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

www.nhtsa.gov

DOT HS 811 470b **August 2011**

Crash Warning Interface Metrics

Task 3 Report: Empirical Studies of Effects of DVI Variability

Technical Report Documentation Page

1. Report No. DOT HS 811 470b	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Crash Warning Interface Metrics: Task 3 Report: Empirical Studies of Effects of DVI Variability		5. Report Date August 2011 6. Performing Organization Code
7. Author(s) Emanuel Robinson, Neil Lerner, James Jenness, Jeremiah Singer, Richard Huey, Carryl Baldwin, David Kidd, Daniel Roberts, & Chris Monk		8. Performing Organization Report No.
9. Performing Organization Name and Address Westat 1600 Research Boulevard Rockville, MD 20850		10. Work Unit No. (TRAIS) 11. Contract or Grant No.
12. Sponsoring Agency Name and Address		13. Type of Report and Period Covered Final Report 14. Sponsoring Agency Code
15. Supplementary Notes		

16. Abstract

This report documents the methods and findings of Task 3 under the project "Crash Warning Interface Metrics (CWIM)." The CWIM project has the objective of examining the potential advantages and concerns of Advanced Crash Warning Systems (ACWS), with a particular focus on the driver-vehicle interface (DVI). Task 3 involved new empirical research to address issues of DVI variability for ACWS systems across vehicles. Two experiments were performed. Experiment 1 addressed whether driver response to a forward collision warning (FCW) acoustic alert suffered when the participant switched from a vehicle with one acoustic alert to a different vehicle with a different acoustic alert. After the alert was switched, participants displayed substantially delayed brake reaction times, particularly in one direction of shift. This comparison provides some evidence of a potential negative transfer effect. Experiment 2 investigated whether people who are unfamiliar with ACWS features were able to identify and comprehend status displays for a variety of existing ACWS. Overall, individuals were not particularly accurate in assessing whether an advanced crash warning system was present (more than 40% of these responses were incorrect), but participants were nonetheless confident in their responses. A degree of familiarity with an ACWS (from reading owner's manual materials) improved comprehension slightly, but there was no finding of a systematic trend toward either positive or negative transfer. The final section of the report discusses methodological assessments and implications for each experiment.

17. Key Words Driver-vehicle interface (DVI), advanced crash warning system (ACWS), forward collision warning (FCW), lane departure warning (LDW)		18. Distribution Statement Document is available to the public from the National Technical Information Service www.ntis.gov	
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 72	

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

Acknowledgements

The authors thank Eric Traube, the NHTSA Task Order Manager for this project, and Dr. Eric Nadler with the Volpe National Transportation Systems Center for their guidance and insight during the project. We also thank the Westat and GMU staff who aided in the conduct of the research: Michael Gill, Jeremy Walrath, Shawn McCloskey, Diane Snow (Westat); and Erik Nelson, Andre Garcia (GMU). Finally, we wish to thank the automobile dealership managers (Bill Curry at Capitol Buick, William Gray at Infiniti of Chantilly, and Nina Gee at Darcars Volvo of Rockville) who allowed researchers to access vehicles for purposes of photography or audio recording.

Table of Contents

List of Figures

List of Tables

Executive Summary

This report documents the methods and findings of Task 3 under the project "Crash Warning Interface Metrics (CWIM)." The CWIM project had the objective of examining the potential advantages and concerns of Advanced Crash Warning Systems (ACWS), with a particular focus on the driver-vehicle interface (DVI). Examples of ACWS 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. The larger project also developed recommendations and protocols for methodological aspects of evaluating the ACWS DVI, such as distracter tasks and event scenarios (Lerner et al., 2011). Thus the focus of the project is on identifying the effects of certain warning system features (e.g., warning modality, active intervention in vehicle control) and on establishing common methods and metrics that may be generally applied for evaluating the interfaces in different vehicles. The project does *not* have the goal of proposing a standard interface for that function.

Providing a common a warning interface has both positive and negative potential. The possible drawbacks are significant, so that recommendations should not be made lightly and without a strong empirical or analytical basis. However, despite these concerns, there may be good reasons to establish benchmarks for user comprehension or performance. Driver response to signals that are unfamiliar may be delayed or confused. Safety may be compromised if the user experiences negative transfer between one system and another. A driver who is accustomed to a particular interface in one vehicle may be confused by, react slowly to, or react inappropriately to a warning from a vehicle with a distinctly different ACWS DVI. In fact, the same interface feature could have different meanings in two vehicles. Furthermore, if there is a wide range of possible displays, it would be difficult to keep crash alerts perceptually distinct from other non-emergency displays. The DVI should also convey the status of the warning system to the driver. A driver should be aware of whether a given vehicle has a particular type of warning system (e.g., FCW), have an accurate mental model of how that system operates, and understand whether the system is currently operational. For example, a particular FCW system may only work when the vehicle speed exceeds some threshold, or a LDW may not be functioning because lane markings are inadequate or because there is some sensor or electronic failure in the system. These are complex messages to convey and inconsistency among manufacturers in whether and how such messages are conveyed may lead to driver confusion.

Task 3 attempted to determine the presence and extent of driver problems that may be associated with variability among DVIs. There has been very little research on this issue for ACWS and therefore it is important to determine whether a meaningful problem exists. In previous CWIM project activities, the project team reviewed literature on ACWS, crash warnings, and current practice. Based on these activities, a research plan was developed for two experimental studies that addressed significant gaps in the current literature. Based on the findings of these experiments, taken together with other literature, implications and recommendations are drawn.

Experiment 1 on negative transfer in auditory FCW addressed whether driver response to a FCW warning 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 experiment was conducted in a driving simulator. The safe driving behaviors of

participants were indicative of the validity of the simulator environment, to the extent that participants appeared to be highly motivated to drive attentively and avoid collisions. During the simulator drive, participants were periodically cued to engage in a distracting task. Occasionally a forward event occurred (e.g., sudden slowing of a lead vehicle) that required an emergency braking response. Participants became familiar with a given warning system over the course of two driving sessions in the simulator. In the third session, superficial changes were made to the appearance of the simulator vehicle and the participant was informed that a different vehicle model was now being simulated. For half the participants, the FCW acoustic warning remained unchanged and for the other half (the treatment conditions), the warning was different. The key comparison was in response times to the warnings among drivers with or without a change in the FCW warning.

The FCW alerts used in this experiment were developed as part of the Integrated Vehicle-Based Safety Systems (IVBSS) Initiative. One was developed as an alert for light vehicles (light) and the other was developed as an alert for heavy vehicles (heavy), but otherwise they were developed for the same purpose. Brake response time (RT) was the primary dependent measure. It was calculated as the time between the onset of the collision warning and the first brake depression (i.e., the moment when the participant began to depress the brake).

Participants in all four conditions had faster brake RT times at the last exposure prior to switching to a different warning than they did at their very first exposure, indicating a learning effect from repeated exposures. Once the warning was switched for the treatment group participants, however, brake RTs diverged. Control group participants, who again experienced the same alert that they had experienced earlier, further improved their RTs. Participants who switched from the light warning to the heavy warning had a slight, nonsignificant increase in reaction times, but participants who switched from the heavy warning to the light warning had a significant and dramatic increase in brake RT. This comparison provides some evidence of a potential negative transfer effect.

Of particular interest in the investigation were the similarity ratings between the warning sounds used and other sounds presented in the scenario. Due to the sound characteristics of the individual warnings, the light warning and heavy warning sounds more closely resembled some distracter environmental sounds than others. The average similarity rating between the two alerts was not significantly different from the similarity rating between the light warning and the phone. The similarity rating between the two warnings was however significantly different from similarity rating between the heavy warning and phone. This indicates that participants may have been more likely to confuse the phone ring with the light warning than with the heavy warning, providing a potential explanation for the negative transfer effects obtained for participants who were adapted to the heavy warning sound on the first two days of drivers and then later received the light warning in the switch condition on Day 3.

Experiment 2 on ACWS status display comprehension investigated whether people were able to identify and comprehends status displays for a variety of existing ACWS. Note that unlike the first experiment, this experiment dealt with driver recognition and comprehension of safety system information and not responses to imminent crash warning alerts. The main goal was to assess whether individuals understood what systems were present and were 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 questions about system presence and operational status. The vehicle interior also was presented in several states of operation (i.e., pre-startup, startup, en-route). In addition, a subset of participants was given short versions of vehicle owner's manuals to read before the session, which may or may not have been for the actual vehicle which they were presented. Participants' responses were analyzed for content accuracy, decision time, and accuracy of the location where they looked for particular features or status information.

Overall, individuals were not particularly accurate in assessing whether an advanced crash warning system 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 (correct or incorrect 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 when a participant read the appropriate manual, rather than no manual or another vehicle's manual. In summary, it took participants considerable time to determine whether the vehicle had a particular warning system (mean of 20 seconds across all conditions and questions), but despite this time, they were frequently incorrect.

There were several other main findings from the status display experiment. First, although overall comprehension of warning system status indicators was low, participants were nonetheless confident in their responses. Second, familiarity with the vehicle's manual led to somewhat faster responding in the pre-startup phase, but not other phases. Familiarity improved comprehension somewhat. Third, there was no finding of a systematic trend toward either positive or negative transfer. Overall comprehension rates for the "no manual" and "different vehicle's manual" conditions were quite similar and both were lower than the "same manual" condition. Detailed examination of individual questions, however, suggested that specific instances of negative transfer may have occured. Finally, results show that participants who answered questions about system status correctly did not always select the correct control button or status display 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, or may have been led to the correct answer by an irrelevant cue.

Some of the key findings of the group discussions that were part of the status display experiment that were conducted at the end of each session were: a) unfamiliar acronyms and icons were difficult to understand, b) color can be an effective cue if matched to users' mental models, c) organization of controls and displays is important, d) vehicle interfaces are learnable, and e) instruction manuals are helpful, but are no substitute for experience with the vehicle.

There were also several patterns of design issues that should be noted: a) having a clearly labeled button helped individuals identify the presence of an ACWS, b) using icons instead of acronyms appeared to improve understanding the most, c) presenting system status information in full-word text seems to be more effective in facilitating understanding than using color coded icons.

The findings presented in this report suggest that driver understanding of system status can be impacted positively and negatively based on system design characteristics used to communicate important information about the ACWS. In both cases, rare events (such as an FCW crash alert or a LDW malfunction message) should be self-explanatory to the driver, especially in urgent situations. The current study indicates there is at least a need to further investigate standardization needs and design choices for ACWS.

1 Background and Objective

1.1 CWIM project overview

This report documents the methods and findings of Task 3 under the project "Crash Warning Interface Metrics (CWIM)." The CWIM project has the objective of examining the potential advantages and concerns of Advanced Crash Warning Systems (ACWS), with a particular focus on the driver-vehicle interface (DVI). Examples of ACWS include forward collision warning (FCW) systems and lane departure warning (LDW). ACWS are increasingly common in passenger vehicles and the characteristics of these systems vary considerably among vehicle manufacturers. The project also developed recommendations for methodological aspects of evaluating the ACWS DVI, such as distracter tasks and event scenarios. Thus the focus of the project is on identifying the effects of certain warning system features (e.g., warning modality, active intervention in vehicle control) and on investigating methods and metrics that may be generally applied for evaluating the interfaces in different vehicles. The evaluation methods and metrics are designed to be applicable to whatever specific interface a given vehicle uses for a particular warning function, such as FCW. The project does *not* have the goal of proposing a standard interface for that function. The specific goal of Task 3 was to identify the potential effects of DVI variability across vehicles. Even systems that test adequately by themselves may suffer problems in actual application because users face problems due to the variability in DVI among vehicles.

1.2 DVI variability considerations

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. The concerns related to DVI variability therefore could become more significant as diverse vehicles with such systems proliferate.

One objective for developing CWIM is to insure that systems in new vehicles perform adequately, to at least some acceptable level or benchmark. The intent is to achieve this through the development of proven, repeatable, and efficient test metrics. Even systems that test adequately by themselves, however, may suffer problems in actual application because users face problems due to the variability in DVI among vehicles. 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. The concerns related to variability among DVIs therefore will become more prominent as diverse vehicles with such systems proliferate.

One approach to address this concern might be standardization of some aspects of the DVI. Standardizing a warning interface, however, has both positive and negative potential. The possible drawbacks are significant, so that recommendations for standardizing should not be made lightly and without a strong empirical or analytical basis. Some of the concerns with standardizing the ACWS DVI include the following:

- A standard may constrain what industry can do, which may limit innovation
- Technology advancements may suggest new and better approaches, not compatible with the standard
- The standard may ultimately be inconsistent with aspects of future in-vehicle environments (e.g., new types of displays)
- Each manufacturer may have a different suite of warning systems and system features, and a single approach may not be optimal for all manufacturers or all drivers
- A standard may oppose manufacturers' interests in product differentiation and conflict with the given esthetic approach of a given vehicle. Another approach of equal effectiveness to the standard might be reasonable, yet not allowed

However, despite these concerns, there may be good reasons for promoting some common features for the ACWS DVI interface. Driver response to signals that are unfamiliar may be delayed or confused. Safety may be compromised if the user experiences negative transfer between one system and another. A driver who is accustomed to a particular interface in one vehicle may be confused by, react slowly to, or react inappropriately to a warning from a vehicle with a distinctly different ACWS DVI. In fact, the very same signal could have explicitly different meanings in two vehicles. Furthermore, if there is a range of possible displays, it may be difficult to keep crash alerts perceptually distinct from other non-emergency displays. The DVI must also convey the status of the warning system to the driver in order to achieve its function. A driver should be aware of whether a given vehicle has a particular type of warning system (e.g., FCW), have an accurate mental model of how that system operates, and understand whether the system is currently operational. For example, a particular FCW system may only work when the vehicle speed exceeds some threshold, or a LDW may not be functioning because lane markings are inadequate or because there is some sensor or electronic failure in the system. These are complex messages to convey and inconsistency among manufacturers in whether and how such messages are conveyed may lead to driver confusion.

1.3 Objectives of Task 3

Task 3 provided the opportunity to conduct new empirical research to determine the presence and extent of driver problems that may be associated with DVI variability. There has been very little research on this issue for ACWS and therefore it is important to determine whether a meaningful problem exists. Previous CWIM project activities included a review of literature on ACWS, crash warnings, and current practice. Based on these activities, a research plan was developed for two experimental studies that addressed significant gaps in the current literature. These two experiments are described in overview in Section 2, and in detail in Sections 3 and 4. Based on the findings of these experiments, taken together with other literature, implications and recommendations are drawn. Section 5 discusses the implications of the study methods and findings.

2 Task 3 Overview

2.1 Prior project tasks

Tasks 1 and 2 of this project provided a basis for the planning and conduct of the Task 3 experiments. Task 1 continued information gathering from a previous project and focused largely on the methodological issues surrounding a common evaluation protocol to be used in assessing the DVI for various ACWS applications. Part of this work also included obtaining feedback on issues and practices from the automotive industry and other stakeholders, through a Federal Register notice and subsequent site visits. Task 2 also included a review of literature and current practice related to the use of auditory warnings for ACWS. In addition to examining formal literature on the topic, the project also surveyed sound and voice messages currently being used in production vehicles. An inventory of in-vehicle sounds from a variety of vehicle systems was also recorded for a sample of 13 vehicle models. These included alerts associated with potential collision situations (e.g., FCW, LDW, parking aid, adaptive cruise control – ACC) as well as other sounds in the vehicle environment (e.g., seat belt warning, lights on, door open, navigation, cell phone). Findings from these initial tasks helped to identify key research issues and specific systems and displays for use in the Task 3 experiments.

2.2 Procedural overview: Experiment on negative transfer with auditory FCW warning

This experiment addressed whether driver response to a FCW warning 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 experiment was conducted in a driving simulator. During the simulator drive, participants periodically engaged in a distracting task. Occasionally a forward event occurred (e.g., sudden slowing of a lead vehicle) that required an emergency avoidance response. Participants became familiar with a given warning system over the course of two driving sessions in the simulator. In the third session, superficial changes were made to the appearance of the simulator vehicle and the participant was informed that a different vehicle model was now being simulated. For half the participants, the FCW acoustic warning remained unchanged and for the other half, the warning was different. The key comparison was in response times to the warnings among drivers with or without a change in the FCW warning. Details of the procedure are in Section 3.

2.3 Procedural overview: Experiment on ACWS system status display comprehension

This experiment addressed how well people comprehend status displays in vehicles with quite different display strategies and whether familiarity with one vehicle's systems was helpful or interfered with understanding another vehicle's ACWS status. Unlike the FCW auditory warning experiment, this experiment did not deal with the immediate driver response to an imminent crash warning. Rather, it dealt with the driver's understanding of the status of the ACWS: Is a given warning function (e.g., FCW) present in this vehicle or not?; Is the function presently active ("on" or "off")?; Is the function currently fully operational? A driver who misunderstands these issues may fail to recognize a particular warning or may adopt a driving style that is based on a false assumption about the warning system. The experiment visually simulated the vehicle interior that a driver sees (dashboard, steering wheel, and console displays). Vehicles were selected to demonstrate a range of systems, based on the number and types of ACWS functions and their interface approaches (e.g., interface layout, use of icons, text, acronyms). For a variety of status display scenarios, information was collected on what the participants understood, how confident they were in their understanding, how long it took to process the information, and where they looked for system presence or status information. Prior to the experiment, some participants read owner's manual sections relevant to the vehicle they would see during the experiment, others read owner's manual sections for a different vehicle, and some read nothing. The experiment allowed an assessment of viewer comprehension as a function of interface design and system familiarity (manual conditions). Details of the procedure are in Section 4.

3 Negative Transfer with Auditory FCW Experiment

The research plan was based on the findings of the Task 2 Literature Review, DVI Survey, and Sound Inventory. The objective was to examine transfer effects for auditory FCW alerts to help determine whether standardization of such warnings might provide benefits.

The purpose of the empirical phase was to test whether driver reactions to a FCW event change when triggered by a different auditory alert than the one previously learned. This sequence of events represents the situation in which a driver with a vehicle equipped with a FCW system drives another vehicle equipped with a different FCW system (e.g., a rental car). The research prediction was that presenting a novel FCW alert to participants who had previously learned a different alert would result in slower responding compared to the participants in the control conditions who experience the same alert that they had learned.

The fundamental approach was to use a simulated driving context to create an association between forward collision events and a FCW auditory alert before changing the auditory alert to determine if responses were affected.

Whereas it was desirable to maintain a relatively low rate of FCW events over multiple drives, the practical requirement of creating sufficient association between the auditory alert and FCW events required an artificially high event rate. Given proper simulation conditions (e.g., realistic and varied collision event types, other traffic demanding the driver's attention), this approach can maintain experimental validity without sacrificing efficiency of data collection (Green, 2008). Therefore, the selected approach is one that balanced gathering enough response behavior while minimizing an artificially high number of FCW events.

Due to practical constraints, the experiment only used FCW events and alerts and did not include lane departure events or warnings. It is likely that transfer findings with the FCW alerts will apply to other auditory warnings.

To create a reasonable context in which participants encountered a different FCW system from which they were accustomed, the cover story for the experiment was one of testing how drivers handled various driving environments, tasks, and distractions when switching between vehicles. GMU only has one driving simulator so we used methods of changing the superficial features in the driving environment to make the supposed vehicle change salient to the participants. For example, one such change is the manufacturer "badge" emblem on the steering wheel (e.g., Ford then Toyota) and a different license plate. Further, participants were specifically told that they were now driving a different simulated vehicle and that they could expect its handling characteristics to be different (though the handling characteristics actually did not change).

Pilot testing results indicated that this manipulation was sufficient to induce the belief that the simulated vehicle had changed. A number of participants noted that the vehicle appeared to have different steering properties despite the fact that no changes to the steering gain were actually made. No adjustments were made to any of the other vehicle dynamics (i.e., braking dynamics were not changed).

3.1 Design

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.

The experimental design was a between-subjects manipulation of the test phase: Control (no alert change) versus treatment (alert change). Two auditory-only alerts, a light vehicle warning and a heavy vehicle warning were selected for use in this study. The alerts are described in detail in Section 3.2.4.

The alert pairings were counterbalanced so that half the participants experienced the light vehicle warning first and the other half experienced the heavy vehicle warning first. Detailed descriptions of these warnings are presented in Section 3.3.3. Table 1 outlines the alert conditions along with the number of participants in each cell.

Table 1. Summary of experiment conditions

Control	Treatment	Total
Light→Light n = 15	Light→Heavy n = 15	n = 30
Heavy→Heavy n = 15	Heavy→Light n = 15	n = 30
n = 30	n = 30	N = 60

Participants engaged in six 20-minute drives over three sessions (two drives per session). With training and post-experiment questionnaires, participants spent approximately four hours in the experiment. The first four drives constituted the learning phase and Drives 5 and 6 constituted the test phase (see Table 2 below).

Table 2. Summary of experiment phases

	Learning Phase		**Test Phase**
Day:	**1**	**2**	**3**
Activity:	Training + D_1+D_2	D_3+D_4	D_5+D_6 + Post
Duration:	1.5 hours	1 hour	1.5 hours

For all participants, the first session began with a training drive to allow participants to get accustomed to the simulator and the various distraction tasks they were asked to complete. The training drive lasted approximately ten minutes. Each subsequent drive lasted approximately 20 minutes (separated by a break within session). Each session was separated by one to three days. A total of six FCW events were presented across the four learning phase drives (Drives 1-4) with a minimum of one event in each drive. This approach allowed for some mitigation against expectation and anticipation of events in each drive. Vehicle "change" occurred in session 3. The sole FCW event in Drive 5 had the different auditory alert (treatment condition). Drive 6 had two FCW events, both with the changed alert (treatment condition). For the control condition, there was no change in auditory alert despite all other vehicle "changes."

Participants experienced all auditory signals (including environmental sounds) via laptop computer prior to participation to familiarize them with each sound, with the exception that participants only experienced the FCW alert that they would receive in the learning phase, so that participants in the treatment conditions (who receive a different FCW alert in the test phase) would not receive any pre-exposure to the alert that they would experience in the test phase. The meaning of each sound (e.g., cell phone ring tone) was defined.

3.2 Method

3.2.1 Participants

A total of 67 participants were enrolled in the investigation. Six discontinued the experiment due to simulator adaptation syndrome. One participant was screened out because of age outside of the target range. None failed to return for all three sessions. The target number of participants, 30 in the control and 30 in the treatment condition (60 total), was achieved. Of these, 28 were male (16 in the control group and 12 in the treatment group). Age ranged between 19 and 64 with a minimum of five years driving experience. The average age for the 60 participants with complete data sets was 27.5 with a range of 19 to 64, SD=9.1. The average driving experience was 10 years.

3.2.2 Scenarios

Three scenarios were used for the six drives, with participants driving through each scenario and back (two drives) in each session. Each drive was approximately 15 miles with a posted speed limit of 45 mph throughout, which participants were instructed to obey. Each drive took 20 minutes on average to drive. Each 20-minute drive 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 drive in a safe manner as they would normally during the drives.

The three scenarios had different superficial characteristics (e.g., buildings, trees, and layout), but they were matched in terms of key components of proportion rural highway and suburban/commercial, number of traffic control devices, number of turns, etc. All scenarios took place in dry, daylight conditions. Participants were guided through the drives via turn-by-turn guidance presented aurally and visually through an in-vehicle display. More details about the route guidance system are provided in the distraction task section below.

3.2.3 FCW events

The FCW events are based on those used in the Integrated Vehicle-Based Safety Systems (IVBSS) project. Figure 1, from the *IVBSS Human Factors and Driver-Vehicle Interface (DVI) Summary Report* (Green et al., 2008), diagrams each of the event scenarios. The following four event types were selected:

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.

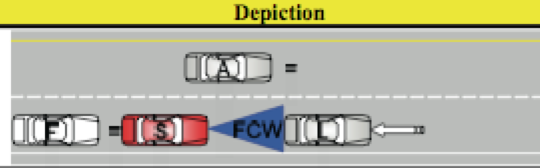
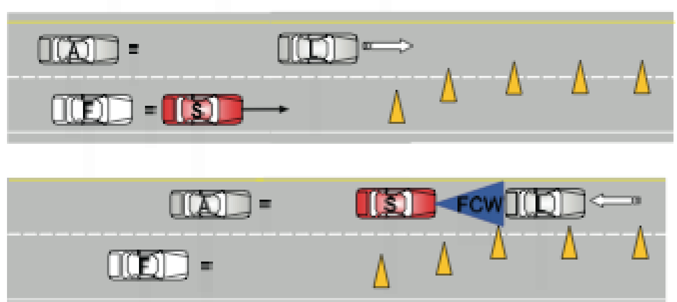
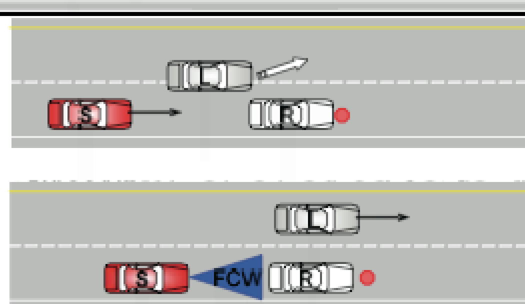
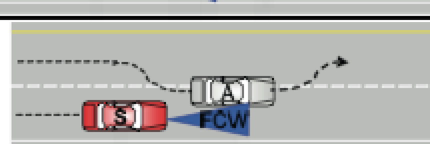
Scenario	Depiction
1-FCW Lead vehicle (L) suddenly decelerates.	
2-FCW Subject vehicle (S) changes lanes due to construction zone and lead vehicle suddenly decelerates to trigger FCW.	
3-FCW The lead vehicle changes lanes to reveal a stopped vehicle.	
4-FCW Cut-in.	

Figure 1. FCW event diagrams (from Green et al., 2008)

Mirrored versions (e.g., cut-in from right rather than left; work zone closes left lane rather than right) of each of the four event types above were used to generate the required nine FCW events across the six drives. In addition, different vehicle models were used for the surrounding vehicle types in each event (including LVs). The ninth event was the lead-vehicle sudden braking event with only changes to the surrounding vehicles. This event was selected due to its simplicity and fewest "tip-off" cues like a vehicle approaching from behind in the adjacent lane.

Corresponding non-events were also included throughout each scenario to help mitigate anticipation. For example, the subject vehicle drove through work zones (with barrels narrowing to one lane) where the lead vehicle did NOT brake suddenly as it did in the corresponding FCW event. Similarly, there were times when the LV changed lanes abruptly but no stalled vehicle was present. It is important to note that the purpose of this experiment did not include trust in system reliability so the simulated FCW system was 100% reliable.

In total participants experienced 36 planned events (9 collision events and 27 non-collision events). All participants received the 36 events in the same order thus all 9 of the FCW events were in the same order. The order of the 36 events was randomly assigned across drives with the following constraints: consecutive collision or non-collision events were not the same, 3 FCW events occurred each day, 4 collision and non-collision events occurred per drive, a lead vehicle event occurred on drive 1 of session 1, drive 3 of session 2, drive 1 of session 3, and drive 3 of session 3. The objective was to have all participants experience lead vehicle braking events for their three critical exposures to the warning (first exposure, pre-switch, and post-switch) because each different type of FCW event was expected to lead to slightly different patterns of responding, and using all lead vehicle deceleration events for analysis was expected to maximize the comparability of the data across participants and exposures. Events were not counterbalanced across participants because software limitations did not allow road objects (construction cones, stalled cars) to be dynamically placed into the roadway during runtime.

If a programmed lead vehicle deceleration event did not occur or failed to elicit a warning during one of the three critical event exposures (the first drive in the first session, last drive in the second session, the first drive in the third session, or the third drive of the last session) then the experimenter triggered a manual lead vehicle event later in the drive by pressing a button on his/her computer. Experimenters triggered the manual event when the participant was responding to a distraction task prompt (see Section 3.2.5) to maximize the likelihood that the participant would be distracted and would therefore experience the FCW alert. Whereas in the programmed lead vehicle deceleration events, vehicles were placed to the side of the participant's vehicle to discourage swerving, in the manual lead vehicle deceleration events there was no adjacent traffic. When the experimenter triggered the lead vehicle deceleration, the lead vehicle decelerated at the same rate as in the programmed events.

A pilot study was conducted to determine the alert timing algorithm. An algorithm was defined to issue an alert when time to collision with a lead vehicle is less than 1.5 seconds. The algorithm struck a balance so that alerts would not occur too early (i.e., before participants could identify a threat) or too late (i.e., when a collision is already unavoidable).

3.2.4 Alerts and other auditory signals

As part of Task 2, an inventory of in-vehicle sounds from a sample of production vehicles was created to include system warnings and alerts (FCW, LDW, parking aid, ACC), as well as more general sounds associated with or commonly found in the vehicle environment (seat belt warning, lights on, door open, cell phone, etc). Two auditory alerts were selected based on this inventory and those from the IVBSS study (Green et al., 2008): The IVBSS light vehicle warning and the IVBSS heavy truck warning. These warnings 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 completely different automobiles. Both FCW alerts were presented at 85 decibels (dB) against a background road noise average 62 dB (modulated by traffic present in the driving scene).

The IVBSS light vehicle warning was faster (shorter inter-pulse interval and shorter pulse duration) than the IVBSS heavy truck warning. Warnings with more pulses per second are perceived as more urgent (Campbell et al., 2007; Edworthy, Loxley, & Dennis, 1991; Green et al., 2008; Haas & Edworthy, 1996; Hellier, Edworthy, & Dennis, 1993; Marshall, Lee, & Austria, 2007; Patterson, 1982).

The IVBSS heavy warning incorporated a fundamental frequency of 600 Hz with one harmonic at 1800 Hz or two frequencies within a single burst, whereas the IVBSS light vehicle warning consisted of a fundamental frequency of 1500 Hz with five harmonics for a total of six frequencies with a single burst. Using multiple frequencies increases noticeability and perceived urgency (Campbell et al., 2007). Table 3 summarizes the characteristics of both alerts.

Table 3. FCW alert characteristics

Alert Characteristic	IVBSS Light Vehicle FCW Alert (Light)	IVBSS Heavy Vehicle FCW Alert (Heavy)
Tone	Abstract	Abstract
Frequency modulation	None	Two-tone
Frequencies	1500 Hz, 4500 Hz, 7500 Hz, 10500 Hz, 16500 Hz, 19500 Hz	600 Hz, 1800 Hz
Pulse duration	50 ms	320 ms
Burst duration	700 ms	320 ms
Interburst interval	10 ms	2 ms
Interpulse interval	30 ms	0 ms
Onset ramp	5 ms	none
Offset ramp	20 ms	none
Number of bursts	2	3
Pulses per burst	7	4 (beginning with 1800 Hz then pulses of both frequencies)
Warning duration	1300 ms	1300 ms

Alerts were presented through five speakers simultaneously: three in front of the driver and two behind. The simulated FCW system had an auditory alert only. No visual component was involved. Whereas all FCW systems currently on the market use both visual and auditory components to their alerts, using different visual indicators along with the different auditory alerts would have presented a confound, and thus the decision was made to include only the auditory component.

It was deemed important to have additional acoustic information present in the driving environment so that the FCW alerts would not be the only significant change in acoustic information, and therefore highly conspicuous. The following sounds were included during each drive:

- Road noise – became louder as speed increased, 62 dB while at the prescribed speed of 40 mph.
- Distraction task auditory feedback – Each button "press" resulted in auditory feedback in the form of a tone or click (dissimilar to other sounds in the scenario). The feedback sound also provided the participant an indication that his or her button press had been registered.
- Cell phone ring that participants were required to silence with the press of a button on the touch screen (i.e., no conversation task) – An existing cell phone ring tone was used and was presented at 70 dB. The ring tone was a simple, non-melodic tone, with a fundamental frequency of 600 Hz. The cell phone ring occurred three or four times per drive and did not coincide with the distraction task, FCW event, or route guidance instructions.
- Route guidance and verbal instructions (e.g., BEEP, "Right turn ahead") – presented at 70 dB.
- Siren as police cruiser passed by – infrequent (once per drive); presented at approximately 70 dB.
- Other vehicles honking – infrequent (i.e., one or two times per drive); presented at approximately 70 dB.
- Check engine alert – in addition to the check engine light in the dashboard alert cluster, an auditory chime was sounded to alert the participant to check engine.

These sounds were designed to provide a rich acoustic environment, with some having abrupt onsets like the FCW alerts and cell phone ring and others being more continuous (e.g., road noise and distraction task sounds). The auditory FCW alerts and other key sounds presented within the vehicle are provided for playback as .wav sound files in Table 4.

Table 4. Sounds presented during simulated drives*

FCW Alerts	Other In-Vehicle Sounds	External Sounds
Light vehicle alert.wav	Phone ring.wav	Car horn.wav
Heavy vehicle alert.wav	Oil Warning.wav	Police siren.wav

*Sounds are playable in Microsoft Word; double-click a sound icon in the table to play it

3.2.5 Distraction task

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 touch screen device was a 7-inch LCD display that was located to the right of the steering wheel, at the top of the center stack, which is a common location for touch screen LCD displays in vehicles. Participants were required to visually attend to the display when performing the task using the touch screen interface and therefore to glance away from the road.

The distraction task was the same for all drives, despite the “vehicle” changes implemented for Drives 5 and 6. Participants were told that the vehicles had the same aftermarket touch screen system. Also, the majority of the trials in the distraction task were not paired with FCW events in order to reduce the likelihood that participants would learn to anticipate a FCW event whenever they began to perform the distraction task.

The distraction task was implemented within the simulator’s current software and met the desired distraction task characteristics. The task is a variation of the “Simon” task in which participants listen to a sequence of directions presented aurally (e.g., “Up, Down, Left, Left, Up”) and then are required to repeat the sequence with button presses. See Figure 2 below.

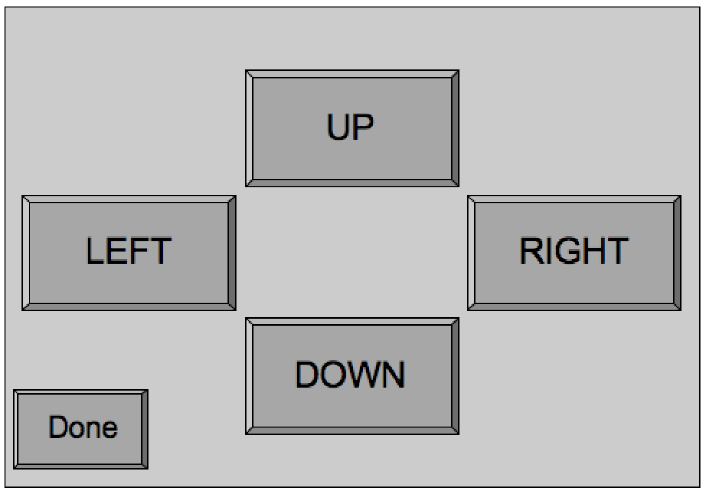


Figure 2. Simon task touch screen interface

After participants were cued to begin the Simon task, they had ten seconds to complete the task. Fifteen seconds after the Simon task was completed, participants were cued to begin the next Simon trial. This pattern continued throughout each drive. The Simon task imposed a memory load and required a sequence of button presses. Participants also felt an urgency to complete each trial immediately so they would not forget the cued sequence. It also has the advantage of providing additional acoustic information in the environment. Another benefit of this task was that it imposed a relatively continuous demand for attention and therefore kept participants distracted for a considerable proportion of time. Participants were given the following instructions on how to perform the Simon task:

> “*You will hear a sequence of button directions. For example, you could hear “UP, LEFT, UP, RIGHT.” After the sequence is complete, you will need to replicate the same sequence of button directions by pressing the appropriate buttons in order on the touch screen, followed by the “Done” button. Please answer as quickly as possible while maintaining accuracy. Your task performance will be measured based on time to replicate the sequence and accuracy.”*

3.2.6 Route guidance instructions

Route guidance instructions were presented aurally and visually. As participants approached a turn, they heard a verbal instruction (female voice). For example, “Turn right at the next light.”

Visual instructions (see Figure 3 below) showing an arrow indicating left or right turn was inset in the lower left or right corner of the forward driving display, respectively. The location in the forward display mimicked a head-up display presentation style.

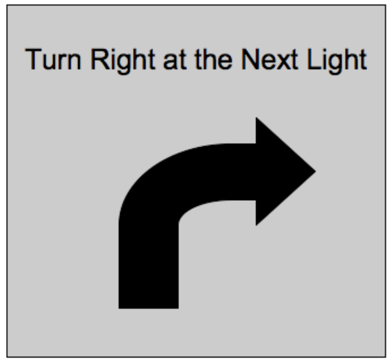


Figure 3. Route guidance visual instruction example

Only one auditory guidance instruction per turn was provided. Each trip could have up to two turns. The instructions occurred approximately 150 meters (nearly 500 ft) prior to each turn. The visual instruction remained through each turn, disappearing immediately following the turn. If participants missed a turn, the experimenter would direct them to turn around and resume the route.

3.2.7 Familiarization protocol

In an experiment such as this one, the process of familiarizing participants with the sounds and tasks was deemed critical. The following protocol was used to expose participants to the sounds and tasks prior to the experimental drives.

After participants provided informed consent, the experimenter provided an overview of the driving and distraction tasks, and explained that participants would complete multiple drives over three days, and that they would change vehicles on the third day. 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. In particular, participants were told that the experiment was aimed at examining how driving performance might change as a function of driving different vehicles. As such, they were instructed that they would be "driving" the simulator as a Ford Focus at first and then it would be changed to a similarly equipped Toyota Corolla.

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.).

While seated in the simulator, participants were introduced to the sounds they would encounter while driving. These sounds were presented statically—while not driving the simulator. Each sound was introduced by the experimenter and then played twice (a few seconds each). Participants were then played the following six sounds and asked to identify their source:

- FCW alert: *only* the alert that the particular participant would experience in the first phase of the study (i.e., either light or heavy, but not both)
- Route guidance verbal instructions (i.e., "Turn left (right) at next intersection")
- Horn honk
- Police siren
- Cell phone ring
- System alert chime (check engine, oil warning)

Any participant who could not correctly identify all six sounds repeated the sound training until all six sounds could be identified. After completing the sound exposure, participants were introduced to the subsidiary Simon distraction task. After instructions, participants were provided with approximately two minutes (roughly eight trials) of Simon task practice while seated in the simulator, but not while driving.

Once participants had been trained on the Simon task they were introduced to the driving task, including the route guidance system. The turn-by-turn and auditory instructions were demonstrated prior to the familiarization drives. Participants took three familiarization drives, each of which lasted approximately four minutes. The first familiarization drive consisted of driving the simulator and following the lead vehicle and route guidance instructions. The second familiarization drive added the Simon task to the first drive. The third familiarization drive added cell phone rings, honks, and sirens to the second drive.

3.2.8 Sound comparison task

Following the final drive of the experiment all participants performed a sound comparison task to gauge the perceived similarity of each of the warning tones, variations of the warning tones and other sounds used in the simulation environment. The comparisons can be split into two categories, those made between warning tones and warning tone variations (39 comparisons total,) and those made between warning tones and selected environmental sounds (12 comparisons total).

Comparisons between warning tones and warning tone variations were made in order to assess the extent to which tone properties such as frequency spectra (fundamental frequency/harmonics) and temporal dynamics (number of tone bursts/burst duration) contribute to perceived tone similarity. Comparisons between warning tones and the selected environmental sounds (siren, car horn, oil warning, phone ring) were made in order to assess the likelihood that these environmental sounds may have been mistaken for the warnings tones presented during the experimental drives.

The matrix of comparisons presented in Appendix A3 illustrates the sound pairs that participants rated. The value in each cell signifies how many times each comparison was made. The majority of comparisons were given once, but a subset of critical comparisons was given twice to assess comparison reliability. Sound comparison data from 15 participants were removed from analysis due to poor assessed reliability in the critical comparisons.

Sound stimuli were presented using a Visual Basic 2008 program written specifically for this experiment. The order of the 51 comparisons was shuffled (randomized without replacement) between participants. Likewise, the presentation order of the two stimuli in each comparison pair was randomized for each trial. Stimuli, normalized to a sound pressure level of 70 dB, were presented via headphones (Sennheiser HD 280 Pro). After each sound pair presentation, participants were instructed to rate their similarity by adjusting a slider ranging from "Very Dissimilar" to "Very Similar" (see Figure 4). These rankings were coded on a numeric scale ranging from 1 (very dissimilar) to 7 (very similar.)

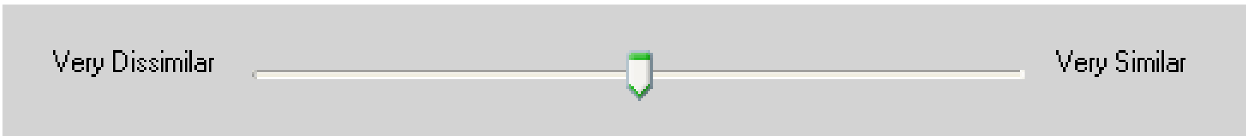


Figure 4. Slider used for similarity ratings

3.3 Findings

3.3.1 Event exposure

As described in the Method section, a number of different forward collision scenarios were developed and utilized in this negative transfer investigation. Although each participant was exposed to nine potential forward collision events, only their 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) were used for analyses. Note that 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. Also note that because some potential forward collision events did not result in FCW alert actuation, participants did not necessarily receive each of the three key exposure events during the same drives or with identical event types. Table 5 illustrates the frequency of exposure to each event type by the time period during which it was experienced for all participants.

Table 5. Frequency of exposure by event type – overall

Event Type	First exposure (learning 1)	Pre-switch exposure (learning 2)	Post-switch exposure (test)
Lead vehicle decelerates	39	28	28
Lead vehicle decelerates (manually triggered)	12	15	16
Cut in	0	5	12
Construction zone lead vehicle decelerates	8	0	3
Stopped-vehicle reveal	0	12	0
Total	**59**	**60**	**59**

The table shows that the majority of exposure events were lead vehicle braking events. Specifically, 149 of the 178 (84%) exposure events were lead vehicle braking events (lead vehicle, manually triggered lead vehicle, and work zone). FCW events were programmed so that all three critical exposures to the FCW alerts would be lead vehicle deceleration events, but because not all events resulted in FCW alerts (as described above), 16 percent of critical exposure events were of other types. Although it was expected there would be a total of 180 of these events (3 x 60 subjects), two observations were missing. These missing observations were due to the fact that two participants did not have all three event types (a first, pre-switch, and post-switch exposure to the collision warning). Specifically, one subject's first exposure to a warning was also his or her pre-switch exposure. For the current purposes we classified this observation as a pre-switch exposure. Similarly, one subject did not have a post-switch exposure to the collision warning and therefore could not be included in subsequent analyses. Frequencies of exposure by event type were comparable in control and treatment conditions.

An additional issue resulting in some level of data loss was the observation that 12 participants received the "reveal" event during the pre-switch time periods. As discussed in Section 3.5, this exposure event led a number of participants to respond to the event by steering around the stalled vehicle rather than braking. This collision avoidance response, though valid, could not be equated with brake response behavior and therefore could not be included in subsequent analyses.

A breakdown of the frequency of event type by exposure (first, pre-switch, post-switch) as a function of the session of drive (session 1, 2, or 3) is presented in Appendix A4. Note that the first exposure to a warning occurred for all participant during drives in session 1; the pre-switch exposure to a warning, with two exceptions, occurred in session 2; and the post-switch exposure occurred in session 3.

3.3.2 Warning response

3.3.2.1 Warning response profiles

Brake response time (RT) was the primary dependent measure. It was calculated as the time between the onset of the collision warning and the first brake depression (i.e., the moment when the participant began to depress the brake). Participants were instructed to drive in a safe manner as they would normally, and care was taken to recruit participants from a wide age range. The safe driving behaviors of participants were indicative of the validity of the simulator environment, to the extent that participants appeared to be highly motivated to drive attentively and avoid collisions. This safe driving, however, also resulted in many trials in which participants reacted to an emerging threat before the FCW alert was presented, and these trials were discarded.

As previously mentioned, 84 percent of the pre-switch and post-switch exposure events were lead vehicle braking events (lead vehicle, manually triggered lead vehicle, and work zone). These event types were analyzed together in order to maximize statistical power. Trials consisting of brake RT less than 200 ms were eliminated from the analyses as these indicated that participants may have begun to respond by releasing the accelerator prior to the onset of the warning. The 200 ms criterion was selected because due to perceptual and motor limitations, it is unlikely that a participant could have responded to an alert in less than 200 ms (in other words, it is likely that the participant began to react before the alert was issued). Trials in which participants responded to the alert without braking (e.g., swerved) were also removed from analysis. Use of these criteria resulted in 86 remaining observations (48% of the total of 178 events).

Pearson's chi-square tests were conducted to determine if there were any trends in the distribution of discarded trials that might have influenced the data set. A separate chi-square test was performed for each of the three critical exposures (first, pre-switch, post-switch), comparing the number of valid and discarded trials in each of the four experimental conditions. Specifically, the test looked at whether there were any systematic differences between the warning conditions for trials that were considered valid (Brake RT > 200 ms) versus those that were discarded. Results indicated that the distribution of discarded trials was fairly consistent between experimental groups for each of the three critical exposures (first: $\chi^2=1.17$, $df=3$, $p=.76$; pre-switch: $\chi^2=1.70$, $df=3$, $p=.64$; post-switch: $\chi^2=2.38$, $df=3$, $p=.50$).

Brake profiles from these observations as a function of time of exposure are illustrated graphically in Figure 5, where each line represents the mean brake input for participants in each experimental condition. The figure shows that, compared to other conditions, participants in the Heavy→Light condition had slower and more prolonged responses to the alert at the post-switch exposure in response to when they first hear the alternate warning. Pedal input profiles, including both brake response and accelerator input for each of the three alert exposures, separated by warning group can be found in Appendix A5.

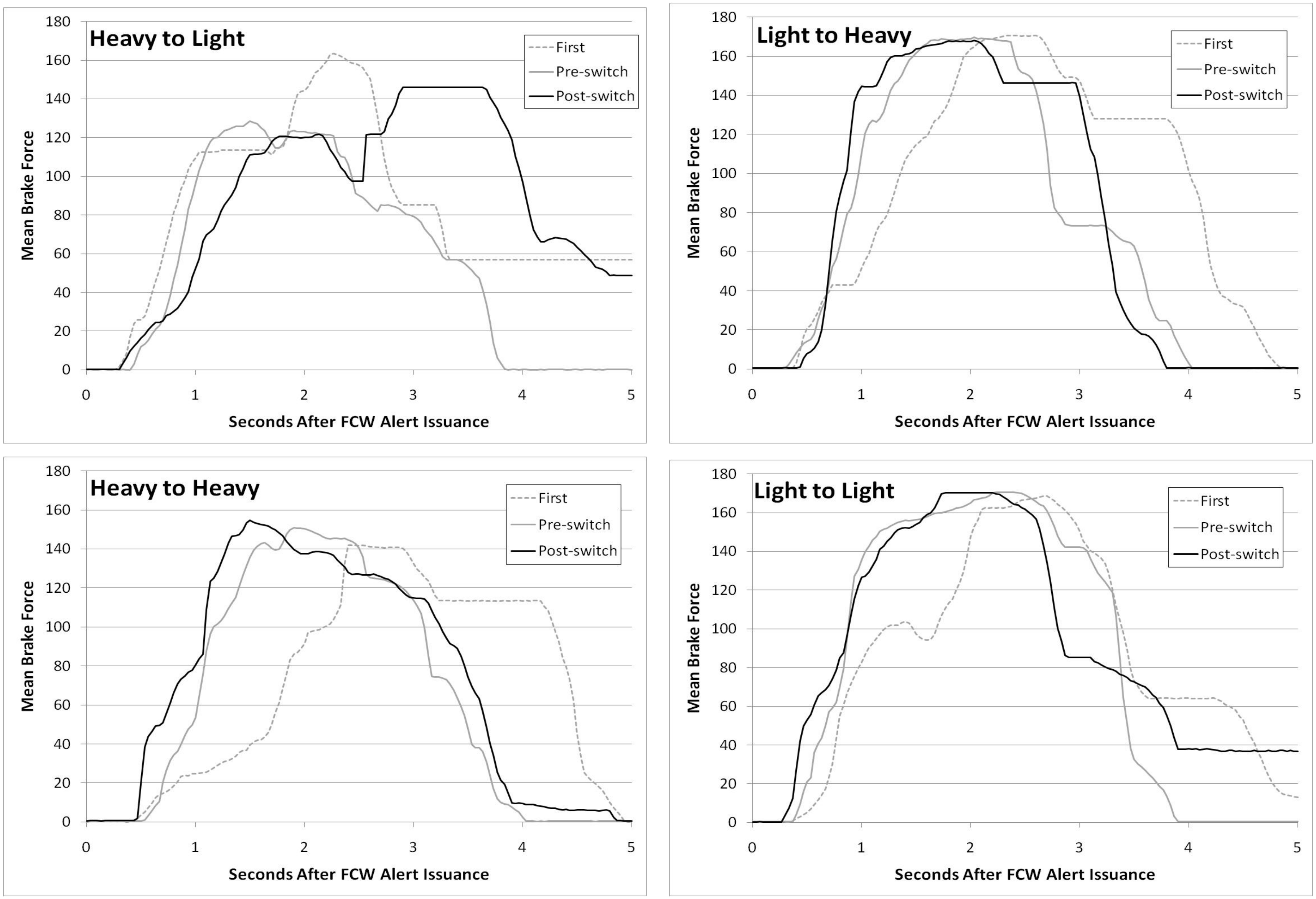


Figure 5. Mean brake pedal input profiles for each experimental condition as a function of FCW event exposure

3.3.2.2 Brake response time (RT)

Brake RT data were subjected to a series of statistical analyses to answer specific questions. The first analysis was a one-way ANOVA by warning sound condition for post-switch brake RT. As illustrated in Figure 6, brake RT was slowest in the Heavy→Light condition, where participants were exposed to the heavy warning in sessions 1 & 2 and then the warning switched to the light warning for the post-switch event in session 3.

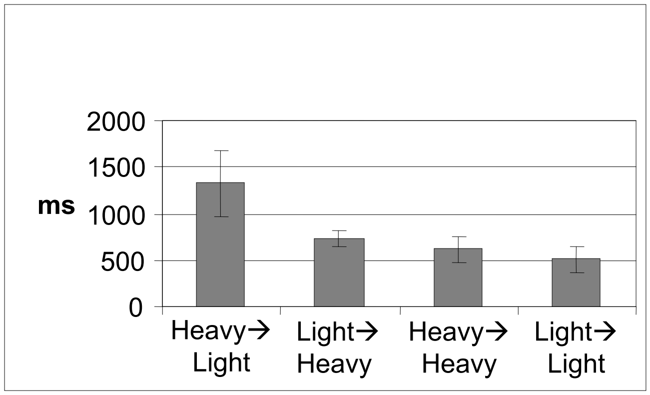


Figure 6. Post-switch brake RT as a function of warning condition group

Note that average brake RTs in the Heavy→Light condition are twice as long as the control conditions. Average brake RT in the Light→Heavy condition is also somewhat slower relative to the two control conditions, but not as dramatically so. This analysis provides some potential evidence for a negative transfer effect. However, a more robust analysis takes into account the change in brake RT as a function of warning condition and exposure (first, pre-switch, post-switch). This analysis is discussed next.

Figure 7 shows participants' mean brake RTs plotted by experimental condition and forward collision event exposure (first, pre-switch, post-switch). As with all other data used in these analyses, trials in which brake RTs were less than 200 ms are not included in the figure data. Brake RTs for participants in the Light→Heavy condition were approximately the same in the pre-switch and post-switch exposures, whereas participants in the two control conditions (Heavy→Heavy and Light→Light) generally responded faster with each subsequent exposure, indicating a benefit of repeated exposures. Participants in the treatment conditions mirrored the improved performance of control group participants, improving their reaction times from the first exposure to the pre-switch exposure (as expected since participants in all conditions the same alert in both exposures), but then diverging from the control groups' patterns of improvement once the alert was switched.

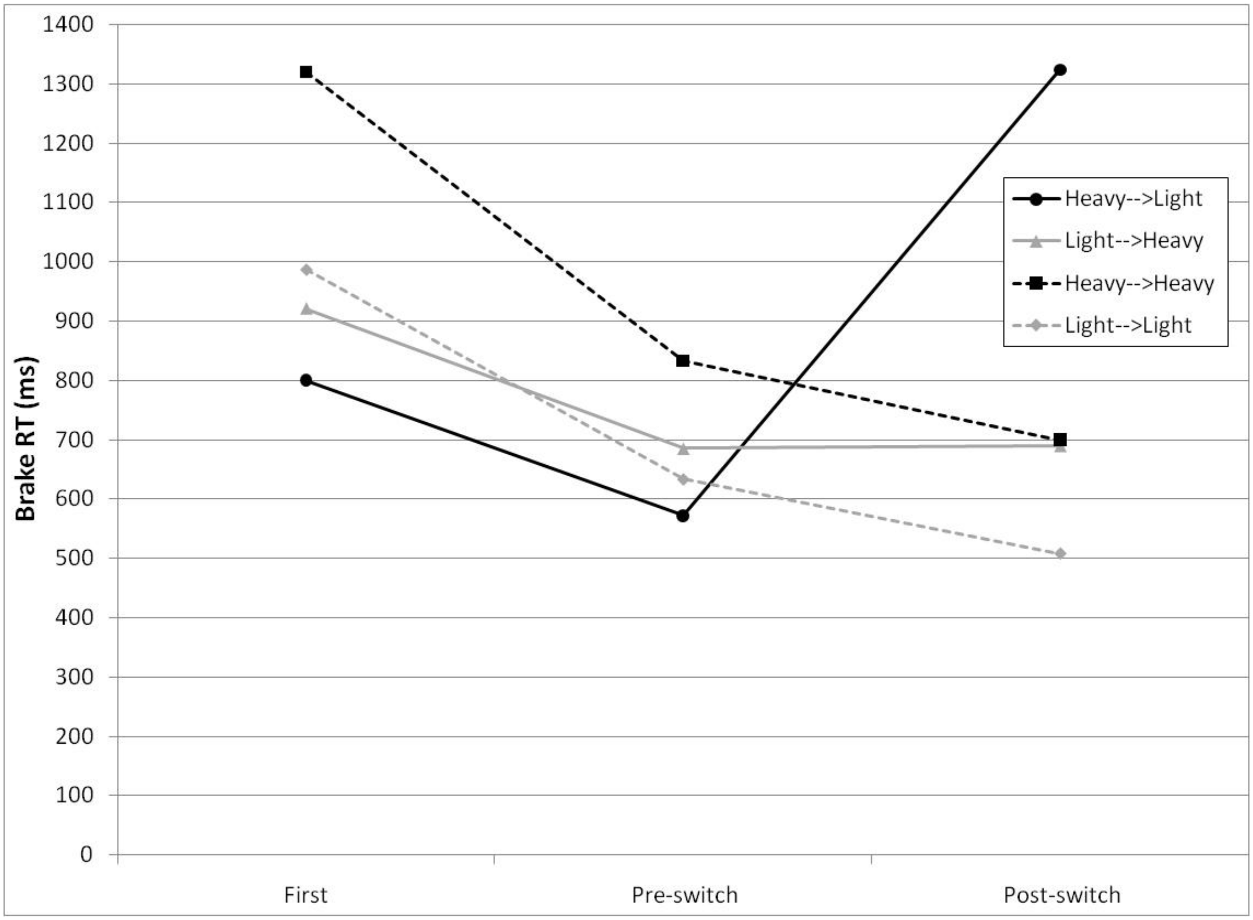


Figure 7. Brake RT as a function of warning condition and exposure time

A mixed repeated measure analysis of brake RT was conducted with warning condition as a between-subjects variable and exposure (pre- and post-switch) as a within-subjects variable. A marginally significant main effect for exposure was observed, $F(1,9)=4.37$, $p=.066$, $\eta^2=.33$. However, this main effect was subsumed by a significant interaction between warning condition and exposure, $F(3,9)=9.62$, $p=.004$, $\eta^2=.76$. Participants exposed to the switch from the heavy warning sound to the light warning sound 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. As will be discussed more thoroughly in a subsequent section, this observed interaction provides relatively strong evidence for the potential existence of a negative transfer effect in at least some conditions.

3.3.2.3 Other measures

Additional measures were examined, namely accelerator release time, temporal headway, and time to collision. However, these metrics provided little information beyond that provided by brake RT. Accelerator release time occurred earlier than brake RT and therefore resulted in many more observations having to be discarded because drivers may have released the accelerator due to anticipation of a developing event rather than the occurrence of a warning. Further, in the current paradigm, presentation of the collision warning was triggered based upon a preset threshold in time to collision and headway distance. Therefore, these measures were not informative for the current purposes of examining warning response.

Another indicator of warning response that was analyzed was the momentum of response. Momentum of response was defined as the time between accelerator release and first brake input. This delta score was analyzed in a one-way ANOVA between warning group conditions. Momentum of response did not differ significantly between the four warning condition groups, $F(3,49)=.72$, $p=.52$.

There were relatively few crashes, with a total of 11 crashes for all participants across all drives. Crash frequency as a function of experimental condition (treatment or control) is provided in Table 6. Crashes were evenly distributed across participants in treatment and control conditions.

Table 6. Frequency of crashes for switch and no-switch conditions by exposure

Exposure	**Condition**	
	Control	Treatment
First	3	3
Pre-switch	0	0
Post-switch	2	3

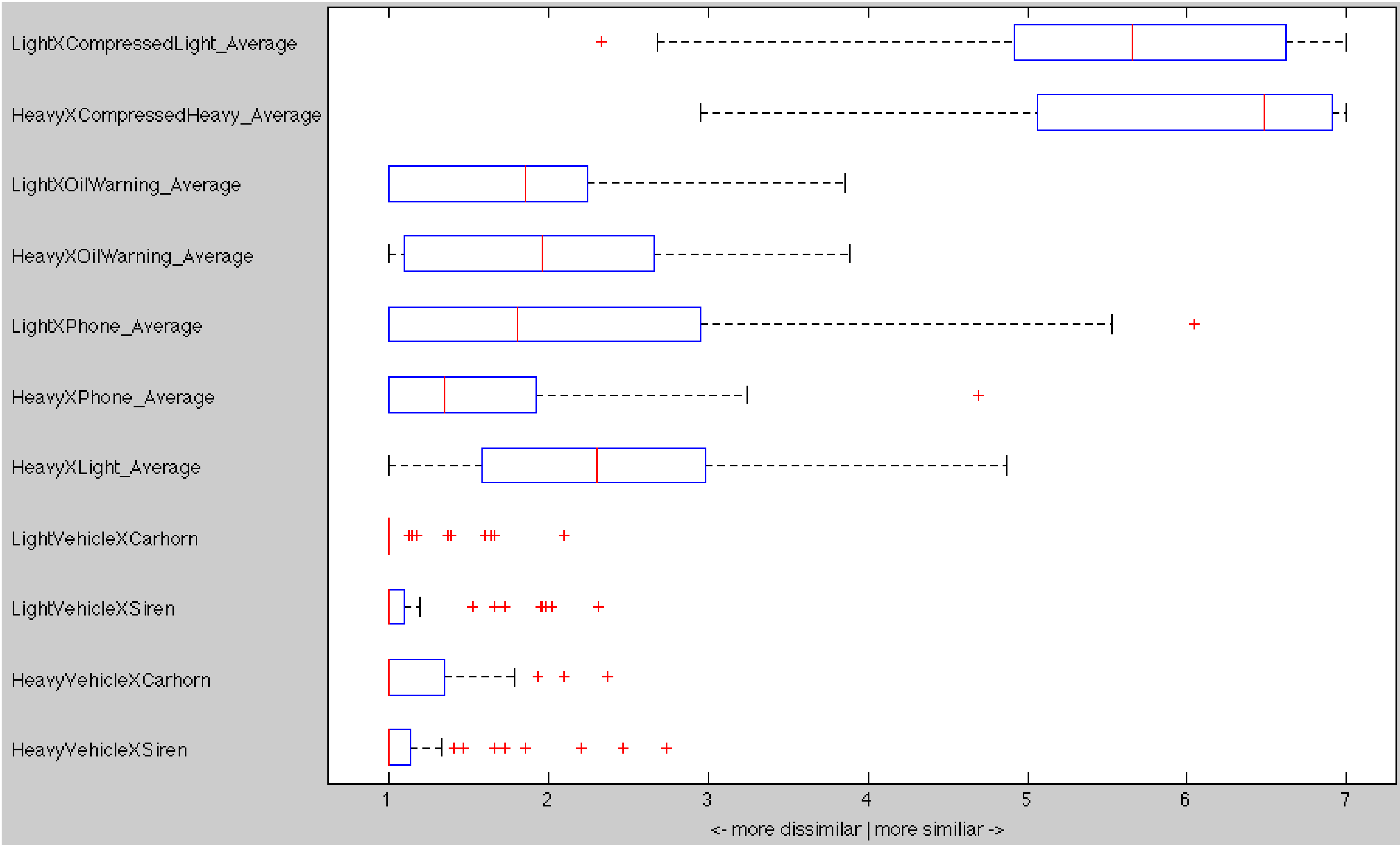
3.3.2.4 Sound comparison results

As previously described in the Method section, on Day 3, following completion of all the 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 purpose of this evaluation was to determine whether various sounds used in this study could have sounded similar to participants, which could potentially cause confusion or delayed responses to alerts. Mean similarity ratings (and standard deviations) for pairs of sounds are provided in Table 7. Note that multiple variations of the original warnings used in the experiment were used as a means of examining the sound characteristics that participants were using to make similarity ratings. Additionally, two pairings of the original warning sounds (light warning compared to the heavy warning) were presented among all other comparisons as a reliability check. Data from 15 participants were eliminated from the final analysis due to low reliability (highly different ratings across multiple presentations). Details of the reliability analysis are in Appendix A6.

Table 7. Mean similarity ratings (and standard deviations) for each sound rating

	Heavy Warning	Light Warning	Heavy 100 ms burst	Heavy, F0 1000Hz	Heavy, F0 1500Hz	Heavy, 8bit, 5000Hz	Light, 150ms burst	Light, 300ms burst
Heavy Warning								
Light Warning	2.35 (0.98)							
Heavy, 100ms burst	4.44 (1.65)	3.08 (1.53)						
Heavy, F0 1000 Hz	5.01 (1.34)	2.80 (1.34)	2.75 (1.24)					
Heavy, F0 1500 Hz	4.69 (1.48)	3.48 (1.56)	2.99 (1.47)	4.87 (1.46)				
Heavy, 8 bit, 5000 Hz	5.86 (1.29)	2.49 (1.18)	3.57 (1.70)	4.61 (1.69)	4.54 (1.56)			
Light, 150ms burst	2.49 (1.13)	4.31 (1.64)	3.00 (1.46)	2.52 (1.16)	3.40 (1.66)	2.46 (1.43)		
Light, 300ms burst	2.83 (1.49)	3.58 (1.57)	2.30 (1.35)	2.56 (1.07)	3.24 (1.51)	2.62 (1.38)	4.51 (1.66)	
Light, 8 bit, 5000 Hz	2.41 (1.23)	5.47 (1.26)	3.61 (1.71)	2.56 (1.07)	2.95 (1.40)	2.94 (1.32)	3.45 (1.58)	2.84 (1.53)

A box plot of individual warning comparisons is shown in Figure 8. Each rectangular box represents the interquartile range of a sound's ratings (i.e., the left edge of the box represents the 25th percentile rating, the middle line represents the 50th percentile rating, and the right side of the box represents the 75th percentile rating). As was expected, participants rated each of the warning sounds to be most similar to variants of the warning sound. Variants were similar sounds shifted along key dimensions such as fundamental frequency, pulse rate, or compression, but that otherwise maintained the primary contour of the warning sound. Median similarity ratings for the original warning sound and compressed variations of the original sound were generally quite high. This provides some support for the feasibility of using a variety of acoustic quality versions of the same sound without great concern for performance consequences. Currently, actual warning sound acoustic quality depends on automotive equipage issues. Some manufacturers have audio systems capable of providing relatively high acoustic quality sound (i.e., 32 bit, 44,100 Hz sampling rate) while others rely on equipment only capable of presenting highly compressed (i.e., 8 bit, 1600 Hz sampling rate) versions of the sounds. Results of the present sound comparison test indicate that, for a given acoustic warning, variations in sound quality across at least the levels examined in the current investigation result in highly similar ratings and are not likely to result in differences in warning response, though this is an issue that could be examined in future investigations.



Note, n=45 because 15 of the 60 participants were eliminated due to poor assessed reliability

Figure 8. Box plot of the sound comparison results

Of particular interest in the current investigation were the similarity ratings between the warning sounds used and other sounds presented in the scenario. As can be seen in Figure 9, due to the sound characteristics of the individual warnings, the light warning and heavy warning sounds more closely resembled some distracter environmental sounds than others. For example, the light and heavy warnings sounds were both rated as somewhat similar to the other sounds utilized in the current investigation, but the similarity rating between the light warning sound and the phone ring was higher than the rating between heavy warning sound and the phone ring. Conversely, similarity ratings were slightly higher for the heavy warning sound and the oil warning than they were for the light warning sound and the oil warning. These relationships are presented graphically in Figure 9 and Table 8.

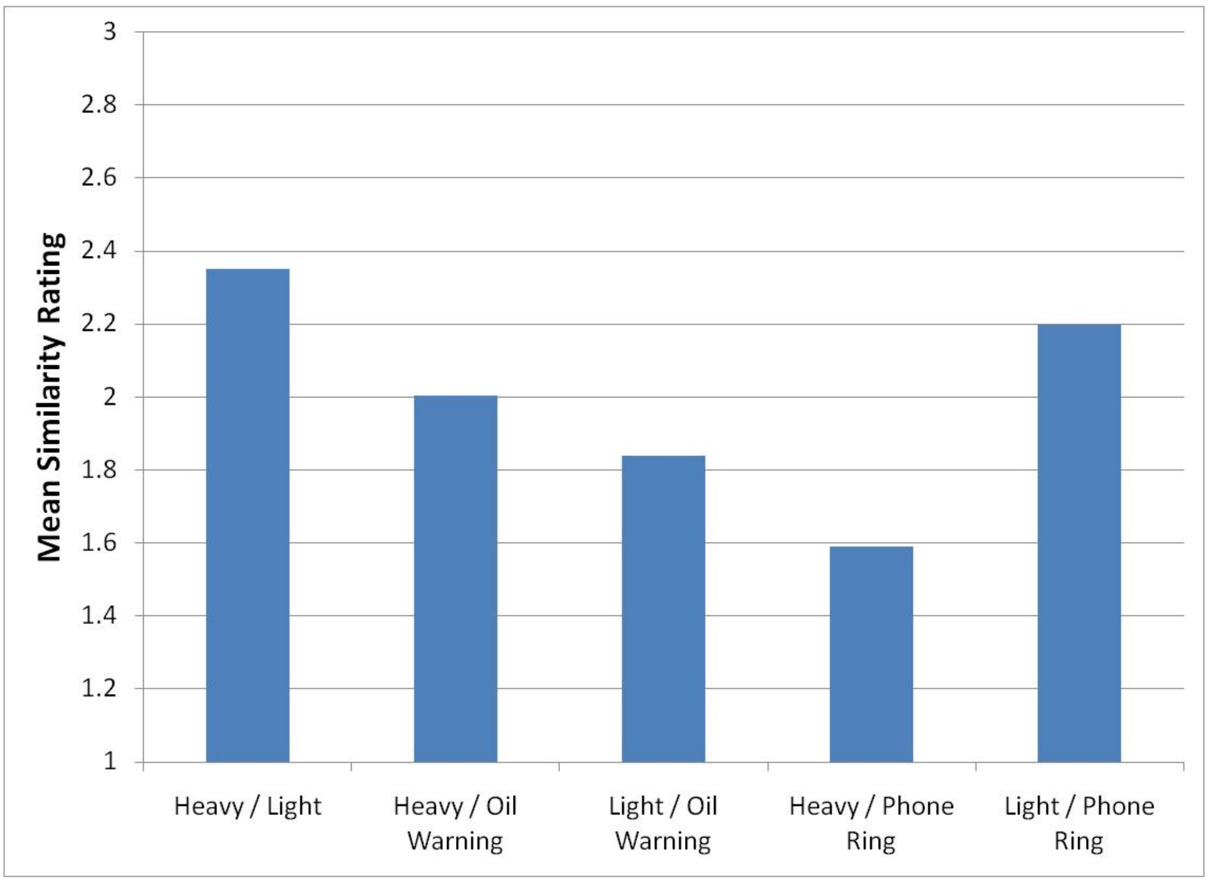


Figure 9. Similarity ratings for key sound comparisons (n=45)

Table 8. Mean similarity ratings for each of the warnings paired with other key environmental sounds

	Heavy warning	Light warning
Light warning	2.3483	
Siren	1.1927	1.1856
Car horn	1.1953	1.0931
Oil warning	2.0028	1.8403
Phone ring	1.5902	2.1992
Compressed version of same sound	5.8629	5.4731

Several of the rankings are not normally distributed. Differences between selected pairs were examined using the non-parametric equivalent of the paired samples t-test, the Wilcoxon signed ranked test. To consider whether the heavy and light warnings differed with respect to confusion with environmental sounds such as the phone ring and oil warning, heavy/oil warning was compared with light/oil warning and heavy/phone ring with light/phone ring. Results of the Wilcoxon test revealed that while the similarity ratings between the heavy and light warning sounds were significantly different from the ratings between the heavy warning and the phone ring, $Z=-4.37$, $p=.001$, the similarity ratings between the heavy and light warnings did not differ from ratings between the light warning and the phone ring, $Z=-1.03$, $p=.31$.

In other words, the average similarity rating between the two warnings was not significantly different from the similarity rating between the light warning and the phone. The similarity rating between the two warnings was however significantly different from similarity rating between the heavy warning and phone. This indicates that participants may have been more likely to confuse the phone ring with the light warning than with the heavy warning, providing a potential explanation for the negative transfer effects obtained for participants who were adapted to the heavy warning sound on the first two days of drives and then later received the light warning in the switch condition in session 3.

3.4 Assessment of methodology

3.4.1 Strengths

The simulated environment appeared to yield high external validity. The simulator was immersive and the motion base provided the necessary vestibular cues to provide a more realistic experience. All six drives participants completed were unique in design and the order of collision events and non-collision events. The three-day, three-session procedure provided a more realistic approximation of how learning occurs in the driving environment than a one-day “cram session.”

There was a low base rate of FCW events with a mixture of non-events to reduce expectancy of a FCW event. The use of four distinct types of FCW events also helped reduce expectancy and provided more realism. When subjects were asked if they knew what they thought the experiment was about during debriefing, only eight out of 60 participants thought that the experiment was examining collision warnings. This suggests that the experiment method was generally effective in obscuring the experiment's focus on FCW alerts.

The two FCW alerts (light and heavy) were within recommended guidelines on warning design (e.g., Campbell et al., 2007) and have been used effectively in past studies. Also, since they were distinct from one another and from other sounds in the vehicle (as evidenced by the sound comparison study), negative transfer effects can be attributed to prior learning interfering with responses to a new warning, not due to similarities between the warnings and other in-vehicle sounds used in the experiment.

The Simon Task created an additional workload for participants, requiring visual, manual, and cognitive involvement, yet it was not so difficult that participants became quickly fatigued, nor did it have a substantially adverse effect on driving performance during baseline (no FCW event) conditions.

3.4.2 Limitations

This experiment did not include a "true baseline" condition in which participants would experience the same driving scenarios and FCW events as other participants, but with no FCW alerts. The inclusion of this condition, while not within the scope of the present effort, would have provided a baseline measure of collision event response against which FCW alert responses could be compared. Lacking this condition, it is not clear how much of an effect the FCW alerts actually had on participants' collision avoidance responses (compared to no alert). This limitation, however, does not weaken any of the findings from comparisons between the two alerts, including the finding of negative transfer.

The four different collision events used in this experiment were not evaluated for comparability (e.g., speed of event detection, perceived threat level, perceived appropriate response). As a result, the ability to compare responses across these events was limited. The experiment design compensated for this limitation by attempting to use the same event type (lead vehicle braking event) for the three critical exposures (first, pre-switch, and post-switch), but this was not always the case because some FCW events did not result in FCW alerts (e.g., because the participant braked in anticipation of a developing threat) or braking responses (e.g., because the participant swerved rather than braked). FCW alerts were based upon headway and time-to-collision to the lead vehicle for the sake of realism. Sounding the warning sooner would have resulted in a larger quantity of usable data (i.e., fewer trials discarded due to participant anticipation of a developing threat), but also would have resulted in an early warning that would be less likely to be associated with the threat of collision and disregarded by the driver. Participants' conservative and safe driving behavior is a good indication of the validity of the simulator methodology, but they seemed to drive quite conservatively and responded to a number of events before the threshold for the collision warning was met. This led to a number of instances where no warning sounded.

This experiment involved the tradeoff of external validity and experimental control. By striving for a "naturalistic" experience in the simulated environment the experiment encountered unexpected difficulties in achieving the experimental design (i.e., identical first, pre-,and post-switch collision events) and analyzing the data (e.g., unbalanced designs, comparing behavioral data across different event types, different exposure rates due to failed collision warnings).

While the Simon Task did successfully distract participants, there was evidence to suggest that participants became more proficient in this task over time. As a result this task may have been less distracting with time. Also, the Simon task prompts were not synchronized with the onset of collision events, thus there was no guarantee that participants were actually distracted (visually or cognitively) at the onset of a collision event. The Simon task prompts occurred frequently throughout the drive, but their distraction was not verified.

While there was some traffic from non-events, however, there was not realistic traffic in this scenario. Non-event traffic would drive 15-20 miles per hour faster than the subject vehicle, and many participants noted this oddity.

Because of the way that the forward vehicle algorithm worked (lead vehicle stopped matching subject vehicle speed after around 45mph) the following distance was not held constant at speeds over 45 mph.

4 ACWS Status Display Comprehension Experiment

This experiment was focused on investigating whether people were able to identify and comprehends status displays for a variety of ACWS. Note that unlike the first experiment, this experiment dealt with drive recognition and comprehension of safety system information and not responses to imminent crash warning alerts. 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. The four key systems investigated were forward collision warning (FCW), lane departure warning (LDW), blind spot warning (BSW), and adaptive cruise control (ACC). 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-startup). In addition, a subset of participants was given short versions of vehicle's operational manual to read before the session, which may or may not have been the actual vehicle which they were presented. Participants' responses were analyzed for comprehension of system presence/status, decision time, and location accuracy. Participants also rated their confidence in their answers. Design, results, and discussion are presented in detail below.

4.1 Method

4.1.1 Design

Participants in this experiment 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, where 1 equals no confidence and 10 equals complete confidence) 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 that he or she sought the information that was used to determine whether a particular system was present or to determine its status.

Formally, the experiment was a fully crossed three-factor experimental design, with the following factors:

- *Vehicle (between-subjects: 3 levels)*: Each participant viewed the interior of one of three vehicles: 2010 Infiniti FX 35, 2010 Buick Lucerne, or 2010 Volvo S80. These particular vehicles were selected in part because they used very different display strategies from each other (e.g., icons, text, acronyms). Also, certain packages of these vehicles ranged in the number of safety systems of interest, with one vehicle having the fewest (two systems) and another vehicle being the most (four systems). All three of these vehicles had option packages that included ACWSs that were not present in the vehicles' base models.
- *Owner's manual familiarity (between-subjects: 3 levels)*: 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. Participants who read a "different" manual were equally likely to receive either of the other two vehicle's manuals.
- *Scenario (within-subjects 3 levels)*: During the experimental session, data were collected in three phases, for which the views 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.

This design allowed the experiment to address the following research questions:

- How well do drivers comprehend the meaning of the ACWS status display in the vehicle, under the various stages of the trip (scenarios)?
    - Do they know whether a given function (e.g., FCW) is present in the vehicle?
    - Do they understand whether the system is currently on and working properly?
    - Do they acquire this information in a timely manner?
    - Is the relevant display information located where people expect to find it?
- How essential is it to have familiarization with the specific vehicle's display features?
- Is there a benefit (positive transfer) or disbenefit (negative transfer) to having familiarity with some other vehicle's display features?

- Are there general approaches or particular aspects of status display design that appear to promote or limit good comprehension?

In considering this experimental design, it is important to understand clearly what the experiment was not intended to address. The experiment was not intended to directly compare different vehicles in terms of how effective their driver interfaces were. Rather, the findings were intended to be “broad brush” contrasts of distinctly different design approaches. Many different status display configurations could have been included in this experiment, and the subset ultimately used may not be representative of the full range of situations nor did it consider the relative importance of different display conditions. Furthermore, some of the findings of “correctness” may be influenced by participant bias in using certain answers. For example, one vehicle did not have a FCW system while the other two did. Participants were reluctant to use the “not applicable” answer category, so if a question concerned whether a FCW system was currently operational, the answers for the vehicle without FCW generally showed a low percentage of correct answers. For reasons such as these, the authors emphasize strongly that this experiment was not designed or intended to provide a systematic evaluation of the particular driver interfaces.

Another point to emphasize is that this experiment dealt with driver recognition of the warning system features present in the vehicle and their current status. The experiment did not address driver response to imminent crash alerts, but rather the context in which these alerts occur. The driver’s understanding of the warning and assistance provided by the vehicle may influence the response to crash warnings as well as other aspects of driving style. Misperceptions may lead to inappropriate driving strategies and slow or confused responses to imminent crash events. This aspect of the ACWS interface has not been the subject of much research. The focus of the experiment is on comprehension, as opposed to response times and avoidance maneuvers.

4.1.2 Vehicles and manuals

2010 passenger vehicles were reviewed in order to select three exemplars for use in this experiment. The three vehicles chosen – 2010 Infiniti FX 35, 2010 Buick Lucerne, and 2010 Volvo S80 – were selected because each included multiple ACWS, they had status display characteristics that were distinct from one another, and they represented a range of manufacturers.

Figure 10 shows an example picture of each of the three vehicle interiors. The figure also illustrates the three general scenario types. In the top panel, the vehicle has not yet been started and no indicator displays are activated. In the center panel, the ignition has been turned on. In the bottom panel, the vehicle is moving at 60 mph.

Broadly speaking, the Buick Lucerne had the fewest relevant safety systems, with only a LDW system and a blind spot warning (BSW) system. Also, the control buttons, which are used to turn features or systems on or off, generally used icons, and the status displays were in text or icons. Next, the Infiniti FX 35 contained three systems of interest (LDW, FCW, and ACC). Also, the Infiniti utilized acronyms for most control buttons, and icons for status information. The Volvo S80 contained the most systems of interest, with four: LDW, FCW, ACC, and BSW. It also used icons for control buttons and a text display for status information. Figure 10 above shows some examples of these interfaces. Appendix B1 shows all the vehicle interior images used in this experiment in a larger size.

Some participants in the experiment read vehicle owner's information provided by the manufacturer. This information either came from the owner's manual or from associated information such as "quick start" guides. The intent was to provide a certain degree of familiarity with a particular vehicle interface. However, the owner's manuals did not necessarily provide the information later queried in the experiment. They were tested on their knowledge to ensure that they had read at least the quick start guide. Owner's manuals for modern vehicles tend to be quite lengthy (400-plus pages) and it was not reasonable to ask participants to read an entire manual. At the same time, the objective was not to orient participants narrowly to the safety systems of interest. Therefore the information that participants were instructed to read prior to their experimental session provided a general review of vehicle features and operations. More detailed information in the full manual was available as an option if the participant wished to read it. Participants assigned to read Infiniti FX35 literature were required to read the vehicle's 46-page quick reference guide, and were also given two relevant chapters of the owner's manual as optional reading. Participants assigned to read Buick Lucerne literature were required to read the 26-page first chapter of the vehicle's owner's manual, which provided an overview of vehicle features and operations, and were told that two relevant chapters of the manual were optional reading. Participants assigned to read Volvo S80 literature were required to read the vehicle's 8-page quick guide, and were also given two relevant chapters of the full owner's manual as optional reading. The length and content of the required reading sections differed between the different vehicle models.

Participants were sent hard-copy documents in advance of their scheduled session and instructed to read at least the required portions prior to arriving for the session. The accompanying letter informed them that they would be quizzed on this material once they arrived at the session. Based on self-report, quiz results, response accuracy during the task, and the subjective impression of the experimenters during the group discussions conducted at the end of the session, participant compliance with reading the materials was excellent, though very few participants read any of the supplementary manual sections that were provided as option reading.

Figure 10. Example interior vehicle photos: Volvo S80 before startup (top), Infiniti FX 35 at startup (center), Buick Lucerne en route (bottom)

4.1.3 Participants

A total of 111 individuals participated in the experiment in May and June of 2010. Participants were approximately evenly divided between males and females, and their ages ranged from 25 to 60. All participants were recruited from the metropolitan Washington, DC area via an advertisement on Craigslist and flyers posted in the area. Recruiting materials stated that "Participants will look at photos of vehicle interiors and answer questions about them," but did not specify the project's focus on status displays. Potential participants were screened via telephone to verify their eligibility. Participants were required to have a current driver's license. 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.

4.1.4 Equipment and photographs

Participants viewed high-resolution photographs of vehicle interior displays shot from the driver's perspective. The monitors through which the photographs were displayed allowed near-full size projection of the images and the screen resolution provided good legibility of text and symbols. The photography and image display were designed to replicate as closely as possible the view that drivers would actually experience when seated in the driver's seat of a given vehicle.

The vehicle interior photos used for this experiment were taken in the vehicles of interest using a Nikon D700 digital single lens reflex camera. The images were 12 megapixels (4256 x 2832) in size, and were captured in a RAW format to allow for the most flexibility in image processing. Many images were taken of each vehicle interior to capture various vehicle states and the presence or absence of various display features (e.g., buttons turned on or off, alert messages illuminated). Each final image used in the experiment was compiled using Adobe Photoshop from multiple original images in order to clearly show all features and displays of interest, and saved as a JPEG file at its original resolution. Symbols and messages of interest that could not be photographed in the stationary vehicle (e.g., certain malfunction indicators) were recreated using image editing software. Final versions of each image used in the experiment were reviewed to confirm that all relevant displays and controls were legible to a viewer seated directly in front of the display. The full set of interior photos used in the experiment is shown in Appendix B1.

The computer displays used for this experiment were Dell UltraSharp 3008WPF widescreen flat panel monitors. These large monitors have a 30-inch diagonal display area and a native resolution of 2560 x 1600 pixels. Each monitor was calibrated to accurately reproduce the experiment images.

Participants interacted with the experiment using handheld touchpads and styli. The touchpads were Mimo 720-S models, with a 7-inch diagonal display and a resolution of 800 by 480 pixels. Participants used the 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.

4.1.5 Procedure

Participants were shown various displays of the vehicle interior, with certain status displays activated in some scenarios. Participants indicated their understanding of the presence or status of various vehicle systems. In phrasing the questions, it was important to avoid descriptive terms that might bias participant's answers. Different manufacturers use different terminologies for systems with similar functions and we did not want to use terms that matched one terminology or display over another. Therefore the research team developed generic descriptive names to use in the questions for all three vehicles. For example, a LDW system was described as "A stay in lane alert system warns the driver if the vehicle starts to drift out of its lane (for example, if the driver is drowsy or distracted"). Participants were given a glossary that defined each generic term that they could refer to during the experiment if they did not understand what a particular feature did. This glossary is shown in Appendix B2. To mask the experiment's focus on advanced safety system status displays, the questions also asked about other features such as passenger airbag displays and stability control, as well as fictional systems (e.g., "enhanced biofeedback reinforcer"). Some questions asked about actual vehicle features that may not have been present in the vehicle. Participants were also informed before beginning the experiment that some questions would ask about systems not present in the vehicle, so that they could not simply assume that a feature was present.

Experiment sessions were conducted at Westat in Rockville, Maryland. Up to three individuals participated per session. All participants in a given session were in the same literature/vehicle condition. Sessions were generally 60 to 90 minutes in duration. Upon arrival, participants read and signed informed consent forms, then, if they were assigned to read vehicle instructional material before their session, completed a brief multiple choice quiz to assess whether they read the assigned materials, and how well they remembered them. The quiz was composed of five multiple choice questions based on recognition of general concepts in the manual. Each participant sat at a computer station facing away from other participants' screens. Participants were not permitted to look at one another's displays. Participants' seating arrangements are shown in Figure 11.

The experimenter explained to participants the general purpose of the research and provided instructions on the tasks that participants would perform. Before beginning the experiment, participants were asked to sit in a normal driving position facing their computer monitors, and their monitor positions were adjusted so that the top of the monitor was approximately at eye level, and the monitor was tilted upward to be approximately perpendicular to participants' lines of gaze. This seating and monitor position allowed participants to view images of vehicle interiors as if they were seated in the driver's seat, with images subtending a field of view comparable to being seated in an actual vehicle.

Once participants were properly situated, the experimenter guided participants step by step through a practice trial to orient them to the task and to allow them to ask questions. Participants then completed a second practice trial at their own paces and without experimenter instruction, but were allowed to ask questions at the end. The practice trials used photos of a vehicle and interior features that were not used during the experimental trials. Once participants were comfortable with the task, the experimental trials began.

Figure 11. Seating position of participants at workstations

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. Details of each block are provided in Appendix B3.

For each block, participants were first informed of the vehicle's state (turned off, turned on and stationary, driving on road), then a photo of the vehicle interior that reflected the vehicle state was shown on the computer monitor. For each block, participants answered a series of questions about vehicle system status. Question order was counterbalanced within each block, with two presentation orders (and scenarios within each phase were also reversed). While participants completed each block at their own paces, the experimenter did not allow participants to begin the subsequent block until all participants were finished with the current block.

Each block included a series of four to ten questions. There were a total of 40 questions across all seven blocks. The questions and answer options appeared on the participant's hand held touch pad. For the pre-startup scenario, all of these questions asked about the presence or absence of a particular system in the vehicle (e.g., "A collision ahead alert system warns the driver when following another vehicle too closely."). 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. When the participant selected the location, a green circle centered on that location appeared on both displays**.** The participant then could confirm the location or select a different location. Figure 12 shows an example of the touch pad display with the location indicated by a green circle on both the touch pad and the large monitor.

Figure 12. Participant using stylus to indicate information location on touch pad

For the three vehicle start-up scenarios, the initial question in each block was “Are any features or functions not working properly?” The participant again selected an answer (“Yes” or “No”) and rated confidence in the answer. Then they were asked “Where was the first place you noticed this information?” and they again indicated the location using the touch pad. Following this general question, the participant was asked a series of questions about specific systems (e.g., “Is the stay in lane alert system enabled and working properly?”). Questions about specific systems were counterbalanced between participants so that half of participants received them in one order and the other half of participants received them in the opposite order. For each question, the participant provided an answer, a confidence rating, and an indication of where they looked (or would expect to look) for this information. For the three en-route driving blocks, the participant was again asked a series of questions about whether specific systems were “enabled and working properly.” For each question, the participant provided an answer, a confidence rating, and an indication of where they felt they should look for this information.

After the final block was completed, the experimenter moderated a discussion with the participants to learn more about their reactions to the features they saw in the photos. The discussion was unstructured, but generally addressed likes, dislikes, confusions, and suggestions for improvement. Discussions were typically 20 to 30 minutes in duration.

4.2 Findings

Given the range of factors and dependent measures used in this experiment, comprehensive treatment of the findings is complex. Therefore, to aid in communicating the results, we have broken this section into three parts, beginning with an overview of findings related to some of the specific research questions that motivated the experiment. Next, some illustrative cases of various outcomes of interest are presented. Then the formal statistical analyses of the data are provided. In addition, further detail on the results is located in appendices to this report.

4.2.1 Summary of findings

This experiment addressed whether vehicle status displays were effectively communicating to drivers. The three vehicles used in the experiment were intended as prototypes of quite different display strategies. For each of the key questions below, we summarize the relevant findings and their implications. Note that we consider "understanding" to encompass accuracy, decision time, and confidence.

4.2.1.1 Do people understand what advanced crash warning functions are present in a vehicle?

Overall, individuals were not particularly accurate in assessing whether an advanced crash warning system 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 (correct or incorrect 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.

Individuals took considerably longer to respond to questions about system presence than system status across vehicles and manual conditions (mean response time was 20 seconds). Note that 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.

In summary, it took participants considerable time to determine whether the vehicle had a particular warning system, but despite this time, they were frequently incorrect.

4.2.1.2 Do people adequately understand the current status of particular systems?

In addition to whether a system was present or not, participants were presented various configurations of system status when the vehicle was started and en-route. As with system presence, participants were not particularly accurate overall (about 60% of responses were incorrect). In fact, participants were less accurate when trying to understand system status than presence. Participants in the Volvo condition displayed better understanding of safety system status than for either Buick or Infiniti.

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. Surprisingly, people tended to be very confident in their responses despite a high rate of errors and long response times.

In summary, general comprehension of status indicators was slow and poor, but participants were nonetheless confident in their responses.

4.2.1.3 Does familiarity with the system (through owner's manual materials) adequately promote understanding of system status displays?

Participants tended to be 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 Volvo 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. In summary, familiarity with the vehicle's manual improved comprehension somewhat, but it was nonetheless not very good overall.

4.2.1.4 Does familiarity with one system (through owner's manual materials) influence comprehension (positively or negatively) in a different vehicle?

One question was whether having familiarity with a different vehicle's manual led to improved comprehension (positive transfer) or worse comprehension (negative transfer) relative to no familiarity . There was no finding of a systematic trend toward either positive or negative transfer. Overall comprehension rates for the "no manual" and "different vehicle's manual" conditions were quite similar (41% versus 43%, respectively) and both were lower than the "same manual" condition (54%). Detailed examination of individual questions, however, suggested that specific instances of negative transfer may have occurred.

4.2.1.5 Is information located where drivers look for it and expect to find it?

Results show that 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. For instance, participants may have assumed that a vehicle had an ACC feature if they saw information about cruise control in general. 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. As noted previously, participants' generally high confidence in wrong answers may reflect such errors in interpretation.

4.2.1.6 *Are there observed issues with the use of acronyms, icons, color coding, etc.?*

There were several design issues:

- 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 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. 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 issue.
- 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. This was not the case 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.

Many of the design issues that became apparent in the experiment results were also raised and elaborated upon in the group discussions held with participants at the end of the experiment session. Key discussion group findings are presented in Section 4.3.5, and a detailed discussion summary is presented in Appendix B5.

4.2.2 Example cases

In this section we will highlight several cases which provide interesting comparisons across vehicles by driver phase and display format.

First example: Volvo participants had relatively high accuracy across all three phases, in contrast to the Buick and Infiniti participants showing a drop in understanding during the startup phase (and then a recovery during the en-route phase). An interesting pattern seems to occur with the Infiniti. Performance is relatively high in the pre-startup phase (56 percent), where participants would need to recognize the presence of the system through the "LDW" control button. Performance then drops drastically in the startup condition. This may be due to two issues: a) when the system is not functioning properly, the only indication is an orange icon of a car and lane lines in the speedometer (note that this orange icon is also the same color as a light on the LDW control button), b) when the system is functioning, the only indication is an orange light on the LDW control button which participants mentioned during discussions as seeming to indicate a malfunction ("green for go" versus "red for stop"). Figure 13 shows these features. Once participants entered the en-route driving phase, performance increased to the highest level yet. This may be the result of a green icon on the dash representing the system functioning properly.

Green is usually understood as an indication of proper functioning (as also mentioned in the discussions with participants) and may have been a better cue.

Figure 13. Infiniti LDW control button (left) and orange status indicator icon (right)

Second example: There was a particularly strong performance in the pre-startup phase for participants in the Infiniti condition. Indeed, almost all participants were able to identify the presence of an ACC feature in the Infiniti before startup (M=97%), which indicates a very clear display of this information regardless of prior exposure to the manual. What seems to be unique about the pre-startup display of the Infiniti was a button labeled with "CRUISE ON/OFF" in clear view on the steering wheel (see Figure 14). In general, icons were superior to acronyms on control buttons. This case demonstrates that the right text (i.e., complete words) could be highly effective in improving understanding of system presence.

Figure 14. Infiniti cruise control on/off button

Third example: When focusing on the LDW system, participants were most accurate in the Volvo condition. In contrast, the least accurate participants were in the Infiniti condition, with the Buick condition in between. Both Volvo and Buick utilize text to communicate if there is a malfunction with the system, which may be a reason for the higher accuracy of participants with those vehicles." Figure 15 shows these displays.

Figure 15. LDW malfunction indications (clockwise from top left: Volvo, Buick, Infiniti)

Fourth example: There was a particularly long mean decision time (over 25 s) during pre-startup for Infiniti participants. There are two differences between the Infiniti and the other vehicles that may result in the increased decision time. First, related to location, the Infiniti places the LDW control button in the lower left hand corner next to the steering wheel, and this may be problematic for participants to notice. The Buick also has the LDW control button on the left side, but raised and next to the steering wheel (and also uses an icon). Second, both the Buick and Volvo use icons to mark the LDW controls, whereas the Infiniti uses an acronym. These controls are shown in Figure 16.

Figure 16. LDW control buttons (clockwise from top left: Volvo, Buick, Infiniti)

Fifth example: Although no systematic evidence of negative transfer was found in this experiment, negative transfer did appear to play a role in some cases. Figure 17 shows one such example. The figure shows where participants looked to determine whether a LDW system was present in the Infiniti vehicle in the pre-startup phase. The correct answer was the FCW/LDW control button, which was selected by about 20 percent of all participants. However, the figure shows that nearly 40 percent of participants in the Different Manual condition looked for this information in the dashboard, whereas no participants in the other two Manual conditions looked for this information here. While there is not an obvious explanation for this result, it does present a situation in which participants who read a different manual were especially likely to make an error that other participants (those who read the same manual or no manual) did not make.

Figure 17. Pre-startup Infiniti example showing where participants looked to determine LDW system presence

4.2.3 Detailed results

There are several things to note about the following analyses. First, although participants were asked a variety of questions about real or fake systems and features, the current analyses focus on safety systems most relevant to the CWIM project. Those systems are lane departure warning (LDW), forward collision warning (FCW), adaptive cruise control (ACC), and blind spot warning (BSW).

Second, participants who were asked to read a vehicle's manual before the session were also given a brief quiz and questionnaire at the beginning of the session. The quiz was composed of five multiple choice questions based on recognition of general concepts in the manual. Participants who did not answer at least two correct were removed from any subsequent analyses, under the assumption that they did not comply with instructions and would therefore be problematic to classify (as either manual or no-manual participants). Nine participants' data were removed for this reason (four from the Volvo vehicle condition, three from the Infiniti vehicle condition, and two from the Buick vehicle condition).

Third, there were interactions in several cases, so main effects have to be taken with that caveat. Also, interactions were only analyzed or addressed up to the second order. For space and complexity considerations, higher than second order interactions were not addressed.

Fourth, the three vehicles discussed below are not a direct comparison across vehicles. For example, what may seem like poor performance for Buick compared to the other vehicles when drivers were asked about the FCW system during the startup and en-route phases can actually be the result of the Buick not having an FCW system and participants being reluctant to choose the "not applicable" option. Also, in the startup and en-route scenarios, different configurations of activation and malfunction were used and did not always map across all three vehicles (partially as a result of each vehicle not having every system present). Consequently, one should not take the following results as an indication of a particular vehicle's performance, but rather focus on comparisons of display type, manual presence, etc.

Fifth, given the exploratory nature of this study, significance criteria are presented at either p <.05 level of the p<.10 level. The latter level, while not being traditional in certain academic disciplines, is useful in the current case to provide the reader with cases that may indicate an interesting pattern which would be missed under a more stringent criterion. The reader will be reminded that p <.05 is a convention used in some fields, but a higher or lower level p value can be utilized depending on the need for stringency (e.g., medical trials of a dangerous drug versus personality comparisons). In the current context, a slightly less stringent p value cutoff is used to allow for the examination of borderline cases that still may yield useful results for future studies.

The results presented in the sections that follow are organized by key dependent variable: Response accuracy, confidence ratings, and decision time. Each section begins with a table showing ANOVA results for key main effects and two-way interactions. Results are then described in further detail. Because of the large number of analyses conducted, only select findings are described in these sections. See Appendix B4 for the full set of ANOVA analyses for key safety systems, including additional details about analysis method.

4.2.3.1 Comprehension Accuracy

Table 9 below shows p values for ANOVAs conducted using participants' accuracy in understanding the status displays of each of the four key safety systems, and overall across all safety systems. The data used for these analyses include participants' task responses for each of the seven blocks of the experimental session. Blank cells in the table indicate p>.10. Mean percent correct responses by vehicle and manual condition are shown in Figure 18. The figure shows that despite overall low performance, participants in all vehicle conditions were more likely to provide correct responses about the FCW status displays if they had read the same vehicle's owner's manual.

Table 9. Accuracy ANOVA significance levels by safety system

Effect	All safety systems	LDW	FCW	BSW	ACC
Phase	*	*	*		*
Vehicle	*	*	*	*	*
Manual	*	*	+		
Vehicle x Phase	*	*	*		*
Phase x Manual					
Vehicle x Manual			+	+	

** = p<.05;   + = p<.10*

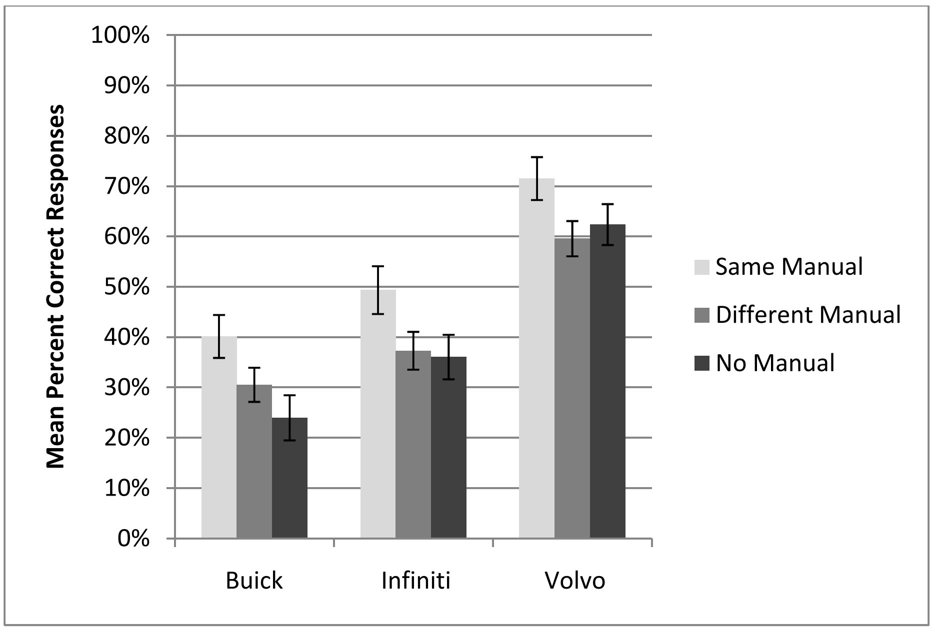


Figure 18. Mean percent correct responses for all systems (with standard error bars)

Both Vehicle and Manual effects were robust for all safety items. Participants in the Volvo condition were the most accurate (M=65%), followed by the Infiniti condition (M=41%) and the Buick condition (M=32%), which was at chance. Note that, overall, participants were not accurate, regardless of prior exposure to the manual.

There was an expected Manual effect, with participants gaining the most from having prior exposure to the manual of the vehicle they were asked questions about ("same" M=54%). Having no manual (M=41%) or a different manual (M=43%) resulted in poorer performance. Although there was a benefit to having the same manual as the vehicle being asked about, performance was still rather low across all vehicles regardless of manual exposure. In other words, having a manual helped, but not enough for participants to develop a good understanding of safety system presence or operation.

Participants were more accurate in the pre-startup phase (M=57%), where they were asked about the presence or absence of a system, than the other two phases (startup and en-route; M=37%

and M=43%, respectively) which included questions of system activation and proper functioning. Note that the pre-startup phase contained 3 response options per question and the remaining two phases contained 4 response options per question.

A large part of the significant Vehicle X Phase interaction is the result of Buick participants having a continual drop in accuracy from pre-startup, to startup, to the en-route phase. In contrast, the Infiniti and Volvo only show a drop after the pre-startup phase, but then level out.

In order to make the most direct comparison across phases, vehicles, and manual conditions, it is necessary to focus on the one system that was present in all vehicles and asked about during all phases: LDW. Figure 19 charts response accuracy for all LDW items across vehicle and manual conditions. In contrast to all other safety systems under consideration in this experiment, when focusing on the LDW system, participants were more accurate in both pre-startup (M=71%) and en-route (M=72%), but less accurate during startup (M=53%). This may indicate that the control button labeling in the pre-startup phase and the function/activation indicators in the en-route phase were easier for participants to understand. For example, in the pre-startup phase, each vehicle had a clearly labeled button (using either acronyms or icons) to indicate the presence of an LDW system.

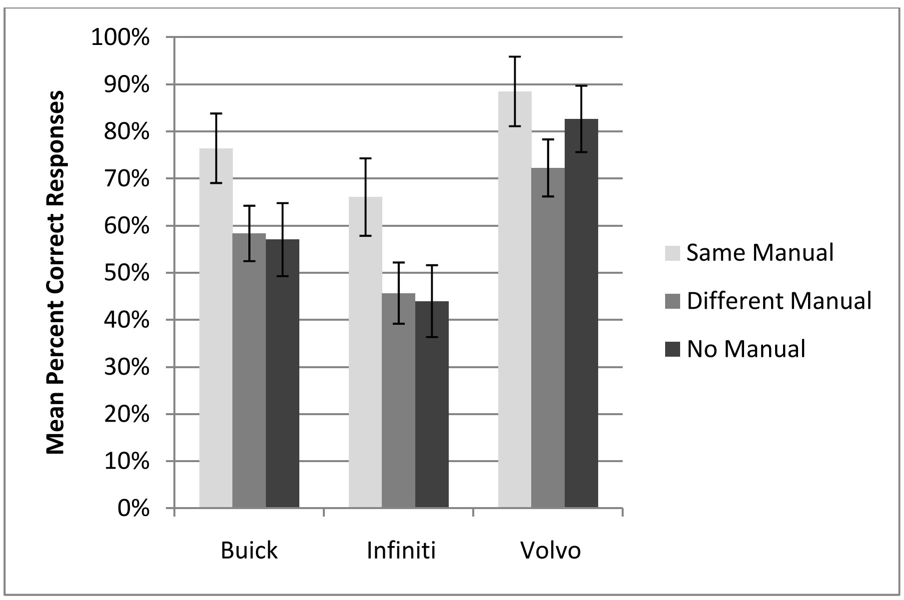


Figure 19. Mean percent correct responses for LDW items only (with standard error bars)

There is a similar pattern of main effects and the same interaction for the LDW system as when all safety systems were analyzed. For example, having the vehicle's matching manual (M=77%) is more beneficial than having no manual (M=59%) or another vehicle's manual (M=61%). Interestingly, performance overall is higher for the LDW system than across all systems.

The Vehicle main effect has a different pattern. Similar to the entire set of safety items discussed above, participants were most accurate in the Volvo condition (M=81%). In contrast, the least accurate participants were in the Infiniti condition (M=52%), with the Buick condition in between (M=64%). As shown in

Table 9, these differences were significant ($p<.05$). Both Volvo and Buick utilize text to communicate if there is a malfunction with the system, which may be a reason for the higher accuracy of participants with those vehicles.

The Vehicle X Phase interaction for LDW is driven by the Volvo participants having high accuracy across all three phases, in contrast to the Buick and Infiniti participants showing a drop in performance during the startup phase (and then a recovery during the en-route phase). An interesting pattern seems to be occurring with the Infiniti. Performance is reasonably high in the pre-startup phase, where participants would need to recognize the presence of the system through the "LDW" control button (M=59%). Performance then drops drastically to M=26% in the startup condition. This may be due to two issues: a) when the system is not functioning properly, the only indication is an orange icon of a car and lane lines in the speedometer (note that color of this icon is similar to the color of the "on" indicator light on the LDW control button, which could be perceived as contradictory information), and b) when the system is functioning, the only indication is an orange light on the LDW control button which some participants mentioned during discussions as seeming to indicate a malfunction ("green for go" versus "red for stop"). When asked where they found the information that let them know whether the system was on and functioning properly, participants overwhelmingly clicked on the location of the orange LDW icon, which indicates that participants understood that the icon indicated LDW system status, but that they did not understand its meaning (see Figure 20). Once participants began the en-route phase, performance with the Infiniti status displays increased to its highest level (M=70%). This may be the result of a green icon on the dash representing the system functioning properly. Green is usually understood as an indication of proper functioning (as also mentioned in the discussions with participants) and may have been a better cue.

Participants in the Infiniti condition demonstrated particularly low performance when asked about the BSW system (M=11%). This was probably due in large part to Infiniti not having a BSW system, and participants being reluctant to choose the N/A option. But the low performance is not just a matter of participants' reluctance to choose the N/A option during the start-up and en-route phases, because even in the pre-startup phase (where N/A is not an option) where participants are being asked about system presence, performance was quite low (M=21%).

The effect of Manual was not significant. Participants performed poorly (the highest mean, and for matching manuals, was only M=43%).

The results of ACC were similar to LDW and FCW, only with much lower overall performance regardless of manual condition. Participants were better able to identify the presence of ACC (first phase M=60%), than its proper functioning in the later phases (M=26% and M=29%). The Vehicle effect is driven by extremely low performance in the Buick condition (M=09%), which is likely the result of the Buick not having an ACC system and participants being reluctant to respond "not applicable" (although performance is still quite low even in the pre-startup phase, M=14%, which is a simple question about presence/absence). The significant Vehicle X Phase interaction is also somewhat driven by this poor performance in the Buick condition, but is also a result of particularly strong performance in the pre-startup phase for Infiniti. Indeed, almost all participants identified the presence of an ACC in the Infiniti before startup (M=97%), which indicates a very clear understanding of the message "CRUISE ON/OFF," though it is not clear if participants were able to differentiate ACC from basic cruise control.

Figure 20. Image showing where participants clicked to indicate where they found information about LDW functionality, with most clicks clustered on orange LDW icon in speedometer

4.2.3.2 Confidence

Participants' confidence ratings across all safety systems increased significantly (p<.001) from pre-startup (M=7.8) to startup (M=8.2) to en-route (M=8.4). Although not significant, participants seem to be slightly more confident in the Volvo condition (M=8.5) than either the Buick or Infiniti conditions (M=7.8 and M=8.1, respectively). Also not significant, but trending, participants are slightly more confident with the same manual (M=8.4) than with a different manual (M=8.2) or no manual (M=7.8).

Table 10 below shows significance levels for all main effects and two-way interaction effects conducted on participants' confidence ratings for each of the four key safety systems, and overall across all safety systems. The data used for these analyses include participants' responses for each of the seven blocks of trials. Blank cells in the table indicate p>.10. Mean confidence ratings by vehicle and manual condition are shown in Figure 21. The figure shows that confidence overall is particularly high. In fact, the levels of confidence do not reflect the performance levels demonstrated in the earlier section. Drivers were more confident than their knowledge warranted (regardless of vehicle phase or manual information).

Participants' confidence ratings across all safety systems increased significantly (p<.001) from pre-startup (M=7.8) to startup (M=8.2) to en-route (M=8.4). Although not significant, participants seem to be slightly more confident in the Volvo condition (M=8.5) than either the Buick or Infiniti conditions (M=7.8 and M=8.1, respectively). Also not significant, but trending, participants are slightly more confident with the same manual (M=8.4) than with a different manual (M=8.2) or no manual (M=7.8).

Table 10. Confidence ANOVA significance levels by safety system

Effect	All Safety Items	LDW	FCW	BSW	ACC
Phase	*	+	+	*	
Vehicle		*	*		
Manual		+			
Vehicle x Phase	+	*		*	*
Phase x Manual	+	*			
Vehicle x Manual					

** = p<.05;   + = p<.10*

The Vehicle X Phase interaction for all systems is driven by participants in the Volvo condition becoming more confident after the vehicle is started. That may be due to clear text information displayed at the startup and en-route stages, but only icon buttons displayed as an indication of system presence in the pre-startup phase. This shift is much less apparent in the Buick and Infiniti.

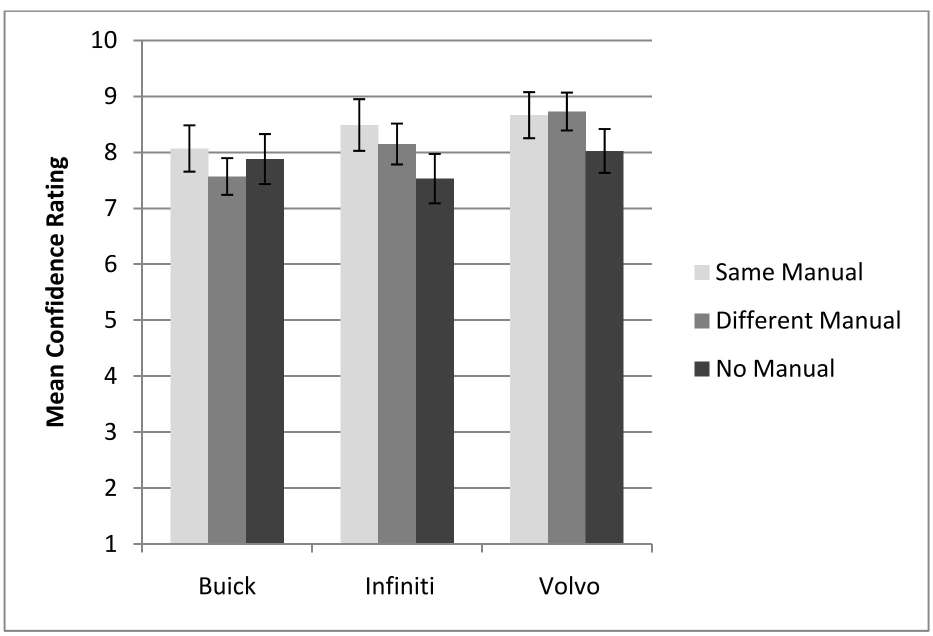


Figure 21. Mean confidence ratings for all systems, where 1 = no confidence and 10 = complete confidence (with standard error bars)

The Phase X Manual interaction seems to be the result of participants being more confident in the pre-startup phase when the manual matches, then when having no manual or a different manual. Once again, the effect of manual may be weakened in the later phases for vehicles which present clear text information about system functioning (e.g. Volvo).

LDW systems were present in all three vehicles. Figure 22 charts mean confidence ratings across vehicle and manual conditions for all LDW items. The figure shows that Volvo drivers are the most confident (M=8.9) and Infiniti drivers are the least confident (M=8.0), although all conditions showed quite high levels of confidence. There was a significant trend for Phase, where people became increasingly confident in their answers from pre-startup to startup to en-route. There is also an effect of Manual, where participants are most confident with the manual that matches the vehicle for LDW items (M=8.8), and least confident with no manual (M=7.9).

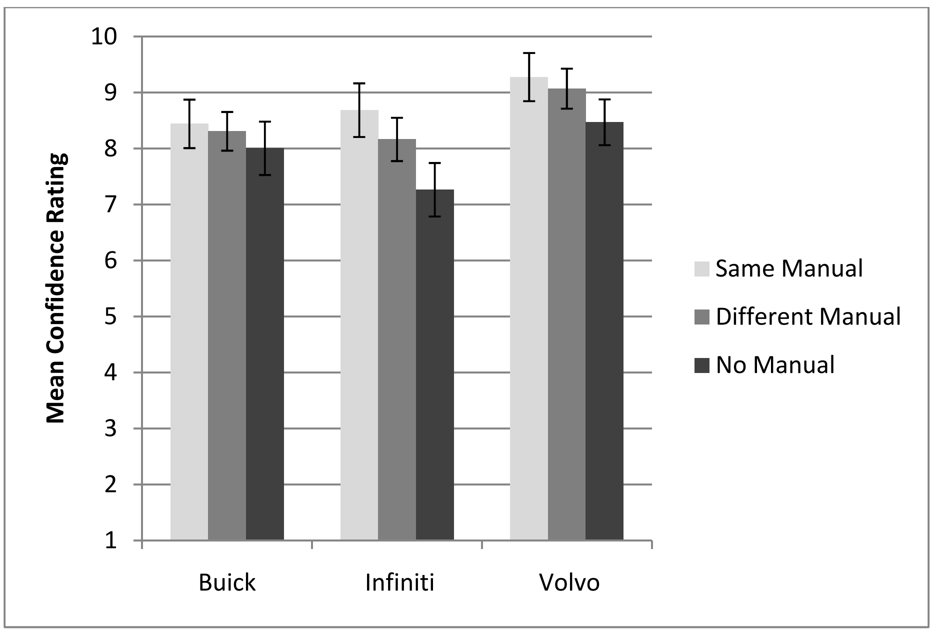


Figure 22. Mean confidence ratings for LDW items only (with standard error bars)

The significant Vehicle X Phase interaction when the results are limited to those pertaining to LDW is driven by changing behavior in the Infiniti condition. For Buick and Volvo, confidence stays consistently high across all three phases. In contrast, confidence builds across phases for the Infiniti, with the lowest confidence during startup (M=7.5) and the highest confidence during the en-route phase (M=8.7). The high confidence in the en-route phase mirrors the higher accuracy of participants in that phase and is one of the few times that participant confidence matched performance.

The significant effect of the Phase X Manual interaction is the result of higher confidence for participants who had a matching manual (M=9.2) than participants who have a different manual (M=8.1) or no manual (M=7.5) in the pre startup phase. This difference shrinks during the startup and en-route phases, which is likely a result of other cues displayed after the vehicle was started (e.g., descriptive text in the Volvo's status displays).

4.2.3.3 Decision Time

Table 11 shows significance levels for ANOVAs conducted using participants' decision time for each of the four key safety systems, and overall across all safety systems. Any decision times less than 2 seconds or greater than 57 seconds were excluded from the following analyses (which corresponds to the bottom 2.5% and top 2.5%, respectively). The data used for these analyses include participants' response times from each of the seven blocks of the experimental session. Blank cells in the table indicate p>.10. Mean decision time by vehicle and manual condition are shown in Figure 23. The figure shows that mean response times were fairly consistent across vehicle and manual conditions, with somewhat longer times in the Infiniti/No Manual condition and slightly shorter times in the Volvo/Same Manual condition.

Table 11. Decision time ANOVA significance levels by safety system

Effect	All Safety Items	LDW	FCW	BSW	ACC
Phase	*	*	*	*	*
Vehicle	*	*	*	*	*
Manual			*		
Vehicle x Phase	*	*	*	*	*
Phase x Manual	*		+		
Vehicle x Manual					

** = p<.05; + = p<.10*

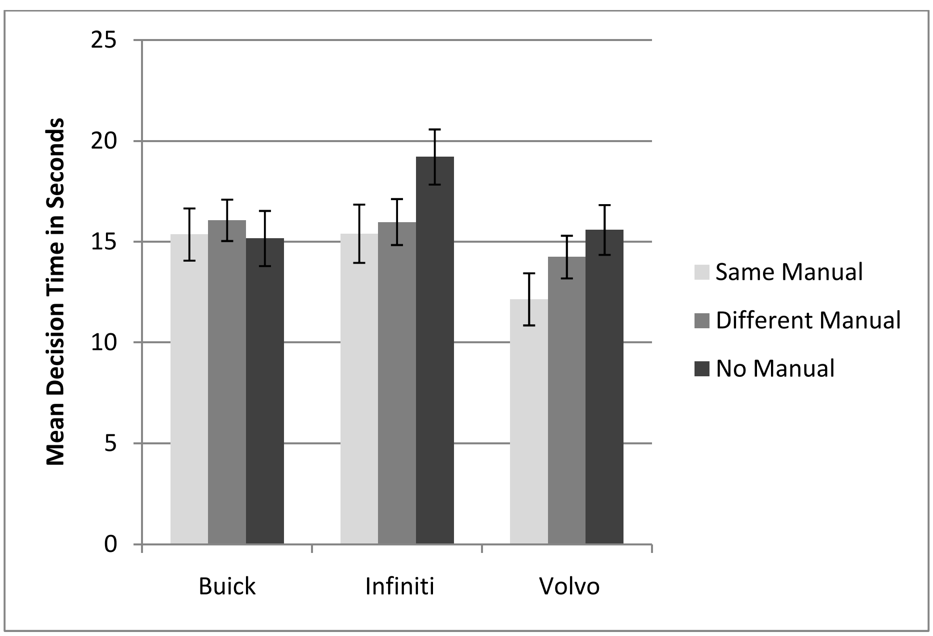


Figure 23. Mean decision time in seconds for all systems (with standard error bars)

Participants took significantly longer to decide in the pre-startup phase (M=20.4 s) than in the startup (M=14.9 s) or en-route phases (M=11.1 s). There was a significant difference between decision times for participants in the Infiniti and Volvo conditions (M=16.9 s and M=14.0 s, respectively). This parallels the relative accuracy for these vehicle conditions. There is not a significant main effect of Manual, although the decision time means show a trend in the expected direction of showing that participants without a manual responded slower than participants with the matching manual.

The significant Vehicle X Phase interaction resulted from the larger decrease in decision time for Infiniti and Volvo from pre-startup to startup.

The significant Phase X Manual interaction is driven by a larger decision time in the pre-startup condition for those who did not have a manual (M=22.1 s) compared to those who did have a manual (M=18.3 s). This difference in decision time was not evident in the startup and en-route phases. Furthermore, while the Phase X Manual interaction was significant when results for all evaluated safety systems were combined, there were no significant results ($p<.05$) when each safety system was analyzed individually.

Similar to the findings from all safety systems, the LDW results (see Figure 24) show that Phase has a main effect with reduced decision time for participants in the latter two phases of the experiment. This is driven by a reduction in decision time from pre-startup (M=16.5 s) to startup (M=11.9 s). The significant main effect of Vehicle is due to the Infiniti participants taking longer to reach a decision (M=16.2 s) than participants in either the Buick (M=12.0 s) or the Infiniti (M=10.2 s).

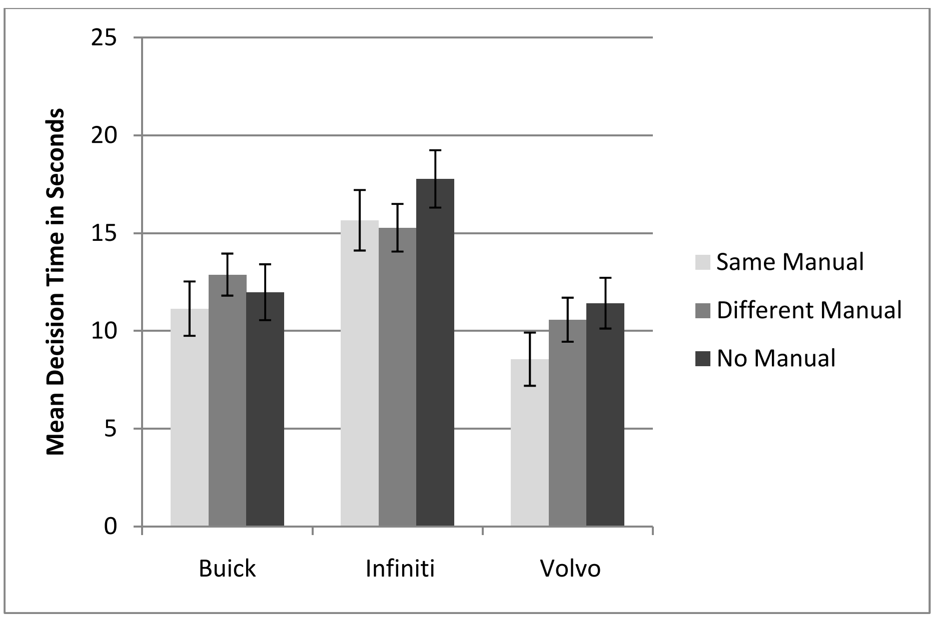


Figure 24. Mean decision time in seconds for LDW items only (with standard error bars)

The Vehicle X Phase interaction for LDW items is driven by a particularly long mean decision time (M=25.6 s) during pre-startup for Infiniti participants that then drops sharply to become more in line with decision times in the other vehicles. There are two differences between the Infiniti and the other vehicles that may have resulted in the long decision time for Infinity during startup. First, the Infiniti LDW control button is located below and to the left of the steering wheel, and this may be problematic for participants to notice. The Buick also has the LDW control button on the left side next to the steering wheel, but higher than the Infiniti (and it also uses an icon). Second, both the Buick and Volvo use icons to mark the LDW controls, whereas the Infiniti uses an acronym.

There is a similar pattern of results for both BSW and ACC. There is a slightly different pattern for FCW, where having a manual has a significant impact on decision time. Participants who had a matching manual took the least amount of time (M=13.6 s) compared to participants with different manuals (M=15.1 s) and no manual (M=17.5 s). There was also a marginal Phase X Manual interaction, driven by longer decision times in the pre-startup condition for those individuals with no manual or a different manual (p=.06). This difference decreased in the startup and en-route phases, likely as a result of additional cues (e.g., text descriptions in the Volvo) which may have helped participants make quicker decisions (possibly in addition to task exposure).

4.2.4 Location where participants found/sought system status information

After participants provided their answers for each item (system presence/status, confidence rating), they were asked to identify the location in the displayed vehicle photo where they found or sought the information that they used as the basis for their answers. Scatterplots that indicate the locations where participants looked for information provide some insights regarding their answers and expectations. The location findings generally appear to reflect the accuracy findings: For items that had relatively high mean accuracies, participants also tended to more accurately identify where the feature or status information was located. One consistent revelation from these plots is that when participants are unsure or incorrect about where to find status information, they most often look for it in the dashboard. Figure 25 shows an example of this. In the example, the LDW system is not active, as indicated by the unlit indicator lamp on the LDW button. While many participants correctly identified this button as the location that indicates LDW system status, many others appeared to expect the status to be presented somewhere on the dashboard.

Figure 25. Buick startup image showing locations where participants looked for LDW status information

4.2.5 Group discussion key findings

This section describes some of the key findings of the group discussions that were conducted at the end of each experimental session. A more complete summary of the group discussion findings organized by vehicle is presented in Appendix B5.

- *Unfamiliar acronyms and icons were difficult to understand.* Participants reported that many unfamiliar acronyms and icons were difficult to interpret without prior knowledge of their meanings. Many icons were seen as ambiguous. For example, icons for LDW, FCW, and ACC were often mistaken for one another or for other systems such as traction control. While acronyms were generally difficult to understand, some participants noted that it is even more difficult in the many cases where acronyms used in vehicles are not even true acronyms (e.g., BLIS for BLind Spot Information System in Volvo). Many suggested that icons and acronyms be made more intuitive and standardized across vehicles. (Note that while the Society of Automotive Engineers (SAE) has developed a standard that specifies symbols for use on vehicle controls, indicators, and tell-tales (SAE J2402, 2010), this standard does not currently address symbols for ACWS features.) Some participants also suggested that words should be used instead of, or in addition to, icons and acronyms.
- *Color can be an effective cue if matched to users' mental models.* Participants were accustomed to green meaning good or functional and red meaning bad or nonfunctional. The meanings of other colors were generally more difficult for participants to interpret. In the Infiniti displays, the colors used were ambiguous or counterintuitive to some participants. Specifically, the use of red as a primary color in the display theme, and the use of orange for the FCW/LDW control button light and the LDW icon confused some participants.
- *Organization of controls and displays is important.* Many participants commented that given the large number of displays and controls in the vehicles, it is important for manufacturers to organize and prioritize information. Participants suggested that similar types of information, such as error/warning messages, should be located in the same place. Some participants also suggested that non-critical information such as outdoor temperature should not be presented on the dashboard, which should be reserved for critical information, especially since the dashboard displays were perceived to be overcrowded with information. Collocating controls for similar features (e.g., advanced safety systems) was also suggested to improve interface intuitiveness and usability. Participants had mixed reactions regarding the placement of ACWS controls on the steering wheel.
- *Vehicle interfaces are learnable.* Despite initially being overwhelmed by the number of advanced vehicle features and lack of intuitive interfaces, many participants felt that with some practice, they could learn to use the vehicles' features. Participants generally had positive overall reactions to the vehicle interfaces.
- *Instruction manuals are helpful, but are no substitute for experience.* Most participants who read quick start guides before their experiment sessions felt that this material helped them to understand vehicle features. Even those who read about vehicles different than the ones they saw in the experiment tended to feel this way. Nonetheless, participants generally felt that while they picked up general concepts from the manuals, there is no

substitute for actual experience to really be able to understand interfaces as complex as those in the vehicles used for this experiment.

4.2.6 Assessment of methodology

This experiment used an approach similar to an enhanced cognitive walkthrough to investigate drivers' understanding of in-vehicle systems and displays, as well as the usefulness of reading owners' manuals (Bligard & Osvalder, 2007; see Mahatody, Sagar, & Kolski, 2010 for a recent review of various cognitive walkthrough methods). This approach is particularly suited for investigating users' knowledge of a system or possible user error. Note that unlike most cognitive walkthrough methods, the current approach did not allow the users to interact with the system (e.g., pushing buttons and receiving feedback). Instead, users were asked questions about knowledge based on vehicle states displayed in photographs. This section discusses the advantages and limitations of this methodology.

- *Experimental stimuli and setting.* The stimuli for this experiment were high resolution displays of photographs of the actual interiors of three production vehicles. This methodology allowed participants to make ratings based on realistic displays, and appeared to be quite successful in replicating the in-vehicle environment, but external validity could have been better ensured by placing participants within actual vehicles or a realistic simulation environment. Although participants experienced the vehicle displays as they would appear in various vehicle states, all displays were static and participants could not interact with them in any way. Interaction and trial-and-error may be ways for drivers to acquaint themselves with the vehicle's features. Furthermore, many displays would change under different conditions, and such changes may give drivers cues as to how the features actually work. For instance, many participants did not understand that some ACWS only operate above a certain speed threshold, but if they had the opportunity to drive the actual vehicle, they might begin to make the connection between speed changes and system status display changes. Another limitation of the use of existing vehicle displays as stimuli was that individual display and control characteristics could not be manipulated or directly compared against one another without the presence of numerous confounding variables. For example, the Phase variable confounded exposure with the displayed features. Future implementations of this method could, however, include such comparisons if multiple variations of a particular display are compared while controlling for confounds. It should also be noted that generic names for each ACWS were used so that the same descriptive name could be used for all vehicles, even though most had their own proprietary names for features. The use of generic names may have accounted for some degree of confusion about system presence, though the inconsistent identification of analogous features by various manufacturers is likely to lead to similar difficulties. Another limitation is that because this experiment only used the displays and instructional materials from three vehicle models, the generalizability of these findings is limited. Finally, as participants experienced these systems in a lab setting without any driving task, the attentional demands these displays might place on a driver remain unclear as do their safety effects. Future research using this methodology could be done in a driving simulation environment if such measures are desired.
- *Exposure to manuals.* The current experiment found that participants were not particularly enlightened after exposure to information materials (i.e., owner's manuals

and quick start guides). It should be noted that participants were not able to read the short manuals while also interacting with the vehicles. This may have affected learning. Though the experiment found that participants overwhelmingly did appear to have read the required sections before their sessions, they may have lacked appropriate context and motivation to do so, as they knew that they would not actually operate a vehicle during the session. Furthermore, participants were only required to read the vehicle quick start guide, and each vehicle's quick start guide had varying amounts of information about the key systems evaluated in this experiment. On the other hand, being motivated to read the short manuals (because it was required for the experiment and because participants were aware that they would be quizzed to ensure that they read them) may have actually augmented comprehension beyond what is found in the general population. Consequently, it is not truly possible to evaluate the information materials for quality and conveying information, so future research should develop and evaluate prototype informational materials (print and/or video). Finally, although one objective of this research was to identify transfer effects caused by different combinations of manual material read and actual vehicle experienced, it is not clear how the limitations of the manual reading task may have influenced transfer.

- *Participant characteristics*. Participants were selected to participate only if they drove vehicle models without ACWS features from slightly older model years that were manufactured by companies other than the three used in this experiment. While this was important to ensure that participants did not have previous exposure to relevant features, the sample might not be reflective of the individuals who would buy or drive vehicles that include ACWS features. Many participants indeed noted that the vehicle displays they experienced during the experiment were far more complex than their own. The effects seen in this experiment might therefore represent naïve drivers and not drivers who have experience driving with any ACWS.

5 Implications of Study Methods and Findings

The two studies that composed the current report were drastically different in approach, medium of presentation, design, and analysis. Regardless, there was a common theme focusing on the implications of DVI variability in auditory alerts and visual display status 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. Consequently, implications are discussed separately for each experiment. The final report from the broader CWIM research effort (Lerner et al., 2011) provides additional discussion, synthesis and implications drawn from these studies and other studies of driver response to, and comprehension of, ACWS.

5.1 Negative transfer with auditory FCW experiment implications

Experiment 1 focused on auditory alerts in crash imminent situations. Based on the findings of Experiment 1, there are several implications regarding variability of auditory alerts in forward collision warning systems (and in ACWS in general):

- *Transfer problem*: There is potentially a substantial transfer problem with auditory alerts in a FCW system 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. These data suggest that signal parameters can influence transfer effects between vehicles. Further research is needed to determine the specific characteristics of alerts that can influence transfer of learning, and to expand this line of research to include other types of ACWS alerts (e.g., LDW, BSW).

- *Familiarity*: There was also a rather large familiarity effect in Experiment 1. 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 exposures, the difference was approximately 130 ms. Not surprisingly, this indicates that if people come to recognize a familiar sound from general experience as a driver or passenger, their responses would be faster the next time an alert is presented. In the present experiment, 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 finding of a very large effect for one transfer direction and a smaller effect for another transfer direction indicates that transfer effects are 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, such as cell phone rings, 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)?
        - This experiment does not indicate the critical parameters, other than to note that the shift to the heavy warning was more favorable than the shift to the light warning. The two warnings differed in a variety of parameters, including frequency modulation, primary frequency, pulse duration, burst duration, onset and offset ramps, pulses per bursts, number of bursts, and intervals between pulses and between bursts as well as in their similarity to the other auditory alerts (e.g., phone ring) that were presented.
    - 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)?

- *Methodology and evaluation protocol*: There are also several implications for further methodological refinement with the current task, design, and development of evaluation methods for the auditory ACWS transfer experiment:
    - Scenarios: what scenarios failed, what scenarios were successful, and how can the scenarios improve?
        - About half of forward collision event trials had to be removed from analyses, either because participants responded to the event without braking (e.g., by swerving) or because brake responses occurred before event onset or within 200 ms of event onset, indicating that they may have begun to brake before hearing the FCW alert. This indicates the need for an improved scenario to make the task more distracting and a proper implementation of the event so that the event coincides with engagement in the distraction task (especially so that there is not the need for manual experimenter triggering, which occurred occasionally in the auditory ACWS transfer experiment). Also, there was not realistic traffic and non-event traffic would drive 15-20 miles per hour faster than the participant vehicle. As indicated by Green (2008), a complex simulated roadway environment with realistic surrounding traffic can improve the validity of the simulator method, especially when presenting potential collision events at an unrealistically high frequency.
        - All sounds were generated from the same group of speakers and lacked directionality. Real world sounds would come from the entire environment and have many more possible combinations. Future research should investigate additional ecologically valid sounds and sound locations.
    - Distraction task: what are the strengths/weaknesses of the Simon task; and in what ways can it be improved?
        - As noted above, in about half of the staged forward collision events, participants were not sufficiently distracted to successfully implement the alert scenario. The response pad that was used to enter the response directions in sequence during the Simon distraction task always had the response directions conform to cardinal directions (e.g., "up" was at the top of the layout). It is possible that some participants noticed this and used a strategy of "feeling" where the response buttons were without having to look away from the roadway while entering in their responses. There was evidence that participants became more proficient at the task over time. An improved version of the Simon distraction task could present a similar directional response pad, but with each presentation the physical location of the directions to be pressed would change (creating something of a Stroop effect). For example, "up" would be at the bottom of the layout for one trial and the top for another trial.

5.2 ACWS status display comprehension experiment implications

Based on the findings of Experiment 2, there are several implications for ACWS status display interfaces and information:

- *Overall comprehension*: There is a comprehension problem with vehicles containing unfamiliar ACWS systems, indicated by low comprehension rates for system presence and status, 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.
    - Participants who read quick start guides prior to the experiment also displayed low comprehension. Of course, reading the short owner's manuals without the car present may have reduced the value of this information. Furthermore, participants were not instructed to pay any particular attention to ACWS features mentioned in the materials, and some of the materials provided little information about relevant ACWS features. Conversely, reading the materials with motivation to attend to them in detail, as provided in this experiment, is not what most of the population do.
- *Manual information*: Reading manufacturer-provided information was somewhat helpful, but the problems remained. Also, having read information about a <u>different</u> vehicle did not generally provide benefit. The limited improvement was vehicle-specific.
- *Variability in understanding*: For any particular state of any particular safety system, there may be wide variability among vehicles in terms of how well people understand the situation. Although this research did not evaluate safety effects, it is possible that poor comprehension could lead to safety issues if a person has a system but is not aware of it or mistakenly believes that a system is present/operational and tries to rely on it while driving at high speeds (e.g., lane departure warning). To date, the project team is not aware of any research that has directly investigated the safety effects of ACWS comprehension issues.
- *Design issues*: A variety of design issues may have affected comprehension and decision time.
    - Displays were not always consistent with population stereotypes about where to find the information or how it is color coded. For example, participants had a mental model of green being an indication of a properly functioning system, and red indicating error/disabled/malfunction. These population stereotypes were not always compatible with vehicle design in the case of certain icons or buttons.
    - Acronyms were difficult for people to use and varied from vehicle to vehicle.
    - Complete words as labels or in text displays yielded the highest comprehension, regardless of vehicle familiarity from reading the owner's manual.
    - It should be noted that this experiment was not designed to directly compare different vehicles' approaches to particular messages. The results do not imply that one particular manufacturer's approach is better than another's. However, the large differences among vehicles in participants' abilities to correctly answer the questions indicate that there should be some means of screening ineffective cases.
- *Implications of DVI variation among vehicles*: This experiment found general comprehension problems and lack of transfer from reading the manual of one vehicle to understanding the status displays of another. Potential areas for improvement are noted below:
    - Consistent terminology (e.g., text or acronyms) for particular warning functions may improve comprehension. Common terminology and acronyms exist for some

functions, such as antilock brakes (ABS), and other warning systems might benefit from this as well.
  - Consistent icons and of color coding for status may also improve comprehension. These 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 driver-vehicle interface, so location might have to be empirically determined/performance based for the vehicle, rather than there being a single preferred location for all vehicles.
  - Print materials do not seem highly effective in conveying knowledge about system status indications (although the present experiment's procedure was limited in that the participant did not have the vehicle present at the time they read the materials, and some print materials had little information about the ACWS features of interest in this experiment). There is a need for effective quick-overview materials that convey what safety systems are in the vehicle, how status is indicated, and how they operate. 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. 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).
- *Further research questions*: The present experiment investigated individuals' understanding of ACWS status displays. Consequently, there are a variety of potential research questions that are worth pursuing. Several potential research questions are:
  - What are the population expectations for message content, color coding, and display format (e.g., icons, text only, etc.)?
  - What are the population expectations regarding location of status display information?
  - Would there be a better understanding of comprehension and system confusion/advantages if individuals were recruited who actually drove one vehicle versus another instead of viewing photographic displays of the ACWS interface?
  - Would comprehension of the status displays change if the participants had become familiar with ACWS by driving an equipped vehicle over an extended period of time?

- *Methodology and evaluation protocol*: There are also several implications for further methodological refinement of the current task, design, and development of evaluation methods for determining the comprehensibility of status displays:
  - It would be informative to experimentally compare specific design strategies. The present experiment selected vehicles based on general design tendencies (e.g., use of text for status versus icons/color), but these designs were not experimentally manipulated. Further investigation would hold other design characteristics constant and focus on manipulating the design factors of interest, thereby providing more cause-effect information regarding comprehension effects in participants.
  - The current experiment found that people were not particularly enlightened after exposure to information materials (e.g., quick start guides), although this may

have occurred in part because the materials did not always include relevant information. It should also be noted that participants were not able to read the short manuals while also interacting with the vehicles. This may have influenced learning. On the other hand, having incentives to read the short manuals closely may have actually augmented comprehension beyond what is found in many individuals who do not read any materials before operating a vehicle. Consequently, it is not truly possible to evaluate the information materials for quality and conveying information, so another investigation could develop and evaluate prototype informational materials (print and/or video).

- The status display experiment used photographic mockups of actual vehicle interiors with near life size displays. Although this seemed sufficient for responding, ecological validity would have been increased if the current task took place within actual vehicles.

6 Conclusion

Task 3 provided the opportunity to conduct new empirical research to address implications of DVI variability for ACWS features. The effort attempted to determine the presence and extent of driver problems that may be associated with DVIs and the variation among them in different vehicles. There has been very little research on this issue for ACWS and therefore it is important to determine whether a meaningful problem exists. Based on the findings of these experiments, taken together with other literature, implications were drawn and are discussed in detail. The findings suggest that driver understanding of system status can be affected positively and negatively based on the design approach used to communicate important information about the ACWS. Rare alerts (such as an FCW crash warning or a LDW malfunction message) should be self-explanatory to the driver, especially in urgent situations. The current study indicates there is a need to further investigate the effects of DVI variability on driver behavior and comprehension.

References

Bligard, L.O., & Osvalder, A.L. (2007). An analytical approach for predicting and identifying use error and usability problem. In *HCI and Usability for Medicine and Health Care, Lecture Notes in Computer Science, 4799,* 427–440.

Campbell, J.L., Richard, C.M., Brown, J.L., & McCallum, M. (2007). *Crash warning system interfaces: Human factors insights and lessons learned* (HS 810 697). Washington, DC: National Highway Traffic Safety Administration.

Edworthy, J., Loxley, S., & Dennis, I. (1991). Improving auditory warning design: Relationship between warning sound parameters and perceived urgency. *Human Factors, 33*(2), 205-231.

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., & 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.

Haas, E.C., & Edworthy, J. (1996). Designing urgency into auditory warnings using pitch, speed, and loudness. *Computing & Control Engineering Journal (August 1996)*, 193-198.

Hellier, E.J., Edworthy, J., & Dennis, I. (1993). Improving auditory warning design: Quantifying and predicting the effects of different warning parameters on perceived urgency. *Human Factors, 35*(4), 693-706.

Lerner, N, Jenness, J., Robinson, E., Brown, T., Baldwin, C., & Llaneras, R. (2011). *Crash warning interface metrics. Task 12 final report.* Washington, DC: National Highway Traffic Safety Administration.

Mahatody, T., Sagar, M., & Kolski, C. (2010). State of the art on the cognitive walkthrough method, its variants, and evolutions. *International Journal of Human-Computer Interaction, 26(8),* 741-785.

Marshall, D.C., Lee, J.D., & Austria, P.A. (2007). Alerts for In-Vehicle Information Systems: Annoyance, Urgency, and Appropriateness. *Human Factors, 49*(1), 145-157.

Patterson, R.D. (1982). *Guidelines for Auditory Warning Systems* (CA Paper 82017). London: UK: Civil Aviation Authority.

SAE International (2010). *Surface vehicle standard J2402: Road vehicles – Symbols for controls, indicators, and tell-tales.* Society of Automotive Engineers.

DOT HS 811 470b
August 2011

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

NHTSA
www.nhtsa.gov