CRASH WARNING INTERFACE METRICS: EVALUATING DRIVER-VEHICLE INTERFACE CHARACTERISTICS FOR ADVANCED CRASH WARNING SYSTEMS

Emanuel Robinson
Neil Lerner
James Jenness
Westat
United States
Eric Traube
National Highway Traffic Safety Administration
United States
Timothy Brown
University of Iowa
United States
Carryl Baldwin
George Mason University
United States
Robert Llaneras
Virginia Tech Transportation Institute
United States
Paper Number 11-0252

ABSTRACT

The Crash Warning Interface Metrics (CWIM) project addressed issues of the driver-vehicle interface (DVI) for Advanced Crash Warning Systems (ACWS). The focus was on identifying the effects of certain warning system features (e.g., warning modality) and on establishing common methods and metrics that may be generally applied for evaluating DVIs in different vehicles. The project did *not* have the goal of proposing standard interfaces for particular warning functions, but it did consider implications for design. The project included analytical activities and five experiments. Each experiment investigated the effects of ACWS DVI on driver behavior or comprehension using a different methodology. An objective of these studies was to determine the appropriateness of the various methodologies for use in subsequent human factors research on ACWS DVIs. Implications were discussed for methods to evaluate DVIs including driving scenarios, research participant characteristics, pre-familiarization with the warning system, the distraction task, the participant's task and associated expectancies, accommodating user settings and options, the use of comparison benchmarks, and issues in the treatment of data. Key research needs were identified for carrying the work of this project forward, including research related to ACWS modality, ACWS design, and CWIM assessment methods.

INTRODUCTION

Background

ACWS use sensors to assess potential or emerging hazard situations and provide warning information to drivers. Example systems include forward collision warning (FCW) and lane departure warning (LDW). ACWS are increasingly common in passenger vehicles and the characteristics of these systems vary considerably among vehicle manufacturers. In particular, the means by which the warning information is conveyed to the driver – the DVI – varies in many respects, including the warning sensory modality, display location, information content, coding, sensory attributes (e.g., intensity), temporal aspects, and active intervention in vehicle control. Given the potential diversity of DVIs for analogous ACWS functions in different vehicles, the question arises as to whether all of the alternative implementations are reasonably effective and also whether this diversity may cause safety problems.

The National Highway Traffic Safety Administration (NHTSA) initiated a program to provide for systematic evaluation of the DVI for ACWS functions. This paper gives an overview of Phase 2 of the CWIM project, conducted as part of this programmatic effort. Three broad issues related to the ACWS DVI were encompassed within the project: (1) a consideration of alternative display modalities for conveying the warning (specifically for FCW and LDW applications); (2) methods for a

common evaluation procedure for assessing a DVI; and (3) implications for DVI conventions.

Warning mode A variety of alternative display modes may be used to present ACWS messages to the driver. Auditory alerts and visual displays (lights, icons, text) are the most common modes. However, haptic signals of various sorts have received recent attention and are now found in some vehicles. Haptic signals include seat vibration, steering wheel vibration, brake pulse, seatbelt pre-tensioning, and accelerator counterforce. Acoustic icons and voice messages are alternative acoustic displays that may be used in place of beeps and other sounds that do not have inherent meaning. In addition to display modes, some ACWS include an active component. An "active" warning includes automatic partial control of a vehicle's behavior (e.g., direction, speed) through steering/braking. This automatic action may itself serve as a warning cue, and may promote driver responses that aid in crash avoidance, in addition to any direct safety effects from the vehicle response itself. Examples of active systems include an FCW system that applies crash-imminent braking and an LDW that applies a steering correction when the vehicle is drifting out of lane. Knowledge about driver response to such active interventions is limited. Furthermore, current commercial examples are typically moderate in terms of vehicle control aggressiveness, and they appear intended as aids to driver actions rather than autonomous collision avoidance. Active warnings are of particular interest both because of their potential to promote improved driver response and because of the possibility that they may induce inappropriate driver reactions or poor consumer acceptance (e.g., if a system is incompatible with users' expectations). Examples of inappropriate driver reactions include overcorrection in steering, strong lateral acceleration, severe braking, startle responses, and driver confusion. In order to devise evaluation methods and design guidance that remain appropriate as active systems evolve and become more common, it is important to understand how drivers respond to these types of ACWS. Therefore, the CWIM project included empirical evaluation of passive and active warning modes.

Methods for DVI evaluation While it is important to have effective DVIs for ACWS functions, a consensus means of *evaluating* a given system does not exist. The field lacks a valid, practical, consensus method for determining the efficacy of a DVI for a particular ACWS application. A set of specific research methods, dependent measures, and analysis methods could provide valid, reliable, and repeatable assessments. Such a consensus set of methods is what is meant by CWIM.

The metrics considered in this project are directed at the evaluation of operational (commercial or prototype) ACWS, rather than as techniques to be used in earlier design stages. The metrics might be applied in various ways, such as evaluating the performance of the ACWS DVI (quantitatively and/or against established criteria), comparing the performance of alternative systems, providing a basis for consumer information (e.g., the type of information useful for the New Car Assessment Program), or supporting regulatory or safety actions.

This project developed initial suggestions for a range of CWIM factors. The process of developing and establishing consensus for CWIM is complex. Various manufacturers use different modalities and display types, so a common metric must be able to encompass any type of interface. Since a particular ACWS may be integrated as part of a system of warnings, the method must have a reasonable means of testing a given function in isolation, without penalizing the system by removing important context. Not all nominally similar safety functions operate in the same manner; for example, some warnings may only operate within certain speed thresholds. Some vehicles may provide advance information or alerts, prior to the situation in which the actual crash warning occurs; the means of incorporating this aspect into a test protocol is not obvious. Some ACWS include limited active intervention in some aspect of vehicle control (e.g., partial braking, counter-steering). This complicates the use of vehicle control or driving outcome measures as indices of the effectiveness of the DVI. Any evaluation method will have to specify the driving scenario(s) in which the warning occurs, yet the relative effectiveness of two interfaces may depend on the specific scenario used. Finally, the metric is intended to be applied to operational systems, and these may not be readily available or may employ proprietary algorithms not easily adapted to test methods such as driving simulators. Thus while there are important advantages to a common evaluation method, there are challenges in accomplishing this.

Design convention considerations Although the goal of this project was not to standardize any particular warning interface, there may be some benefits of conventions. Drivers may come to be familiar with the DVI in their personal vehicles, but as ACWS become more ubiquitous, drivers may confront unfamiliar interfaces when they use rental vehicles, share vehicles, or acquire a new vehicle. They may have false assumptions about vehicle functions and displays or may react slowly or inappropriately to emergency events. However, design conventions for a DVI may also have significant drawbacks and should not be proposed

without a strong basis. For this reason, the CWIM project included empirical research that investigated design convention issues.

Project Overview

The project involved a combination of empirical research and analytic activities. The initial efforts of the project were analytical. This work examined research literature, crash analyses, current practice for DVI design and evaluation, and expert/stakeholder feedback. This defined needs, options, and preliminary suggestions for use in the subsequent project activities. Subsequently, five empirical experiments were conducted. These included three experiments that compared various crash warning interface modes and examined the methods used to evaluate them. There were also two experiments that addressed various aspects of DVI comprehension and potential issues related to display variability. Each of the five experiments used a different methodology. One objective of this project was to determine the appropriateness of the various methodologies for use in subsequent human factors research on ACWS DVIs. While each experiment was developed to provide findings that could stand alone, some comparisons can be made between the various study methodologies. The findings of the analytical and empirical efforts were then considered in the development of suggestions for ACWS DVI evaluation, crash warning interface design, and design convention needs. This project overview paper focuses on the empirical studies and the implications for DVI evaluation. Greater detail on all aspects of the work may be found in the project final report (Lerner et al., 2011).

LDW WARNING MODE EXPERIMENT

The objective of this study was to determine how readily drivers are able to use LDW to improve lane recovery and crash avoidance, and in particular how this is related to warning modality and active warning strategies used by the LDW system. Active warnings (e.g., active countersteer) were of particular interest, both because they presumably have greater potential to promote rapid vehicle control responses and because their potential to induce inappropriate driver reactions is not well understood. The study also addressed driver acceptance issues. A system that is not well accepted by drivers may be disregarded or disabled and would therefore not be effective. Finally, through development of the experimental protocol, this study addressed issues surrounding the best approaches for evaluating the driver interface of LDW systems.

Method

The experiment was conducted in the National Advanced Driving Simulator (NADS-1) at the University of Iowa. A two-lane bi-directional rural highway with 3-meter lanes used in the study was representative of the most common roadway departure crash scenarios according to Najm, Koopmann, Boyle, and Smith (2002). The roadway database was designed so that it had both long two-lane highway straight-aways as well as a variety of left and right curves. The drive was approximately 30 minutes in duration.

The most common crash types and ones that are generally the most injurious and fatal were chosen for examination in the study: a) Vehicle drifts off road to the right, b) Vehicle drifts over the centerline, with oncoming traffic, and c) Vehicle fails to keep lane in a left curve entry. Each participant was exposed to these three scenarios while they were periodically distracted by a secondary task. Participants were also exposed to a false alarm in which the LDW alert activated while driving through a construction zone.

This study compared driver responses to passive and active LDW warnings and to a control condition in which no warning was given about an impending lane departure. Passive LDW warnings included an acoustic alert and a tactile alert (steering wheel vibration). Active warnings included a weak active countersteer and a stronger active countersteer. The form and magnitude of all alerts tested were within the range of alerts on production vehicles now sold in the United States or elsewhere, and on pre-production vehicles tested by NHTSA. Data were obtained from 90 participants (18 participants in each of five different LDW system groups, including one control group that did not experience any LDW).

Participants were instructed to perform a variety of secondary tasks while driving including a visual/manual "bug task" which distracted them from the forward roadway long enough that a lane departure could occur unnoticed. The bug task required that the participant turn and reach into the back seat to trace the path of a simulated insect on a touch screen display. To ensure that an LDW was obtained, a bias in the steering was triggered to nudge the car to the desired side of the lane during a distraction event. This was initiated based on driver engagement with the bug task. To mask the drift, no motion cues were provided.

Results

Although specific findings varied somewhat across the range of dependent measures included in this study, the general outcome was that all four warning conditions were superior to the baseline control condition and that frequently the "weak torque active LDW" condition performed best (although not always statistically significantly so). For example, the weak torque warning was significantly better than all or all but one other treatment for measures of maximum lane exceedance, severity of initial steering angle, total amount of time spent out of lane, and number of inappropriate behaviors elicited.

Overall, the impact of the warning modes on the number of inappropriate behaviors observed from driver showed that the highest number of inappropriate behaviors was observed from those participants who did not have an LDW. There were significantly fewer inappropriate behaviors for participants who experienced either of the active systems. The number of subjects who fully departed their lane, or ran off road, was significantly less for the strong torque warning and significantly greater for the auditory warning.

For the false alarm scenario, in which there was no actual lane deviation, the strong torque differed from the other warning conditions in that drivers with this system responded with greater, though unnecessary, vehicle control actions. The strong torque group had significantly earlier steering responses, greater peak steering rates, acceleration, and jerk.

Even though the active torque LDWs appeared to be more effective than passive alerts in minimizing lane departures, participants felt that they were more problematic. Participants were not asked to directly compare different warnings in this study. However, those participants who experienced the weak torque rated that warning as less effective in capturing their attention as compared to other participants' ratings of other LDW warnings. The group experiencing the auditory warning found it more effective at capturing attention as compared to other participants' ratings of either of the active warnings. Furthermore, the passive systems were viewed by the participants as being more helpful than the active systems. Participants who experienced a passive warning felt that the system was more easily interpreted than those who experienced an active warning. Participants who experienced a passive warning also felt that the system was more reliable than those who experienced an active warning.

FCW WARNING MODE EXPERIMENT 1

This experiment compared driver responses to two different FCW systems (passive versus active) on two different crash scenarios and a false alarm event. The passive FCW driver interface incorporated a head-up display (HUD) and an auditory alert. The active FCW used a brake pulse to alert the driver by exerting momentary activation of the brakes. Specifications for these warnings were developed in consultation with NHTSA. The two crash scenarios used in the study were a decelerating lead vehicle and a stopped lead vehicle. The experiment was conducted in the NADS-1 simulator. The roadway environment was similar to that used in the LDW study.

Method

To support this research on the effectiveness of FCW system warnings, it was necessary to use a distraction task that would reliably and repeatedly insure that the driver's eyes are off road for several seconds prior to the forward collision events. Because drivers can use peripheral vision to monitor the roadway, it was essential to direct the driver's gaze away from the forward view. To achieve this, the same simulated bug task used in the LDW study was used here.

Thirty-two participants experienced one of the two FCW systems; 16 other participants in a third group (baseline) did not have any FCW system. For each forward collision event, measures of initial vehicle control, inappropriate responses, and lane recovery were recorded. Following the drive, the driver's acceptance of the FCW system was assessed.

Results

There were no statistical differences (at the $p<0.05$ level) in response time among the conditions, although there was a trend across the reaction time measures used. When looking at initial response to the event, drivers in the baseline condition took longer to release the accelerator than drivers in the warning conditions relative to their initial engagement in the distraction task. There were also trends towards faster performance for responses relative to the time the alert was issued. On average, drivers responded by releasing the accelerator and applying the brakes 375 ms sooner with a warning than without. When applying the brakes, there were significant differences in both the level of braking and the maximum deceleration achieved by the driver. Peak brake pedal force was less forceful for drivers with the brake pulse than for drivers in the baseline and auditory/visual warning conditions. Also, these drivers achieved a maximum brake pressure that was

36% less than was achieved in the other conditions. Drivers in the brake pulse condition achieved a peak deceleration level that was 15% less than for drivers in the other two conditions. These differences in braking response did not translate into differences in collisons. Overall, there were no differences between groups in the number of participants who avoided collisions for either crash scenario.

Participants in each FCW group with alerts reported that they easily understood why the alert was presented, that the system successfully caught their attention, and that the alert was easy to see-and-hear or feel. The passive auditory/visual alert was rated significantly easier to interpret than the active brake pulse.

FCW WARNING MODE EXPERIMENT 2

This experiment, conducted on a test track at NHTSA's Vehicle Research and Test Center, compared driver responses to FCW systems that used either a HUD visual alert, an auditory beeping alert, a seatbelt tensioning device, or a some combination of two or all three of these alerts. See Forkenbrock et al. (2011) for the complete technical report on this experiment.

Method

Each of 64 participants was randomly assigned to one of eight groups. Participants in the first group experienced no FCW alert while participants in the other seven groups experienced one of seven different possible combinations of FCW alerts. The primary objective of the study was to develop a protocol suitable for evaluating FCW DVI effectiveness on the test track. A second objective was to compare the effectiveness of a small set of FCW alerts using the protocol that was developed.

Adult participants were recruited from the general public. Each participant experienced only one FCW event and had no exposure to the FCW system prior to experiencing the event. Each participant was asked to follow a lead vehicle while attempting to maintain a constant headway. Feedback on current headway was provided to the driver on a visual display. The participant was also asked to perform a secondary task which involved diverting their attention away from the forward roadway toward a visual display mounted near the back of the front passenger seat. After performing this task several times and driving back and forth across a straight test track, the distraction task was performed for a final time. During the final distraction task, while the participant was looking away from the roadway, the lead vehicle was abruptly steered out of the travel lane, revealing a stationary vehicle (a realistic-looking, full-size, balloon car) in the immediate path of the participant's vehicle. At a nominal time-to-collision of 2.1 seconds from the stationary vehicle, the FCW alert was presented to the driver.

Results

All eight participants in the baseline group with no FCW alert collided with the balloon car. Similarly, all eight participants who received only the HUD alert collided with the balloon car and 7 of 8 participants who received only the beeping alert collided with the balloon car. Among the various FCW alert combinations tested, it was apparent that FCW systems that included the seatbelt pre-tensioner as an alert were more effective in helping drivers avoid a collision than other FCW driver interfaces that did not include this alert. Approximately half of the participants who received the seatbelt tensioning alert (alone or in combination with other alerts) avoided colliding with the balloon car. The results of this study showed that the seatbelt pre-tensioner was effective at causing the driver to disengage from the secondary task (ending their visual commitment to the secondary task) and directing the driver's eyes back to the forward roadway in time to respond to the stationary vehicle in their travel lane. The timing of the protocol was such that many participants collided with the balloon car even though they had initiated some evasive maneuver. In addition to the outcome variable (collision/avoid) several other dependent measures were recorded in this study, such as timing variables (e.g. time from FCW activation until the driver's eyes were back on forward roadway), brake application timing and force, steering responses, speed of participant's vehicle at time of collision, etc.

IMPLICATIONS OF WARNING MODE EXPERIMENTS

The results of the three studies show that warning systems that included active and haptic warnings generally were more effective than systems that did not include these features (e.g., auditory and visual warnings). The FCW test track experiment showed that a haptic seatbelt pre-tensioning alert may be effective by helping to physically reorient the driver toward the forward roadway. Results from the LDW study suggests that there may be an optimal balance between having an active torque warning that is too weak and one that is too strong. A strong active warning signal will not necessarily be as effective as a weaker active warning and may lead to greater inappropriate driver responses to false alarms. Also, based on the results of the LDW study it is important

to note that there may be a mismatch between driver perception of effectiveness and effectiveness as measured by the performance measures.

FCW NEGATIVE TRANSFER EXPERIMENT

This experiment on negative transfer in auditory FCW addressed whether driver response to a FCW alert during simulated driving suffered when the participant switched from a familiar vehicle with one acoustic alert to a different vehicle with a different acoustic alert. A substantial decrement in response times after the vehicle change would suggest that there is a lack of transfer from one warning system to the other. The method and findings of the experiment are briefly described below. Additional details are in Robinson et al. (2011).

Method

A total of 60 licensed drivers completed the study. Of these, 28 were male. The average age was 27.5 with a range of 19 to 64, SD=9.1. The average driving experience was ten years and the minimum driving experience was five years.

The experiment consisted of two phases: the learning phase and the test phase.
- The *learning phase* was used to create the association between a particular auditory alert and various FCW events.
- The *test phase* was used to assess whether participant reactions to FCW events changed when exposed to a different auditory alert.

In the learning phase, half of participants were familiarized with one alert and half were familiarized with a different alert. In the test phase, half of participants received the same alert again (control groups) and half of participants received the alternative alert (treatment groups). Table 1 summarizes the four experimental conditions.

Table 1.
Summary of experiment conditions

Control	Treatment
Learning: Light Test: Light n = 15	Learning: Light Test: Heavy n = 15
Learning: Heavy Test: Heavy n = 15	Learning: Heavy Test: Light n = 15

Two auditory-only alerts, one designed for light vehicles (Light) and one designed for heavy vehicles (Heavy), were selected for use in this study. The Light alert had a fundamental frequency of 1500 Hz and a fast pulse rate. The Heavy alert had a fundamental frequency of 600 Hz and a slower pulse rate. The alerts were developed as part of the Integrated Vehicle-Based Safety Systems (IVBSS) project (Green et al., 2008). These warnings were selected because they were proven effective in field tests and were representative of the types of warnings that might be used in a collision warning system, yet sufficiently distinct in sound and acoustic parameters to give the impression that they were from two different automobiles. Both FCW alerts were presented at 85 decibels (dB) against a background road noise average of 62 dB (modulated by traffic present in the driving scene).

Participants completed a total of six simulated drives over the course of three experimental sessions (two drives per day). Each drive took about 20 minutes and consisted of rural highway and suburban/commercial segments with various traffic control devices and contexts (e.g., stop signs, traffic signals, curves, construction zones). Other traffic was present at low density throughout. Participants were instructed to obey the posted 45-mph speed limited drive in a safe manner as they would normally. The three scenarios were matched in terms of key components of proportion rural highway and suburban/commercial, number of traffic control devices, number of turns, etc. Participants were guided through the drives via turn-by-turn guidance presented aurally and visually through an in-vehicle display.

Throughout each drive, participants were required to perform a subsidiary task on an in-vehicle touch screen device, which was designed to increase the likelihood that participants' eyes would be directed at the touch screen rather than the roadway when a FCW event occurred. The task was a variation of the "Simon" task in which participants listened to a sequence of directions presented aurally (e.g., "Up, Down, Left, Left, Up") and then were required to repeat the sequence with button presses. Fifteen seconds after each Simon task was completed, participants were cued to begin the next Simon trial. This pattern continued throughout each drive.

Four FCW events based on those used in the IVBSS project (Green et al., 2008) were used in this experiment:
1. Lead vehicle (LV) suddenly brakes with another vehicle blocking a steering evasive maneuver.
2. Work zone lane reduction with LV sudden braking.
3. LV evasive maneuver to reveal stalled vehicle with another vehicle blocking a steering evasive maneuver.
4. Cut-in and sudden brake. Another vehicle coming up behind the cut-in vehicle blocks the participant's option to swerve.

Mirrored versions (e.g., cut-in from right rather than left) of each of the four event types, plus an additional lead-vehicle sudden braking event were used to generate a total of nine FCW events across the six drives.

Participants were instructed that the experiment was designed to investigate safe driving in the midst of various in-vehicle devices and tasks along with different driving situations. Prior to the first experimental drive (on Day 1), each participant was seated in the simulator and was introduced to its features and capabilities (including the "systems" in the vehicle for the purposes of the experiment (i.e., navigation system, FCW system, etc.). Participants were also introduced to the environmental sounds they would encounter while driving (e.g., car horn, cell phone ring) and the FCW alert that they would hear during the learning phase of the experiment. Participants were also introduced to the distraction task and route guidance instructions that they were to follow. Participants then completed a 10-minute practice drive to gain familiarity with the simulator and experimental procedures.

Two drives were completed on each day. For the first two days, participants experienced one of the two FCW alerts (Light or Heavy) whenever a FCW event occurred. On the third day, participants in the treatment conditions experienced the alternative FCW alert, while participants in the control conditions again experienced the same FCW alert that they experienced during the first two days.

Results

Although each participant was exposed to nine potential forward collision events, the data from about half (52%) of staged events were discarded because participants either may have begun to respond to the developing collision event before the FCW alert was triggered (i.e., participant began braking less than 200 ms after alert activation) or because they responded to the alert without braking (i.e., by swerving into another lane). Using the remaining data, three critical FCW event exposures were defined for each participant, and these critical exposures were used for all analyses. The critical events were the first exposure (the first time a participant heard the FCW alert), pre-switch exposure (the most recent event during which the FCW alert sounded before it was switched), and post-switch exposure (the first event in which the collision warning sounded after the switch). The terms "pre-switch" and "post-switch" are inclusive of both treatment and control condition participants, even though participants in the control conditions received the same alert in both time periods.

Brake response time (RT) was defined as the time between FCW alert issuance and the participant's first brake input. Figure 1 shows the mean brake RT for participants in each experimental condition by alert exposure period. The figure shows that while participants in the control conditions improved their performance in each successive period, participants in the treatment conditions displayed impaired responding, particularly in one of the two conditions. A mixed repeated measures ANOVA of brake RT was conducted with warning condition as a between-subjects variable and alert exposure (pre- and post-switch) as a within-subjects variable. The main effect for exposure was not significant at the $p<.05$ level, $F(1,9)=4.37$, $p=.066$. However, the main effect was subsumed by a significant interaction between warning condition and exposure, $F(3,9)=9.62$, $p=.004$. Participants exposed to the switch from the Heavy warning to the Light warning took significantly longer to respond to the post-switch event, relative to the time they took to respond at the pre-switch exposure and relative to the other three groups.
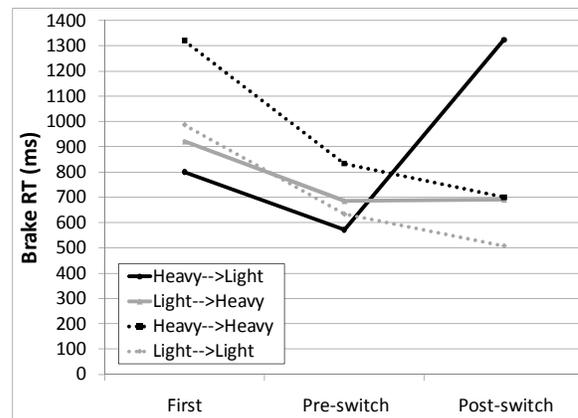

Figure 1. Mean brake RT as a function of warning condition and exposure period.

Following completion of all six drives, participants were asked to rate how similar pairs of sounds were on a scale of 1 to 7, with 1 being "very dissimilar" and 7 being "very similar." The similarity ratings were generally low, and the Heavy and Light warnings were both rated fairly consistently when compared to other sounds in the experiment, with the exception that the phone ring tone was perceived to be more similar to the Light warning than the Heavy warning. The relative similarity of the Light warning and the phone ring could have led to some confusion between these sounds, and therefore have played a role in the significant performance decrement observed among participants who switched from the

Heavy warning to the Light warning, though there is no direct evidence to support this interpretation.

ACWS STATUS DISPLAYS EXPERIMENT

This experiment focused on investigating whether people were able to identify and comprehends status displays for ACWS systems. The systems investigated in this experiment were LDW, FCW, blind spot warning (BSW), and adaptive cruise control (ACC). The main goal was to assess whether individuals understood what systems were present or operational in a vehicle and whether prior exposure to that vehicle's operational manual (or another vehicle's manual) affected that knowledge. Participants were presented with high-resolution images of a vehicle's interior, and then asked about system presence, operational status, and control button locations. The vehicle interior also was presented in several states of operation (e.g., pre-drive). Additional details are in Robinson et al., 2011.

Method

Participants viewed high-resolution images of a vehicle interior and answered questions regarding the presence or status of various vehicle safety systems. Data were collected regarding the following variables:
- *Comprehension* was defined as a correct response to a question about system presence or status.
- *Decision time* was defined as the amount of time (in seconds) taken by a participant to answer a question about system presence or status.
- *Confidence* was a participant's subjective rating (on a scale of 1 to 10) of their confidence that their response to a question about system presence or status was correct.
- *Location* was the position within the image of the vehicle interior that a participant indicated he or she sought the information that was used to determine whether a particular system was present or to determine its status.

The experiment was a fully crossed three-factor experimental design, with the following factors:
- *Vehicle (between-subjects)*: Each participant viewed the interior of one of three vehicles: 2010 Infiniti FX 35, 2010 Buick Lucerne, or 2010 Volvo S80. These vehicles were selected in part because they used different display strategies from each other (e.g., icons, text, acronyms). The vehicle models also ranged in number of safety systems.
- *Owner's manual familiarity (between-subjects)*: Prior to the experimental session, participants read sections of the owner's manual or related

manufacturer-provided material for one of the three vehicles, or did not read any manual. This resulted in cases where the participant had familiarization (through the manual) with the vehicle they subsequently viewed in the experiment, cases where the participant had familiarization with a different vehicle than the one they saw in the experiment, and cases where the participant had no familiarization with any of the vehicles. Participants were not informed until they arrived for their sessions which vehicle interior they would experience.
- *Scenario (within-subjects)*: During the experimental session, data were collected in three phases, for which the images of the vehicle interior represented three situations: prior to starting the vehicle, after starting the vehicle (but before driving), and during driving. The state of the displays and the particular questions asked were appropriate to the particular scenario.

A total of 111 licensed drivers from the general population participated in the experiment. To ensure that participants had limited familiarity with the features of interest in this experiment, individuals were only selected to participate if they did *not* currently or recently drive a vehicle of model year 2006 or later, and if they did *not* currently or recently drive a Buick, Infiniti, or Volvo vehicle of any model year.

Participants viewed high-resolution photographs of vehicle interior displays shot from the driver's perspective. The monitors through which the photographs were displayed had a 30-inch diagonal display area and 2560x1600 pixel native resolution, which allowed near-full size projection of the images and good legibility of text and symbols. Participants used handheld touch pads to select answers to yes/no questions, to select ratings of confidence, and to indicate where in the interior of the vehicle they looked for particular features.

Up to three individuals participated per session. If participants were assigned to read vehicle instructional material before their session, they first completed a five-question multiple choice quiz to assess whether they read the assigned materials, and how well they remembered them. Each participant sat at a computer station facing away from other participants' screens. The experimenter then introduced participants to the experiment task and guided them through two practice trials.

The experimental trials were organized into seven blocks. Each block of trials was based on a particular image of the vehicle interior. The first block was for the scenario where the vehicle was turned off (prior to vehicle ignition). The next three blocks were for scenarios where the vehicle was started, but not yet

moving. Each of these three photographs showed a different configuration in terms of what ACWS features and displays were on, off, or not functioning properly. The final three blocks were for scenarios where the vehicle was in motion. Again, each of the three photographs showed a different configuration in terms of what features and displays were activated.

There were a total of 40 questions across all seven blocks. The questions and answer options appeared on the participant's touch pad. For the pre-ignition scenario, all of the questions asked about the presence or absence of a particular system in the vehicle. The participant selected an answer ("present" or "not present"). Participants were asked to respond to each question as soon as they decided on an answer. The data collection system recorded the response time from the presentation of the question to the answer selection and confirmation. Once the answer was selected, the participant then rated their confidence that their selected answer was correct using a 10-point scale. Following that, the participant was asked "Where did you look for this information?" A photo of the vehicle interior was displayed on both the large monitor and the touch pad, and the participant used the stylus to point to the location on the touch pad. Figure 2 shows a participant viewing a vehicle display and responding on the touch pad.

Figure 2. Participant using stylus to indicate information location on touch pad.

Results

The primary dependent variables analyzed using mixed repeated measures ANOVAs in this study included accuracy of responses to questions, confidence in answers, and decision time. This section highlights some key findings of these analyses. Detailed results and statistical analyses are presented in Robinson et al. (2011).

Overall, individuals were not particularly accurate in assessing whether an ACWS was present (more than 40% of these responses were incorrect). This was consistent across all vehicles, systems, and whether or not the participant read a manual (same or different vehicle's manual). There was some variation by vehicle: Volvo and Infiniti participants were better at identifying system presence. Also, there was slightly better accuracy in identifying system presence (for some systems) when a participant read the appropriate manual, rather than no manual or another vehicle's manual. Mean percent correct responses across all questions as a function of vehicle and manual condition are shown in Figure 3.
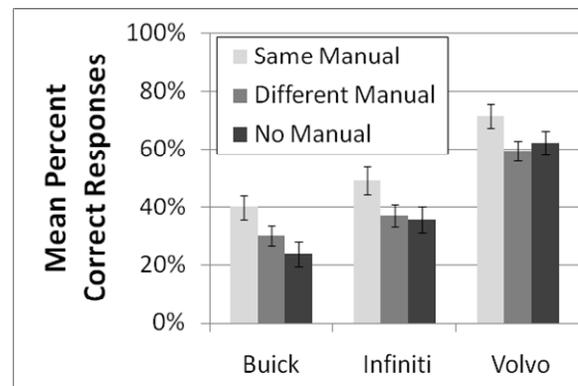

Figure 3. Mean percent correct responses for all systems (with standard error bars).

Individuals took considerably longer to respond to questions about system presence than system status across vehicles and manual conditions. Participants were always asked about system presence first, so much of this extra processing time may have been the result of the need to gain overall familiarization with the displays. Participants in the Infiniti condition took the longest to respond about system presence. Also, as expected, having the same vehicle's manual resulted in shorter decision times in all conditions.

As with system presence, participants were not particularly accurate when assessing system status (about 60% of responses were incorrect). In fact, participants were less accurate when determining system status than system presence. Participants in the Volvo condition displayed better comprehension of safety system status than those in the Buick and Infiniti conditions.

In addition to the low overall accuracy levels, participants also took a relatively long time to recognize the status of a safety system, even under static display conditions. Although participants were

already familiar with the layout due to earlier trials asking about presence of a system, decision times were long (about 15 s for startup scenario and 11 s for en-route scenario, on average) when asked about system status regardless of vehicle or whether or not there was prior exposure to the vehicle's manual. Participants tended to be very confident in their responses despite a high rate of errors and long response times.

Participants were somewhat more accurate when asked about the presence or status of a system if they had prior exposure to that vehicle's manual. Having a manual led to improved accuracy, but even with a manual, overall comprehension, as indicated by accuracy, was still rather low. In particular, participants in the Volvo condition who were given the manual had significantly higher accuracy than the other two vehicle conditions. Familiarity with the vehicle's manual led to somewhat more accurate and faster responding in the pre-startup phase, but not other phases. There was no finding of a systematic trend toward either positive or negative transfer based on the manual that participants read.

Results indicate many participants who answered questions correctly did not select the correct location where they should have looked to determine the correct answer, which suggests that some participants who answered correctly may have done so by intuition or chance, may have recalled some information from a manual without being aware of location, or may have been led to the correct answer by an irrelevant cue. When participants selected the wrong location, it was often on the dashboard, which suggests that when in doubt, participants expected to find status information in this area. Conversely, some participants who answered incorrectly actually did look at the correct location to choose their answer, which suggests that these participants found the correct cue, but misinterpreted it.

Having a clearly labeled button (e.g., LDW or an LDW icon for all three vehicles) helped individuals identify the presence of a safety system. In contrast, not having a clearly labeled control button but rather a generic menu button can make system identification much more difficult (e.g., BSW in the Buick, which is discussed specifically in the fifth example case in the next section). In addition, using icons or full-word text instead of acronyms appeared to improve understanding substantially. Volvo and Buick both used icons for LDW systems, and as a result seemed to produce faster recognition of system presence than was observed in the Infiniti, which used acronyms. It should be acknowledged, however, that this experiment was not designed to formally address this.

Presenting system status information in full-word text form seems to be more effective in facilitating understanding than using color coded icons if the color coding is not intuitive to drivers. The Volvo vehicle used the most text to communicate information, and also had the highest overall understanding by participants. In contrast, the Infiniti relied on color coded icons, and showed the lowest performance, perhaps because the meanings of colors were in some cases ambiguous or counterintuitive to participants. In addition, the presence of text messages seemed to ameliorate the effects of not having a manual or having another vehicle's manual. The opposite was true for icon color codes.

The Infiniti's placement of control buttons in the lower left hand corner below the steering wheel may be problematic for participants to notice. Decision times were longer for the Infiniti in the pre-startup phase, indicating longer search and recognition times.

IMPLICATIONS FOR DESIGN CONVENTION

The FCW negative transfer experiment and the ACWS status display experiment were drastically different in approach, presentation medium, design, and analysis. Regardless, there was a common theme focusing on the implications for design convention in auditory alerts and visual displays of ACWS systems. The first experiment focused on auditory alerts and potential transfer effects from one alert to another. Similarly, the second experiment investigated understanding of visual displays and potential transfer effects. Design convention implications are discussed separately for each experiment.

FCW Negative Transfer Experiment

Transfer problem There is potentially a transfer problem with FCW auditory alerts as presented in the current experiment, but its dimensions and conditions are not clear. In one direction of shift (heavy to light warning), the slowing of the brake response was more than 700 ms, which is quite large. In the other direction (light to heavy warning), the response time did not change, whereas participants in both control conditions improved their reaction times to the already-familiar warning by about 130 ms.

Familiarity There was a large familiarity effect in the experiment. The brake response time reductions across successive sessions, from first exposure to the post-switch trial (for the control groups), were about 500-600 ms. Even from the session 2 pre-switch to the session 3 post-switch conditions, the difference was approximately 130 ms. This indicates that if people come to recognize a familiar sound from general experience, their responses are faster the next time the alert is presented. This also indicates a

potential benefit of consistency of auditory alerts in ACWS.

Potential design convention benefits Participants experienced FCW alerts up to nine times over the course of a three day experiment, whereas in the course of actual on-road driving, these alerts are likely to be experienced much less frequently, which is likely to result in very different patterns of learning and familiarization. Under normal driving conditions, it may not be reasonable to assume that familiarization will occur quickly, so it is important to ensure that alerts lead to quick and proper responding regardless of prior experience.

Further research questions The transfer and familiarity effects suggest that some form of design convention could improve driver response time. The finding of a very large effect for only one transfer direction indicates that transfer effects may be highly dependent upon alert characteristics. For example, the asymmetry in transfer effects may be related to the similarity of the warning to other sounds that occur in the environment or perhaps to reactions to particular features of the alerts (e.g., spectral characteristics). However, this is an empirical issue that cannot be resolved with the current design and data. Several potential research questions are:

- What factors or components cause some auditory warnings to be more effective when there is a shift from the expected sound?
- What sound features could be used to maintain transfer (e.g., temporal pattern, primary frequency, tonal quality)?
- Would there be a better understanding of transfer effects and familiarity if individuals were recruited who actually drove one vehicle versus another?
- Would negative transfer effects occur in naturalistic circumstances (e.g., if the participants had become familiar with ACWS by driving an equipped vehicle over an extended period of time)?

ACWS Status Display Experiment

Overall comprehension There is the potential for a comprehension problem with vehicles containing unfamiliar ACWS systems, indicated by low comprehension rates and slow response times across all three vehicles used in the status display comprehension experiment. People unfamiliar with the systems had difficulty identifying system presence, operational status, and location.

Manual information Reading manufacturer-provided information was somewhat helpful, but the problems remained. Also, having read information about a *different* vehicle did not generally provide benefit. The limited improvement was vehicle-specific. It should be noted that the information read by participants, and the context in which it was read, were not typical of how individuals read owner's materials (if typically read at all), so the effects of manual information on ACWS comprehension should be interpreted with that caveat.

Potential design convention benefits Given the comprehension problems seen in this experiment, limiting the variability of some aspects of the vehicles' status displays may provide benefits. Potential areas for design convention are noted below:

- Use of standard terminology for particular warning functions, so that text or acronyms based on them are consistent.
- Use of standard icons and color coding for status. Color codes or icons should be congruent with drivers' mental models (e.g., green indicating activated or properly functioning systems).
- Status information should be located where people expect to see it. It is not clear to what extent this expectancy will be related to other aspects of the DVI, so location might have to be empirically determined/performance based for each vehicle, rather than there being a single preferred location for all vehicles.
- Effective quick-overview materials that convey what safety systems are in the vehicle, how status is indicated, and how they operate could help to enhance comprehension. Visual demonstrations might be appropriate and could be provided through web sites or other digital means. Because many drivers do not read owner's manuals, it might be beneficial to develop materials that are enticing or interesting to drivers, or to provide a demonstration or tutorial at the point of vehicle purchase. It may be reasonable to provide some criteria for such materials to help ensure system comprehension. Manual materials could be streamlined and presented in a way that is an easy reference (similar to the tabular format used in the drug industry).

DVI EVALUATION METHODOLOGY

Objective and Scope of the CWIM Application

In considering CWIM procedures, it is important to keep in mind the intended role of CWIM testing. The methods are intended to assess the DVI for a particular warning function in commercial or near-production systems. The intent is specifically to have a common method for evaluating the DVI of a commercial system.

Evaluating the DVI is not the same as quantifying or rating the performance of the safety system itself. The DVI is only one component of the system.

Therefore CWIM is not specifically concerned with how well a system addresses a crash situation, but more narrowly with how well the DVI conveys its status and the relevant crash-imminent information to the driver. How quickly and accurately does the driver perceive the threat and respond to the warning display, and does the interface elicit appropriate actions? The CWIM suggestions are focused on this goal.

Other issues related to scope include the following:

- Currently, CWIM is focused on the immediate driver response to warnings in potential crash situations. Not addressed are longer term influences on driver behavior and performance, such as automation complacency.
- Related to the concern above, the focus is also on the response to a particular warning function display, not on broader aspects of safety system performance. For example, driver response may be influenced by the frequency with which false alarms or nuisance alarms occur. This is important for assessing a system, but is not part of the DVI evaluation. Likewise, the effectiveness of a specific warning may depend on how well the particular function is integrated into the broader system of functions and information displays within the vehicle.
- The CWIM suggestions also must be tempered by practical considerations. It would not be feasible for a common evaluation procedure to experimentally manipulate all of the many factors that might interact with ACWS DVI performance (e.g., number of event scenarios included, roadway types, driver impairment, weather conditions, types of distraction). Some narrowing to a common set of conditions that will be practical for ACWS DVI assessment is required. Also, requirements in terms of unique facilities, costs, and practicality must be considered.

Specific Methodological Issues and Suggestions

Ten key factors were addressed in consideration of a testing protocol. For some of these, preliminary suggestions could be put forth. For others, resolution is not yet achieved and there is disagreement over potential approaches. Greater discussion of the issues and options related to each of these factors may be found in the project final report (Lerner et al., 2011). A brief presentation of the ten factors follows:

Driving scenario Two general aspects of the driving scenario must be kept consistent if common results are to be expected across testing sites. One is the general character of the roadway, such as the number and width of lanes, speed limits, presence of other traffic, type of setting (e.g., urban, rural), environmental conditions, and so forth. The other aspect is the dynamics of the potential crash event(s) under which driver response to the DVI will be evaluated. To some extent, the decisions about the driving scenario will be determined by whether a driving simulator or test track facility is used, since safety and practicality considerations limit what might be done on a test track. As a general principle, the general character of the roadway should not be more complex than is required by the event scenario. For LDW and FCW events, as with most other warning functions, there are a limited number of pre-crash scenarios that account for a large portion of crash outcomes (e.g., Najm, Smith, and Yanagisawa, 2007). Specific criteria for event scenarios are described in Lerner et al. (2011), but the general suggestion is to include a limited number of scenarios based on their prominence in crash statistics. Exceptions to this may be required on pragmatic grounds for test track procedures.

A related concern is the number of ACWS activations that research participants are exposed to. In naturalistic driving, ACWS activation is likely to occur very rarely, but in an experimental context it may be necessary to induce an artificially high number of activations, exposing participants to one or more activations over a relatively short period of driving. It is not clear how the absolute number of activations, or the rate of activations over time, influence participants' expectations, driving behavior, or reactions to warnings. There are, however, some principles of simulation study design that may mitigate the effects of an unrealistic ACWS activation frequency. Based on recent ACWS simulator research studies with relatively high rates of ACWS activation, Green (2008) provides five recommendations for maintaining validity: 1) use real world crash data as a basis for crash scenarios; 2) use a large number of potential crash scenarios; 3) include between three and five other vehicles in surrounding traffic that require the participant's attention; 4) use real on-road data to select crash parameters (e.g., closing rate); and 5) design scenarios to minimize the number of unusable trials (e.g., warning not issued or participant not sufficiently distracted).

Participants There are various alternative strategies as to how the sample of research participants should be composed. The approach favored here is that the participant sample be based on a relatively stable and homogeneous portion of the typical driving public. It specifically excludes special groups based on diminished capabilities or risky actions or populations defined by consumer attributes. This is most consistent with the goal of a

common CWIM methodology, which is to compare DVI "A" with DVI "B" in a stable, repeatable manner. Additional criteria for participant selection are provided in Lerner et al. (2011).

Distracting the driver ACWS are intended to support the driver in recognizing emerging hazards. The primary purpose of systems such as LDW and FCW is to alert the driver who is distracted or otherwise unlikely to detect the event on their own. Evaluating the DVI for these systems therefore should include an appropriately distracted driver. The means of distracting the participant is a key part of any common CWIM methodology. In normal driving for most people, relatively long glances away from the road are quite rare and are difficult to predict. Therefore the experimental method must have some means of inducing appropriate visual distraction at known times. Distraction is a complex issue and there are many forms of distraction. The choice of a distraction task is currently an issue in research and standards efforts and is a matter of some contention. The CWIM project (Lerner et al., 2011) identified ten criteria for defining an ideal distraction task. In practice, there are tradeoffs and it may not be possible to optimize all of them.

Warning system context An ACWS functions within the context of the particular vehicle that it is designed to support. The warning occurs in the context of other safety functions, displays, and communications within the vehicle, and may occur as part of a progressive warning strategy or be related in some way to a parallel safety-relevant system (e.g., adaptive cruise control). Therefore the question arises as to how to deal with the warning system context. The purpose of the CWIM evaluation is not to quantify the effectiveness of the safety system in crash avoidance, but more specifically the ability of the DVI to convey the appropriate information and induce the appropriate driver response. Therefore a particular DVI may be evaluated on a stand-alone basis within the framework of a given vehicle and driving context, even if it may occur in a particular vehicle within the context of earlier informational messages or lower level alerts. Obviously, performance might be better if these related messages were present. However, the earlier alerts might not always be sufficient; if they were, the imminent crash warning would not be required. Therefore the CWIM protocol should test the worst-case situation where the driver has not taken account of other messages and is responding only to the imminent crash warning itself. If resources allow, it may be of interest to include within-context testing as well, but the primary context for evaluating the warning display should be a stand-alone presentation.

Familiarity with the technology Driver response to a warning depends to some degree on the driver's familiarity with the ACWS. At one extreme, a person may not realize the technology for a particular warning capability exists. Or, they may not realize that the particular function is present in the vehicle they are driving. Or, they may understand that it is present but have no idea what it looks, sounds, or feels like. They may or may not have familiarity with other commercial products that fulfill a similar function. At the other extreme, they may be highly experienced with the specific system present in the vehicle they are driving. The experiments described above on FCW negative transfer and status display comprehension suggested that familiarity may influence response time or accuracy. Therefore, the question arises as to what degree of familiarity with the technology participants should have under CWIM procedures. One perspective is that a totally naïve driver represents the "worst case" and therefore should be the basis for the evaluation. Another suggests that this is neither a representative nor fair basis for testing a particular DVI. According to this view, drivers may be assumed to at least be aware that a warning function is present in their vehicle; furthermore, they will only be totally naïve to the look, sound, or feel of the display once, and after that, all future driving will be done with some awareness of the system. Some types of ACWS warnings may be expected to occur with some frequency (e.g., lane departure or blind spot warnings) while others could be quite rare, so drivers may have less familiarity with the ACWS DVI in their own vehicles. While the most appropriate degree of system familiarization remains open to debate, and may depend upon specific research objectives or methods, our general suggestion is to provide a limited and controlled degree of pre-exposure, particularly if the study design includes repeated exposures to an ACWS. Ideally, a study might incorporate both naïve and familiarized participants. Further research investigating the effects of various familiarization schemes (including no-familiarization) may help to refine this recommendation.

Participant expectancy The nature of the participant's driving task and the expectancies engendered by the procedures are a critical concern. The intent is to impose the potential crash situation on drivers who are driving in their normal manner and are not anticipating the probable occurrence of an emergency event. The instructions defining the purpose of the experiment *from the participant's perspective* are critical. Participants should not have any indication that the researcher's interest is specifically with crash warning systems. Instructions

to participants and associated materials should not promote this perception. As much as possible, the procedure should foster the feeling that drivers can simply behave in their normal manner. Lerner et al. (2011) describe a general procedure to meet these needs, based on the premise that participants believe they are going to experience a new prototype vehicle that includes a variety of innovative design features. This allows some familiarity with the ACWS alert without focusing participant attention specifically on crash warnings.

Accommodating user settings and options
Manufacturers may design systems that allow the user to select or program various aspects of the system response, or systems that adapt to the characteristics or performance of the driver. Thus there may be user-controlled or dynamic variance in DVI characteristics such as display intensity, display type, triggering criteria, or timing of displays. If a display attribute is adjustable in some dimension, what setting should be used for CWIM testing? Should the procedure use the most conservative setting, the least conservative, a mid-point, a default setting, a setting selected by the research participant, or some combination of these? The recommendation here is to use the default or mid-point setting, because it carries the implication that this is the "normal" option and any deviation from this is the user's responsibility.

Comparison conditions
If CWIM methods are meant to evaluate the effectiveness of a DVI, there is the question of "effective compared to what?" Is comparison made to a benchmark value, control condition, or "standard" interface? Is the evaluation to be taken in absolute or relative terms? Is the assessment quantitative or a pass/fail decision through comparison with some criterion? Our recommendation is that CWIM evaluations include a relevant control condition(s) and that pass/fail outcomes result from comparison to that condition. Ultimately, it would be most desirable to define some absolute performance levels for a particular dependent measure, based on a sufficiently large study to define this threshold empirically. However, this is not feasible until some standard metric is agreed upon, a threshold is established through adequately large empirical efforts, and the measure is shown to be highly reproducible across different evaluation sites. Until absolute metrics have been adequately demonstrated, the performance of a given DVI must be made on a relative basis, compared to a benchmark condition included in the same evaluation study. The suggestion here is that CWIM evaluations of a DVI include two benchmark conditions within the same study. One of the benchmarks is a "no warning" control condition. The other is a fully-specified "basic" DVI. These two benchmarks would define thresholds for three levels of performance: (a) no benefit; (b) basic effect (adequate); (c) superior. The comparison with the no warning control condition is desirable because poorly designed DVIs may have no appreciable beneficial effect and in some cases may even prove worse than no warning at all. Furthermore, the control condition may provide a confirmation of the appropriate urgency of the potential crash scenario and the distraction procedures. The comparison with a basic standard comparison DVI is useful because merely showing an improvement relative to a no-warning control condition is a very minimal basis for evaluating a DVI. If a simple and common type of warning is shown to have some beneficial effect, the CWIM procedure should determine whether a given DVI is similar to, worse than, or superior to this basic display. We suggest a comparison signal be drawn from major previous research (e.g., CAMP) or from a detailed survey of parameters in current warning signals in production vehicles. The "basic" DVI should be an exemplar of typical vehicle warnings, but should not be identifiable as uniquely the display of any specific manufacturer's product. An additional advantage of having these benchmark conditions is that it will permit "calibrating" comparisons across testing locations or testing times.

Treatment of data
Lerner et al. (2011) recommended that common analytic methods not rely solely on comparisons of central tendency. Example data were provided to illustrate how differences may be more pronounced in the tails of response distributions, and these might be obscured with a focus on central tendency.

General test method
Because CWIM requires a highly repeatable measurement system, this implies the use of either driving simulator or test track methods. Actual on-road driving with the ACWS may be valuable, but it does not provide the control needed for a formal assessment tool to compare DVIs. Both driving simulators and test tracks have advantages and limitations and both are potentially useful for CWIM testing. Simulators provide the flexibility to program a wide range of potential crash scenarios and environments without exposing participants to any real hazards. However, test track driving provides a real driving environment and does not require the use of advanced simulation facilities.

CONCLUSIONS

The five experiments described in this paper investigated human factors issues related to ACWS DVI warning mode, comprehension, and warning variability between vehicles. Each experiment used a

different methodology, and the methodologies themselves were assessed to determine their appropriateness for future human factors research on ACWS DVIs, and to define the key issues and decisions that must be addressed.

The three studies on ACWS warning modes generally found that haptic and active warnings led to quicker and more appropriate responding than visual and auditory alerts, when participants were visually distracted. The simulator study on DVI variability found that participants' responses to a FCW warning are slowed when they switch from a familiar auditory warning to an unfamiliar auditory warning. The laboratory study on ACWS DVI comprehension found generally low comprehension of ACWS presence and status among naïve participants, and only a slight benefit when provided with some familiarization through owner's manual materials.

While each experiment was intended to provide findings that could stand on their own, some comparisons can be made between the various experiments' methods. On the whole, the study found that there was no one best method, but rather many options, each of which has advantages and limitations depending on the objectives of the research. When considering a CWIM test method that could be generally applied to investigate various ACWS DVIs, research methods must balance internal and external validity with considerations such as practicality, repeatability, and cost

ACKNOWLEDGEMENTS

REFERENCES

Forkenbrock, G.J., Snyder, A.A., O'Harra, B.C., Hoover, R.L., Vasko, S.M., and Smith, L.A. 2011. "A Test Track Protocol for Assessing Forward Collision Warning Driver-Vehicle Interface Effectiveness." Washington, DC: National Highway Traffic Safety Administration.

Green, P. (2008). "Developing Complex Crash Warning Simulations for Human Factors Evaluations." In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 1865-1869.

Green, P., Sullivan, J., Tsimhoni, O., Oberholtzer, J., Buonarossa, M.L., Devonshire, J., Schweitzer, J., Baragar, E., and Sayer, J. 2008. "Integrated Vehicle-Based Safety Systems (IVBSS): Human Factors and Driver-Vehicle Interface (DVI) Summary Report." DOT HS 810 905. Ann Arbor, MI: University of Michigan Transportation Research Institute.

Lerner, N., Jenness, J., Robinson, E., Brown, T., Baldwin, C., and Llaneras, R. 2011. "Crash Warning Interface Metrics Task 12 Final Report." Washington, DC: National Highway Traffic Safety Administration.

Najm, W., Smith, J., & Yanagisaw (2007). *Pre-Crash Scenario Typology for Crash Avoidance Research*. DOT HS 810 767. National Highway Traffic Safety Administration, Washington, DC.

Najm, W., Koopmann, J., Boyle, L., and Smith, D. 2002. "Development of Test Scenarios for Off-Roadway Crash Countermeasures Based on Crash Statistics." DOT HS 809 505. Washington, DC: National Highway Traffic Safety Administration.

Robinson, E., Lerner, N., Jenness, J., Singer, J., Huey, R., Baldwin, C., Kidd, D., Roberts, D., and Monk, C. 2011. "Crash Warning Interface Metrics Task 3 Report: Empirical Studies on Standardization Needs." Washington, DC: National Highway Traffic Safety Administration.