study 4; Manual with FOST/Haptics and high lane change practice (same as pilot test 3). Six participants were run, only 5 of which were looking ahead at hazard onset. Of these, there were 4 brake-&-steer and 1 steer only result.

5. Pilot Study 5; Manual driving with FOST/Haptics but with the introduction of event notification as well (i.e., LCA designated in pop-up message during and for 4 seconds after the FOST/haptic display was delivered). Of 10 lane changes, all involved the driver encountering a slower moving lead vehicle. Data collected on 9 participants, 7 of whom were looking up at hazard onset. Results indicated 3 brake-&-steer and 4 steer-only response types. This pattern of results most closely resembled the response patterns observed in the Study 4 but without the detailed explanation provided in training.

6. Pilot Study 6; This study switched to hands-on shared control in autonomous driving. The scenario and HMI were as described in Pilot test 5. Data collected on 10 participants, 6 of whom were looking ahead at hazard onset. Results indicated 5 brake-&and-steer and 1 steer-only response types.

7. Pilot Study 7; Here, the condition was switched to hands-off fully autonomous driving with no FOST/LCA display. However, each participant made 10 lane changes, each after encountering a slower moving lead vehicle. Data was collected on 7 participants, 4 of whom were looking up at hazard onset. Of those participants, all braked-&-steered.

8. Pilot Study 8; This study used hands-off fully autonomous driving without FOST/LCA display and only 2 lane changes during simulator training, each without a lead vehicle ahead. Of 11 participants run, 6 had eyes on road at hazard onset. Results indicated 3brake-only and 2 brake- &-steer response, and 1 steer-only response types. This is the first method to provide brake-only responses like those in Study 1. This condition, therefore, was used as part of the final study to be reported here.

In summary, this study applied a 2x2 experimental design to evaluate the effects of two factors thought to influence driver response to a crash-imminent (i.e., TTC = 1.7 s) FC hazard situation. These factors were the following: Drive Mode (hands-off fully automated versus hands-on shared control automated driving), and Lane Change Adviser State (On versus Off).

8.3 Method

8.3.1 Participants

The sample consisted of sixty-four (64) volunteer Ford Motor Company employees between 19 and 67 years of age (M = 37.22 years, SD = 13.68 years). Age was subdivided into two categories 'younger' and 'older', where 'younger' ranged from 19 to 45 years of age and 'older' ranged from 46 to 67 years of age. A total of 27 females (11 older and 16 younger) and 37 males (7 older and 30 younger) participated in the study. Most of the participants had non-engineering backgrounds (e.g., finance, administration, IT, etc.) and none had prior "surprise event" experience in the Ford VIRTTEX simulator. Additional participant selection criteria included a valid drivers license.

8.3.2 Apparatus

Driving Simulator

See Appendix A for details on the VIRTTEX Driving Simulator. The vehicle dynamics and scenario details for this study were identical to those of Study 3 for the *TTC* = 1.7 s FC hazard condition.

HMI Cluster Display

The base instrument cluster HMI, consisting of speedometer, gear selection, fuel level, etc. was developed by Ford Motor Company as part of another vehicle program. This HMI concept, while not developed as part of this dissertation, defined design constraints on the auditory-visual FOST display that was created for this dissertation. The audio-visual HMI concept of the LCA explained in Chapter 4 was used when the condition was prescribed.

8.3.3 Secondary Tasks

The same set of secondary tasks (adjust climate control; VIRTTEX Numbers Task; iPad Trails task) given in all previous studies (Chapter's 2, 3, and 7) was given here. Drivers in all test conditions were required to complete these secondary tasks. Note that prior to and during the critical event, no secondary task was underway.

8.3.4 Experimental Design

This study was designed as a two factor, one-trial-per-participant, between-subjects design. The two independent factors were Drive Mode and LCA State. The two levels of Drive Mode were: hands- off, fully automated driving versus hands-on shared control automated driving. The two levels of the LCA State were: LCA Off versus LCA On. For both drive modes the LCA, when on, made use of the auditory-visual FOST as described in Ch 4. For the shared control mode the haptic steering display was also part of the LCA HMI. Obviously, the haptic steering display could not be used with hands-off, fully automated driving. Note that this study made use of drivers looking to the road scene as expected for Level 2 automated driving (SAE, 2014).

8.3.5 Response Variables

The dependent variables were driver response time from FC event onset to first response and the nature of response (brake-only, brake-&-steer, steer-only). Driver eye-glance video was also captured but used only to verify whether or not participants were alert and looking toward the road scene at hazard onset.

8.3.6 Simulated Driving Scenario

The simulated daytime drive started at a traffic signal on the edge of a suburban environment with a straight 4-lane undivided road, two lanes in each direction. The drive transitioned to a straight 4-lane rural undivided highway with a posted speed limit of 65 mph (105 kph). The simulated drive was approximately 25 minutes. Road crown was simulated, which caused the host vehicle to drift toward the shoulder after approximately 7 s without steering input by the driver or control by the automated driving system.

There was light traffic density in the oncoming direction with 5-6 vehicles per minute. Light traffic in the adjacent left lane was encountered at a rate of 1-2 vehicles per minute, with those vehicles traveling slower by approximately 15 mph. The slower moving traffic in the left lane prevented the participant from passing this vehicle. A chase vehicle followed the host vehicle at a Time-Headway (TH) of 4 s. A lead vehicle, in the form of a white box truck, traveling between 60 and 70 mph was always present for the LCA off and mostly present for LCA on condition. The reason for speed variation was to expose the participant to longer and shorter THs ranging between 1.5 s and 5.0 s so that the setup for the FC event did not appear unusual.

The simulated driving scenario involved two separate roadways and vehicle interactions depending on experimental conditions (LCA On or LCA Off). LCA On required a lead vehicle driving sufficiently slower than the subject vehicle to activate it. This was accomplished by encountering a white box truck that changed lanes leftward and slowed down to reveal a slower moving passenger vehicle. This occurred approximately two-thirds through the drive. This passenger vehicle was slowing to make a right-hand turn and the LCA activated when the TH was approximately 7 s. The road geometry after the intersection involved a right lane drop intended to guarantee the participant would change lanes to the left. Another slower moving passenger vehicle of a different make and color was encountered shortly thereafter. It was slowing to make a left turn from the left lane and the LCA again activated at a TH of approximately 7 s. The road geometry was such that the right lane opened up again, thus allowing the participant to return to the right lane. Shortly after returning to the right lane the participant eventually encountered another white box truck. This box truck served as the lead vehicle until the FC hazard event. This simulated roadway and vehicle interaction was used for both hands-off, fully automated driving mode and for hands-on shared control driving mode.

For participants with LCA=Off, the simulated roadway was straight throughout. Drivers in this condition were asked to shift to the left lane and then to the right incidentally early in training "to get a feel for driving in VIRTTEX". This was intended to equate the number of lane changes between the two LCA On and Off conditions. In this simulated roadway the white box truck served as the lead vehicle throughout the entire drive up until the FC hazard event.

Both roadways were straight with a construction zone about one-third of the way into the drive that did not require a lane change. This common construction zone was included to afford an opportunity for the driver to experience a manual takeover request from the automation.

Only the $TTC = 1.7\ s$, crash-imminent, hazard condition was tested. The timing of the hazard event was kept identical to that in Study 4. All participants were not distracted at the FC event onset. At the FC hazard event, each participant received the FCW alert. The FCW alert warning sound commenced concurrent with autobraking. Additionally, those with LCA On received FOST only for fully automated driving or FOST plus haptics for the hands-on, shared control, mode.

8.3.7 Procedure

Test participants arrived at the VIRTTEX simulator facility, signed the informed consent form, reviewed a safety video, and were administered a presentation that introduced the automated driving concept. The introductory material explained in general terms how an automated vehicle might operate to manage longitudinal control, lateral control, and car following separation. Included in the training were slides that indicated that the system could not necessarily handle all driving situations. Drivers were told their first responsibility was to safely control the vehicle at all times even when automation was engaged. All participants were instructed to drive in the right-hand lane except when passing.

In this training material the automated driving system was framed as a prototype system that is unable to differentiate between an intentional steering input and an unintentional steering input. For the participants assigned to the hands-on, shared control drive mode, it was explained that for this reason it was required that the participant keep at least one hand lightly on the steering wheel at all times. Participants were told that if they took their hands off for 10 seconds they would receive an audible alert instructing them to “keep hands on wheel”. This was the ruse used to implement a hands-on, shared control, automated driving regime.

The automated driving system was described through a series of screen flows depicting what the participant would see in the instrument cluster as a function of the state of automation. The first illustration described to each participant as what their instrument cluster would look like in manual driving. Once conditions became proper for automation to be engaged, participants would see a pop-up message in the instrument cluster and hear an audible voice prompt telling them so. This was shown in a second illustration. Participants were told that the automation on/off button was a steering wheel toggle button. Once they pressed the toggle button to activate automation, an audible voice prompt gave positive confirmation that automation had been engaged and their instrument cluster would transition to one that depicted to them what automation sees. It was stressed to each participant that if they intentionally or unintentionally turn the steering wheel past a certain point, they would be preemptively given the audible LDW and would see a ’hands on wheel’ depiction show up in the center of their instrument cluster. Participants were told that at this point they need to be actively steering the vehicle but that automation would continue to maintain longitudinal control. Once they re-centered their vehicle in their current lane, or lane of choice, automation would resume lane keeping (i.e., steering control) concurrent with an auditory and visual confirmation via a pop-up stating “lanekeeping resumed”. The final screenflow depicted a takeover request as a message center pop-up with a corresponding audible chime. Participants were told that if at any point they experienced an alert asking them to disengage automation, they could use the brake or the On/Off toggle, at which point they received a positive indication that manual driving was reengaged.

The slides also introduced the secondary tasks that the drivers would need to perform. No mention to participants was made in reference to either the LCA or the discrete haptic steering profile. No mention was made of a forward collision warning or event. After this, the participant was accompanied to the VIRTTEX simulator cab.

After the driver was comfortably seated in the simulator, training began, while stationary, with review of the requested secondary tasks. The ride-along observer, seated in the back passenger seat, then remained silent until the end of the simulator session while the simulator operator continued the driving portion of the training. After static secondary task training, while stopped, participants were asked to put the vehicle into gear, accelerate up to 65 mph, and then drive for a short period of time to become acclimated with driving the simulator.

Shortly thereafter, if the participant was in the LCA = Off condition, they were instructed to make 2 lane changes, one to the left, and once settled in the left lane, a second lane change back to the right lane purportedly for acclimation to the simulator handling characteristics. For the LCA on condition this portion of training was omitted in lieu of lane change practice with LCA later in the drive.

All participants were instructed on how to activate automation. Participants in both the hands-on, fully automated and hands-off, shared control driving modes were provided information on and experienced the ‘maintain control mode’ through a requested lane drift to one side. Then participants were instructed on how to deactivate automation. Additionally, any participant in the hands-on, shared control dive mode was provided information regarding receipt of a hands-off warning if they took their hands off the steering wheel for 10 s. The driving portion of the training continued and involved practicing the secondary tasks while driving.

The VIRTTEX operator in the control room and the ride-along observer in the back seat did not interact with the test participant until the study drive was completed, i.e., after the critical event was over. At that point, the test participant was asked if they recalled receiving a warning of any sort and, if so, could they recall any details about the warning. They also completed a SSQ. Once the simulator was settled, both the participant and observer departed the VIRTTEX dome, and the participant was thanked for his or her participation. The entire experimental session took approximately one hour to complete.

8.4 Results

8.4.1 Response Time Results

One participant failed to respond at all. Therefore, a cell means model analysis of variance was carried out on the response time measures. A simple effect size metric was also calculated for each significant effect. The effect size metric was eta-squared (η^2), defined as the proportion of variance accounted for by a factor. A significant main effect was found for Drive Mode ($F(1,62) = 8.09, p = 0.006, \eta^2 = 0.13$). See Figure 8.1 below. Participants in the hands-on, shared control, automated driving condition responded statistically reliably faster than participants in the hands-off automated driving condition (means of 1.26 and 1.48 s, respectively). The main effect of LCA State was also significant ($F(1,62) = 7.83, p = 0.007, \eta^2 = 0.10$). Participants who received the LCA responded reliably faster than those who did not (means of 1.26 and 1.48 s, respectively). No significant interaction effects were found to exist between Drive Mode and LCA State. This is visually apparent in Figure 8.2 which depicts the distribution of responses by Drive Mode and LCA State.

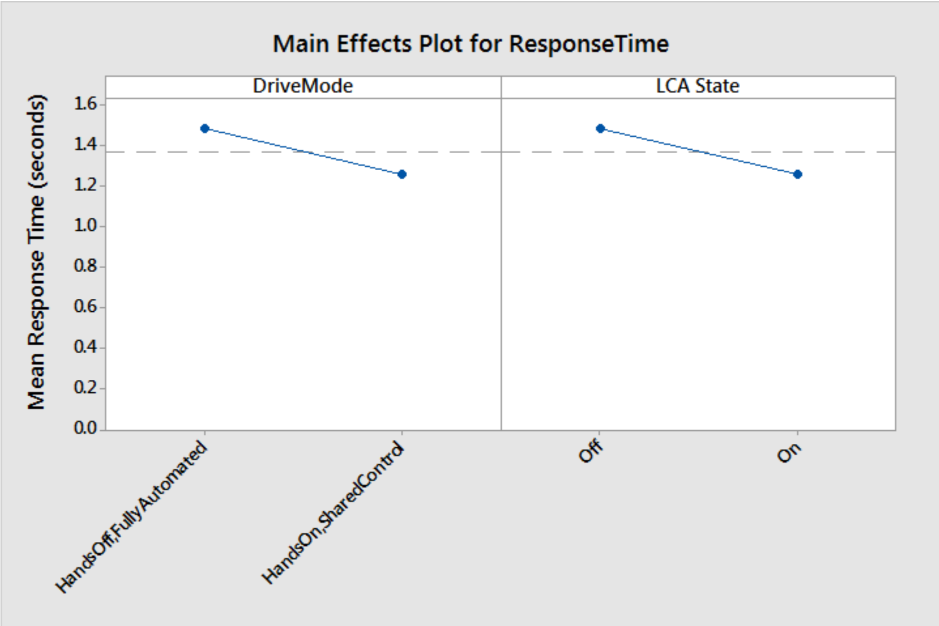

Figure 8.1: Main Effects

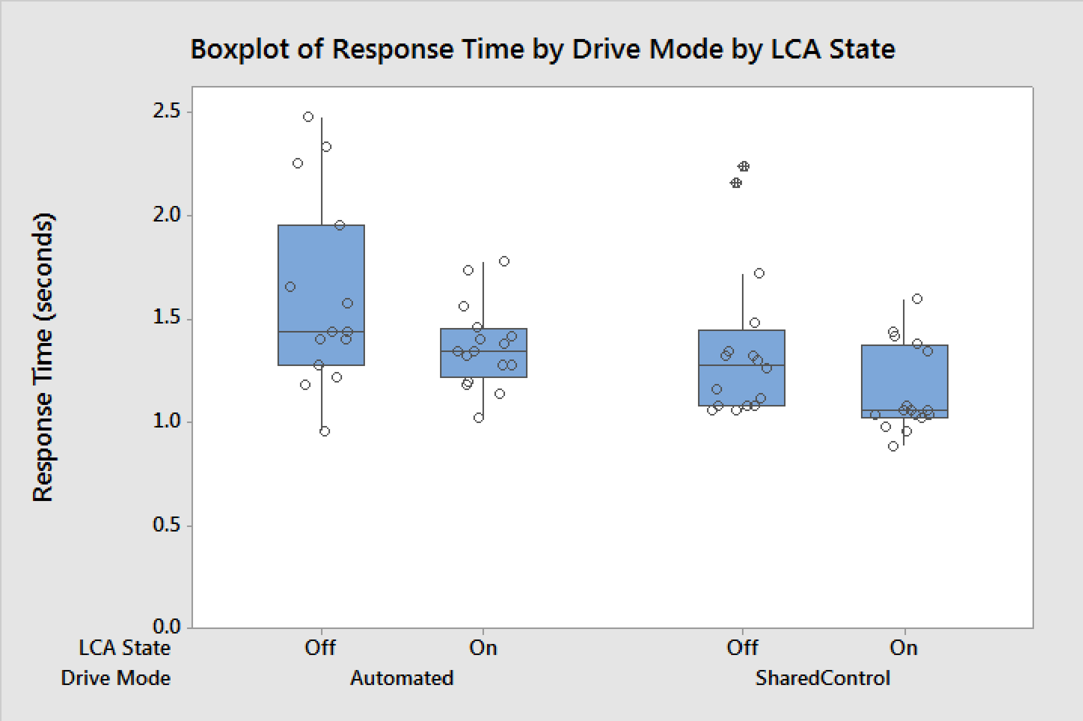

Figure 8.2: Boxplot of Response Times by Drive Mode by LCA State

8.4.2 Response Type Results

A preliminary screening of the response type data revealed one driver who did not respond at all. Thus the response type data analysis was carried out with the remaining 63 participants' responses. Of particular interest was the brake-only behavior and how it might have varied by experimental condition. Therefore, analysis was carried out on response types broken into two categories: 'brake-only' versus 'steer' (i.e., steer-only or brake-&-steer responses). This grouping resulted in Table 8.1 which indicates counts of brake-only behavior divided by total responses per cell. A Q' test (Michael, 2007) was then run that yielded a significant main effect of LCA State On versus Off ($Q' = 4.07, p = 0.044$). The main effect of Drive Mode alone was not significant ($Q' = 0.77, p = 0.380$) nor was there a significant interaction ($Q' = 1.45, p = 0.228$). The results indicate the LCA increased the propensity to steer and reduced brake-only responses across both types of automated driving modes.

Table 8.1: Count of Brake-Only Responses Observed by Condition

	LCA Off	**LCA On**
Hands-Off, Fully Automated Driving	6 / 15	1 / 16
Hands-On, Shared Control Automated Driving	3 / 16	1 / 16

8.5 Discussion

Overall, this study yielded the following results: First, hands-on automated driving was associated with faster response than hands-off automated driving in the face of a sudden need to take over manual driving. Second, the LCA, as a system, did help to achieve faster responding at the FC event onset regardless of automated driving mode. Lastly, the presence of the LCA increased the propensity to steer around the obstacle rather than brake-only. This was true for both automated driving modes. That is, the audio-visual FOST in hands-off automated driving was about as effective as the audio-visual FOST combined with haptics in the shared control automated drive mode. At this point it should be remembered that a haptic steering display is inherently incompatible with hands-off automated driving and could not be provided in this drive mode.

One question is “Did haptics help?”. Figure 8.3 below shows a boxplot of individual response times by LCA for both the hands-off and hands-on automated driving modes. Recall that in hands-off, fully automated driving, participants either received no LCA (only the FCW) at the critical event onset or they received the FCW and LCA with FOST but no haptics concurrently. Participants in the hands-on, shared control automation condition either received no LCA (only the FCW) at the critical event onset or they received FCW and LCA with FOST and haptics. This combination took advantage of the fact that drivers had at least one hand on the wheel at critical event onset. Study 4 indicated no significant differences between these two implementations regardless of whether in manual drive mode or shared control mode and regardless of level of LCA exposure, lane change practice, TTC level, or driver attentional state. In the present study, the Q' analysis of Table 8.1 revealed a significant effect of LCA On or Off, but no interaction and no main effect of drive mode on the proportion of brake-only responses. This provides further evidence that the LCA is effective in increasing the likelihood of steering for both hands-on and hands-off automation even though the latter could not incorporate a haptic steering wheel.

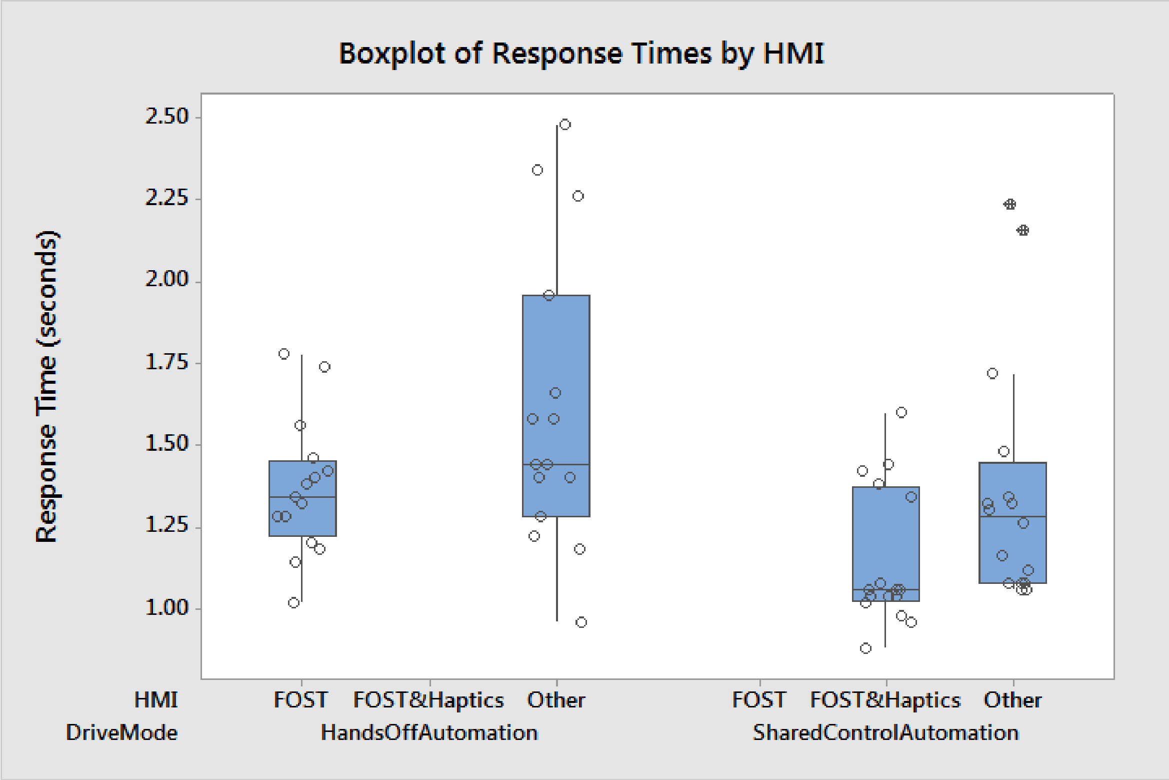

Figure 8.3: Boxplot of Response Times by LCA On or Off

One interesting observation of the FOST and haptics boxplot is the seemingly bimodal response distribution (third boxplot from the left). Based off of the timing of the discrete haptic profile it appears that participants seem to respond or go with the haptics on either the first or second pulse. This was not investigated further and could be the subject of future work. A different question is “Why does LCA help?”. This question will be discussed in the conclusions.

CHAPTER 9 CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH

9.1 Summary and Conclusions

The research problem addressed in this dissertation is how to increase a driver's propensity to steer when advisable rather than brake-only when encountering a forward collision hazard. Such a situation may arise in manual driving or in Level 2 or Level 3 automated driving that requires a manual takeover. Evasive Steer Assist (ESA) can support a successfully executed emergency lane change, but the driver must first steer.

Real world crash data and simulator research indicate that drivers often brake-only when they could or should steer for collision avoidance. Brake-only behavior is more likely under the very conditions that ESA is designed to support. Braking behavior appears to increase as TTC decreases in manual driving. It also appears more likely in manual takeovers of automated driving. Therefore HMI concepts were developed and tested to increase the driver's propensity to change lanes in a forward collision hazard situation.

This research included 5 high-fidelity, moving-base simulator studies. Results of Study 1 showed that 85% of drivers in manual driving exhibited some form of steering behavior for a crash-imminent *TTC* = 1.7 s and participants in hands-off, fully automated driving had a roughly 50—50 response distribution of brake-only versus brake-and-steer responses. Given the reported rear-end crash data and the fact that technology is now leading down the road to automation, this 50—50 response distribution was of interest.

Study 2 was intended to extend Study 1 results over a broader range of TTCs. In addition to the 1.7 s TTC, Study 2 included TTCs of 3.0, 5.0, and 7.0 s presented at the end of a simulated highway drive commute. The results of Study 2 indicated that he response-type results for the 5.0 and 7.0 s TTCs were consistent with prior research by Gold et al. (2013) that drivers were more likely to exhibit increased steering behavior at longer TTCs. Although there was some brake-only behavior at TTC = 3.0 s, all drivers steered around the stopped vehicle in the 1.7 s TTC condition. This last result was not consistent with Study 1 results. This discrepancy was unexpected and may have arisen from a procedural difference that involved more lane-change practice in Study 2, which was included to give test participants experience in the handling characteristics of the simulator. This change in procedure was prompted by pilot testing that indicated people were braking even at longer TTCs. Nonetheless, prior research and crash data indicated that the original Study 1 findings were more consistent with real-life driving. Therefore, the research effort continued on the development of an HMI to facilitate both normal and emergency lane changes which were refined in Study 3. The unexpected consequences of the increased practice led to a methodological investigation in Study 4 that included HMI developed and selected in Study 3.

HMI concepts were developed that might increase a propensity to steer when appropriate. Theories from applied psychology and control engineering were used to design and develop an auditory-visual FOST display and, separately, a discrete proprioceptive haptic steering display. The auditory-visual FOST display might be used in manual driving or hands-off fully automated driving for Level 2 or Level 3 systems. Although the haptic steering display is not applicable to hands-off automated driving, it is appropriate for manual driving or hands-on share-control automated driving to augment the FOST display. A third HMI concept developed for all types of driving is referred to here as an LCA. The LCA recommends lane changes to the driver in normal driving conditions to familiarize him/her with the FOST display and haptics. This takes into account steering as a second-order control task by using a command display in the form of a discrete haptics profile for informing the driver that a lane change is available in crash-probable scenarios and commanding the driver to change lanes in crash-imminent scenarios at highway speed where steering is preferred to braking.

Study 3 was a two-part study conducted to obtain subjective assessments of alternative discrete haptic profiles. This focus was undertaken because of the novelty of the haptic steering display and the fact that it operates on the primary lateral control element (i.e., the steering wheel). In Part 1, pair comparison and scaling methods were used to rank the originally developed profiles on a scale of urgency. The motivation was to pursue the hypothesis that less urgent (i.e., crash-probable) scenarios would favor less urgent profiles and more urgent (i.e., crash-imminent) scenarios would favor more urgent profiles. In Part 2, novel “mirage” methods were applied to evaluate each haptic pattern “in situ”, i.e., at each of the TTCs used in Study 2. Data analysis and the use of MAUT identified a single pattern that was robust for all TTCs tested. Contrary to expectation, multiple patterns at different levels of urgency were not needed. This finding is consistent with psychophysical theory that human preference behavior is often describable by a single-peaked function (Coombs, 1983).

In Study 4, a 2^{6-1} fractional design was undertaken to assess the effects of 6 factors that might influence the driver’s response time or type of response. The 6 factors were drive mode (manual versus hands-on shared control); driver attentional state (alert versus distracted); FOST with or without haptics; high versus low LCA exposure; high versus low lane change practice; and TTCs of 1.7 versus 3.0 s. Response-time results showed significant main effects for drive mode, driver attentional state, and TTC. Response times were faster for manual driving than for shared-control automated driving, when drivers were alert rather than distracted, and in the 1.7 s TTC condition rather than in the 3.0 s TTC condition. These results held regardless of presence or absence of haptics, regardless of level of LCA exposure, and regardless of level of lane change practice. The response time results were consistent with expectations from prior research. ON the other hand, response-type results were surprising in that not a single participant braked-only. All participants did some form of steering. Although these results might suggest that the FOST with or without haptics was highly successful in increased the propensity to steer, this conclusion was not justifiable. For example, only 3 participants even looked at the FOST display during the FC hazard event. In addition, the LCA with FOST and haptic displays were explained in some detail during participant training. This led to a series of pilot tests intended to bring back brake-only behavior that motivated this dissertation from the outset.

Study 5 incorporated the methodology identified in the pilot testing. Specifically, this research was brought full circle in a direct comparison of hands-off fully automated driving and hands-on shared-control automated driving with LCA Off versus On in a 1.7 s TTC FC hazard event and alert drivers as assumed for Level 2 or Level 3 automated driving. Response-time analysis revealed a main effect for drive mode and a main effect for LCA On versus Off. Hands-on shared-control was associated with faster average response times compared to fully automated riving. In addition, the LCA when on was associated with faster responding in both types of automated riving. Since the LCA with hands-off fully automated riving involved the FOST only, this suggest that FOST can be equally effective with or without haptics provided the driver experiences the HMI via the LCA. Response-type results indicate that providing drivers with a recommendation to steer in benign circumstances (LCA On) does increase the propensity to steer and decrease the incidence of brake-only behavior in an emergency.

The lane change adviser was included to provide familiarization with the FOST with or without haptics. This seemed especially important for the haptic display because it was novel, operated on the primary lateral control element, and might be confused for a malfunction if never experienced prior to the critical event. Over the course of the studies, the LCA seems to have had additional effects beyond familiarization. These effects may be referred to as creating a mental set.

Mental set refers to a particular mental network for problem solving (Matsumoto, 2009). The LCA presented the FOST with or without haptics multiple times in the context of car following to suggest lane changes around a slower moving lead vehicle. It appears that this had the effect of expanding the range of options the driver might consider when encountering an obstacle, i.e. to steer rather than to just brake.

Mental set is somewhat distinct from perceptual set which refers to a set of schemas that guide top-down processing of sensory information. For example, Hole (2014) describes perceptual set as an expectation used to decide whether or not other vehicles are present in ambiguous situations. A perceptual set is an expectancy to see things in a particular way. In contrast, a mental set is a schema for thinking about a problem.

Mental set has traditionally been studied as a hindrance to problem solving (Sternberg and Stern- berg, 2016). Mental set can hinder problem solving if the approach is not optimal for a particular instance. A classic example is the Luchins water jar problem from Gestalt psychology. This problem involves measuring out a certain amount of water using 3 different sized jars. A strategy can be developed when solving an initial series of problems that are all isomorphic, i.e., are all amenable to the same approach. However, this series of ‘successes’ can induce a mental set that makes it harder to solve later problems that are not amenable to the same approach but can be solved easily with a different approach. Mental set can make it easier to solve a problem as well. In the case of forward collision avoidance, braking is the most commonly observed solution. In some sense, braking is an example of mental set that hinders problem solving in some instances of rear-end collision avoidance that require steering. What the LCA appears to have done, beyond HMI familiarization, is increase the response repertoire a driver has when encountering an unexpected hazard.

The evidence for this interpretation of the efficacy of the LCA comes from the study outcomes. Results from Study 4 showed that only 3 participants were observed to even glance to the FOST display at FC event onset yet all drivers exhibited some form of steering. In Study 4 all drivers received the LCA with or without haptics, half with high exposure and half with low exposure to it. This study indicated that the steering behavior did not depend on differential levels of lane change practice or amount of LCA exposure. The one common element for all drivers in Study 4 was at least some LCA activation in normal driving as well as at FC hazard onset. In Study 5, recall that the amount of lane change practice was held constant and set at the low level defined in Study 4. This suggests that the differential effect of LCA on or off in Study 5 did not depend on different amounts of lane change practice. Study 5 also eliminated the orientation to the FOST given in

Study 4 driver training and replaced it with event notification in the message center of the instrument cluster. There was opportunity to perceive the FOST with or without haptics in the context of normal lane changes via the LCA. The LCA inherently depends upon a lead vehicle for activation; in this sense, it provides context for steering around an obstacle albeit in benign driving conditions. The difference between Study 2 and Study 4 was the explicit manipulation in the amount of lane change practice with and without a lead vehicle present. The fact that drivers generally steered in the 1.7 s TTC condition also points to a strong mental set that a driver can steer rather than brake-only to avoid an obstacle.

In total, the findings of this dissertation support the following conclusions. First and foremost, the LCA and FOST with or without haptics did successfully change drivers' propensity to steer in high- intensity FC situations. This HMI system may be beneficial for manual driving and Level 2 and Level 3 automated driving. The haptic steering display for hands-on shared control or manual driving was not significantly better in promoting steering than the auditory-visual FOST alone. This appears to arise from a very strong mental set effected by the LCA exposure in normal driving. Beyond this, there appears to be a response time advantage for hands-on shared control as compared to hands-off fully automated driving that merits further research. In addition, research continues on the evolution of the LCA and algorithms to predict when a driver might wish to change lanes.

There are several caveats to the conclusions in this dissertation that should be kept in mind. First, each experiment was conducted in a simulator study and results should be treated as such. Simulator studies, like experiments in general, are of limited duration, involve virtually no exposure to risk, are highly scripted for the intended purpose of the research, and may have demand characteristics that participants experience whether intended or not. Second, although the results show that shared control might be better than hands-off fully automated driving, they cannot and do not answer the question "How fast is fast enough?" from a real-world safety standpoint. In addition, the findings should not be treated as feature-ready but as design guidelines for future Advanced Driver ASsistance (ADAS) research. Finally, it is acknowledged that at least one person did not respond to the FC hazard event in Study 5. This serves as a reminder that an HMI, no matter how robust or well designed, cannot guarantee a successful outcome every time for everyone. The efficacy of any safety system must be proven in real-world driving, over the course of years and millions of miles driven.

9.2 Future Work

An eloquent description of the influence of writing a dissertation and its effect on continuing research is the line from Erlien (2015), "The impact and contribution of a research project is demonstrated by the new and exciting research directions that it inspires." The research described herein is no exception. Some directions for future research are given below.

9.2.1 Lane Change Adviser

The idea of a Lane Change Adviser feature stemmed from the need to create a mechanism for familiarization to the Field of Safe Travel display and the haptic steering profiles. As mentioned in the dissertation, the hypothesis was that increased exposure would lead to familiarity and thus increase the probability of a steering response. Moving forward, the logic required to build an effective LCA must be developed for real-world vehicle applications in a manual, automated, or shared-control driving.

9.2.2 Further HMI Development

The cluster implementation for the FOST display was driven by design directions from another vehicle program. Future research might examine other alternatives such as presenting the FOST display in either a traditional HUD or an augmented-reality head-up display. The AR HUD offers other opportunities for dynamic display, for example a fast-time predictor of where the vehicle will be a short time increment ahead during an evasive maneuver.

9.2.3 Automated Driving and Shared Control

As semi-automated features become more capable, the notion of man-machine systems and sharing of control may become more commonplace. The instances in which shared control might be of most benefit are the topics of continued research. This dissertation provides support for an indication to steer that is distinguishable, directional, and consistent throughout either manual driving or a version of automated driving, i.e., hands-on shared-control. What remains to be discovered is how attractive such a form of automated driving might be and what failure modes (e.g., driver complacency) might be associated with it.

APPENDIX A: VIRTTEX DRIVING SIMULATOR

Each study was conducted in Ford's VIRtual Test Track EXperiment (VIRTTEX) simulator facility. VIRTTEX is a motion-based driving simulator, hydraulically powered in 6 degrees of freedom with bandwidth in excess of 13 Hz in all degrees of freedom. The performance specifications are detailed in Table A.1, below.

Figure A.1: Ford's VIRtual Test Track EXperiment (VIRTTEX) simulator

VIRTTEX is designed to accommodate a full-size vehicle cab, with a model year 2007 Ford Edge used for this study. Tactile, visual, and sound cues are provided to the driver in order to fully immerse drivers into the driving task. Realistic road, wind, and engine sounds are played over a sound system. The vehicle cab includes a steering control loader for accurate feedback of road and tire forces to the driver. The visual system in VIRTTEX is a non-collimated front-projection display system. The display surface is a spherical section with a radius of 3.7 meters. Seven Barco LCoS projectors (Sim7s) are used to form the driving scene on the display surface. Five projectors are used for the forward field-of-view covering 240°*x*39°. Two projectors cover the rear 120°*x*29°, adjusted for appropriate field-of-views in the vehicle cab side mirrors. An additional 45-inch flat-panel LCD monitor with 16:9 aspect ratio is in the rear of the vehicle cab to provide the appropriate field-of-view in the rear-view mirror. A Blue Newt Software image generator (Pixel Transit) running at a fixed 60-Hz rate drives each visual channel. The channels for the Sim7 projectors have a resolution of 2048*x*1536 pixels, and the channel for the LCD monitor has a resolution of 1280*x*1024.

Motion Drive Algorithms (MDAs) are used to transform motion from the host vehicle simulation into the motion capabilities of VIRTTEX (Grant et al., 2003). In general, an attempt is made to reproduce the maximum amount of motion from the vehicle simulation, while staying within the motion capabilities of the simulator. The lateral scaling was changed as a function of the driving condition. For everything except the Forward Collision (FC) event, the lateral motions from the vehicle simulation were scaled by 0.55. The scaling was further reduced to 0.20 for the forward collision event to account for the possibility of the participant steering significantly to the right or left. The 0.55 scaling along with the steering simulation provides realistic on-center steering feel and lateral vehicle control in VIRTTEX (Greenberg et al., 2003).

The longitudinal MDA scaling was changed as a function of vehicle control and driving condition. Except for two conditions, longitudinal accelerations from the vehicle simulation were scaled by 0.60 in order to maximize the accelerations felt by the driver while keeping VIRTTEX within its motion capabilities. The scaling was reduced to 0.20 if the driver applied the brakes during normal driving, and further reduced to 0.10 during the forward collision event.

Table A.1: VIRTTEX Motion Performance Specifications

DOF	**Acceleration**	**Velocity**	**Displacement**
Longitudinal / Lateral	> 0.6 G	> 1.2*m/s*	±1.6*m*
Vertical	1.0 G	1.0*m/s*	±1.0*m*
Pitch / Roll	> 200°/*s*^{2}	> 20°/*s*	±20°
Yaw	> 200°/*s*^{2}	> 20°/*s*	±40°

APPENDIX B: STUDY 4: 2^{6-1} EXPERIMENTAL DESIGN DETAILS

Design plots such as the one in Figure B.1 below show graphically the many characteristics of the design including the factor space the design provides. While not all combinations will be run, as will be explained, this plot shows the magnitude of the design being undertaken. The specific treatment combinations tested are indicated in Figure B.1

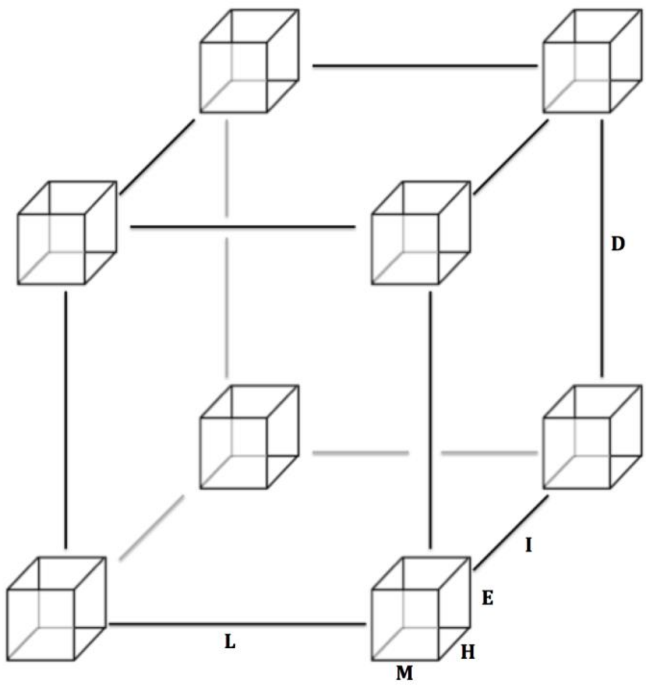

Figure B.1: Resolution VI Design

Treatment Code	Drive Mode (M)	HMI (H)	HMI Exposure (E)	Lane Change Practice (P)	TTC (T)	Distraction (D)
17	-1	-1	-1	-1	-1	-1
2	-1	-1	-1	1	-1	1
3	-1	-1	1	-1	-1	1
20	-1	-1	1	1	-1	-1
5	-1	1	-1	-1	-1	1
22	-1	1	-1	1	-1	-1
23	-1	1	1	-1	-1	-1
8	-1	1	1	1	-1	1
9	1	-1	-1	-1	-1	1
26	1	-1	-1	1	-1	-1
27	1	-1	1	-1	-1	-1
12	1	-1	1	1	-1	1
29	1	1	-1	-1	-1	-1
14	1	1	-1	1	-1	1
15	1	1	1	-1	-1	1
32	1	1	1	1	-1	-1
1	-1	-1	-1	-1	1	1
18	-1	-1	-1	1	1	-1
19	-1	-1	1	-1	1	-1
4	-1	-1	1	1	1	1
21	-1	1	-1	-1	1	-1
6	-1	1	-1	1	1	1
7	-1	1	1	-1	1	1
24	-1	1	1	1	1	-1
25	1	-1	-1	-1	1	-1
10	1	-1	-1	1	1	1
11	1	-1	1	-1	1	1
28	1	-1	1	1	1	-1
13	1	1	-1	-1	1	1
30	1	1	-1	1	1	-1
31	1	1	1	-1	1	-1
16	1	1	1	1	1	1

Figure B.2: Resolution VI Treatment Combinations

Power Analysis

In order to design for detection of a large effect, a proper power analysis must be conducted. For a 2^{6-1} Resolution VI experimental design, all 6 main effects and 15 2-way interactions are assessed with $1df$ in the numerator. Assuming $N = 64$ with 2 replicates per each of 16 treatment combinations or cells and 3-way and high-order interactions are assumed negligible, the common error $df's$ are $64 - (6 + 15) - 1 = 42$. Cohen (1988) provides an adjusted-n formula for use in determining statistical power by means of tables. Specifically, the adjusted $n(n')$ for a fixed-effects multi-factor design is given as:

$$n' = \frac{df_{denom}}{df_{numerator} + 1} + 1$$

where… df_{denom} = the degrees of freedom for the error term or denominator of the F-test or denominator $df = 42$; $df_{numerator}$ = the degrees for the numerator of the F-test or numerator $df = 1$ for each effect; and n' is the adjusted sample size used to enter into the power tables provided in Cohen (1988) or $n' = 22$ for this experimental design.

If we assume that $\alpha = 0.05$; $df_{numerator} = 1; n = 22$; and a 'large' effect size of $f = \left(\frac{1}{2}\right) d = 0.4$, then the estimated power is given as $Power = 0.75$ (per Table 8.3.12 in Cohen (1988), p. 311).

BIBLIOGRAPHY

Adams, L. D. (1994). Review of the literature on obstacle avoidance maneuvers: August 1994.

Adams, L. D., Flannagan, M. J., and Sivak, M. (1995). Obstacle avoidance maneuvers in an automobile simulator. Technical report.

Allen, R. W. (1995). The driver's role in collision avoidance systems. In *Workshop on Collision Avoidance Systems*.

Beckman, B. (1998). *Physics of Racing Series*. Stuttgart West.

Beruscha, F., Augsburg, K., and Manstetten, D. (2011). Haptic warning signals at the steering wheel: A literature survey regarding lane departure warning systems (short paper).

Bhise, V. D. (2011). *Ergonomics in the automotive design process*. CRC Press.

Blommer, M., Curry, R., Swaminathan, R., Tijerina, L., Talamonti, W., and Kochhar, D. (2017). Driver brake vs. steer response to sudden forward collision scenario in manual and automated driving modes. *Transportation research part F: traffic psychology and behaviour*, 45:93–101.

Box, G. E., Hunter, J. S., and Hunter, W. G. (2005). *Statistics for experimenters: design, innovation, and discovery*, volume 2. Wiley-Interscience New York.

Bradley, R. A. and Terry, M. E. (1952). Rank analysis of incomplete block designs: I. the method of paired comparisons. *Biometrika*, 39(3/4):324–345.

Cacciabue, P. C. (2013). *Guide to applying human factors methods: Human error and accident management in safety-critical systems*. Springer Science & Business Media.

Campbell, J. L., Richard, C. M., Brown, J. L., and McCallum, M. (2007). Crash warning system interfaces: human factors insights and lessons learned. *NHTSA report# HS*, 810:697.

Chambers, J. M. (1983). *Graphical methods for data analysis*.

Chelst, K. and Canbolat, Y. B. (2011). *Value-added decision making for managers*. CRC Press.

Chen, K., Chan, A., and Ma, Q. (2014). Cell phone feature preferences among older adults: A paired comparison study. *Gerontechnology: international journal on the fundamental aspects of technology to serve the ageing society*.

Cicchino, J. B. (2016). Effectiveness of forward collision warning systems with and without autonomous emergency braking in reducing police-reported crash rates. *Arlington, VA: Insurance Institute for Highway Safety*.

Cliff, N. (1996). *Ordinal methods for behavioral data analysis*. Psychology Press.

Cliff, N. and Keats, J. A. (2003). *Ordinal measurement in the behavioral sciences*. Psychology Press. Cohen, J. (1988). Statistical power analysis for the social sciences.

Coombs, C. H. (1983). *Psychology and mathematics: An essay on theory*. University of Michigan Press.

Dang, T., Desens, J., Franke, U., Gavrila, D., Schäfers, L., and Ziegler, W. (2012). Steering and evasion assist. In *Handbook of intelligent vehicles*, pages 759–782. Springer.

David, H. A. (1963). *The method of paired comparisons*, volume 12. DTIC Document.

Duarte, M. E. C. and Rebelo, F. (2007). Virtual reality in the study of warnings effectiveness. In *International Conference on Ergonomics and Health Aspects of Work with Computers*, pages 189–198. Springer.

Dunn-Rankin, P., Knezek, G. A., Wallace, S. R., and Zhang, S. (2004). *Scaling methods*. Psychology Press.

Edworthy, J. and Adams, A. (1996). Warning design: an integrative approach to warnings research.

Engstrom, J. and Aust, M. L. (2011). Adaptive behavior in the simulator: Implications for active safety system evaluation.

Erlien, S. M. (2015). *Shared vehicle control using safe driving envelopes for obstacle avoidance and stability*. PhD thesis, Stanford University.

Fairbanks, R. J., Fahey, S. E., and Wierwille, W. W. (1995). Research on vehicle-based driver status/performance monitoring: Seventh semi-annual research report. Technical report.

Gibson, J. J. (1960). The concept of the stimulus in psychology. *American Psychologist*, 15(11):694. Gibson, J. J. (1966). The senses considered as perceptual systems.

Gibson, J. J. and Crooks, L. E. (1938). A theoretical field-analysis of automobile-driving. *The American journal of psychology*, 51(3):453–471.

Gold, C., Damböck, D., Lorenz, L., and Bengler, K. (2013). take over! how long does it take to get the driver back into the loop? In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 57, pages 1938–1942. SAGE Publications Sage CA: Los Angeles, CA.

Grant, P., Blommer, M., Cathey, L., Artz, B., and Greenberg, J. (2003). Analyzing classes of motion drive algorithms based on paired comparison techniques. *DSC North America, Dearborn (Michigan)*.

Grant, P. R., Blommer, M., Artz, B., and Greenberg, J. (2009). Analysing classes of motion drive algorithms based on paired comparison techniques. *Vehicle system dynamics*, 47(9):1075–1093.

Greenberg, J., Artz, B., and Cathey, L. (2003). The effect of lateral motion cues during simulated driving.

Proceedings of DSC North America, pages 8–10.

Greenwald, A. G. (1972). On doing two things at once: time sharing as a function of ideomotor compatibility. *Journal of experimental psychology*, 94(1):52.

Guilford, J. P. (1954). Psychometric methods.

Hart, S. G. and Staveland, L. E. (1988). Development of nasa-tlx (task load index): Results of empirical and theoretical research. *Advances in psychology*, 52:139–183.

Hoffmann, E. R. and Chan, A. H. (2011). Alternative approaches to the design of four-burner stoves. *Ergonomics*, 54(9):777–791.

Hole, G. J. (2014). *The psychology of driving*. Psychology Press.

Iida, Y. (2009). The number of circular triads in a pairwise comparison matrix and a consistency test in the ahp (¡ special issue¿ operations research for performance evaluation). *Journal of the Operations Research Society of Japan*, 52(2):174–185.

Kelley, C. R. (1968). *Manual and automatic control*. J. Wiley & Sons.

Kendall, M. (1975). Rank correlation methods (4th edn.) charles griffin. *San Francisco, CA*, page 8. Kendall, M. G. and Smith, B. B. (1940). On the method of paired comparisons. *Biometrika*, 31(3/4):324–345.

Kennedy, R. S., Lane, N. E., Berbaum, K. S., and Lilienthal, M. G. (1993). Simulator sickness question- naire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology*, 3(3):203–220.

Knezek, G. A. (1978). *Circular triad distributions with applications to complete paired comparisons data*. PhD thesis.

Kusano, K. D. and Gabler, H. C. (2011). On-scene determination of driver crash causation and avoidance maneuvers in rear-end collisions. In *Proceedings of the 3rd International Conference on Road Safety and Simulation*.

Lechner, D. and Malaterre, G. (1991). Emergency manuever experimentation using a driving simulator. Technical report, SAE Technical Paper.

Lerner, N. (1996). Preliminary human factors guidelines for crash avoidance warning devices. Technical report.

Lerner, N., Jenness, J., Robinson, E., Brown, T., Baldwin, C., and Llaneras, R. (2011). Crash warning interface metrics final report. Technical report.

Lewis, J. R. (1989). Pairs of latin squares to counterbalance sequential effects and pairing of conditions and stimuli. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 33, pages 1223–1227. SAGE Publications.

Lorenz, L., Kerschbaum, P., and Schumann, J. (2014). Designing take over scenarios for automated driving how does augmented reality support the driver to get back into the loop? In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 58, pages 1681–1685. SAGE Publications.

Matsumoto, D. (2009). *The Cambridge dictionary of psychology*. Cambridge University Press.

Maxwell, A. (1974). The logistic transformation in the analysis of paired-comparison data. *British Journal of Mathematical and Statistical Psychology*, 27(1):62–71.

Mazzae, E., Barickman, F., Forkenbrock, G., and Baldwin, G. (2003). Nhtsa light vehicle antilock brake systems research program task 5.2/5.3: Test track examination of drivers collision avoidance behavior using conventional and antilock brakes. *Washington, DC: National Highway Transportation Safety Administration*.a

Michael, G. A. (2007). A significance test of interaction in 2xk designs with proportions. *Tutorials in Quantitative Methods for Psychology*, 3(1):1–7.

Michell, J. (1999). *Measurement in psychology: A critical history of a methodological concept*, volume 53. Cambridge University Press.

Newman, J. R. (1956). *The world of mathematics*, volume 1.

Niebel, B. W. (1976). Motion and time study. homewood, Illinois: Richard d. Irwin. *Inc. l976*.

Pepping, G.-J. and Grealy, M. A. (2012). *Closing the gap: The scientific writings of David N. Lee*. Psychology Press.

Petermeijer, S. M., Abbink, D. A., and de Winter, J. C. (2015). Should drivers be operating within an automation-free bandwidth? evaluating haptic steering support systems with different levels of authority. *Human factors*, 57(1):5–20.

Poulton, E. (1981). Human manual control. *Comprehensive Physiology*.

Rice, R. S. and Dell'Amico, F. (1974). An experimental study of automobile driver characteristics and capabilities. Technical report, DTIC Document.

Ross, R. T. (1934). Optimum orders for the presentation of pairs in the method of paired comparisons. *Journal of Educational Psychology*, 25(5):375.

Rumar, K. (1990). The basic driver error: late detection. *Ergonomics*, 33(10-11):1281–1290. Schinka, J. A., Velicer, W. F., and Weiner, I. B. (2003). *Research methods in psychology*. Wiley.

SAE On-Road Automated Vehicle Standards Committee. (2014). Taxonomy and definitions for terms related to on-road motor vehicle automated driving systems.

Schmidt, R. A. and Wrisberg, C. A. (2008). *Motor learning and performance: A situation-based learning approach*. Human Kinetics.

Singer, J., Lerner, N., Kellman, D., and Robinson, E. (2015). Crash warning interface metrics: Warning and message perception under ambient noise conditions laboratory experiments: Report no. dot hs 812 207. *Washington. Zugriff am*, 10:2015.

Smiley, A. (2015). *Human factors in traffic safety*.

So, Y. (1995). A tutorial on logistic regression. *Cary, NC: SAS White Papers*. Sternberg, R. J. and Sternberg, K. (2016). *Cognitive psychology*. Nelson Education.

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological review*, 34(4):273.

Tijerina, L., Blommer, M., Curry, R., Swaminathan, R., Kochhar, D., and Talamonti, W. (2015). Repeated measures testing of driver collision warning. Technical report, SAE Technical Paper.

Tijerina, L., Johnston, S., Parmer, E., Pham, H. A., and Winterbottom, M. (2000). Preliminary studies in haptic displays for rear-end collision avoidance system and adaptive cruise control system applications. Technical report.

Tillemans, S. (2011). *Development of an instrument for data collection in a multidimensional scaling study of personal Web usage in the South African workplace*. PhD thesis, Stellenbosch: Stellenbosch University.

Tsukida, K. and Gupta, M. R. (2011). How to analyze paired comparison data. Technical report, DTIC Document.

Ulusoy, M. and Sipahi, R. (2014). Towards the design of a human-machine interface via disturbance adaptive control: An analogous machine-to-machine system. In *Haptics Symposium (HAPTICS),* *2014 IEEE*, pages 431–435. IEEE.

Victor, T., Dozza, M., Bärgman, J., Boda, C.-N., Engström, J., and Markkula, G. (2014). Analysis of naturalistic driving study data: Safer glances, driver inattention, and crash risk.

Wickens, C., Lee, J., Liu, Y., and Gordon-Becker, S. (1997). Automation. *Flight to the Future: Human Factors in Air Traffic Control. National Academy Press, Washington, DC*.

Wogalter, M. S., Conzola, V. C., and Smith-Jackson, T. L. (2002). Research-based guidelines for warning design and evaluation. *Applied ergonomics*, 33(3):219–230.

Zwolinski, P. and Sagot, J.-C. (1998). A simulation approach to the design and evaluation of auditory interfaces in a highspeed train driving cab.

ABSTRACT

HUMAN-MACHINE INTERFACE DEVELOPMENT FOR MODIFYING DRIVER LANE CHANGE BEHAVIOR IN MANUAL, AUTOMATED, AND SHARED CONTROL AUTOMATED DRIVING

by

WALTER JOSEPH TALAMONTI

May 2017

Advisor: Dr. R. Darin Ellis

Major: Industrial Engineering

Degree: Doctor of Philosophy

Rear-end crashes are common on U.S. roads. Driver assistance and automated driving technologies can reduce rear-end crashes (among other crash types as well). Braking is assumed for Forward Collision Warning (FCW) and Automatic Emergency Braking (AEB) systems. Braking is also used for Adaptive Cruise Control (ACC) and in automated driving systems more generally. However, steering may be advised in an emergency if the adjacent lane is clear and braking is unlikely to avoid a collision. Steering around an obstacle when feasible also eliminates the risk of becoming the new forward collision hazard. Driver assist technology like Evasive Steer Assist (ESA) and Level 2 or Level 3 automated driving systems might facilitate emergency lane changes but may require the driver to manually initiate the maneuver, something which drivers are often reluctant to do.

A Human-Machine Interface (HMI) might advise the driver of a steerable path when feasible in forward collision hazard situations. Such an HMI might also advise a driver of normal lane change opportunities that can reduce travel time, increase fuel efficiency, or simply enhance the driving experience by promoting ‘flow.’ This dissertation investigated the propensity of drivers to brake-only versus steer in both manual and automated driving situations that end in a high-intensity forward collision hazard. An audio-visual Field of Safe Travel (FOST) instrument cluster display and a haptic steering wheel HMI were developed to advise drivers in both discretionary and emergency situations of a lane change opportunity. The HMI was tested using a moving base simulator in manual driving, in fully autonomous driving, and in shared-control autonomous driving during a simulated highway commute that ended in a high-intensity forward collision hazard situation. Results indicated that a) driver response was affected by the nature of the automated

driving (faster response in hands-on shared control versus hands-off fully autonomous driving); b) exposure to the HMI in normal lane changes both familiarized the driver with the HMI and introduced a mental set that steering was also a possibility rather than braking only; c) and that drivers used their direct vision to determine their response in the emergency event. A methodological issue related to mental set was also uncovered and resolved through screening studies. The final study brought the dissertation full-circle, comparing hands-off fully automated driving to hands-on shared control automated driving in the context of either providing some or no exposure to the developed Lane Change Adviser (LCA) system concept. Results of the final study indicated that shared control lies somewhere between that of manual driving and hands-off fully automated driving. Benefits were also shown to exist for the LCA system concept irrespective of whether the discrete haptic profiles are included or not. The discrete haptic profiles did not statistically reliably increase response times to the forward collision hazard event, although they do show a trend toward decreasing response variability. This finding solidified the fact that by implementing a system for benign driving that aids in establishing a mental set to steer around an obstacle may actually be beneficial for rear-end crash scenarios.

This dissertations contributions include a) audio-visual Field of Safe Travel (FOST) display concepts; b) discrete haptic steering display concepts; c) a paired-comparisons scaling of urgency for haptic displays applied while driving; d) a new “mirage scenario” methodology for eliciting subjective assessments in the context of a forward collision hazard, briefly presented then removed, without risk of simulator sickness, and e) a methodological lesson for others who wish to investigate semi-automated and automated driving interventions- that they must manage driver mental set carefully.

AUTOBIOGRAPHICAL STATEMENT

Walter Joseph Talamonti received his B.S. degree in Electrical Engineering from the University of Michigan—Dearborn and his M.S. degree in Biomedical Engineering, sub-specializing in Neuro-functional Magnetic Resonance Imaging (fMRI), from Wayne State University in 2009. Walter is currently a Ph.D. candidate in Industrial & Systems Engineering at Wayne State University. His research interests include Human Factors, Human Performance Modeling, Artificial Intelligence, Simulation, and Controls.

During his study at Wayne State University, Walter made a number of technical presentations at SAE and HFES. Papers he co-authored have been published in Transport Psychology Part F, and papers he authored have been accepted for publication in the journal of Applied Ergonomics.

Walter is also a full-time employee at Ford Motor Company where he holds the position of Research Engineer. There he conducts human factors research in areas of Human-Machine Interface (HMI), driver warning systems, and autonomous driving. He has authored and co-authored a number of U.S. Patents and patent applications and is a member of HFES, SAE, and IEEE.