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

NHTSA
www.nhtsa.gov

DOT HS 812 208

November 2015

# Driver-Vehicle Interfaces for Advanced Crash Warning Systems: Research on Evaluation Methods and Warning Signals

## Disclaimer



Suggested APA Format Citation:

Lerner, N., Singer, J., Huey, R., Brown, T., Marshall, D., Chrysler, S., … & Chiang, D. P. (2015, November). *Driver-vehicle interfaces for advanced crash warning systems: Research on evaluation methods and warning signals.* (Report No. DOT HS 812 208). Washington, DC: National Highway Traffic Safety Administration.

## Technical Report Documentation

| 1. Report No. | 2. Government Accession No. | 3. Recipient's Catalog No. |
|---|---|---|
| DOT HS 812 208 | | |
| **4. Title and Subtitle**<br>Driver-Vehicle Interfaces for Advanced Crash Warning Systems: Research on Evaluation Methods and Warning Signals | | **5. Report Date**<br>November 2015<br>**6. Performing Organization Code** |
| **7. Authors**<br>Neil Lerner, Jeremiah Singer, Richard Huey, Timothy Brown, Dawn Marshall, Susan Chrysler, Rose Schmitt, Carryl L. Baldwin, Jesse L. Eisert, Bridget Lewis, Ana I. Bakker, Dean P. Chiang | | **8. Performing Organization Report** |
| **9. Performing Organization Name and Address**<br>Westat, Inc.<br>1600 Research Boulevard<br>Rockville, Maryland 20850-3129 | | **10. Work Unit No. (TRAIS)**<br>**11. Contract or Grant No.**<br>DTNH22-11-D-00237 |
| **12. Sponsoring Agency Name and Address**<br>United States Department of Transportation<br>National Highway Traffic Safety Administration<br>1200 New Jersey Avenue SE.<br>Washington, DC 20590 | | **13. Type of Report and Period Covered**<br>Final Report / September 2011 - August 2014<br>**14. Sponsoring Agency Code** |
| **15. Supplementary Notes**<br>Eric Traube (COTR/TO) | | |

**16. Abstract**

The Crash Warning Interface Metrics (CWIM) program addresses issues of the driver-vehicle interface (DVI) for advanced crash warning systems. This report summarizes the methods and findings of 16 experiments grouped under three main research areas: (1) *Research toward collision warning and lane departure warning protocol development* explored methodological aspects of warning DVI evaluation, including basic method (test track, simulator), incentive structure, distraction task, alert timing, training and pre-exposure to systems, and simulator motion fidelity. (2) *Research on variability among warning signals* included three main topic areas. It continued an earlier line of research investigating the potential for negative transfer of learning when switching between vehicles with different forward collision warning (FCW) DVIs. It also included a series of experiments on categorical perception of warnings and alerts, which cumulatively defined signal parameter criteria that effectively delineate alarms from less urgent messages. Finally, this research area investigated acoustic signal detectability and perception under varied ambient noise conditions during real driving. This experiment found that signal detectability, perceived urgency, and perceived meaning can be substantially impaired under conditions of elevated ambient noise. (3)*Temporal aspects of interference from other in-vehicle messages* included two driving simulator experiments that investigated the effects of a non-urgent alert occurring before an urgent crash warning on driver behavior. The temporal gap between alert and warning was systematically varied, as were the modalities of the early alert and the warning. The first experiment involved undistracted drivers; the second involved drivers who were distracted at the time of a critical event. The experiments found that non-urgent messages that preceded warnings did not necessarily slow responding to the warning, though longer time gaps tended to result in faster reaction times than shorter time gaps. The experiments also provided some evidence that warning response is impaired when the early alert and warning share the same perceptual mode. Implications for DVI design were drawn from the findings of these experiments. Formal FCW and LDW evaluation protocols were developed as methods for assessing DVIs for production or near-production systems.

| 17. Key Words | 18. Distribution Statement |
|---|---|
| human factors, driver vehicle interface (DVI), advanced crash warning system (ACWS), connected vehicle (CV), distraction, forward collision warning (FCW), lane departure warning (LDW) | Document is available to the public from the National Technical Information Service www.ntis.gov. |

| 19. Security Classif. (of this report) | 20. Security Classif. (of this page) | 21. No. of Pages | 22. Price |
|---|---|---|---|
| Unclassified | Unclassified | 141 | N/A |

**Form DOT F 1700.7** Reproduction of completed page authorized.

# Acknowledgements

The authors thank the following individuals who contributed to the research efforts described in this report. From Westat: Michael Gill, Jeremy Walrath, Daniel Kellman, Luis Romero, and Diane Snow. From George Mason University: Andre Garcia. The authors also thank Eric Traube (NHTSA contracting officer's technical representative for this project) for his guidance and many contributions to the work, as well as Eric Nadler (Volpe National Transportation Systems Center) for his technical review and insights.

# List of Acronyms and Abbreviations

| | |
|---|---|
| ACWS | advanced crash warning system |
| ANOVA | analysis of variance |
| CD | compact disk |
| CV | connected vehicle |
| CWIM | crash warning interface metrics |
| dB | decibels |
| DRI | Dynamic Research, Inc. |
| DVI | driver-vehicle interface |
| EMRT | evasive maneuver response time |
| FCW | forward collision warning |
| ft | feet |
| g | gravitational force |
| GLM | general linear model |
| GMU | George Mason University |
| HFCV | Human Factors for Connected Vehicles |
| Hz | Hertz |
| IBI | interburst interval |
| IPI | interpulse interval |
| IEC | International Electrotechnical Commission |
| LCD | liquid crystal display |
| LDW | lane departure warning |
| m | meter |
| ms | millisecond |
| NADS | National Advanced Driving Simulator |
| NHTSA | National Highway Traffic Safety Administration |
| OEM | original equipment manufacturer |
| s | second |
| SDLP | standard deviation of lane position |
| SOA | stimulus onset asynchrony |
| SPL | sound pressure level |
| TTC | time to collision |
| UI | University of Iowa |
| V2I | vehicle-to-infrastructure |
| V2V | vehicle-to-vehicle communication |
| VRTC | Vehicle Research and Test Center |

# Table of Contents

**Appendices (under separate cover):**

# List of Figures

# List of Tables

# Executive Summary

The Crash Warning Interface Metrics (CWIM) program addresses issues of the driver-vehicle interface (DVI) for Advanced Crash Warning Systems (ACWS). ACWS are increasingly common in passenger vehicles and the characteristics of these systems, including the DVI, vary considerably among vehicle makes and models. A DVI for crash avoidance situations must be rapidly and reliably detected. It must be quickly comprehended and should not be confused with other messages. It must draw driver attention to the threat. It must result in a rapid, appropriate, and accurate vehicle control response. Given the diversity of ACWS and their associated DVIs, there is a need for research on the design and evaluation of ACWS DVIs.

This report documents the methods and findings of research conducted under the NHTSA project "Crash Warning Interface Metrics – Phase 3." This project builds upon the findings of previous CWIM work, which was conducted under an earlier contract. The objectives of this Phase 3 project included the following:

- Develop test protocols for FCW and LDW driver interfaces that can be used to evaluate production or near-production products
- Provide implications for crash warning interfaces, based on CWIM research findings
- Identify concerns with inconsistency of warning interface features among vehicles and means of addressing these concerns

To achieve these objectives, three parallel lines of research (encompassing 16 distinct experiments) were conducted:

- Research toward FCW and LDW evaluation protocol development
- Research on variability among warning signals
- Research on temporal aspects of interference from other in-vehicle messages

<u>Research toward FCW and LDW evaluation protocol development</u>

This line of research included five experiments on evaluation protocol features; an experiment investigating ACWS research methods in the Connected Vehicle (CV) vehicle-to-vehicle communication (V2V) context; and an investigation of methods to conduct LDW DVI evaluations on a test track or actual roads. This research was methodological in nature, addressing how to assess DVIs rather than how to design them.

The first experiment replicated a previous test track evaluation of driver response to surprise FCW events (Forkenbrock et al., 2011) using the NADS-1 simulator. The objective of this experiment was to determine whether drivers' visual commitment to a distraction task or FCW event outcome differed between the two methods. Participants in the simulator study were randomly assigned to one of three warning modality conditions (no warning, audio, seat belt tensioner). Results showed that findings between the test track and simulator methodologies were generally comparable, and that the more controlled environment of the simulator allowed for somewhat more precise control over FCW event timing.

The second experiment investigated the effects of various incentive structures and distraction tasks on drivers' commitment to the distraction task. Researchers hypothesized that increasing financial incentives for completing secondary tasks would increase commitment to the secondary task. Providing no incentive might lead to insufficient commitment, but a high incentive might lead to too much incentive, leading drivers to ignore crash warnings. Three incentive structures were used, although total potential compensation was the same for all groups: high base pay with low additional financial incentives, low base pay with high additional financial incentives, and fixed-sum compensation with no incentive component. Four distraction tasks were used: a number recall task, a "bug" task (manually tracing a moving target on a screen), a message reading task, and a menu selection task. All of the distraction tasks were designed to encourage the driver to look away from the forward roadway for a long enough time to allow the FCW event to occur. Distracting drivers' attention from a developing hazard for long enough for a warning to occur is critical to be able to evaluate driver responses to a DVI. The number recall distraction task had the longest mean time from task start to end of first visual commitment as well as low variance between participants, which suggests that it is effective in distracting drivers during the initiation of an FCW event. The number recall task also resulted in relatively high lane position stability, which is beneficial in maximizing experimental control and consistency of FCW event staging. Results showed that there was no statistically significant effect of incentive on drivers' visual commitment.

The third experiment examined the effects of training and familiarity with FCW and LDW systems on warning response. Researchers hypothesized that providing training and familiarity might lead to lower variability in warning responses, which would be beneficial for protocol sensitivity to differences in alert characteristics. Providing participants with awareness that the crash warning systems were present in the car through a pre-drive training presentation resulted in no statistically significant differences in engagement with the secondary task or driver response for the lane departure events. Prior exposure to the warning systems tended to result in longer visual commitments to the distraction task, which is methodologically useful in allowing more time for crash warning events to materialize. Regarding the use of more than one crash warning event during a participant's session, results suggest that participants' engagement with the distraction task is little changed after the first LDW event, but that participants' responses to warnings are improved for the second event. For FCW events, drivers are often more cautious after the first event.

The fourth experiment investigated the effects of warning onset timing. Results suggest that late LDW onset is advisable for DVI evaluation purposes to minimize nuisance alarms, which negatively influence participants' impressions of the system. Warning timing had little effect on participants' behavior during FCW events.

The fifth experiment investigated the effects of three levels of motion fidelity (full, partial, none). Full motion fidelity provides accurate vestibular motion feedback to drivers, whereas partial motion fidelity provides limited vestibular feedback. For LDW events, the no-motion condition resulted in shorter secondary task engagements, but larger lane exceedances. The partial-motion condition also resulted in shorter secondary task engagements, but lane exceedances were smaller than in the no-motion condition. For FCW events, a number of interactive effects were observed, which makes definitive determinations difficult. Based upon the results of this study, partial or no motion is likely acceptable for evaluations of relative DVI performance, but the complex interactions in FCW events point to the need to match the real world as closely as possible.

The sixth experiment evaluated driver response to warnings when participants were engaged in a distraction task based on a prototypical Connected Vehicle (CV) task. This task involved a visual display located near the forward field of view and did not require long glances away from the roadway. This task was found to be insufficiently distracting to allow for FCW and LDW events to develop unnoticed, and was therefore unsuitable for use as a distraction task for ACWS DVI evaluation.

The final effort in this line of research was an exploration of methodologies to allow the evaluation of LDW DVIs on a test track or road. Various approaches were explored and implemented, but none were able to produce lane departures (that would be unnoticed by a distracted driver) quickly and reliably enough to suggest LDW evaluation protocols using on-road methods.

Research on variability among warning signals

In an earlier phase of the CWIM research program (Robinson et al., 2011), substantial "negative transfer" (slower response) was observed among distracted drivers when they became familiar with one auditory FCW "A1" and then encountered an unfamiliar auditory FCW "A2" in a "different" simulated vehicle. In a follow-on experiment conducted in the present CWIM Phase 3 research effort, the evidence was less straightforward. Some evidence for negative transfer was revealed; however, it was primarily observed in only one of the two main switch conditions where there were relatively few other sounds in the environment. Still, the most successful crash avoidance response was observed when FCW sounds remained stable across three days of exposures.

In addition to the negative transfer experiment, a series of investigations was conducted to determine the key characteristics of a warning sound that enable the sound to be distinctly perceived as representing a highly urgent FCW. Three methodologies were implemented to specifically examine the range of key parameters that result in a majority (85%) of listeners perceiving a given sound as an urgent collision warning.

Sounds derived from a survey of existing OEM FCWs and notifications were compiled. Each sound's acoustic properties (e.g., spectral frequency components, pulse duration, onset and offset, harmonics, interpulse interval) were analyzed. From this inventory, as well as existing published guidelines (e.g., Campbell, Richard, Brown, & McCallum, 2007), a list of key parameters was developed and examined in the following experiments.

In one set of experiments participants sorted each sound into its perceived category (e.g., alarm, status alert or social notification). Regression techniques were used to identify key characteristics that accounted for a substantial amount of the variance in category classifications. Cutoff scores for each of these key parameters were then established by examining the values of each characteristic that resulted in a high probability (defined as 85% in the research plan) of classification as an alarm (warning plus alert). Specially, a sound that had a "peak to total time ratio" (ratio of the time that the sound is at full intensity to the entire pulse duration including onset and offset) of greater than or equal to .7, an interburst interval (IBI) of less than or equal to 125 ms, at least 3 harmonics and a base frequency of greater than or equal to 1000 Hz was classified as an alarm over 90% of the time. Meeting only the first of these criteria – ratio greater than or equal to .7 -- resulted in an alarm classification rate of 70%. These criteria fit classification performance well. However, only one sound actually met all four criteria and the criteria were established by the same data. Therefore, two additional sort tasks were completed with additional sounds and new participants. Results of these sort experiments indicated that the criteria and cutoffs established in the first experiment predicted FCW classification equally well with over 90% classification accuracy in alarm category if all four criteria were met.

Additionally, using alternative methodologies (method of adjustment with single and multiple parameters) yielded results that confirmed several of the parameters (base frequency, presence of at least three harmonics). Though not specifically manipulated, the Ratio parameter was also met by the resulting sound created in each of the method of adjustment procedures. A final experiment in the series examined the predictive capabilities of these criteria while participants were engaged in a simulated driving task and concurrently performing a distraction task consisting of a visual version of an n-back task. The criteria had strong predictive capabilities under these conditions. Sounds that met all four of the criteria were classified as alarms more than 90% of the time. Further, when exposed to sounds meeting all four criteria participants were able to make classifications faster than when sounds met three or fewer of the criteria.

The next area of investigation within this line of research was an experiment to evaluate auditory alerts in real ambient noise conditions. This research is described in greater detail in a separate stand-alone report (Singer, Lerner, Walrath, & Gill, 2014). For this experiment, 15 warning and alert sounds were selected for investigation. Sounds included tonal signals and voice messages, which were presented at approximately 65 dB (A-weighted). A subset of these sounds was also presented at 75 dB. The experiment was a three-factor design, with one between-groups factor (vehicle type) and two within-groups factors (interior noise condition, acoustic signal). Three different vehicles were used in the experiment in order to provide a representative range of vehicle types: (1) a small car, (2) a larger sedan, and (3) an SUV. Participants drove the test vehicle on a limited-access highway at approximately 60 mph. While driving, they were occasionally presented with an alert sound. All sounds were presented under three different ambient noise conditions: (1) baseline, (2) front windows down, and (3) music playing. Upon hearing a sound, participants pressed a button and then rated the sound on noticeability, perceived urgency, message categorization, and, for voice messages, intelligibility.

Results show significant main effects for ambient noise condition and for alert sound for noticeability, urgency, and response time. While all alerts were detected at high rates under the baseline (quiet vehicle interior) condition, detection was impaired by the presence of music, and even more so when the front windows were open. Sounds presented at the higher 75 dB level, however, were generally detected even in the noisier conditions. Ratings for alert noticeability and urgency were also lower under the music on and windows open conditions. Ambient noise condition also affected alert categorization. Alerts that were likely to be categorized as urgent crash warnings under baseline conditions were less often categorized as urgent crash warnings under conditions of elevated ambient noise. Importantly, there were also interactions between ambient noise condition and alert sounds, indicating differences in how well sounds of similar loudness tolerated interference from noise. In other words, ambient noise conditions affected different alert sounds in different ways.

Research on Temporal Aspects of Interference from other In-Vehicle Messages

The two experiments conducted in this line of research investigate the effects of the time gap between an early non-urgent alert and a FCW, known as stimulus onset asynchrony (SOA), and the modalities of the early alert and the FCW. Previous research suggests that an early alert in close proximity to a warning can impair reactions to the warning (Wiese and Lee, 2004; Hibberd, Jamson,, & Carsten, 2013). Early alerts were either verbal or tonal, and FCWs were either auditory or haptic. Both experiments were conducted in a driving simulator. The first experiment involved drivers who were not explicitly distracted. The second experiment followed the same general procedure, but drivers were engaged in a distraction task when lead-vehicle braking (i.e., forward collision) events were initiated.

The experiments found that non-urgent messages that shortly preceded FCWs did not necessarily slow responding to the FCW. In many cases, the preceding message actually improved the speed of returning attention to the road or initiating braking in response to the FCW. However, observed effects were complex and depended on the mode of the warning, the mode of the preceding alert, and the duration of the interval between the two signals. Haptic FCW signals led to faster responding than tonal signals and this relative advantage was generally more pronounced when there was only a brief (150 ms or 300 ms) interval between the preceding alert (verbal or tonal) and the FCW. When the FCW was auditory (pulsing tone), responding was generally faster when the preceding alert was verbal rather than tonal at the briefer (150 ms or 300 ms) intervals. These relative performance results provide some support for theories of resource competition, whereby tasks that share the same perceptual and/or response mode compete more than when these modes are distinct. However, the anticipated finding of absolute decrements in performance due to preceding tasks at brief intervals (150 ms, 300 ms) was not observed. Performance was generally enhanced, relative to no preceding message or to shorter intervals, when the interval was 1 second.

Overview of DVI Evaluation Protocols

A key product of this project is a set of preliminary evaluation protocols for DVIs of FCW and LDW systems. The purpose of the protocols is to define well-specified, repeatable procedures that will allow a common method for evaluating the DVI of a production or near-production system. The intent is not to assess a vehicle's warning system, per se, but specifically the driver interface portion of that system. The protocol efforts under the CWIM project were intended to evaluate the ability of the DVI to engender rapid and appropriate driver responses, regardless of other aspects of the system. The protocols were developed for auditory and haptic signal evaluation. While visual signals may be included as components of warnings, they are generally not appropriate for single-mode presentation of time-critical warnings (Campbell et al., 2007), and were therefore not included in this research, given the wide range of important research issues related to auditory and haptic signals. The protocols developed in this project represent a combination of direct empirical evidence from project research; other literature; and expert judgment. Because the protocols have not yet been subjected to peer review, notice and comment, and/or direct testing, these specifications are preliminary.

Three separate, though related, protocols were developed: An FCW driving simulator protocol; an LDW driving simulator protocol; and a FCW test track protocol. The protocols themselves are provided as a separate deliverable product of this project, under separate cover.

The three protocols share a common broad approach, but numerous procedural details must be specified if the procedure is to be repeatable by a range of organizations. The protocols spell out these specific details, with supporting rationale. Each of the three protocols includes the following sections:

- Participants: age, gender, sample size, inclusion/exclusion criteria
- Method: ruse, recruitment, incentive, training
- Apparatus: simulator/test track, secondary task, headway display (if applicable)
- Simulator/test track scenario: event timing and choreography, driving environment, practice and exposure to DVI
- Post data collection procedure: video coding, secondary task interaction
- Data handling: dependent measures, data reduction, data analysis

Implications for the Design and Implementation of Crash Warning DVIs

A number of implications for ACWS DVI design were derived from the findings of the various experiments under the CWIM project. It is important to note that these implications were drawn solely from the CWIM project findings and are not intended to serve as explicit recommendations or guidance. The various implications are grouped under a set of design goals for a particular topic. The format is that there is first a topic heading, followed by a high-level design goal. This is followed by a set of specific CWIM project findings related to the general goal. After the findings, there is a section on "rationale and discussion." This section explains the empirical and analytic basis of the conclusions. It also provides a discussion of the limitations, caveats, additional research needs, and other issues associated with the conclusions.

The implications are grouped under the following topic headings:

- Compatibility of auditory alert features with intended message meaning categories
- Maintenance of auditory alert meaning under in-vehicle noise conditions
- Haptic signal applicability for FCW
- LDW alert modality
- Warning signal response disruption by preceding non-urgent alerts
- Conveying information on the presence and status of ACWS

# 1 Background and Objectives

This report documents the methods and findings of research conducted under the NHTSA project "Crash Warning Interface Metrics – Phase 3." As the title suggests, it continued a line of work begun under a previous contract. Findings on the earlier phase of this CWIM line of work may be found in three reports:

- *Crash Warning Interface Metrics Final Report*, DOT HS 811 470a (Lerner, Jenness, Robinson, Brown, Baldwin, & Llaneras, 2011)
- *Crash Warning Interface Metrics Task 3 Report: Empirical Studies of Effects of DVI Variability*, DOT HS 811 470b (Robinson et al., 2011a)
- *Crash Warning Interface Metrics Task 3: Report Appendices*, DOT HS 811 470c (Robinson et al., 2011b)

## 1.1 Background

The CWIM program addresses issues of the driver-vehicle interface (DVI) for Advanced Crash Warning Systems (ACWS). ACWS have the potential to improve driver performance and reduce the frequency and severity of common crash situations. ACWS are vehicle systems that use sensors to assess potential or emerging hazard situations and provide crash warning information to the driver. In some cases the system may also initiate some vehicle control action (e.g., braking, steering). Forward collision warning (FCW) and lane departure warning (LDW) are among the more common ACWS and are the focus applications within this project. However, additional ACWS functions are available (e.g., pedestrian collision warning, lane change/blind spot warning) and emerging technologies, such as V2V and V2I Connected Vehicle (CV) concepts (http://www.its.dot.gov/connected_vehicle/connected_vehicle.htm), will permit additional types of warnings (e.g., emergency electronic brake lights). ACWS are increasingly common in passenger vehicles and the characteristics of these systems, including the DVI, vary considerably among vehicle makes and models.

The success of any warning system will depend in part on the properties of the DVI and how drivers respond to it. The DVI refers to the displays and controls through which the driver and the system interact. Important attributes include the warning mode (e.g., visual, auditory, haptic); warning display content; signal conspicuity; display location; timing; reliability (e.g., minimal false alarms and missed detections); active intervention in vehicle control aspects; and relation to (and integration with) other systems and displays. The DVI can also include information about the status of warning systems, so that the driver comprehends what safety functions are present in the vehicle and their current operational status (e.g., malfunction). ACWS features vary considerably among vehicle makes and models along all of these dimensions. Furthermore, the automobile industry is developing various new and innovative interfaces that go beyond the traditional dash or console visual displays and tonal signals. For example, these include haptic signals (e.g., seat vibration), head-up displays, voice commands, and light panels, among others.

Given the diversity of DVI attributes, the question arises as to whether drivers perform equally well, or at least adequately, in terms of response time and appropriateness with any particular DVI. Addressing this question is complicated by the fact that driver response to a particular ACWS DVI is not a function of the DVI alone. For example, systems dealing with the same function (e.g., FCW) may differ in terms of operational aspects, such as triggering criteria or operational conditions. Some systems may include cautionary alerts or status displays related to the warning function. More generally, a given function operates within a broader system of warnings, alerts, communications, and displays. This could influence how the driver responds to any particular element of that system. Vehicle cockpits also differ in terms of layout, complexity, and interfaces. Vehicle handling and operational performance may also influence crash avoidance. Some vehicles also include automatic braking, emergency brake assist, or other automated features that aid in crash avoidance. Given all of this, it is evident that although the ACWS DVI is a critical element of a warning system, it is only one aspect of that system. This distinction is important, since this project is focused on the DVI. It is not intended to compare any particular systems in terms of predicted crash reduction benefits. Rather, it is concerned with assuring that the DVI component of a system leads to comprehension and appropriate driver actions.

To be considered effective, a DVI for crash avoidance situations must have a number of attributes. It must be rapidly and reliably detected. It must be quickly comprehended and should not be confused with other messages. It must draw driver attention to the threat. It must result in a rapid, appropriate, and accurate vehicle control response. The assessment of these factors is complicated, and to date there is no widely accepted procedure for evaluating DVIs and determining their effectiveness.

<u>Important Note</u>**: Because this project focused specifically on the DVI rather than the ACWS as a whole, DVI effectiveness, as defined in this report, does not directly refer to crash reduction or mitigation. Rather, effectiveness refers to the ability of the DVI to influence attention, comprehension, or behaviors in a way that is appropriate for the hazard context.**

Another potential concern with the effectiveness of ACWS DVIs comes from the diversity of the DVIs themselves. As noted above, vehicles may differ greatly in terms of the DVI for a common function. Displays may differ in modality, location, content, timing, and other aspects. It is possible that the DVI for any particular vehicle may meet good design criteria and test well on its own yet there may be driver performance problems due to the diversity itself. Drivers accustomed to a warning in one vehicle may react slowly or inappropriately when they encounter a different DVI in another vehicle. It is even conceivable that the same display (e.g., a given tonal signal) may have different meanings in different vehicles (Robinson et al., 2011). Because imminent crash warnings may only occur rarely and under urgent circumstances, drivers may not experience differences until a critical event occurs. The potential for negative transfer therefore exists: Experience with one system's DVI may subsequently interfere with rapid and appropriate responding in a vehicle with a different DVI. Therefore when considering the design and evaluation of ACWS DVIs, potential problems emerging from DVI diversity also need to be considered.

## 1.2 Previous CWIM Activity

As noted above, the work conducted under this project continued a general line of work initiated previously. Previous work included the following:

- A review of literature related to methods and measures
- Development of a taxonomy of response metrics for crash warning interface dependent measures
- Solicitation of industry, stakeholder, and public feedback on project objectives and approaches based on a Federal Register notice, internal meetings with representatives of various NHTSA offices, meetings with industry groups and individual companies
- Initial identification of criteria for selecting evaluation methods and procedural factors that need to be addressed by a protocol

- Driving simulator experiments using preliminary evaluation methodologies to compare warnings in different sensory modalities
- Driving simulator research examining potential negative transfer effects when participants encounter a change in warning interface when driving a "different" vehicle than the one they were made familiar with
- A laboratory simulation of various vehicles' crash warning system status display interfaces to identify user problems in the comprehension of advanced safety system features and status

This previous work is documented in the reports cited in Section 1.0. In addition to the research carried out under the previous contract, a complementary project, conducted by NHTSA's Vehicle Research and Test Center (VRTC), demonstrated a test track methodology for use as the basis of an FCW interface evaluation protocol (Forkenbrock et al., 2011).

Among the notable findings and concerns that emerged from this previous work were the following:

- There is mixed automotive industry support for a standardized evaluation protocol. There is strong industry resistance to standardizing the interface design itself, largely because auto makers want the ability to design systems and DVIs and they see fit. There is also industry concern about evaluation of a specific warning function (e.g., FCW) within a broader system context.
- "Non-traditional" interfaces have promise for crash warning applications and a protocol procedure will need to be capable of encompassing all of these displays and sensory modes, including various haptic stimuli.
- It is necessary to divert driver visual attention away from the forward roadway so that the measured response is to the warning and not to visual detection of the emerging event that is being warned about. However, it is difficult to identify secondary tasks which meet all desirable criteria and that reliably result in removing gaze from the forward view for an adequate amount of time. Candidate distraction tasks were identified.
- "Straw man" evaluation protocol characteristics were produced based on existing literature, practice, and expert judgment. This exercise highlighted a number of key unresolved issues and shortcomings in current ability to generate empirically-based and defensible protocol requirements.
- Specific evaluation protocol methodological issues identified for resolution included the following:
    - Driving scenario
    - Participant sample characteristics
    - Participant familiarity with the technology
    - Means of distracting the driver
    - Controlling driver expectancy
    - Dealing with warning system context
    - Accommodating user settings and options
    - Comparison conditions and benchmarks
    - Treatment of data
    - General test method
- Negative transfer effects potentially may be quite large. The negative transfer experiment described in Robinson et al. (2011) was conducted over three sessions. In the first two sessions, participants became familiar with an FCW interface. In the third session,

in which the simulated vehicle was "changed" to an ostensibly new vehicle, the FCW sound changed for half the participants and remained the same for the other participants. Two FCW sounds were used. The delay in responding to an unfamiliar FCW sound was very large in one direction of transfer (on the order of about a half-second slower) but much smaller in the other direction. The basis for this difference, and the robustness of the finding, is not clear.

- People have generally poor comprehension of vehicle displays and controls related to ACWS and other vehicle safety systems. Displays generally do not convey well the existence and mode of operation of vehicle safety features and do not convey well the status (on/off, currently operational) of functions.
- Manufacturer-provided materials are generally not adequate to convey information related to advanced safety system operation and displays, even if they are read by the consumer.
- The negative transfer findings, as well as the problems of driver comprehension of functions and status, suggest a possible benefit for at least minimum specifications or boundary conditions in the absence of a standardized warning interface.

## 1.3 Objectives

The objectives of this project included the following:

- Develop test protocols for FCW and LDW driver interfaces that can be used to evaluate production or near-production products
- Derive implications for design principles that promote effective crash warning interfaces, where effectiveness refers to the effect of the warning DVI on driver attention, comprehension, and behavior rather than direct measures of crash reduction or mitigation.
- Identify concerns with inconsistency of warning interface features among vehicles and means of addressing these concerns

The research conducted to meet these objectives therefore was designed both to evaluate the effects of DVI features on driver performance and to evaluate the effects of various research methods and procedural elements. The scope of this research was limited to human factors considerations of in-vehicle warning DVIs. Broader issues of ACWS performance and parameters were not directly addressed.

# 2 Summary of CWIM Phase 3 Research Activities

Under the current project, sixteen research experiments were carried out covering a range of issues. Some of these explicitly carried forward work begun under a previous contract and some were new lines of work. For convenience, these experiments are grouped under three broad categories:

- FCW and LDW evaluation protocol development
- Dealing with variability among warning signals
- Temporal aspects of interference from other in-vehicle messages

Table 1 lists all of the experiments, with summary information. Taken together, the experiments were intended to provide methods for evaluating commercially available or near-production crash warning interfaces, providing guidance on aspects of crash warning interface design, and providing insight on factors to maintain warning effectiveness under real-world application conditions (e.g., within the context of broader messaging systems, within temporal proximity to potentially distracting alerts, and within noisy vehicle cab environments).

Given the large number of experiments included in the project, this final report is not able to provide extensive detail on the methods and analyses for individual experiments. The sections on each experiment focus on major methodological aspects and key findings. Greater detail on individual experiments may be found in separate appendices, as indicated in the Table of Contents.

The FCW and LDW evaluation protocol development experiments addressed the goal of developing well-specified test procedures for the objective evaluation of production-level warning system interfaces, with a focus on auditory and haptic signals. These experiments were not directed at methods for designers to use in developing systems but rather at formal procedures for evaluating interfaces for systems that are already developed (production or near-production vehicle systems). Five experiments were conducted to provide an empirical basis for a number of basic protocol aspects, such as participant incentive structure or driving simulator motion feedback requirements. A critical goal of this line of research was to choose a distraction task and related protocol features that would result in valid, reliable, and repeatable distraction from the roadway for a long enough time to allow a threat to develop and a warning to be presented without the driver first becoming aware of the threat. All of these studies were conducted using the NADS-1 driving simulator at the University of Iowa. Another experiment, also conducted using the NADS facility, considered how the protocol might be adapted for application in the CV environment (in which V2V and vehicle-to-infrastructure [V2I] communications, as well as information from other sources, might result in a more information-rich environment for the driver). In addition to testing warnings independently, it is important to test them in this information-rich context while participants engage in realistic secondary tasks. A final study in the "protocol" line considered whether it might be feasible to adapt the simulator procedures for LDW interface evaluation to an on-road methodology. There are certain advantages to on-road methods that could be important for LDW but the reliable inducement of lane departures without driver awareness is necessary for a well-controlled evaluation. This study attempted to develop practical on-road methods.

The group of experiments dealing with variability among warning signals was based on the likely case that there will not be a single, standardized FCW (or other) warning interface common to all vehicles. This presents concerns about whether some minimum requirements may be desirable and whether the variability among vehicles itself may be a problem. The negative transfer experiment continued research from a previous contract. The concern is that even if various manufacturers each design "effective" warning interfaces, the driver confusion or hesitation associated with differences among systems could lead to delayed or inappropriate driver responses to a warning. The set of experiments on auditory warning perceptual space dealt with the warning signal parameters that result in the perception that a particular sound is, or is not, subjectively experienced as an urgent warning. The intent was to provide a basis to ensure that even if different signals are used in different vehicles, they are naturally interpreted as being the "same" in terms of general meaning. The final experiment that dealt with variability among warning signals considered the in-vehicle noise conditions under which real-world warnings might be experienced. Not all warnings that appear effective under moderate vehicle noise listening conditions may be equally effective under higher noise conditions. This experiment investigated driver perception of warnings during actual on-road driving under several background noise scenarios.

The two experiments on near-simultaneous safety alerts and crash warnings concerned the timing of warning signals relative to other non-warning alerts. The two experiments used similar methods. In a driving simulator, the time interval between a non-warning message and a warning signal was varied from very brief (150 ms) to relatively long (1000 ms) and the effects on driver response were compared to a "no prior message" baseline. The first experiment did not include a specific distraction task to distract the driver at the moment the messages occurred. The second experiment did include a visual distraction task. The experiments included different early message modes (verbal or tonal) and different FCW modes (auditory or haptic).

This overview of the experiments conducted within this project provides a context for interpreting each of the individual experiments. Sections 3, 4, and 5 provide further detail on the methods and findings of each experiment. Section 3 describes the experiments related to protocol development. Section 4 describes the experiments related to variability among warning signals. Section 5 describes the experiments on near-simultaneous safety alerts and crash warnings. More extensive reports on individual experiments are provided as appendices under separate covers.

**Table 1. Summary of research conducted**

| | Experiment | Method | Site |
|---|---|---|---|
| **Research Toward FCW and LDW Evaluation Protocol Development** | | | |
| Basic Protocol Features | | | |
| | 1. Simulator comparison with test track methodology | Driving simulator | U of Iowa-NADS |
| | 2. Secondary task and incentives | Driving simulator | U of Iowa-NADS |
| | 3. System training and exposure | Driving simulator | U of Iowa-NADS |
| | 4. Alert timing and modality | Driving simulator | U of Iowa-NADS |
| | 5. Simulator motion feedback | Driving simulator | U of Iowa-NADS |
| Evaluation in the Connected Vehicles (CV) Context | | | |
| | 6. Crash warning evaluation with CV distractions | Driving simulator | U of Iowa-NADS |
| On-Road Lane Departure Method | | | |
| | 7. On-road LDW protocol method assessment | Instrumented vehicle | Westat |
| **Research on Variability among Warning Signals** | | | |
| Negative Transfer Among Vehicles | | | |
| | 1. Auditory FCW negative transfer | Driving simulator | GMU |
| Auditory Warning Perceptual Space | | | |
| | 2.1 Perceptual sort of auditory alerts | Perception lab | GMU |
| | 2.2. Adjustment of alerts – single parameter | Perception lab | GMU |
| | 2.3. Adjustment of alerts – multiple parameters | Perception lab | GMU |
| | 2.4. Validation sort based on key parameters | Perception lab | GMU |
| | 2.5. Rapid categorization of alerts under divided attention | Desktop simulator | GMU |
| Signal Effectiveness under Noise Conditions | | | |
| | 3. Warning perception in ambient noise environments | On-road | Westat |
| **Research on Temporal Aspects of Interference from other In-Vehicle Messages** | | | |
| Driving Simulator Research on Near-Simultaneous Safety Alerts and Crash Warnings | | | |
| | 1. Research with undistracted drivers | Driving simulator | DRI |
| | 2. Research with distracted drivers | Driving simulator | DRI |

# 3 Research toward FCW and LDW Protocol Development

One of the key objectives of the CWIM program was to conduct research to refine and validate methodological and procedural aspects of ACWS DVI evaluation to aid in the development of testing protocols. Two lines of research are described in this section: 1) a series of experiments investigating the effects of manipulating methodological aspects (Experiments 1-6), and 2) an exploration of on-road methods for testing LDW DVIs (Experiment 7).

## 3.1 Basic Protocols

This section describes a series of five experiments that evaluated methodological aspects of potential FCW and LDW evaluation protocols. For a protocol to yield reliable, repeatable results, numerous procedural details must be specified. These experiments provide an empirical basis for selecting a number of important methodological features.

Independent of the present project, other NHTSA-funded research was carried out at the Vehicle Research and Test Center (VRTC) that evaluated a test-track methodology for evaluating FCW DVIs (Forkenbrock et al., 2011). The methods and conclusions from that study provided a starting point for the research conducted under the present project. The five experiments reported here were conducted using the NADS-1 driving simulator at the University of Iowa.

The VRTC test track procedure had the participant drive an extended route while following a lead vehicle and maintaining a constant following headway. Periodically during the drive that participant was required to engage in a secondary (distraction) task that required the driver's eyes to be diverted from the forward field of view. The task was monitoring a changing-number display located to the driver's right. The participant engaged in this task several times with no critical event occurring. On a subsequent occurrence, however, the lead vehicle abruptly moved to the left lane, revealing a stopped vehicle in the lane directly ahead of the participant. This was actually an inflatable "balloon vehicle" which could be struck safely. The VRTC study included a variety of FCW signals (between subjects) intended to alert the distracted driver to the hazard ahead. Various dependent measures were used to assess the effects of the warnings. Forkenbrock et al. (2011) recommended a metric based on the time from the onset of the FCW to the instant the driver reestablishes a forward facing view.

The experiments reported here addressed various aspects of this general procedure. The first experiment provides a direct comparison of driving simulator versus test track findings. The second experiment compares alternative secondary (distraction) tasks and incentive structures. The third experiment examines the effect of training and exposure regarding the presence and characteristics of the particular warning function. The fourth experiment addresses the timing of alerts. The fifth experiment considers how the degree of simulator motion capability influences the experimental findings.

The same general procedure was used across all experiments and took approximately 1.5 to 2 hours to complete. This procedure included the following steps:

- a phone screening to determine eligibility;
- informed consent;

- verification of driver's license;
- demographic survey;
- training before the drive;
- simulator drive; and
- completion of questionnaires.

Dependent measures examined across the experiments are defined in Table 2. Taken together, this set of experiments provides support for a number of decisions regarding methodological details for a common evaluation protocol for evaluating FCW and LDW DVIs.

**Table 2. Summary of dependent variables**

| Measure | Description |
|---|---|
| Time to begin visual commitment | Time from task prompt to the point where the participant had turned their head away from roadway |
| Duration of first visual commitment | The time from the beginning of visual commitment to the first interruption, or end of visual commitment |
| Duration of visual engagement | Total time of visual commitment to distraction task after subtracting glances to front scene |
| Duration of visual commitment | Time from beginning of first visual commitment to the distraction task to end of last visual commitment including any glances back to the roadway |
| Time from start of task to end of first visual commitment | Time from when the task was first displayed, regardless of when the participant looks to the display, to when participant returned their glance to the roadway the first time |
| Time to end of visual commitment from warning | The time the warning began or would have begun until the first instant that the participant's vision returns forward |
| Speed at conclusion of visual commitment | Subject vehicle speed at conclusion of visual commitment |
| Glances back to roadway | Count of full glances back to roadway during distraction task |
| Minimum time to collision | The minimum time to collision (TTC) with the stopped vehicle after the alert |
| Adjusted minimum TTC | The adjusted minimum TTC with the stopped vehicle after the alert with positive values indicating minimum TTC and negative values indicating how much sooner the driver would have needed to respond based on collision velocity and deceleration profile |
| Brake reaction time | Time from alert to brake pedal depress |
| Speed of subject vehicle | Speed of subject vehicle at specified point |
| Standard deviation of lane position | Standard deviation of subject vehicle lane position over the duration of event |
| Lane departure | subject vehicle crosses lane boundary |
| Duration of lane exceedance | Amount of time subject vehicle was outside lane boundary |
| Area of lane exceedance | A measure that takes into account the lateral and longitudinal distances that the vehicle is out of lane |
| Maximum lane exceedance | The maximum lateral distance that the leading edge of the subject vehicle is out of the lane |

| **Measure** | **Description** |
|---|---|
| Steering reaction time | Time from alert to steering response |
| Collision | An instance in which the subject vehicle strikes the lead vehicle during a critical braking event |

### 3.1.1 Experiment 1: Comparisons of Testing Environments - Test Track & Simulator

Introduction

The aim of this experiment was to assess the comparability of the testing environments by comparing results from a VRTC test track study examining FCW interfaces (Forkenbrock et al., 2011) with a replication of that protocol on the NADS-1 full-motion, high fidelity simulator. This study replicated the test track drive from the VRTC effort in a simulator driving scenario on NADS-1. It used a road course that provided a parallel driving environment on two-lane rural roadways and left curves with a single "lead vehicle revealed" event in the latter portion of the drive. The distraction task and the incentive structure from the VRTC test track protocol were replicated for these studies. The VRTC incentive structure was scaled down to meet institutional review board requirements aimed at minimizing coercion associated with incentives.

The design focused on the ability to compare general aspects of driver behavior among a simulated two-lane road, a simulated test track, and actual test track data from Forkenbrock et al. (2011). The goal of this experiment is first to determine the extent to which a simulated road course can produce similar data as from a test track, and then to determine if a simulated road course, which has greater face validity than a test track, could replicate the results from simulated test track. Although alert modality was also included as a variable, the sample size was not sufficient to discriminate between these alternatives. Rather, they provided an appropriate range of alert modes across which driver behavioral effects could be compared.

There were two main hypotheses for this study:

- There will be differences in visual commitment or outcome between the VRTC test track course versus the NADS-1 simulated test track.
- There will be differences in visual commitment or outcome for the simulated test track course and simulated road course for the NADS-1 simulator.

Method

This study utilized a 2x3 factorial with between-subject variables: platform and alert modality. Platform was at two levels: simulated test track and simulated road course. Additionally, published data from the VRTC test track study were used for comparison purposes. FCW Alert modality had the levels of no warning, audio, and seat belt tensioner. These alert modalities represented a subset of the data that was previously collected on the test track as part of prior CWIM research (Forkenbrock et al., 2011). Thirty-six participants (six per cell) completed all study procedures successfully. Participants were 35-55 years of age with an even gender split. The VRTC test track study used sixty-four participants, 25-55 years of age, with an even gender split.

As the aim of this effort was to compare data to previously collected data from the test track, common dependent measures were used. Data from the simulator were reduced to provide the same measures used on the test track. The primary dependent measures considered were visual commitment to the task and driver performance measures: Specifically, time to end of visual commitment, time to end of visual commitment from warning, speed at conclusion of visual commitment, speed of the subject vehicle, minimum TTC between the subject vehicle and the stopped lead vehicle, and brake reaction time (see Table 2). No feedback about whether a collision had occurred was directly given to participants. Participants may have visually perceived a collision if one occurred, however there were no kinematic cues indicating collisions and the drive ended following the collision event.

To accommodate the tasks, two displays were mounted in the simulator cab for this study: The number recall display located to the participant's right and the headway display, which provided headway in feet, located near the participant's line of vision (see Figure 1). Participants were told that the goal of the research was to examine headway maintenance and were instructed to maintain a headway (gap) of 110 feet ± 15 feet, and to repeat the numbers from the display when instructed to do so.

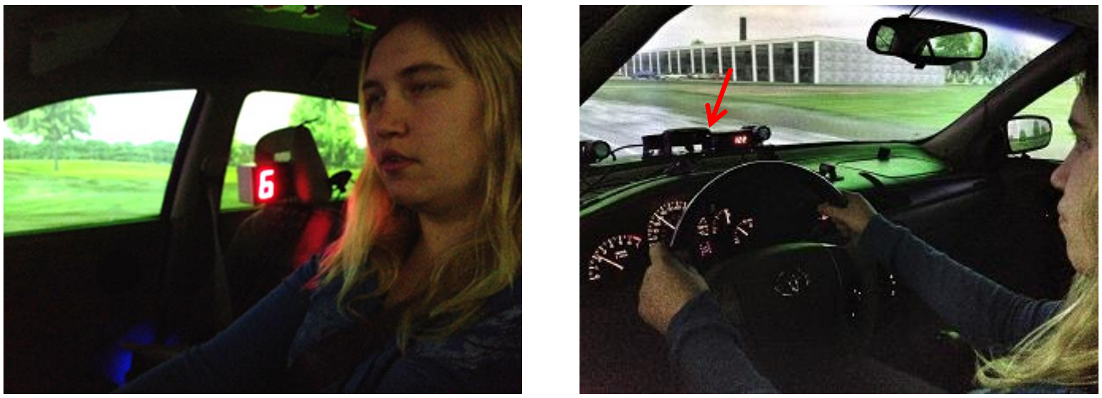


**Figure 1. Number recall task display (left) and headway maintenance display (right)**

Participants were compensated with a base pay of $17.50 with incentive compensation of up to $27.50 for a total maximum compensation of $45. The incentive was based on the participant's ability to correctly complete the number recall tasks and maintain the specified headway. The maximum incentive for headway maintenance was $8.45 and the maximum incentive for the number recall task was $19.05. This incentive structure paralleled the structure used by Forkenbrock et al. (2011). The absolute value of the payments was lower in the present study, but the relative value of each component was the same.

Two driving scenarios were used for this study: test track and road course. The test track was a recreation of the VRTC test track with a 3600-ft two-way straightaway with turnarounds at both ends. The road course was a series of four long straightaways connected by 90-degree left curves. Each scenario had a practice segment followed by four segments in which participants maintained headway, engaged in the number recall task, and received feedback. Throughout the four segments, participants followed a lead vehicle at a nominal headway of 110 feet with a nominal speed of 35 mph. During the fourth segment, a "lead vehicle revealed" event was triggered. This refers to a scenario in which the vehicle that the participant is following abruptly changes lanes to reveal a stopped vehicle in the lane ahead of the participant. The choreography of the lead-vehicle-revealed event followed the timeline laid out in Figure 2. This figure lays out two timings that are used across these studies, an early and a late warning; however, only the late warning timing was used for this experiment, in which the warning is issued at a TTC of 2.1 s.

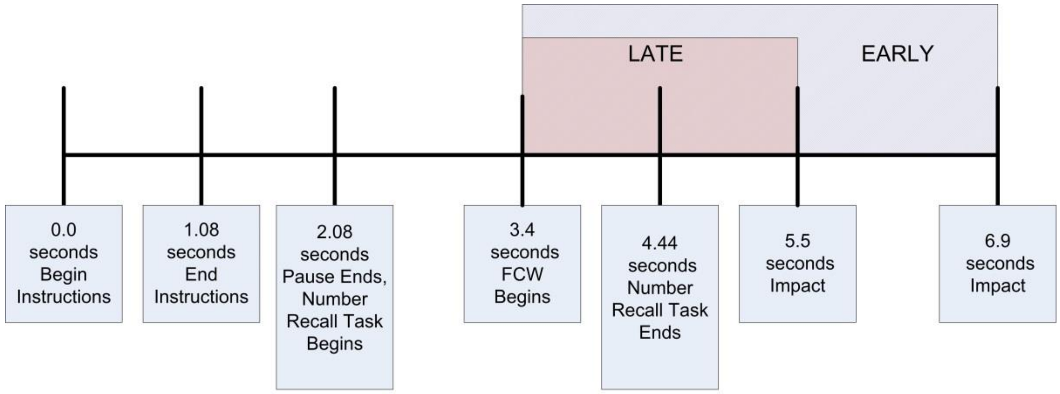


**Figure 2. FCW event timelines**

<u>Key Results</u>

The similarity of the choreography between the test track and the simulated test track was first examined. This involved examining the speed of the subject vehicle and the minimum TTC between the subject vehicle and the stopped lead vehicle at critical points during the task. Overall, there was a good correspondence between the simulated test track and the VRTC test track data. To illustrate this correspondence, subject vehicle speeds at four critical points in the process are presented in Table 3 for the test track and simulator. The observed speeds in the NADS on the virtual test track closely resemble the values from VRTC. The mean values differ by less than 0.5 mph across the four points in time. One important item to note relates to the speed at conclusion of visual commitment on the NADS based on eye-tracking. The initial plan was to utilize eye-tracking to identify the beginning and end of visual commitment automatically; however, after reviewing the data, it is clear that the placement of the display was causing difficulties for the eye tracker. As a result, manual video coding was employed that included dual coding each event. Two independent coders reduced each drive video, which was next verified by a third coder who had not previously coded any of the drive video and who settled any discrepancies. The distance and minimum TTC to the stopped vehicle results showed similar patterns of conformity; however, the values from the simulator were able to more precisely match the target values, on average. For example, the minimum TTC when random numbers were presented for the final event was expected to be 3.4 s but had a mean of 3.11 s on the test track and 3.33 on the NADS.

**Table 3. Speed across critical points during lead vehicle revealed event**

| Description | Subject Vehicle Speed (mph) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Task Instruction | | Random Numbers Presented | | FCW Alert | | Visual Commitment Concludes | | |
| | VRTC | NADS | VRTC | NADS | VRTC | NADS | VRTC | NADS Eye Tracking | NADS Video Coding |
| Min | 33.0 | 31.0 | 31.1 | 32.3 | 30.8 | 31.3 | 30.8 | 0.1 | 30.6 |
| Max | 37.5 | 38.4 | 38.1 | 37.6 | 38.3 | 38.0 | 38.2 | 34.3 | 39.4 |
| Mean | 35.2 | 35.0 | 35.2 | 34.8 | 35.1 | 34.6 | 34.9 | 24.6 | 34.5 |
| Std Dev | 1.1 | 1.6 | 1.3 | 1.2 | 1.4 | 1.4 | 1.5 | 8.5 | 1.7 |
| Median | 35.2 | 35.0 | 35.2 | 34.8 | 35.2 | 34.4 | 35.1 | 27.0 | 34.1 |
| Expected | 35 | 35 | 35 | 35 | 35 | 35 | 35 | 35 | 35 |

Given that the execution of the scenario was similar between the two environments, a comparison of the evaluation of how the two approaches impact evaluation of crash warning DVIs can be made. The primary outcome measure was time to end of visual commitment from warning. When comparing the end of visual commitment relative to the warning or time the warning would have gone off, the data from NADS show faster mean response times and less variance than the VRTC data (see Figure 3). For example, without a warning, drivers on the test track responded in 1.4 s on average, but in 1.0 s on the simulator. This may reflect the fact that drivers in the NADS study began their visual commitment sooner with drivers in the simulator beginning in 0.9 to 1.0 s after the instruction started, but on the test track drivers didn't begin their visual commitment until 1.4 to 1.5 s. This resulted in similar visual commitment duration between the two environments indicating that total time on the secondary task potentially plays a role in when drivers disengage from the task.

When examining brake reaction times, the means from the NADS-1 study were shorter on average for the auditory and no warning conditions, but about the same for the seat belt tensioner condition (see Figure 4). When looking only at the NADS data, no difference in mean brake reaction time is observed, although VRTC found a slight improvement from no warning for the seat belt tensioner and a slight delay in reaction from the auditory alert.

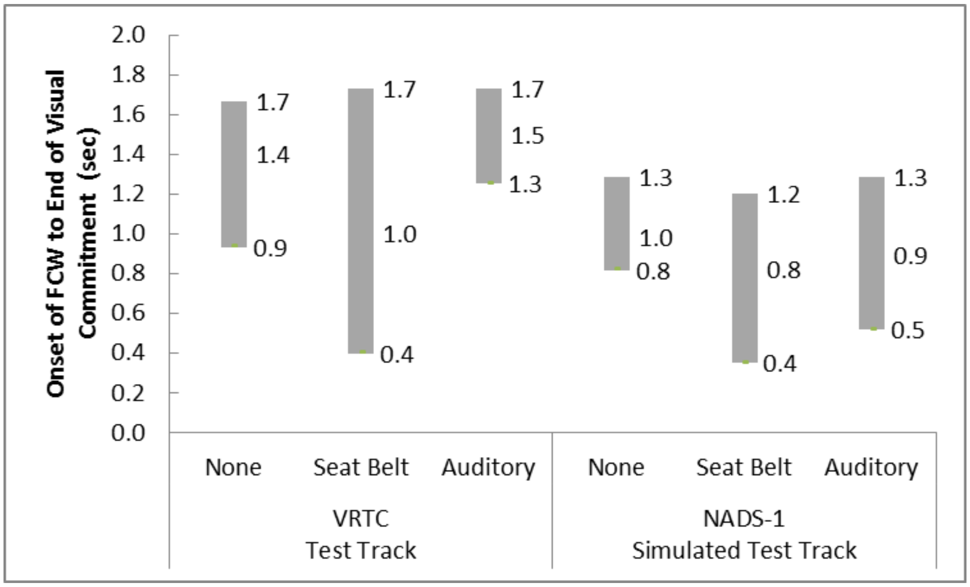


**Figure 3. Range and mean time to visual commitment end from warning**

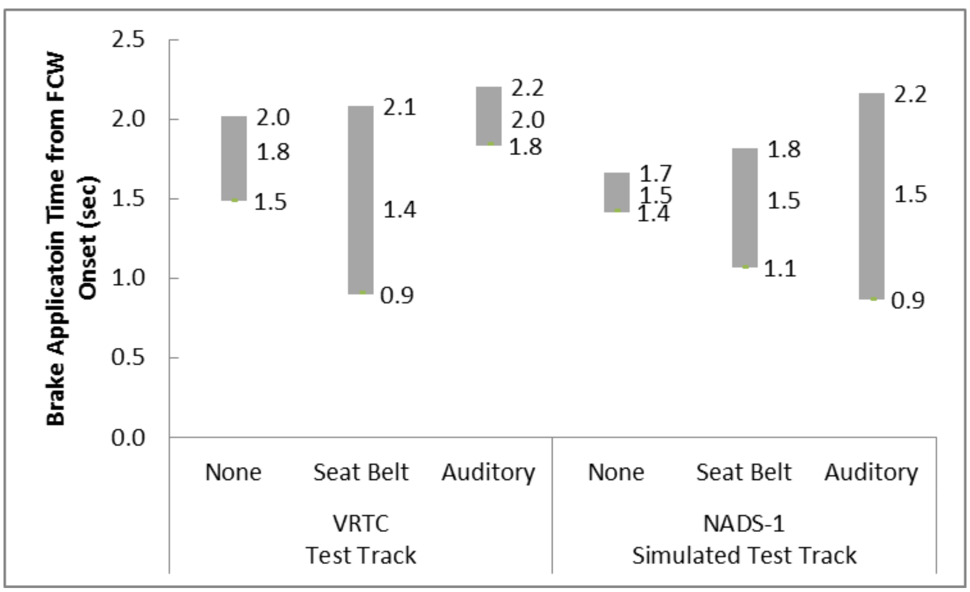


**Figure 4. Range and mean brake reaction time**

Next, time to end of visual commitment between the virtual test track and the virtual road course was examined through an analysis of variance (ANOVA) using the SAS general linear model (GLM) procedure. There was no significant difference between the road course and the test track ($p=0.7972$) nor any interaction with warning type ($p=0.3419$). Figure 5 shows the distribution of the data for each warning type on each course along with the means. There was no main effect of warning type ($p=-.1251$).

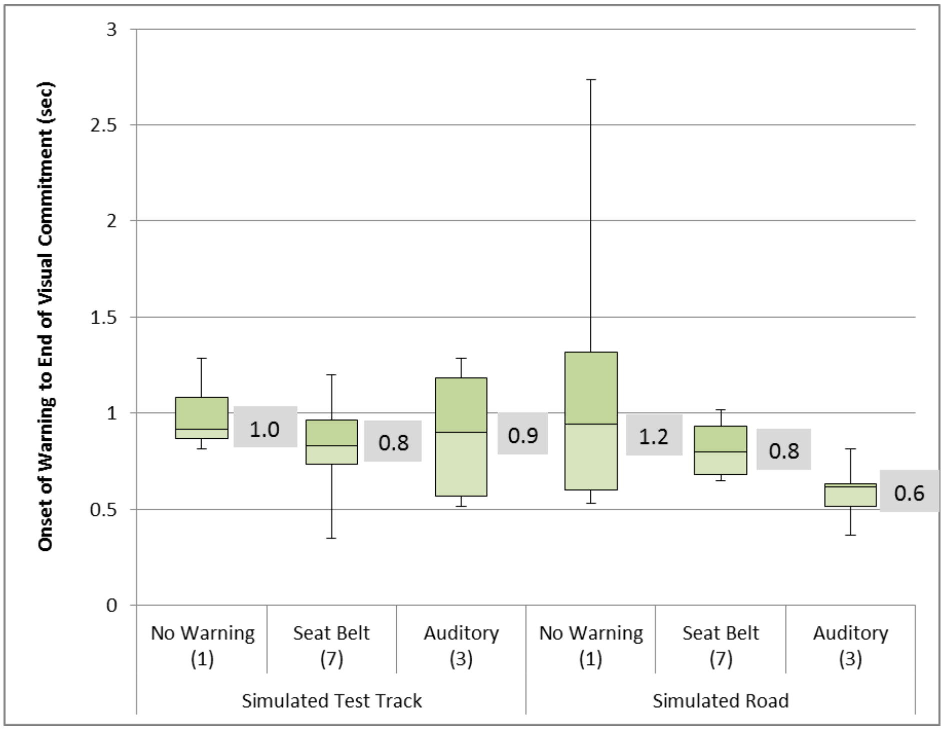


**Figure 5. Box plot comparison of time to end of visual commitment on simulated test track and road courses**

<u>Discussion</u>

Overall, the performance of participants on the simulated test track was quite comparable to the performance on the VRTC test track. From a choreography standpoint, the advantages of the simulator actually resulted in a more precise execution of the interaction between the subject and other vehicles. Participants in the simulator were faster to engage in the numbers task and faster to stop it in reaction to the alerts, but overall visual commitment duration remained the same between the two environments. Response measures showed similarities between the test track and the simulator, particularly for the no warning and seat belt tensioner conditions. The results suggest that protocols developed on the test track should be able to be translated into the simulator. It also suggests that the reverse would also be true, with simulator scenarios translated to the test track.

When comparing the simulated test track and simulated road environments, results were generally comparable. This was particularly true for the no warning and seat belt tensioner conditions. Although there was no statistical difference, there appeared to be a trend across several measures for a more cautious response from participants with the auditory warning on the road course compared to the test track environment. Overall, these results point to an opportunity to translate events between road courses and test track in some circumstances if care is taken to match the timing of the events and adjust the protocol to provide a parallel context for the driver relative to the driving situation.

The main aim of this study was to examine how test track and road protocols compare. For FCW events, both test track and road routes seem to be useable for the evaluation of the DVIs for these systems. In light of the desire to examine these system DVIs in a more natural driving environment and the desire to include lane departure events for subsequent studies, a simulated road course may be the best option for comprehensive evaluation of crash warning interfaces.

### 3.1.2 Experiment 2: Secondary Task and Incentive Considerations

Introduction

The aim of this study was to examine different distraction methods across incentive systems to determine the most effective approach to distracting the participant sufficiently to initiate imminent crash events. To be an effective distraction for the evaluation of a DVI for crash warning system, the task must produce a glance away from the roadway that is of sufficient duration and predictability that either the FCW event can unfold, or for the driver's vehicle to be diverted out of the lane without noticing. Additionally, this study aimed to determine the role of financial incentive on driver willingness to disengage with secondary tasks when presented with a surprise alert. The protocol included four distraction tasks, three incentive structures, and two alert modalities. The scenarios were designed to hold FCW alert onset constant relative to engagement with the task. No training on FCW or the alert was provided to participants prior to their drive. Participants were not exposed to the FCW alert during their drive prior to the single FCW event at the end of the drive. The only purpose of the FCW events was to evaluate the participant's willingness to disengage from the distraction task based on secondary task incentive. Additionally, no LDW system was present. This is further described below.

There were three primary hypotheses for this study:

- There will be significant differences between distraction tasks in terms of engagement and degree of visual commitment.
- There will be significant differences in engagement with the distraction task across the levels of incentive.
- For the FCW event, drivers reach the end of visual commitment more quickly when no financial incentive is provided compared to the two conditions with financial incentives.

Method

This study combined two experimental designs into a single simulator experiment. The first experimental design was a 3x4 mixed factorial with three levels of incentive structure as the between-subject variable and four levels of distraction task as the within-subject variable. The second experimental design was a 3x2 between-subject design presenting two levels of alert modality, audio and seat belt tensioner, at each of the three levels of incentive structure. The final distraction task in the drive was always the number recall task and was paired with a decelerating lead vehicle event and an FCW alert. Twenty-four participants completed all study procedures successfully with eight participants per cell for the first experimental design and four participants per cell for the second experimental design (see Table 4).

**Table 4. Secondary task and incentive study experimental conditions**

| Incentive Structure | Experimental Design 1 Distraction Task: Number Recall | Bug | Message Reading | Menu Search Selection | Experimental Design 2 Final Event Alert Modality: Audio | Seat Belt Tensioner |
|---|---|---|---|---|---|---|
| High base | | 8 participants | | | 4 participants | 4 participants |
| Low base (VRTC) | | 8 participants | | | 4 participants | 4 participants |
| No incentive (NADS) | | 8 participants | | | 4 participants | 4 participants |

Three incentive structures were used, although total potential compensation was the same for all groups: high base pay ($27.50) with low additional financial incentives (up to $17.50), low base pay ($17.50) with high additional financial incentives (up to $27.50), and a fixed sum compensation ($45.00) with no incentive component.

Four distraction tasks were presented as within-subject variables: number recall, bug, message reading, and menu selection. Consistent with Forkenbrock et al (2011), each task had a duration of 2.36 seconds to allow for the events to be triggered while the driver was looking away and for the response to be attributable to the warning.

- Number Recall Task – This task was located at least 90 degrees to the right of the participant's forward facing position. One second after receiving the instruction to begin, five random single-digit numbers were presented for 472 ms each. The participant was to repeat them aloud in the correct order to the experimenter following the task (Forkenbrock et al., 2011).
- Bug Task – This task required that the participant turn and reach into the back seat to trace the path of an animated insect on a touch screen display that was located to the rear of the left portion of the passenger seat headrest, at least 90 degrees to the right of the participant's forward facing position. A message indicated when it was time to begin the task. A red X appeared on the screen to orient the participant's finger to the display. 1.5 s after the message concluded, the insect began to buzz to start the task and continued to buzz for the duration of the task (2.36 s). A trail behind the bug was colored green, yellow, or red to provide visual feedback on performance. This task was an adaptation of the insect used by Lerner et al. (2011) and Brown et al. (2011).
- Message Reading Task – This task was presented to the driver on a touch-screen located in a head-down position to the right of the center console, ensuring that the driver's gaze would be directed down away from the front window. A message indicated the start of the task and the appearance of the text. The text was similar to that of an email. Timing was consistent with the number recall task. The task for the driver was to read aloud as much of the text as possible in the time available. When the task concluded after 2.36 s, the screen was cleared, but the participant could continue to recite any of the text that they had read before the task ended.

- Menu Selection Task – Drivers were presented with a visual distraction task on a touch-screen that mimicked a generic interaction with in-vehicle devices such as mp3 players or navigation systems. The driver was presented a two-word target phrase and asked to locate an appropriate match from the four listed on the screen. This task was located on the same display used for the message reading task such that the driver's gaze was directed down away from the front window. The phrases appeared 1.5 s after the prompt concluded. A correct match was when any of the words in the presented phrase appeared in the same location in the selected phrase. During the drive, a message indicated when to begin the task. When the match phrase was located, the driver pressed the screen location for the selection. The phrases were presented on the screen for a total of 2.36 s. If the initial selection was incorrect, the participant could try again if there was time remaining (Lees et al., 2007).

The driver distraction tasks for future use were chosen keeping three criteria in mind. The first criterion was chosen based on the conclusion by the Drivers Focus-Telematics Working Group (June 26, 2006) that the system should allow the driver to leave at least one hand on the steering control. The second criterion was to steer the driver's gaze away from the front view in order to instigate a "surprise" imminent, threatening situation. In order to do this, the driver's gaze needed to be averted continuously or with minimal disruption(s) for the duration of the task. Klauer et al. (2006) defined and classified relative risk of secondary tasks as simple, moderate, and complex. The 100-car naturalistic driving database (Dingus, 2008) used a basic odds ratio equation to determine that a score of 1.0 indicates a crash/near crash risk is equivalent between the primary and secondary task. In this study, glance durations and mean glance duration were determined for activities conducted while driving. Table 5 compares the distraction tasks and displays the equivalents of demand, odds ratio category, and mean glance duration levels. The tasks are shown in Figure 6.

**Table 5. Distraction task justification**

| Distraction Task | Gaze Location | Demand | Odd Ratio Category | Mean Glance Duration |
|---|---|---|---|---|
| Number Recall | Back Seat | Visual Verbal | Moderate Task | 1.04 |
| Message Reading | Front Head Down | Visual Verbal | Complex Task | 1.45 |
| Bug | Back Seat | Visual Manual | Complex Task | 1.40 |
| Menu Selection | Front Head Down | Visual Manual | Moderate Task | 1.14 |


**Figure 6. Secondary tasks: Number recall task (upper left), bug task (upper right), message task (lower left), and menu selection task (lower right)**

Three displays were mounted in the simulator cab for this study: a monitor to display the number recall and bug tasks located to the participant's right, a display for the message reading and menu search task located down and to the right of the participant, and a headway display located near the participant's line of vision.

Three different training presentations were used for this study – one for each of the incentive conditions. Each provided detail on the distraction tasks to be performed. They differed in the description of the incentive pay, which was based on the experimental conditions for each participant.

Additionally, participants practiced with the distraction tasks before going into the simulator to become comfortable performing each task. Participants were prompted to engage with each of the tasks consistent with the method used in the simulator. They were asked to practice until they could perform each task successfully. Participants were told that the goal of the research was to evaluate several new in-vehicle equipment designs and technologies, when in fact their interaction with the distraction tasks and response to a surprise FCW warning were being evaluated.

One driving scenario with two orders of distraction tasks was used for this study. The scenario had a practice segment followed by four segments in which participants maintained headway and engaged in the distraction tasks. The order of the distraction tasks was blocked using a Latin Squares approach to balance the presentation of the distraction tasks across the drive. Throughout the drive segments, participants followed a lead vehicle at a nominal headway of 185 feet with a nominal speed of 55 mph. Participants were told that headway maintenance was the ability to maintain a specified distance from their vehicle to the vehicle in front of them. For the final event, which was always associated with the number recall task, a lead vehicle began to rapidly decelerate while the driver was engaged in the task, resulting in an FCW alert. The alert was timed to occur one second after the first number appeared. The incentive structure is provided in Table 6.

**Table 6. Incentive structure**

| | Base only | High base | Low base |
|---|---|---|---|
| Headway (4) | N/A | $1.35 | $2.12 |
| Number Recall (5) | N/A | $0.61 | $0.95 |
| Bug (4) | N/A | $0.76 | $1.19 |
| Message Reading (4) | N/A | $0.76 | $1.19 |
| Menu Selection (4) | N/A | $0.76 | $1.19 |

Key Results

There were no significant effects of incentive. What trends were present indicated a potential dis-benefit in terms of driver engagement and disengagement with the secondary task associated with increasing incentive pay, as discussed below. Therefore, the data suggest that no financial incentive for performance should be provided. The number recall task was used successfully in the VRTC study for FCW events, and the present experiment found further evidence of its appropriateness relative to other distraction tasks. Supporting evidence is provided in the following sections on engagement and driving performance.

Engagement is a key consideration in the selection of a distraction task, both in initial engagement with the task (time to begin visual commitment, time from start of task to end of first visual commitment, and glances back to roadway), and total engagement with the task (time to end of visual commitment from warning). Drivers must quickly engage in the task and stay engaged until an alert is presented to allow the orchestration of scenario events. There were no significant effects of incentive on measures of engagement. Engagement measures did differ across the four tasks.

How quickly engagement began following the end of the task prompt is revealed by time to visual commitment (see Figure 7). The thin blue horizontal lines indicate the beginning of the task. The number recall and menu tasks both showed consistent engagement prior to the beginning of the task by the third trial. Once engagement has begun, drivers must stay engaged while scenario events unfold.

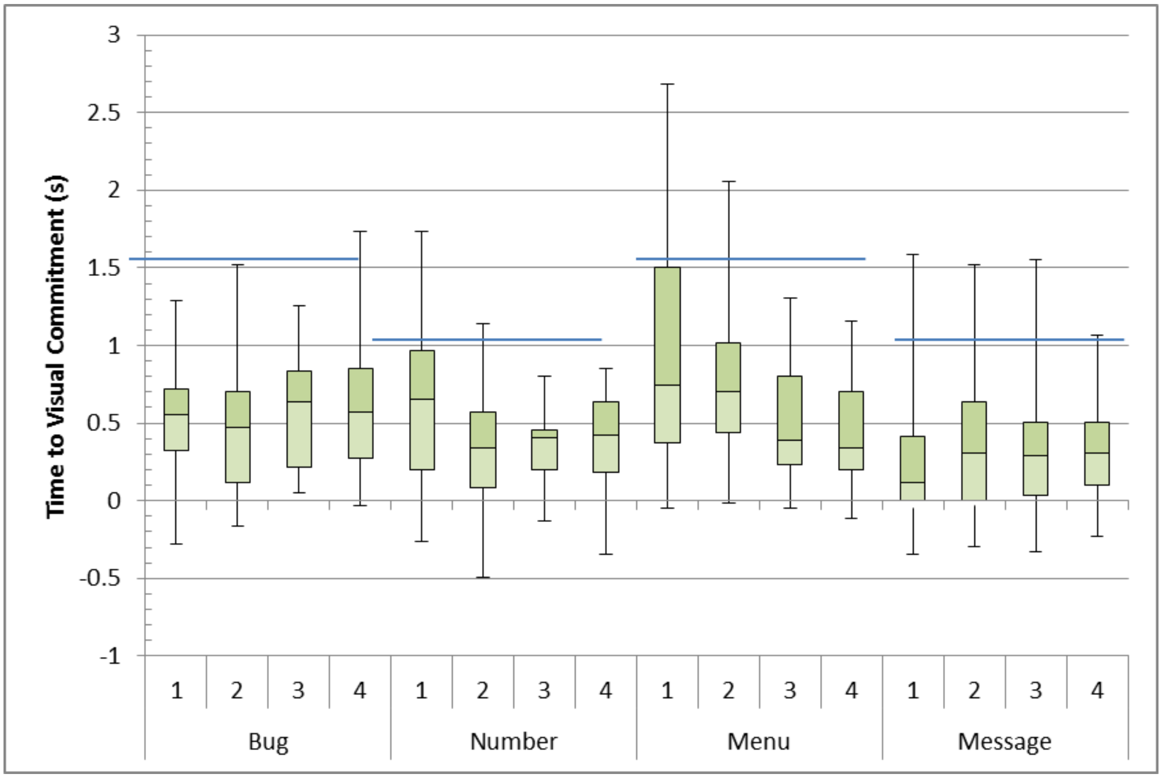


**Figure 7. Time to visual commitment**

The start of the task to the end of the first visual commitment for all four tasks is shown in Figure 8. The red horizontal line represents the time at which the tasks ends. For the number recall task, the median value for all four trials is greater than the end of task time, and for the second, third and fourth trials the 1st quartile value is greater than 2 s which is after the final number has been displayed. It can also be observed that the 1st to 3rd quartile range for the number recall task is much smaller on the last three trials than it is for any of the other tasks. Negative values indicate some participants' first visual commitment began and ended during the one-second pause between the end of the task prompt and when the number display began. This pause was included to allow participants time to turn their head following the prompt.

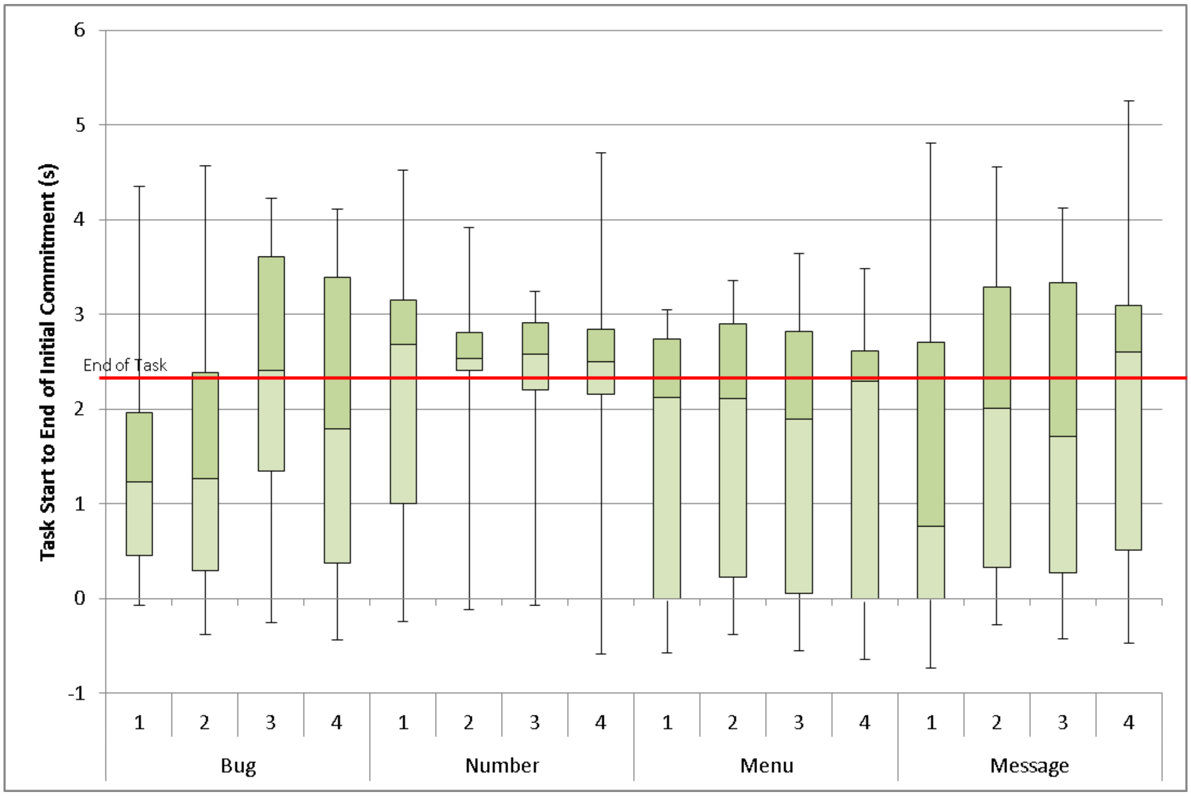


**Figure 8. Time from start of distraction tasks to end of first visual commitments by task and trial**

Some drivers looked back to the forward roadway prior to the end of the tasks as indicated by the number of full and attempted glances back to the roadway during their first visual commitment. The number recall task had the fewest number of full glances back to the roadway. Figure 9 shows the number of times during the first visual commitment to the distraction tasks that the driver moved their visual attention toward the road without ending their visual commitment (i.e., attempted glances). These do not constitute an end to the visual commitment and are of very short durations such that the driver does not get much, if any, visual information from the forward roadway. It should be noted that drivers were more likely to engage in this behavior when performing the number recall task than with the other tasks. Although this is concerning, the frequency was reduced by the third trial and was only 10% greater than the other three tasks (25% versus 15%) when the values stabilize for the third and fourth trials.

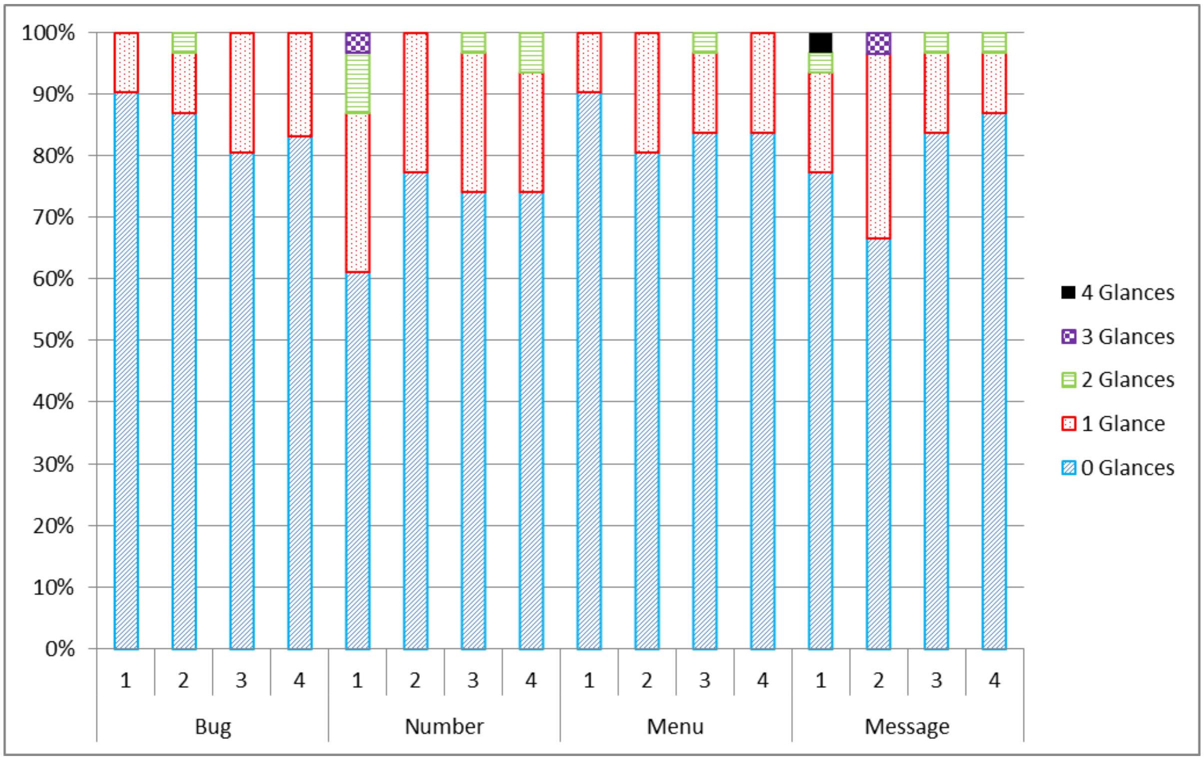


**Figure 9. Proportion of drivers who attempt zero to four glances forward during the first visual commitment**

The effects on driving performance inform whether the same or different distraction tasks should be used in FCW and LDW scenarios. The impact on standard deviation of lane position and percent of drivers who depart the lane during the task are shown in Figure 10 and Figure 11, respectively. Although a significant proportion of drivers departed the lane regardless of the task, the number recall task resulted in the fewest lane departures and smallest standard deviations of lane position. Lane stability is useful in orchestrating both the FCW and LDW events.

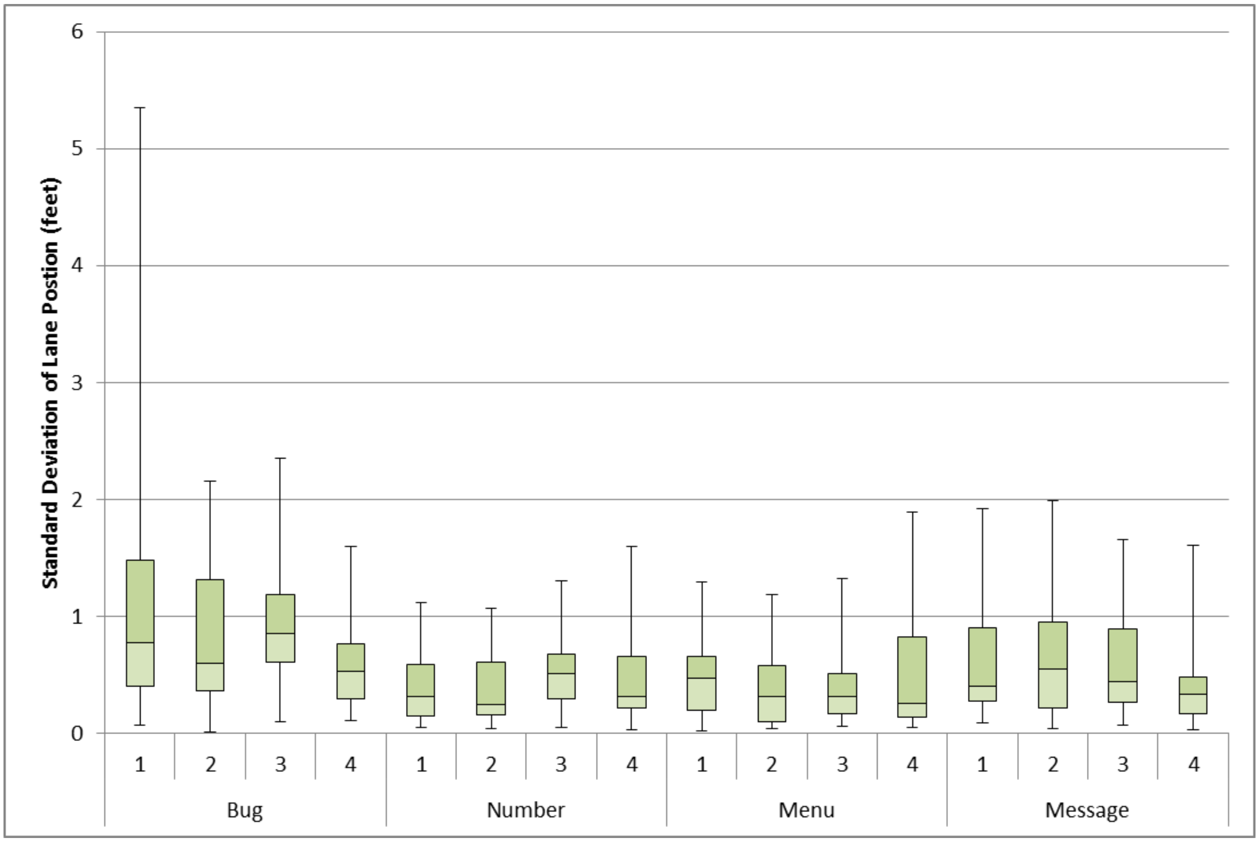


**Figure 10. Standard deviation of lane position by task and trial**

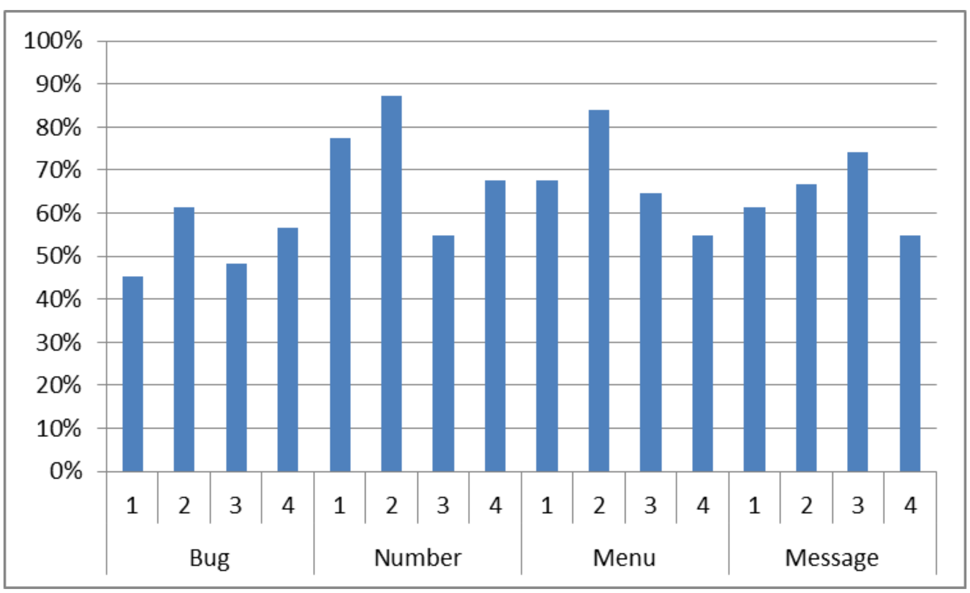


**Figure 11. Proportion of drivers who did not depart their lane**

The mean time to end of visual commitment from the warning is presented in Figure 12 within the box plots for each incentive structure. Although there was no statistically significant difference (p=0.3749), the higher mean (1.7 s) for the low base condition may indicate the higher incentive pay could lead to drivers being less willing to disengage from the task when they have received an alert. The other two conditions, no incentive and high base, had the same mean.

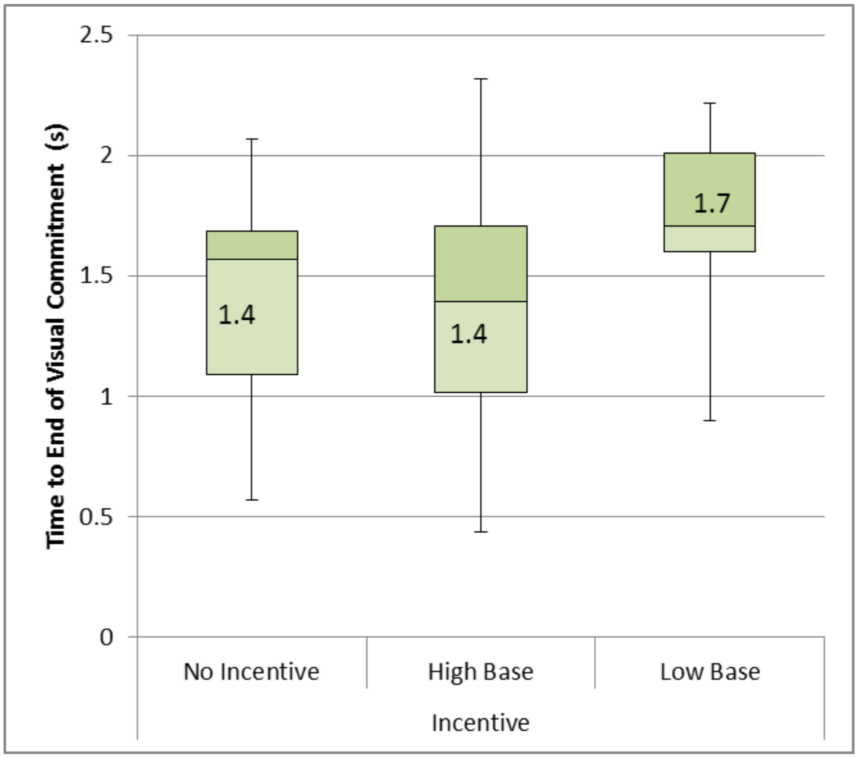


**Figure 12. Time to visual commitment end from warning, by incentive**

<u>Discussion</u>

It is important to remember when considering the distraction tasks that the aim of these tasks is to have the participant engage in an activity that is relatively rare in the real world so that researchers can assess the utility of DVIs for crash warning systems in situations where they are most likely to be needed. To that end, the protocol will require the driver to disengage from the driving task long enough for the lateral and longitudinal events to be initiated; however, the timing requirement is generally longer and more consistent for the longitudinal event. When identifying the best task, researchers must consider not only the mean duration of the task but also the variance and the relationship of the distribution to the time required to initiate events. When comparing the number recall task with the other tasks studied, there are benefits and drawbacks to each. When considering these four tasks, all tasks resulted in drivers beginning the task within the expected window. The number recall tasks performed best when considering the frequency with which drivers' visual commitment ended after the task was complete. Although drivers engaged in the number recall task had more "attempted glances" to the forward roadway without ending their initial visual commitment to the task, this is far outweighed by the longer time from start of task to end of first visual commitment. Additionally, the number recall task leads to the greatest stability in lane keeping and fewest lane departures, which will reinforce with the participant that they can complete the task without looking back to the forward roadway and make it easier to initiate forced lane departures in a simulator environment. The findings of this task suggest that the number recall task is the best fit of those tested and should be used in the simulator protocol.

The use of incentives in driving research has the potential to encourage greater engagement with the incentivized task(s) and to attempt to match some of the inherent motivation that drivers experience in the real world. The risk is that the monetary incentive will instead incentivize drivers to behave in ways that are not consistent with what would be expected in the real world, such as ignoring safety warning alerts to finish a secondary task. When considering incentive structure, longer initial glances reflect a shift in driver willingness to stay engaged with secondary tasks, which by itself is not a significant problem. However, when combined with longer times to end visual commitment in response to a warning, it indicates that the incentive may be causing the driver to ignore the warning which could result in a bias in the evaluation of the warning DVI. Additionally, there was also no significant effect of incentive on headway maintenance. Given there was no significant effect of incentive structure on initial visual commitment and the presence of an incentive discouraged disengagement when the FCW alert was triggered, the no incentive structure offers the best condition for evaluating various warning DVIs. The evidence from this study does not support the idea that incentive improves engagement with the task in such a manner that would improve reliability of collecting data while drivers are distracted without causing drivers to delay response to the alerts.

### 3.1.3 Experiment 3: System Training and Exposure

Introduction

The aim of this experiment was to examine how familiarity with the FCW and LDW systems impacts the evaluation of DVIs for these systems. This was accomplished through two manipulations. One involved different levels of training on the system and associated alert modality. The other involved whether or not participants were exposed to the alert in the vehicle prior to receiving it during an FCW or LDW event. Late alert onset was held constant for FCW and LDW alerts. Three alert modalities were presented to different groups of participants.

There were two main hypotheses for this study:

- There will be greater protocol sensitivity between warning system DVI for drivers with no prior knowledge and no prior exposure.
- The protocol will show more sensitivity for DVI differences during Event 1 than during Event 2.

Method

The experimental design for this study was a 2x2x3x2 mixed between/within-subject design with two levels of awareness, two levels of exposure, three levels of alert modality and two within-subject levels of event. Ninety-six participants completed all study procedures successfully. Two levels of awareness of the systems were provided to participants during the briefing portion of the experimental procedures. The first level was no awareness and the second was training on material similar to that included in the owner's manual of a vehicle that is equipped with an FCW system.

The two levels of exposure were no exposure prior to the warnings and exposure at the beginning of the drive. The choreographed exposure was implemented during the first few minutes of the drive by instructing participants to approach a lead vehicle until they experienced the FCW warning and asking the participants to drive toward the lane lines until they experienced the LDW warning.

The three levels of alert modality were haptic, auditory, and haptic/active – where the vehicle provides inputs to the vehicle controls. For the FCW system, these were seat belt tensioner, auditory (repeated beeps), and brake pulse. For the LDW system, these were steering wheel vibration, acoustic alert from the NADS LDW, and steering wheel torque. For the independent variable event, there were two exposures to the LDW event and up to two exposures for the FCW event.

The number recall task used is consistent with the prior study, described in Section 3.1.2. The headway display was mounted in the same location but switched to an analog gauge display which indicated whether the driver was in the range through two different colored regions (see Figure 13). This change was made to allow drivers to better manage their headway based on prior results. The headway display was no more obtrusive than the eye tracking cameras. Participants were encouraged to adjust their seats to maximize their comfort and roadway visibility.

**Figure 13. Revised headway display location and configuration**

Participants were told that the goal of the research was to evaluate several new in-vehicle technologies, when in fact their response to surprise LDW and FCW events was being evaluated. Participants were instructed to maintain position in the central green zone of the headway display. Two different training presentations were used for this study based upon the experimental conditions. Each provided detail on the distraction tasks to be performed. They differed in the description of systems present in the car based on the experimental conditions for each participant. The simulated road network used for this study was the same as used for the prior study; however, additional task locations were added and all tasks were the number recall task.

The LDW event was similar to the FCW event used in the prior experiments in terms of overall choreography. However, instead of a lead vehicle revealed event, a lateral push was provided to the vehicle to cause it to drift out of the lane while the participant is engaged with the secondary task. The push is provided in such a way as to force a lane departure, but no motion cuing is provided to suggest to the driver that anything is being done to the vehicle.

<u>Key Results</u>

When considering whether to make participants aware of the presence of the alert system before they experience it through training or to provide them with prior exposure to the alert system before the first critical event, the data suggest that participants should have prior exposure but no training.

When looking at response to the LDW, there were no effects associated with awareness of the system or prior exposure to the system. Time to the end of visual commitment was not significantly different for either awareness (p=0.5697) or for exposure (p=0.0985); however there was a significant three-way interaction between exposure, gender and warning type (p=0.0413). As can be seen in Figure 14, there is no consistent gender pattern across the warning types and exposure. When looking at the warnings, prior exposure results in greater differentiation between conditions for the males, but little difference for females. An ANOVA using the SAS GLM procedure was used for these analyses.

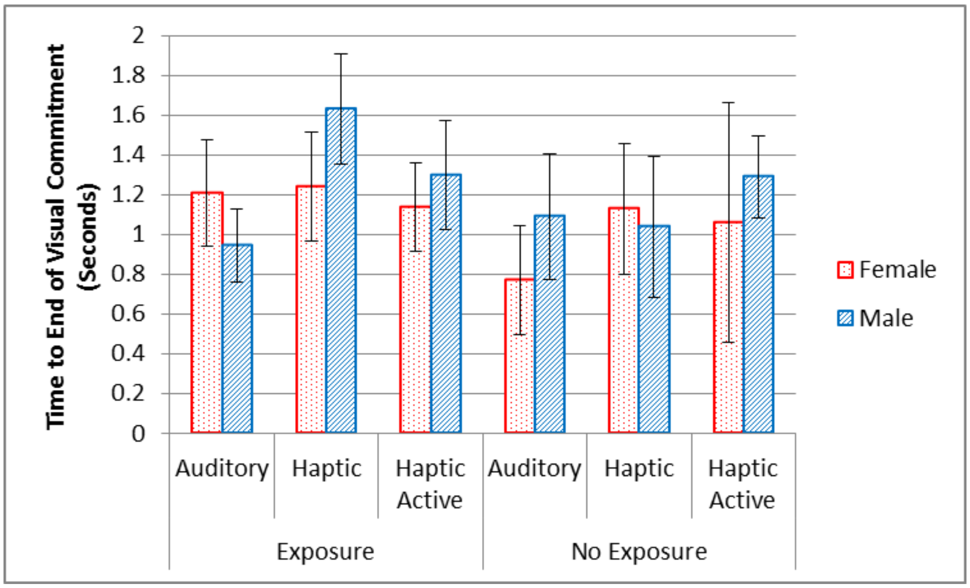


**Figure 14. Interaction of exposure, gender and warning type on time to the end of visual commitment for LDW**

When looking at the other measures that might be influenced by awareness and exposure, there were no significant differences based on awareness of the system for any of the measures of visual commitment or response. However, there were significant differences for exposure. Significant differences were observed for duration of visual commitment (p=0.0077), duration of first visual commitment (p=0.0162), and engagement duration (p=0.0171). As can be seen in Figure 15, commitment durations were longer by 250-300 ms with prior exposure to the alert.

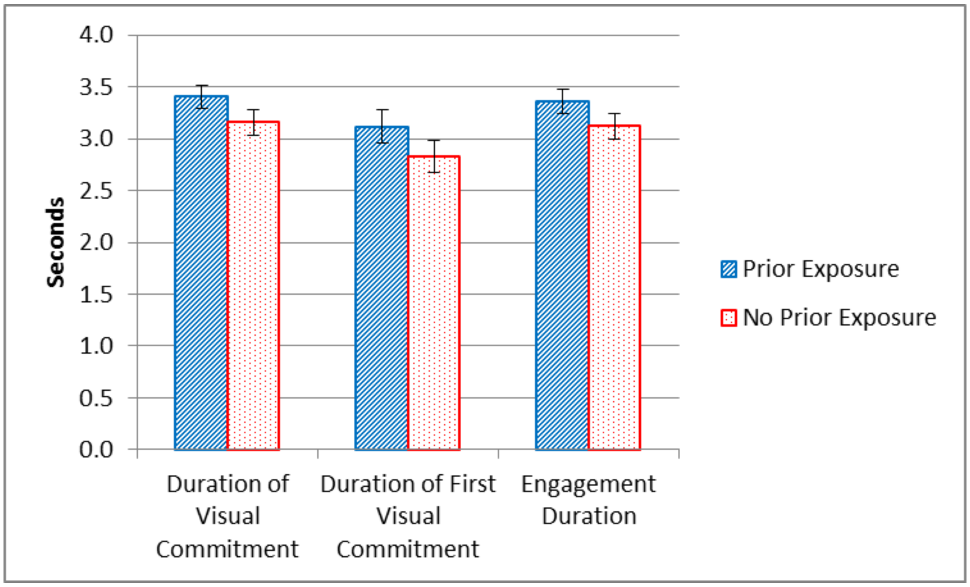


**Figure 15. Visual commitment times for lane departure events for LDW**

When looking at the differences between the warning types, there were no significant differences for duration of visual commitment or engagement duration. When looking at differences associated with gender, there were significant differences associated with duration of visual commitment ($p=0.0439$) and engagement duration ($p=0.0343$). As shown in Figure 16, males were willing to commit to the tasks longer than females by 160 and 180 ms, respectively.

When looking at differences associated with event, there were significant differences associated with outcome. Differences were found for area of lane exceedance ($p=0.0049$) and maximum lateral exceedance ($p=0.0011$). As shown in Figure 17, the second lane departure event resulted in outcomes of lower severity.

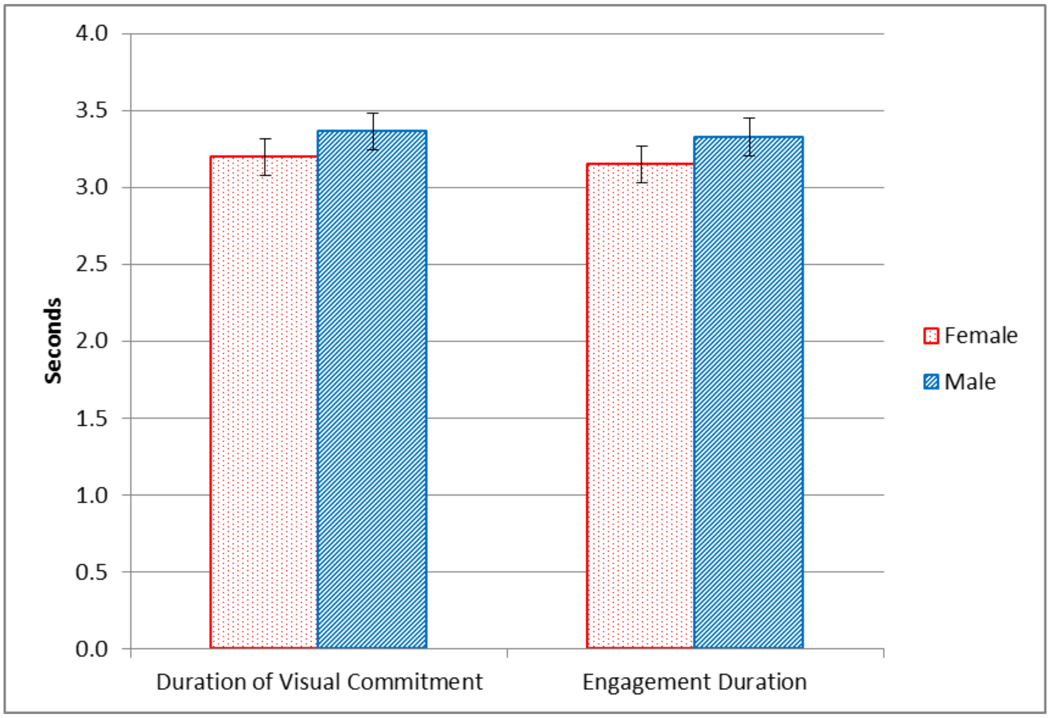


**Figure 16. Commitment times for lane departure events by gender**

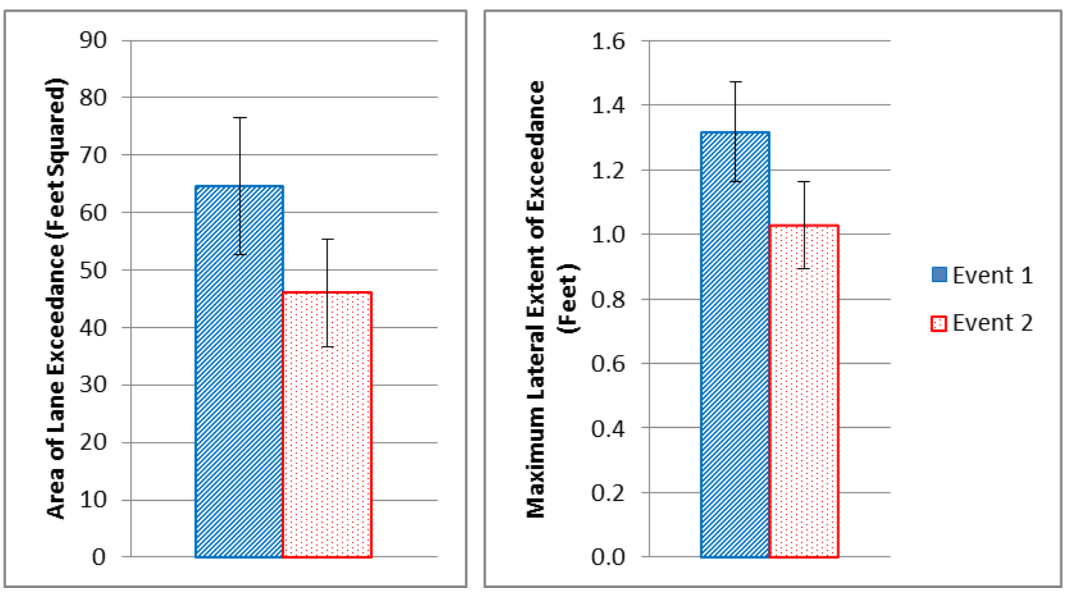


**Figure 17. Outcome for lane departure events by event in terms of area of lane exceedance (a) and maximum lateral exceedance (b)**

When looking at response to the alert for forward crash events, there were two significant effects associated with awareness of the system and none associated with prior exposure to the system. Brake reaction time to the alert was significantly different for awareness (p=0.0082) but not for exposure (p=0.7419). Figure 18 shows that participants who were aware of the system responded approximately 171 ms faster than participants who were not aware of the system.

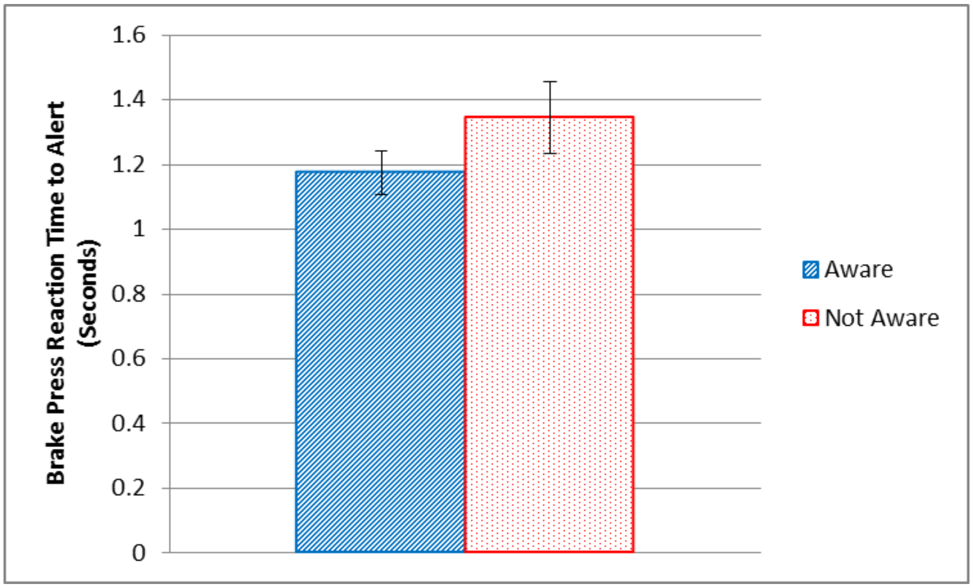


**Figure 18. Main effect of system awareness on brake reaction time**

When looking at the other measures that might be influenced by awareness and exposure, there was a significant difference relative to awareness for minimum TTC (p=0.0480) (see Figure **19**), but no significant differences for exposure (p=0.6599). For minimum TTC, participants who were aware of the alert had significantly greater minimum TTCs.

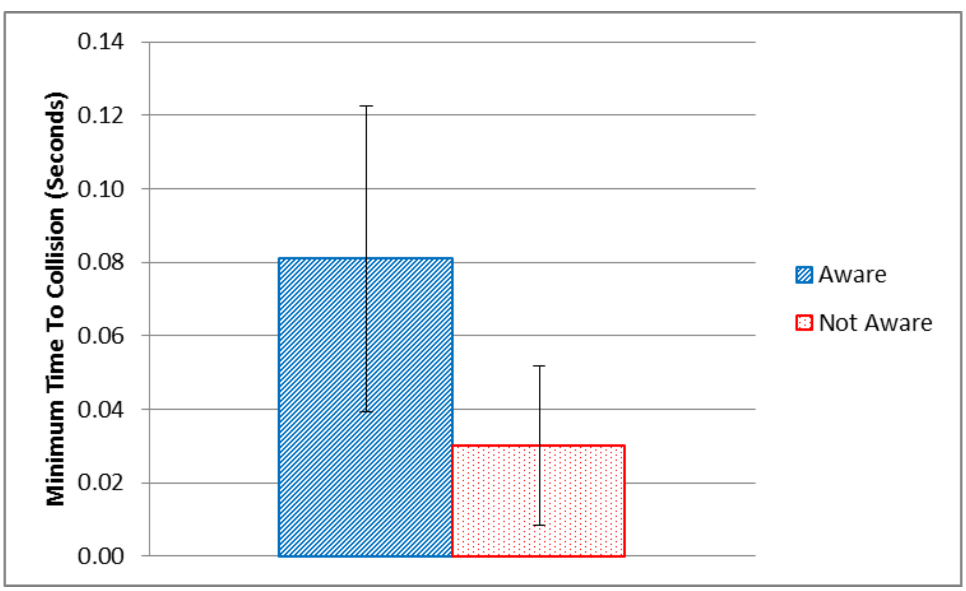


**Figure** 19**. Main effect of awareness on minimum TTC**

When looking at the differences between the warning types, there was a significant difference for the primary measure of time to end of visual commitment (p=0.0366) but not for brake reaction time (p=0.1275). There were also significant effects for duration of visual commitment (p=0.0480), duration of first visual commitment duration (p=0.0461), and engagement duration (p=0.0432). As can be seen in Figure 20, visual commitment durations were shorter for the haptic warning condition than for the active haptic warnings.

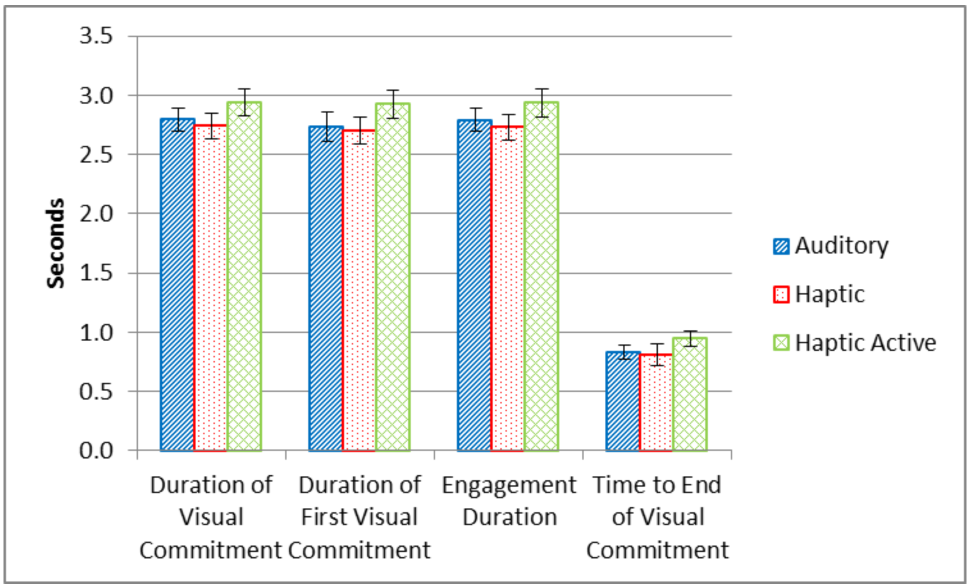


**Figure 20. Main effect of warning type on visual commitment for forward crash events**

There were no significant main effects of gender. However, there was a significant interaction between awareness of the system and time to begin visual commitment (p=0.0072). As seen in Figure 21, females who were aware of the presence of the warning system took significantly longer to engage in the numbers task compared to females who were not aware and both groups of males.

When looking at differences associated with event, there were significant differences associated with visual commitment, response, and outcome. Differences were found for initial visual commitment duration (p=0.0242), time to visual commitment end (p<0.0001), TTC at visual commitment end (p<0.0001), and brake reaction time (p=0.0058). As seen in Figure 22, participants had shorter initial visual commitments (53 ms) and returned their attention to the road quicker (93 ms) for the second forward crash event, and there was a resultant greater TTC at visual commitment end (97 ms). As can be seen in Figure 23, participants were faster to apply the brakes in response to the alert on the second event by 118 ms.

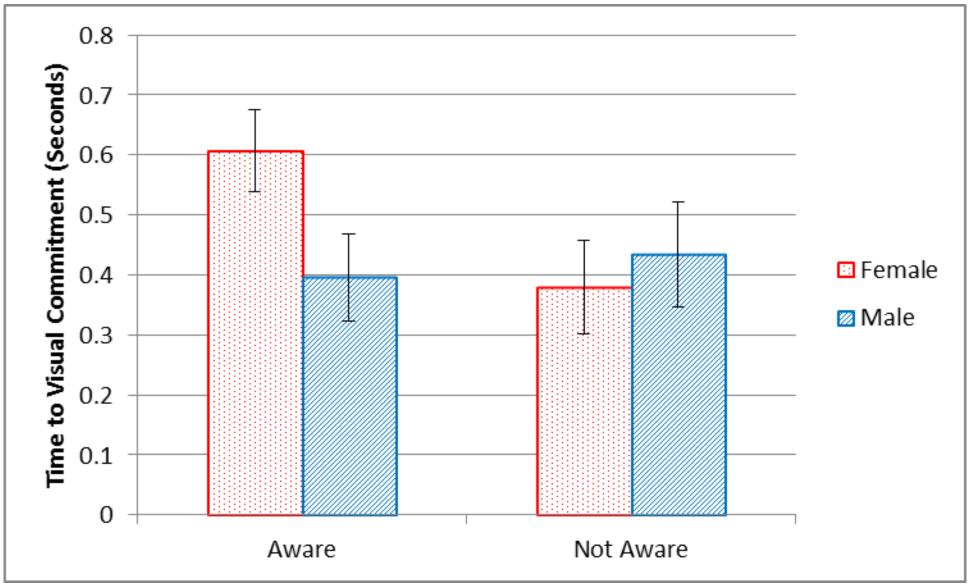


**Figure 21. Effect of system awareness and gender on time to visual commitment**

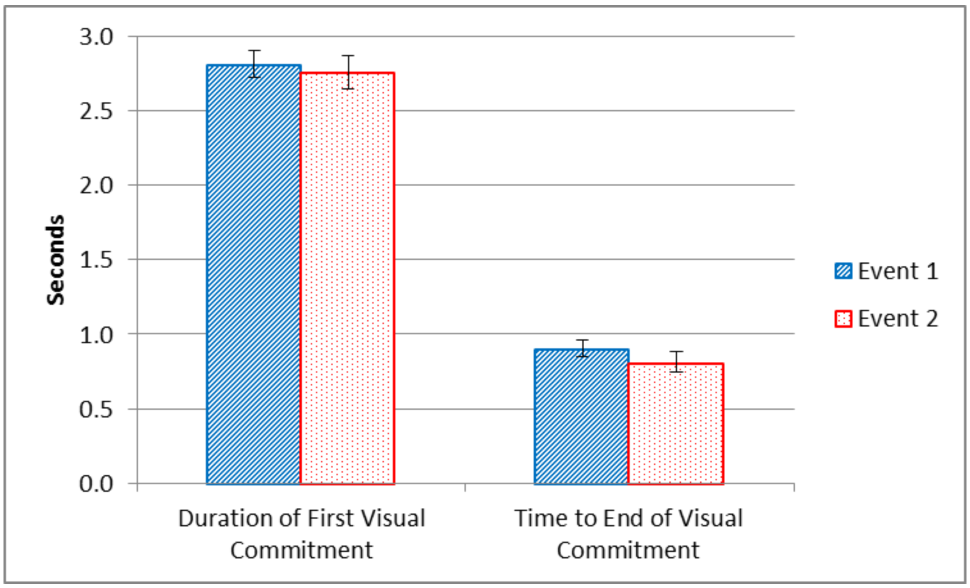


**Figure 22. Effect of repeated event on visual commitment**

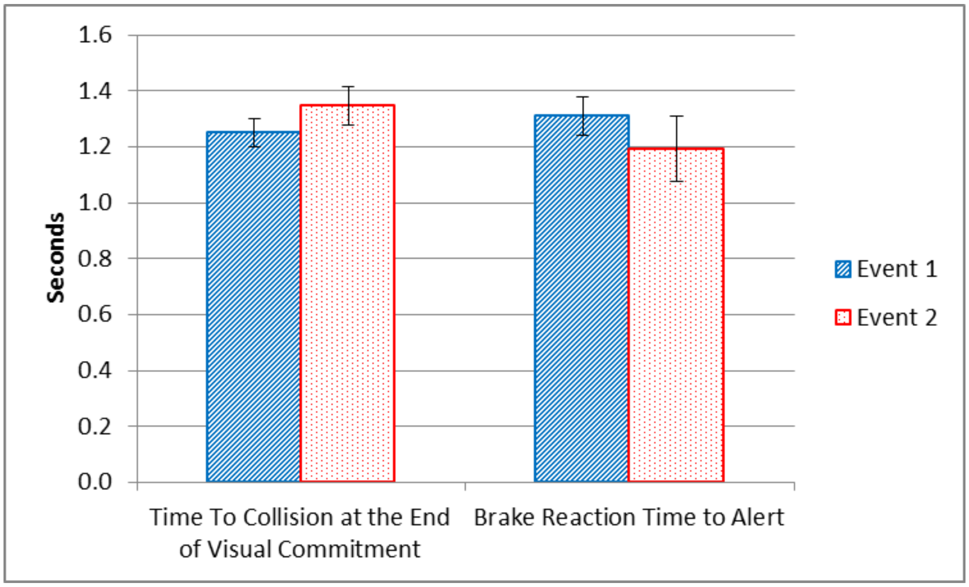


**Figure 23. Effect of repeated event on response in terms of TTC at the end of visual commitment and brake reaction time**

Another consideration is the amount of incomplete data that required replacement which can lead to a less efficient protocol and the potential for sample bias. The number of participants replaced due to lost data for each combination of exposure and awareness is provided in Table 7. These numbers represent how many additional participants needed to be run in each of those conditions to reach the required sample size of 24 participants per cell. Reasons for replacement include visual commitment ending prior to warning or headway that was not within appropriate range to trigger event. The condition where there is awareness of the presence of the system but no prior exposure has the greatest number of lost data needing replacement.

**Table 7. Participants requiring replacement due to missed events by Awareness and Exposure Condition**

| | | **Exposure** | |
|---|---|---|---|
| | | No Exposure | Exposure |
| **Awareness** | Not Aware | 7 | 6 |
| | Aware | 13 | 2 |

Discussion

When considering how much familiarity participants should have with the system in order to best assess the effectiveness of the warning system DVI, there are the practical considerations of what occurs in the real world and how different types of familiarity affect the ability to differentiate between DVIs of different effectiveness. The first consideration is largely philosophical as different drivers will have different familiarities with the systems in their vehicle (TIRF, 2012), and cannot be assessed by the data.

Providing participants with awareness that the crash warning systems were present in the car through a pre-drive training presentation resulted in no differences in engagement with the secondary task or driver response for the lane departure events. When considering the forward crash events, the presence of the pre-drive training resulted in greater minimum TTCs and faster brake applications. For neither type of event was there an interaction with type of DVI used for the warning system.

Providing participants with prior exposure to the warning system during the drive resulted in differences in visual commitment for the lane departure events. These differences were that with prior exposure, participants had longer engagement times with the secondary task, but that there was no resultant difference in end of visual commitment in response to the alert. When considering the forward crash events, there were no differences based solely on exposure, but when considering whether it was the first or second event, participants with prior exposure waited longer to end their visual commitment for the first event.

When considering the use of more than one crash warning event in a drive, there is little indication that an additional forced lane departure event has any impact on driver engagement or disengagement from the secondary task for the lane departure events even in the presence of additional unforced/unplanned lane departures associated with performing the secondary tasks, but that there are differences in response which may reflect greater familiarity with what the alert means. For the forward crash events, there are differences in secondary task engagement and response. Participants have shorter initial visual commitment duration and disengage faster from the task in response to the alert for the second event relative to the first event. Additionally, participants respond with a brake application quicker. This may indicate a greater sense of caution after experiencing the forward crash event.

When considering data needing replacement, providing exposure to the alert appears to reduce the amount of data needing replacement, but there is a mixed effect of providing prior awareness of the system.

Regarding the provision of pre-exposure or training on warnings prior to a critical event, knowing about the warning in greater detail seems to allow drivers to respond more quickly, and training tended to result in shorter engagements with the secondary task, though these effects were not particularly strong. While the decision to provide pre-exposure or training might be influenced by research objectives, there is some evidence that not providing pre-exposure or training results in slower warning response and longer visual commitments, which may aid in the staging of critical event trials.

The results of this study point to an advantage in terms of longer commitment times to allow the crash warning events to materialize when prior exposure to the alerts is given at the beginning of the drive.

The results of this study point to the use of more than one lane departure event without concern about the influence of the first event. However, caution is warranted with multiple forward crash events, as the results indicate that drivers might be more cautious after the initial event. It should be noted that for forward crash events, this experiment cannot provide guidance on multiple events where some would be less severe in terms of required driver response than those used in this study.

### 3.1.4 Experiment 4: Alert Timing and Protocol Sensitivity to Systems

Introduction

The aim of this study was to test the protocol elements identified in prior studies across two warning timings, to provide an understanding of how warning timing affects evaluation of the crash warning DVIs. This study examined alert modality and onset for both the FCW and LDW systems through six levels of alert modality and two levels of alert onset.

There were four main hypotheses for this study:

- The protocol will be able to differentiate between warning DVIs and between the DVIs and the baseline no alert condition in terms of initial driver response.
- Earlier warnings will produce a faster response to lane departures and forward collision events.
- There will be no interactive effects between DVI and warning timing – that is that the effectiveness of the DVI will not be dependent upon the warning timing.
- Drivers with a crash warning DVI will respond differently to false alarm events than drivers in the baseline condition who do not experience an alert.

Method

This study had a 2x6 between-subject experimental design. Ninety-six participants completed all study procedures successfully. Two levels of alert onset, early (FCW at 3.5s TTC, LDW at 12 inches from lane line) and late (FCW at 2.1s TTC, LDW at 4 inches from lane line), were presented with six alert modalities: no alert, audio, visual, two haptic, and an active haptic for both the FCW and LDW systems. The two levels of alert onset were used for the FCW and LDW as described in the prior experiments. It should be noted that for the FCW alerts, the timing of the alert relative to the engagement with the task remained fixed as shown in Figure 2. In practical terms this means that the stopped lead vehicle is at placed at different distances from the driver for these two warning timings. Six levels of alert modality for both the FCW and LDW systems were used. For FCW these included; an auditory alert comprised of repeated beeps; a visual alert comprised of HUD lights; a haptic alert comprised of a seat belt tensioner; a haptic alert comprised of seat vibration; an active/haptic alert comprised of a brake pulse; and a no alert condition. For LDW these included: an auditory alert comprised of an acoustic tone; a visual alert comprised of amber lights in the side view mirrors; a haptic alert comprised of steering wheel vibration; a haptic alert comprised of seat vibration; an active/haptic alert comprised of steering wheel torque; and a no alert condition. . The apparatus, scenarios, number recall task, and headway display remained the same as used in System Training and Exposure study described above.

Additionally, two false alarm events were evaluated to assess the extent to which DVI impacts whether the participant responds to the alert versus responding to the driving situation in response to the alert.. The false alarm event for the lane departure system involved the presence of old markings on the roadway for a lane shift. The participant is engaged with the secondary task and the system warns based on the markings for the old lane shift while the driver's eyes are off the road. The timing of the event is consistent with the forced departure events. The false alarm event for the forward crash system involved the system triggering off a stationary road hazard next to roadway after the lead vehicle passes it. The participant is engaged with the secondary task and the system warns based while the driver's eyes are off the road consistent with the timing for the forward crash events. For both false alarm events, there is no response necessary for the drivers with no warning.

Participants were told that the goal of the research was to evaluate several new in-vehicle technologies, when in fact their response to surprise LDW and FCW events was being evaluated. Participants were provided a training presentation that provided details on the drive, navigation instruction, headway display, secondary task, and general procedures. No training on the warning DVIs was given, as determined by System Training and Exposure experiment results.

Key Results

For LDW events, late warnings may be best to reduce nuisance alarms, but for FCW events the warning timing may not matter greatly and it may be appropriate to use whatever warning timing is associated with the DVI's warning system. The following results form the basis of these recommendations.

*LDW Results*. When looking at response to the alert for lane departure events, there were significant effects for timing and for warning type. There was also a significant interaction between warning onset and modality for steering reaction (p=0.0144) time but not for time to visual commitment end (p=0.2764). As can be seen in Figure 24, visual, auditory, vibrating steering wheel, and steering torque had faster reaction times with late warnings; seat vibration had a faster reaction time with the early warning. In general, later warnings resulted in faster reaction times but not universally. Additionally, all the warning types resulted in better performance than when no warning was provided; however, there is little difference in reaction time between DVIs when the warning is provided early, but greater differentiation with the late warning.

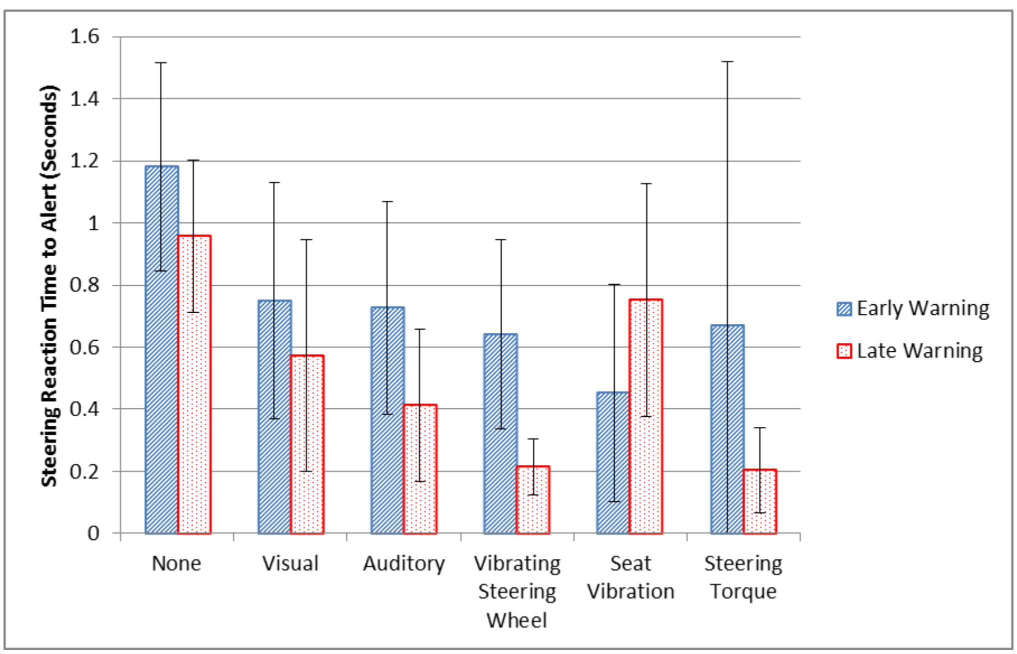


**Figure 24. Interaction between warning timing and warning type on steering reaction time**

When considering the effect of warning timing on time to end of visual commitment, there was a significant main effect (p=0.0006). As can be seen in Figure 25, for late warnings, drivers end their visual commitment faster in response to the warning than when an early warning is provided. For the late warning condition, driver's had been committed away from the forward view longer than with the early warning; and the relative number of total LDW alerts differed between the warning timings as the earlier warning timing resulted in more unforced/unplanned lane departure alerts across the drive.

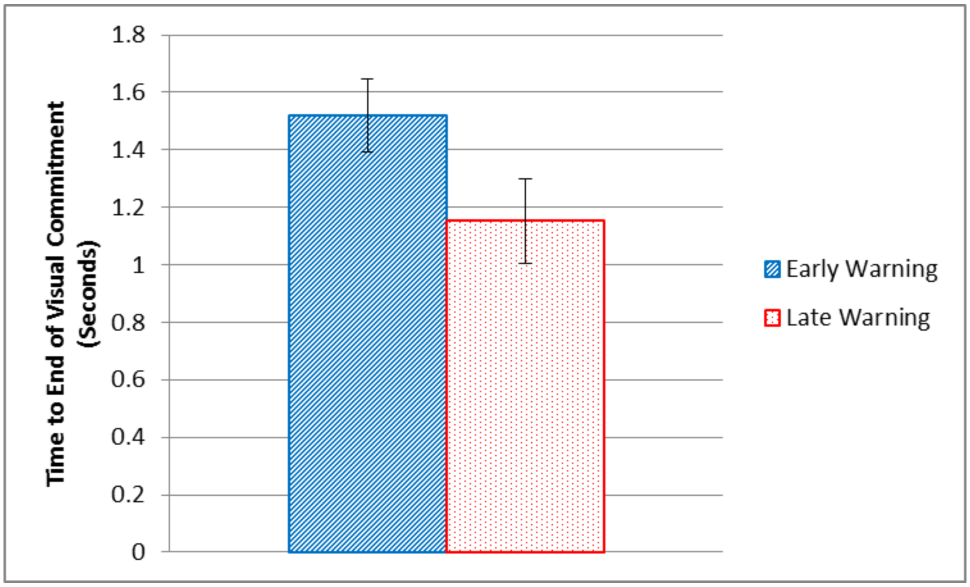


**Figure 25. Effect of warning timing on time to end of visual commitment**

Additionally, there was one main effect for gender. There was a significant difference for time to end of visual commitment (p=0.0006). Female drivers ended their visual commitment faster than males in response to the alert on average (see Figure 26).

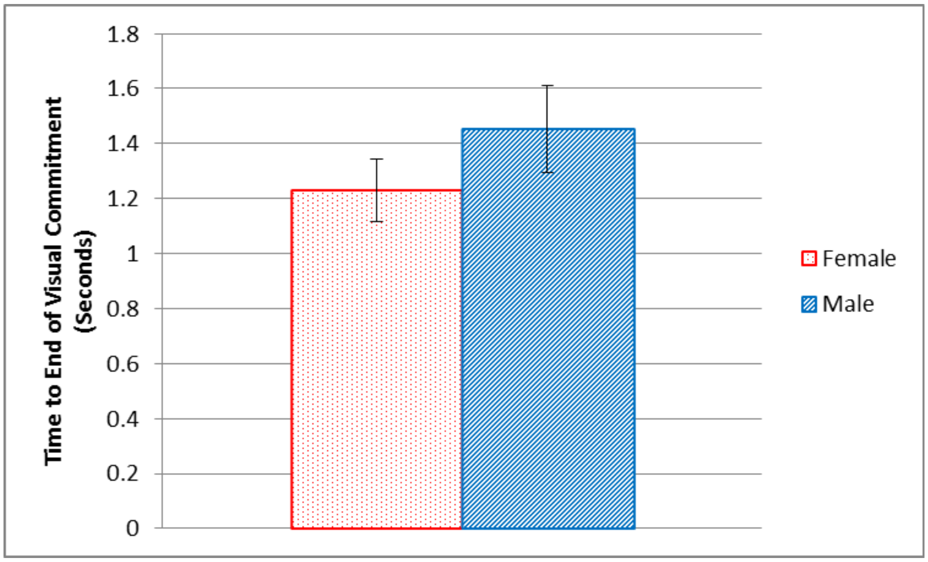


**Figure 26. Effect of gender on time to visual commitment end**

*FCW Results*. When looking at the two primary endpoints for response to the alert (time to visual commitment end and brake reaction time) for forward crash events, there was a significant effect for timing but not for warning type. There were no interactions directly between timing and warning type for either measure. There was a significant effect of warning timing for brake reaction time (p=0.0280), but not for time to end of visual commitment (p=0.2114). As can be seen in Figure 27, late warnings resulted in faster brake reaction times compared to early warnings. There were no significant effects for warning type for either time to end of visual commitment (p=0.0518) or brake reaction time (p=0.5366).

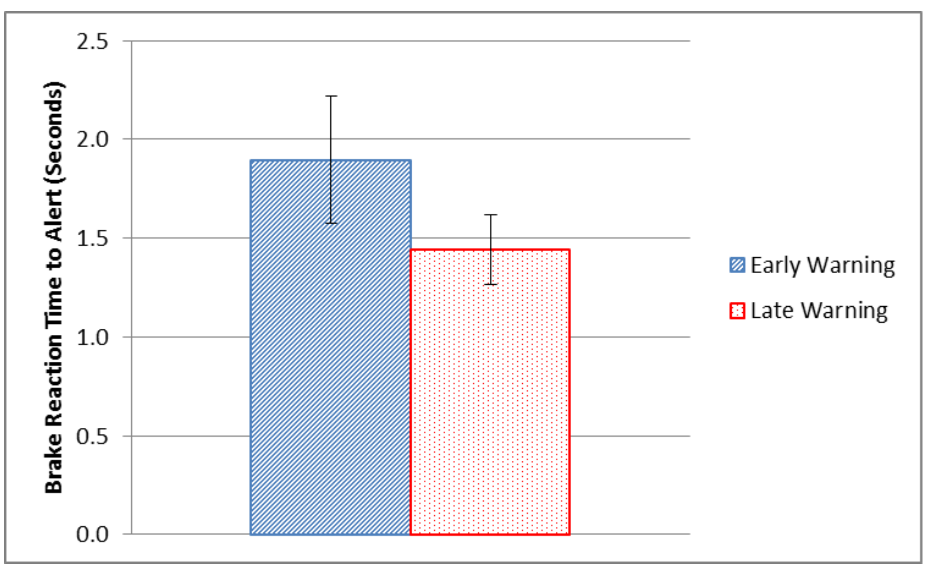

**Figure 27. Effect of warning timing on brake reaction time**

There was a three-way interaction between timing, warning type, and gender for several variables. Significant differences were found for visual commitment duration (p=0.0187), initial visual commitment duration (p=0.0066) and engagement duration (p=0.0156) as well as for adjusted minimum TTC (p=0.0351). Figure 28 shows the relationship for adjusted minimum TTC. Negative values indicate how much sooner the driver would have needed to respond to avoid a collision based on deceleration at collision and difference in velocity at collision. As can be seen, late warnings tend to result in smaller values, and males have smaller values than females in all but three cases. The very large negative values are indicative of little response by the driver before the collision occurred.

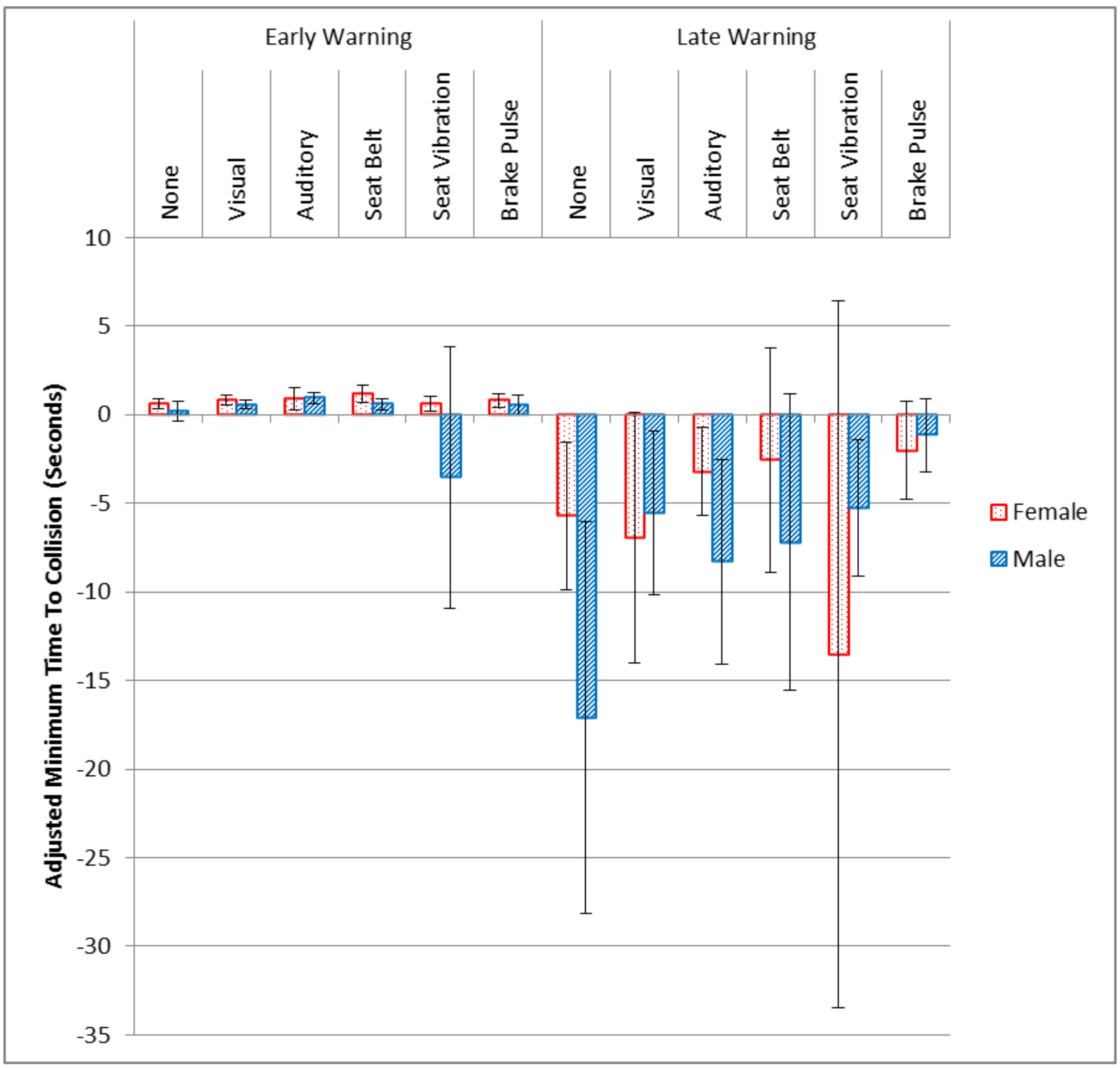


**Figure 28. Effect of warning timing, warning type and gender on adjusted minimum TTC**

As expected, there were main effects for timing and warning type for several additional measures. Timing had significant effects on TTC at visual commitment end (p<0.0001), minimum TTC (p<0.0001), and minimum adjusted TTC (p<0.0001). Differences in TTC at visual commitment end reflect the 1.4 s difference between the warning timings with the early timing resulting in a TTC of 2.4 s and the late timing resulting in a TTC of 1.1 s. Similarly, minimum TTC and minimum adjusted TTC were less in the late warning condition (45 ms and -6.5 s, respectively) compared to the early warning condition (723 ms and 0.4 s, respectively). These results indicate that the expected effects of manipulating time where observed, although as indicated above, the late warning timing showed greater variability within and between warning interfaces (DVIs).

In addition to these measures, there were also significant differences associated with event. There were significant effects of event interacting with warning type on time to visual commitment end (p=0.0092). As can be seen in Figure 29, there is no consistent pattern of improvement from the first event to the second event that would indicate learning. Some warning types such as auditory, seat vibration, and brake pulse showed trends consistent with a learning effect; whereas other the others did not.

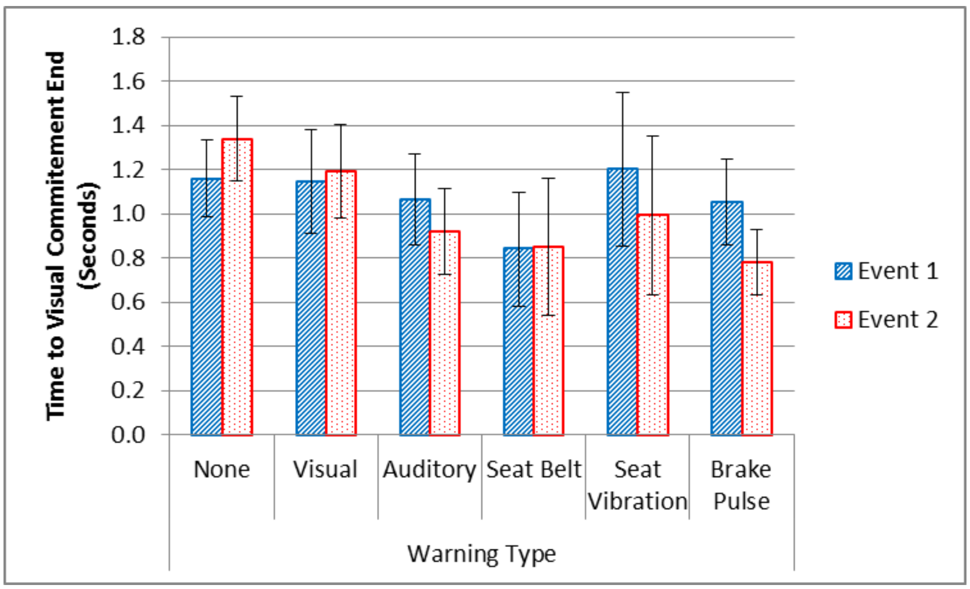


**Figure 29. Effect of event and warning type on time to visual commitment end**

There is also a significant interaction between timing and event for minimum TTC (p=0.0373). As can be seen in Figure 30, minimum TTC is greater for the second event with the early warning, but there is little difference in minimum TTC for the late warning.

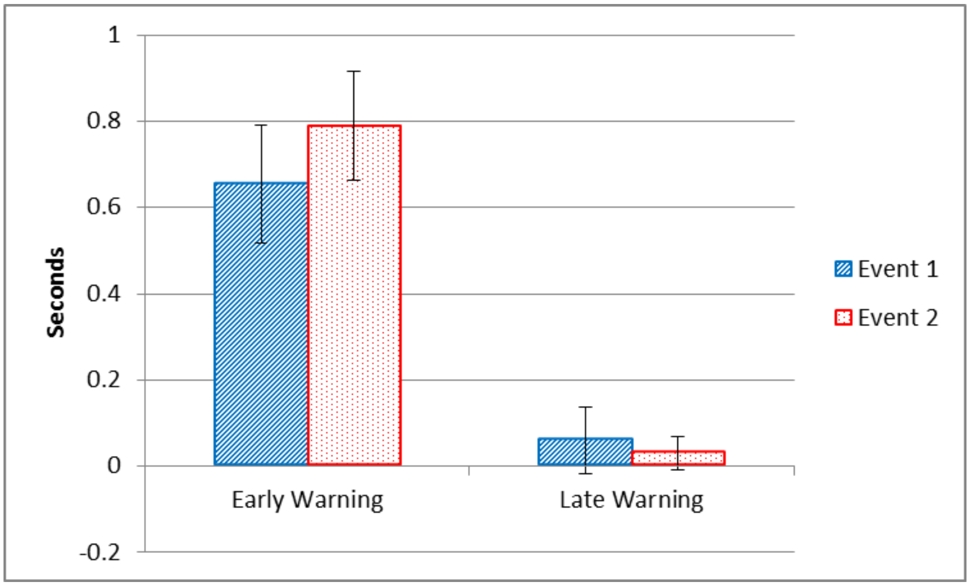


**Figure 30. Effect of warning timing and event on minimum TTC**

*False Alarms.* Because deployed systems will all experience some false alarms which could either degrade system effectiveness if the DVI is viewed as annoying or result in increased risk of crashes due to an unnecessary response, consideration of how drivers respond to a false alarm is important in considering the effectiveness of the warning system DVIs. Time to end of visual commitment and response reaction time (steering for lateral events and braking for longitudinal events) were analyzed for both the lane departure and forward crash events. There was one significant effect of interest for steering reaction time: an interaction between warning type and timing ($p=0.02821$) and a main effect of warning type ($p<0.0001$). The interaction shows no difference between timing for the no warning condition indicating the nominal timing of a steering input associated with general vehicle control (see Figure 31), but significant variability in steering reaction times for conditions that received the false alarm. The auditory alert appeared least likely to trigger an earlier steering response relative to the false alert than would normally have been observed with general vehicle control in the situation. This is more clearly illustrated in Figure 32. Participants with auditory or no warning had similar steering reaction times that were significantly longer than the other warning conditions.

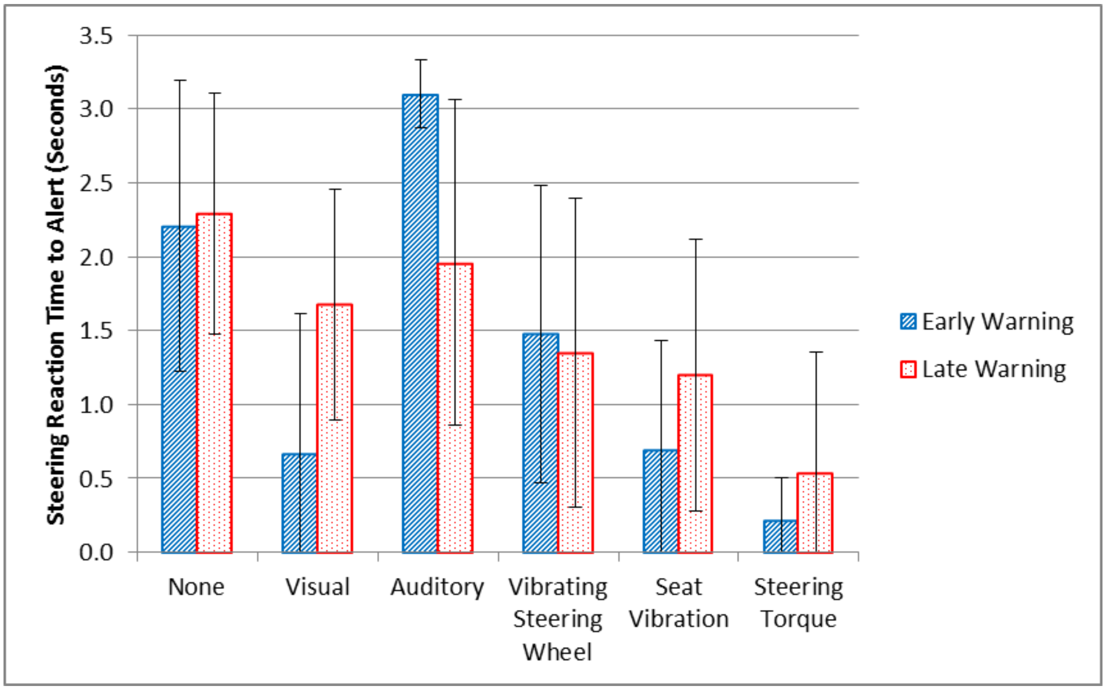


**Figure 31. Effect of warning type and timing on steering reaction time for false alarm**

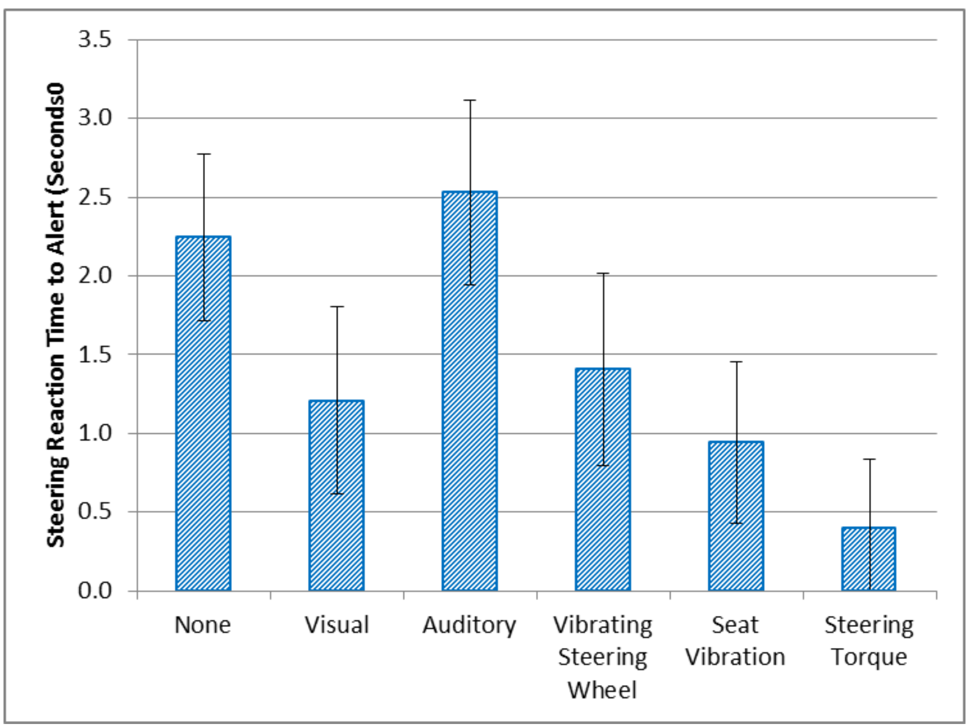


**Figure 32. Effect of warning type on steering reaction time for the LDW false alarm event**

Discussion

When considering the timing of the alerts and the impact on evaluation of DVIs, consideration must be given to the complexities of the interaction between timing and warning type for both lane departure and forward crash events. LDW DVIs seemed to show the greatest effect of warning timing, with the two timings providing different rank ordering of benefit for the DVI tested in terms of initiating a response by the driver. This may be in part because the earlier warning resulted in more alarms not associated with the planned events (occasions without a forced departure) across the drive due to the alarm triggering further from the lane boundaries. The presence of these potential nuisance alarms could have changed how drivers responded to more critical alerts such as the forced departures. This would argue toward a later warning – meaning that alarms would not trigger until the car was closer to the lane boundary - to minimize the impacts of these potential nuisance alarms. For forward crash events, there was no clear indication that one of the timings used in this study would be better than the other for evaluating the DVIs, but instead that it would shift the values based on the timing.

For LDW systems, the data suggest that a later alert is preferable to an very early alert because a very early alert can result in many nuisance alarms. The later warning need not be the one used in a final protocol for evaluating DVIs, but something near that value would be expected to work effectively based on these results. When considering the evaluation of DVIs for FCW systems, warning timing in the range tested in this study would be appropriate as timing had little impact on differences in effectiveness between the DVIs and instead just shifted the values. For the final protocol, this could easily mean that using the timing from the production system may provide an effective comparison so long as they do not differ significantly from the range of timings tested.

The false alarm events provided limited insight to driver response to alarms not related to actual threats to the driver. The FCW false alarm event did not find differences between the warning types indicating that any false alarm effects would be associated with the system rather than the DVI. The LDW false alarms provided no clear differentiation of effect for time to end of visual commitment due to a muddled three-way interaction but did allow differentiation in terms of steering response indicating that for LDW DVI evaluation, the impact of false alarms can be affected by the DVI. Overall, the false alarms used seem effective for the LDW systems but not for the FCW system.

### 3.1.5 Experiment 5: Motion Feedback Considerations

Introduction

The aim of this study was to examine the effect of motion cueing on driver response to crash warning DVI – meaning does the level of simulator motion cuing change how the DVI is experienced by the participant and impact the interpretation of the effectiveness of the DVI. This study investigated the effects of experimental platform by comparing three levels of motion cueing. This was accomplished by comparing a subset of the data from Experiment 4 with data from two levels of lower motion fidelity collected in this experiment: limited motion and no motion. The incentive structure, distraction task, and level of familiarity used during this data collection were those identified previously and used in the Alert Timing and Protocol Sensitivity to Systems study described above.

There were three main hypotheses for this study:

- For DVIs that do not provide haptic inputs, there will be no significant differences in initial driver response across simulator platforms.
- For DVIs that provide vestibular – motion and acceleration - inputs, there will be a significant difference in initial driver response between simulator configurations that provide motion cueing relative to the no motion configuration.
- There will be a significant difference between simulator configurations in terms of minimum time-to-collision and area of exceedance.

<u>Method</u>

The experimental design was a 3x3 between-subject design. Forty-eight participants completed all study procedures successfully. The independent variables were three levels of platform (NADS-1, NADS-1 limited motion, and NADS-1 with no motion) and three levels of alert modality (no alert, auditory, and seatbelt tensioner for FCW and haptic seat for LDW). No new data were collected on NADS-1 with full motion for this study. Data used for comparison was collected in the previous Alert Timing and Protocol Sensitivity study. The apparatus and scenarios remained the same as used in the Alert Timing and Protocol Sensitivity study described above. The number recall task and the headway display used were consistent with the prior study.

Participants were told that the goal of the research was to evaluate several new in-vehicle technologies, when in fact their response to surprise LDW and FCW events was being evaluated. Participants were instructed to maintain position in the green zone of the headway display provided. Participants were provided a training presentation that provided details on the drive, navigation instruction, headway display, secondary task and general procedures.

<u>Key Results</u>

Results of this experiment suggest that full motion cueing, as similar as possible to NADS-1 motion cuing, is beneficial for FCW evaluations. For LDW evaluations, however, limited motion cuing is likely to be adequate. This recommendation is limited to the consideration of relative effects rather than the determination of absolute effects on changes in performance. The following results form the basis of that recommendation.

*LDW Results*. When looking at the two primary endpoints for response to the alert (time to visual commitment end and duration of the lane exceedance) for lane departure events, there was a significant effect for motion but not for warning type for steering reaction time. There were no significant effects for time to end of visual commitment by motion or warning type. There were no significant interactions for these measures. Motion (p=0.0220) had significant impacts on steering reaction time. Figure 33 shows the impact of motion on steering reaction time. As can be seen, the steering response to the alert is delayed when no motion cues are provided. Warning type did not have a significant effect on steering reaction time (p=0.1801). Neither motion nor warning type significantly affected time to visual commitment end (p=0.0840, p=0.0733, respectively). There were no significant interactions for either measure with motion.

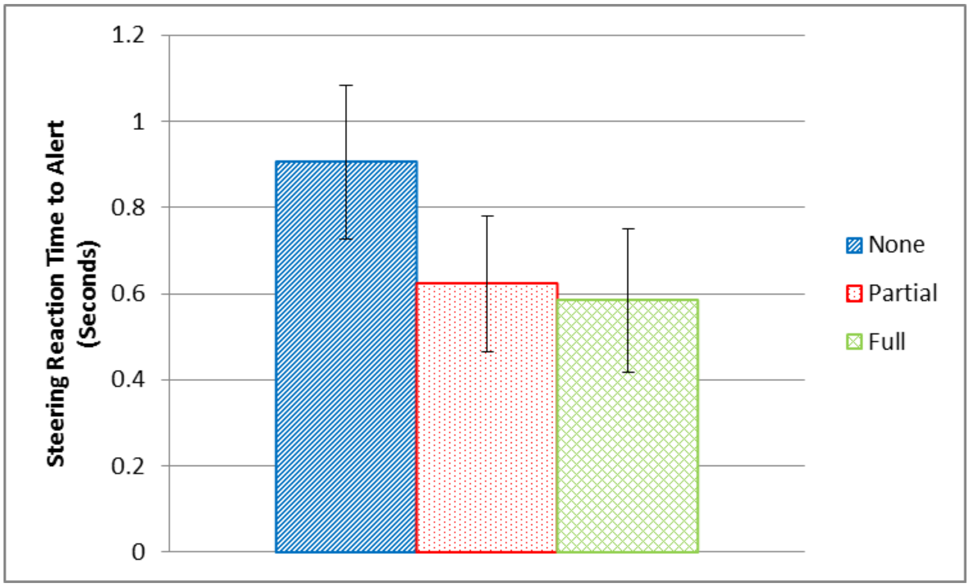


**Figure 33. Effect of motion on steering reaction time**

Motion had several other significant effects on visual commitment and the outcome of the lane departure events. Both the duration of first visual commitment (p=0.0032) and total duration of visual commitment (p=0.0039) were significant: Engagement duration was also significant but aligned with duration of visual commitment and is not reported here. Figure 34 shows the two sets of results. For overall duration, there are significantly shorter commitments with no and partial motion compared to full motion; however, for initial commitment partial motion results in a significantly shorter commitment compared to no motion and full motion. There is also a significant effect of motion on the duration of the lane exceedance (p=0.0185). Figure 35 shows that the duration of the departure is greater with no motion compared to partial motion, but full motion does not differ statistically from either no motion or partial motion.

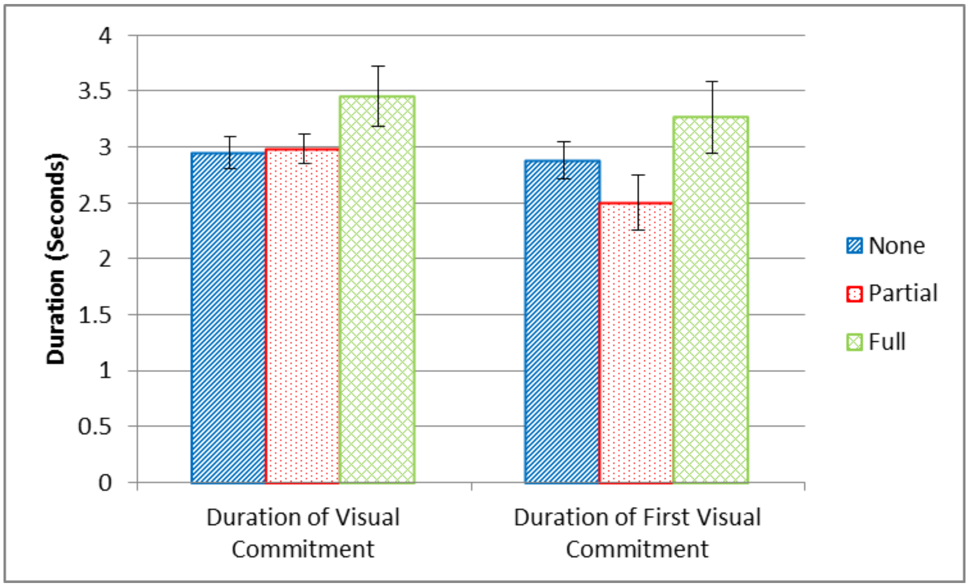


**Figure 34. Effect of motion on visual commitment**

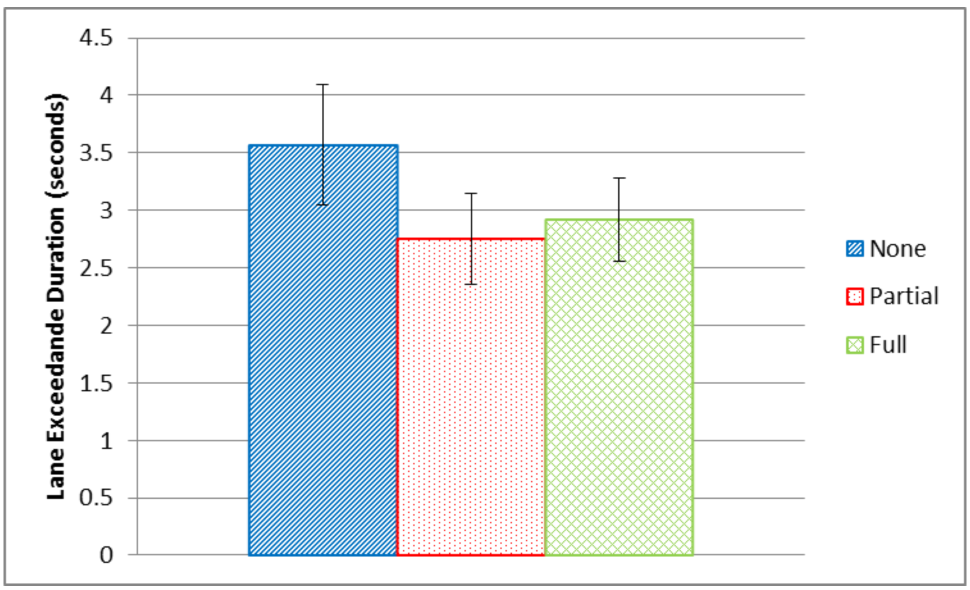


**Figure 35. Effect of motion on duration of lane exceedance**

There was also a significant effect of warning type on visual commitment duration ($p=0.0458$). As can be seen in Figure 36, the auditory alert resulted in a significantly longer commitment duration compared to the seat vibration.

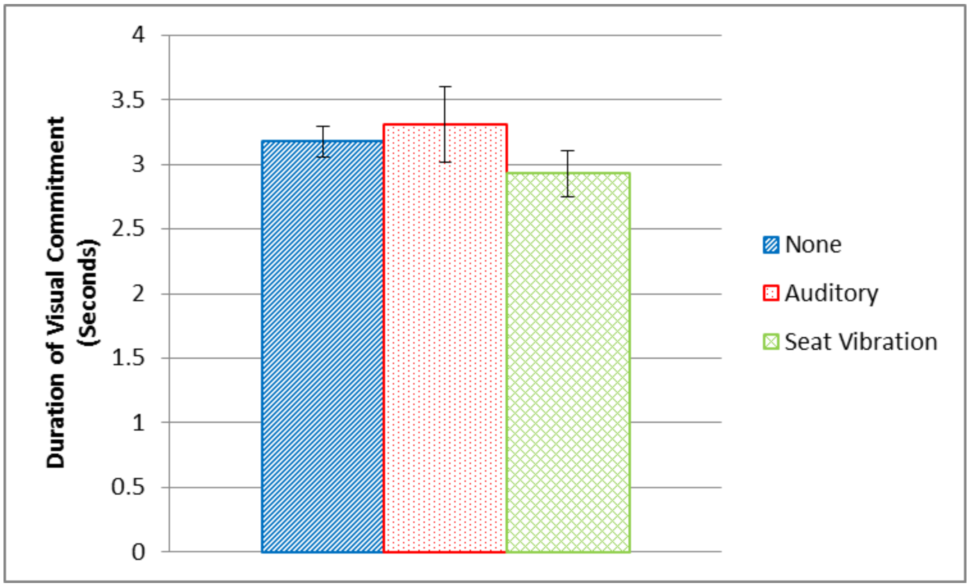


**Figure 36. Effect of warning type on visual commitment duration**

*FCW Results*. When looking at the two primary endpoints for response to the alert (time to visual commitment end and brake reaction time) for forward crash events there are a number of interactive effects that must be considered before the main effects for motion and warning type. There are two significant three-way interactions that include motion and warning type. First, the interaction between motion, warning type, and gender is considered for time to end of visual commitment (p=0.0435) – this interaction was not significant for brake reaction time (p=0.0677). The interaction for time to end of visual commitment is illustrated in Figure 37. As can be seen, the relationship is complex with consistent pattern across motion conditions, although the general trend is for auditory and seat belt tensioner to have shorter times than when no warning is present. Next the interaction between motion, gender, and event is considered. This interaction is significant for both time to visual commitment end (p=0.0187) and brake reaction time (p=0.0342).

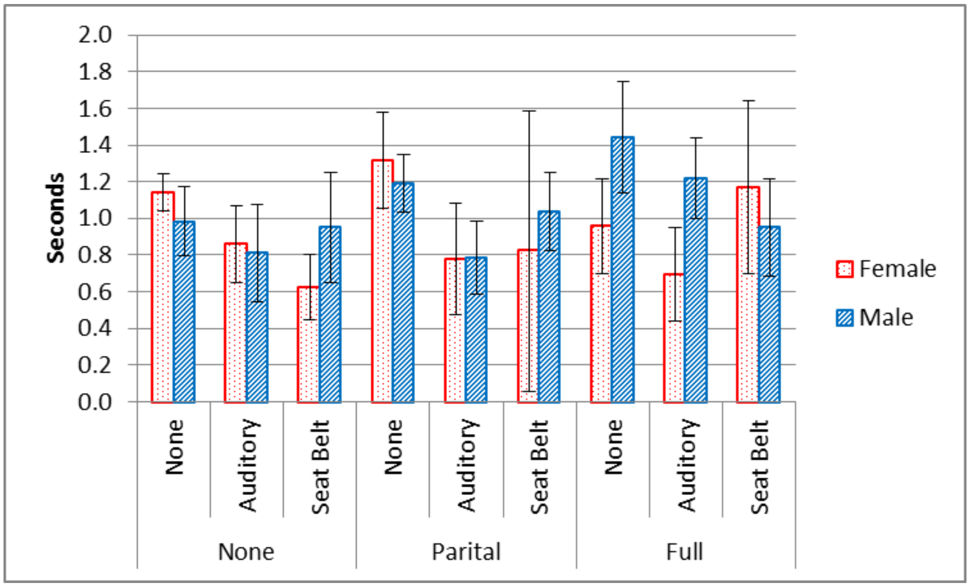


**Figure 37. Effects of motion, warning type and gender on time to end of visual commitment**

When considering the other effects associated with motion and warning type, the other visual commitment measures also have complex three-way interactions between motion, warning type and gender with no clear main effect. The same is true for outcome measures such as minimum adjusted TTC. Due to their complexity and lack of additional insight, they are not presented here.

<u>Discussion</u>

When considering the effect of motion on the evaluation of DVIs for crash warning systems, the primary concern is for interactive effects between motion and warning type. For the lane departure events, interactive effects of this type were not a concern. The main effects of motion largely shifted the distribution but did not change relative effectiveness of the DVIs. When looking at impact of no motion, there tend to be shorter engagements with the secondary task but more severe outcomes (longer duration lane exceedance). When looking at the impact of partial motion, there tend to be shorter engagements and less severe outcomes. It is important to note that one limitation of this study is that although it precisely controls for the study of motion effects, it does not provide insight into how, for example, a simulator with a smaller forward field of view would perform, even if it had the same motion capabilities as NADS-1.

The concern about interactive effects becomes particularly acute for the forward crash events. The plethora of interactive effects associated with motion and/or warning type makes definitive determinations difficult. For example, females tend to have decreasing time to end of visual commitment with motion for auditory warnings, but increasing for seat belt warnings; males are more consistent across motion for seat belt warnings, but longer time to visual commitment end with increasing motion cueing for other warnings. NADS-1 with full motion capabilities is the closest of the three motion configurations to real world driving and absent information indicating similar results with less motion, care must be taken when collecting, analyzing, and interpreting data on DVIs for FCW systems.

Based upon the results of this study, lack of full motion seems to have little effect on relative performance of the LDW DVIs, indicating that partial or no motion is likely acceptable for evaluations of relative DVI performance. The same is not true for FCW evaluations. The complex interactions point to the need to match the real world as closely as possible. This would argue for using the highest fidelity motion cueing available.

### 3.1.6 General Discussion

Research Implications and Methodological Factors

Overall, the key implications of this line of research are that a protocol to evaluate the DVIs for crash warning systems is achievable but that that there are some considerations that will impact how the systems are evaluated. These can be categorized into five broad areas: test platform, general scenario issues, motivation, measures, and sampling.

The first of these areas relates to test platform. It appears that it should be feasible to develop a protocol that could be used on either a test track on in the simulator; however, the constraints of the test track such as its configuration and space available, may limit the types of events that can be triggered. At least for forward crash events, the results appear similar between the environments and between a virtual test track and a virtual road course. For simulators, it appears that there is some flexibility in motion fidelity. There are differences in DVI performance for FCW alerts based upon the motion configuration that need to be considered. Based on these findings, the use of the highest fidelity motion available would be ideal for forward crash events; however, lane departure events seem to work well across motion configurations. It should be noted that the experiments were all done with 360-degree visuals and a full vehicle cab, and the extent to which other simulator configurations would be affected is unknown.

The second area relates to general scenario configuration. For some of this, it is less clear exactly what is the best approach, as it may be dependent upon the context of the overall evaluation. For example, what should researchers expect the driver to know about systems in their vehicle? Do researchers assume that they are aware of how the vehicle they are driving is configured and have some understanding of what the DVI is? For now, the approach to protocol development focuses on maximizing the sensitivity of the protocol and suggests providing exposure but no training. If, however, if it becomes clear that one particular combination of awareness and exposure is predominant, then the protocol should be adjusted accordingly. The other major scenario configuration issue is the number of times to expose drivers to the alerts. It seems that drivers largely find the lane departure events to be the result of their engagement with the task and little differences is seen between events; however, differences between the first and second exposure to the FCW DVI in the context of a crash situation does result in differences, and as such, second exposures should probably be avoided in a final protocol.

The third area relates to motivation. It is difficult in any experiment to replicate inherent driver motivation and that is only made more difficult when considering the motivation that leads drivers to engage in a non-driving task that is of sufficient duration to result in a situation where a crash warning is necessary. The use of financial incentives is an attractive option; however, they do not provide a panacea to the issue of motivation. Participants have different motivations for volunteering to participate in a research study, and for many, the primary motivation is not financial. For those people, the incentive does not help to achieve the performance desired. Additionally, for those who have a financial motivation, they may overcompensate and not respond as expected when presented with a crash warning in the context of a study. This study found some concern to be borne out in the data. As such, the use of financial incentives in this context does not appear to provide any benefit.

The fourth area relates to measures of DVI effectiveness. There is a plethora of measures that can be examined to assess system effectiveness. Because the systems are designed to alert the driver to an impending crash in an effort to get the driver to reengage with the driving task and begin an avoidance maneuver, the primary measures have thus been time to end of visual commitment and response time. It is important to understand that response time is a composite measure that includes not only the time to return visual commitment to the driving task but also the time to initiate the avoidance response. There is some evidence that this composite measure may illustrate that drivers adapt how quickly they begin their response after returning their attention to the road based on the situation. Additional measures and additional ways of combining the measures need further examination.

The fifth relates to sampling. There were several results that showed differences by gender, particularly as it relates to engagement with the secondary task. Females appear to be less willing than males to engage in the secondary task as evidenced by longer times to begin the tasks, and shorter initial visual commitment durations. This is further complicated by the fact that gender interacts with warning type in several cases meaning that the interface that is most effective for males may not be the same one that is most effective for females; although the magnitude of the differences were generally small.

Consideration of Dependent Measures

In reviewing the dependent measures used across the studies, it is important to consider which measures are most appropriate to distinguish differences among DVIs. The aim of the systems being evaluated is to alert the driver to a dangerous situation so that they may return their attention to the roadway and avoid a negative outcome associated with the situation. Measures can therefore be logically divided into two categories: Visual attention and response/outcome. Measures of visual attention relate to how well the DVI performs in getting the driver to disengage with the secondary task and reengage with the driving task. Response/outcome measures relate to how well the DVI performs in terms of promoting appropriate vehicle control actions and preventing or reducing negative outcomes such as collisions and out-of-lane exposure. The response/outcome measures can be further divided into those appropriate for an FCW event and those for an LDW event. Table 8 provides a summary of the statistical significance of the different measures across the final three NADS-1 protocol studies for both FCW and LDW. The table details whether there were statistically significant effects ($p<0.05$), a significant interaction ($p<0.05$), or if there was a main effect trend ($0.05<p<0.10$) related to DVI. Significant main effects are preferentially provided over significant interactive effects, and significant interactive effects are preferentially provided over trends. Additionally, effect sizes relative to the no-warning condition are reported for the Alert Timing and Protocol Sensitivity to Systems and Motion Feedback Considerations Studies in Table 9. This table shows the average and maximum Cohen D effect size for each of the dependent variables, along with an average of those across studies.

It should be kept in mind that the series of experiments was designed to evaluate various methodological options (e.g., alert timing, training/exposure to DVIs) for the protocol and not to actually compare alternative DVIs. Sample sizes and experimental design features were not devised to optimize this discrimination among example DVIs and various measures for comparison. Therefore it should not be surprising that statistically significant effects of DVI were not always obtained for some dependent measures within a given experiment. For this reason, a more liberal "trend" outcome is included in Table 8 along with the "significant" outcome.

When considering visual attention, most of the measures examined showed some level of effect across studies for both FCW and LDW events. Although the measures all share some common elements relating to the visual attention of the driver over a narrow period of time and performed similarly, time to visual commitment end is the measure that is most on point in terms of measuring the effectiveness of the DVI at re-directing driver attention, as it considers only the time from the alert (or when the alert would have occurred) to the driver returning attention to the forward roadway. This measure had significant main effects for two of the three FCW analyses and a significant interaction in the other. For the LDW case, there were significant interactions in two analyses and a trend toward a significant main effect in the other. Additionally, this measure had the largest average and maximum effect size across studies. Of potential concern with this measure is that for Alert Timing Study for FCWs, the interaction between DVI and event (first versus second) showed larger commitment times for the second event versus the first. Specifically, there was a longer time for the no-warning condition on the second instance of the event. As no warning was provided to this group in either case, it is unclear why they exhibited a 180 ms increase while the auditory, seat vibration, and brake pulse exhibited decreases of 145, 207 and 274 ms, respectively, and the visual and seat belt tensioner DVIs exhibit no changes. This may point to some natural variability associated with when and how drivers choose to disengage from distracting tasks. As such, it is important to look beyond the mean to the distribution when considering this measure.

When considering the outcome/response for the FCW events, minimum TTC and collisions did not show significant effects; brake reaction time showed an effect for one of the studies and an interactive effect for another; and adjusted minimum TTC showed significant effects across two studies and an interactive effect for the third. Brake reaction time showed the largest overall effects but failed to show a significant effect or trend for DVI in the Alert Timing and Protocol Sensitivity to Systems Study that was intended to explore the sensitivity of the protocol and measures. Adjusted minimum TTC considers both minimum TTC and collision speed which provides a good summary of safety margin without the range restriction of minimum TTC and collision velocity, and without the binary nature of collisions which effects power. Additionally, it performed well with regard to effect size showing overall effects sizes that were moderate to large.

When considering the outcome/response for the LDW events, area of lane exceedance and maximum lane exceedance showed statistically significant differences associated with DVI for one of the three studies but not the other two; and duration of lane exceedance and steering reaction time showed statistically significant main effects for one study and an interactive effects for another. When considering the effect sizes – the strength of differences between conditions – for these two measures, duration of lane exceedance has a moderate effect size and steering reaction time has a moderate to large effect size. When comparing these two measures they break down into a measure of response (i.e., how quickly steering input is provided) and a measure of outcome (i.e., how long out of the lane). Both provide value and tell a different part of the story; although the outcome measures may provide a better overall perspective on the effectiveness of the response to the DVI.

Overall, although arguments could be made for a variety of measures, time to visual commitment end worked as well as the others and has the value of having been shown to translate well to the test track (Forkenbrock et al., 2011) and would serve well as a primary measure of the effectiveness of returning the driver's attention to the road. For the outcome/response measures for FCW and LDW, the most appropriate measures appear to be adjusted minimum TTC for the evaluation of FCW DVIs and duration of lane exceedance for the evaluation of LDW DVIs. This is particularly true when considering that the protocol will look at shifts in distributions rather than just comparing means in a statistical test.

**Table 8. Summary of statistical significance across dependent measures for final three experiments**

| | FCW | | | LDW | | |
|---|---|---|---|---|---|---|
| | System Training and Exposure Study (No baseline "No warning" Condition) | Alert Timing and Protocol Sensitivity to Systems Study | Motion Feedback Considerations Study | System Training and Exposure Study (No baseline "No warning" Condition) | Alert Timing and Protocol Sensitivity to Systems Study | Motion Feedback Considerations Study |
| Time to Visual Commitment End from Warning | Significant | Interaction | Significant | Interaction | Interaction | Trend |
| Duration of Visual Commitment | Significant | Interaction | Significant | Significant | Interaction | Significant |
| Duration of Visual Engagement | Significant | Interaction | Significant | Significant | Interaction | Interaction |
| Minimum TTC | | | | | | |
| Adjusted Minimum TTC | Interaction | Significant | Significant | | | |
| Collisions | | | | | | |
| Brake Reaction Time | Interaction | | Significant | | | |
| Area of Lane Exceedance | | | | | Significant | |
| Duration of Lane Exceedance | | | | | Significant | Interaction |
| Maximum Lane Exceedance | | | | | Significant | |
| Steering Reaction Time | | | | | Significant | Interaction |

**Table 9. Summary of average and maximum Cohen D effect size across dependent measures and experiments focused on comparing alerts to baseline**

| | FCW | | LDW | | |
|---|---|---|---|---|---|
| | Alert Timing and Protocol Sensitivity to Systems Study | Motion Feedback Considerations Study | Alert Timing and Protocol Sensitivity to Systems Study | Motion Feedback Considerations Study | Across Studies |
| Time to Visual Commitment End from Warning | 0.68/1.13 | 0.83/0.99 | 0.34//0.63 | 0.00/0.31 | 0.46/0.77 |
| Duration of Visual Commitment | 0.54/1.07 | 0.56/0.69 | 0.35/0.67 | 0.15/0.50 | 0.40/0.73 |
| Duration of Visual Engagement | 0.52/1.02 | 0.54/0.69 | 0.35/0.61 | 0.00/0.29 | 0.35/0.65 |
| Minimum TTC | 0.45/0.68 | 0.05/.016 | | | 0.25/0.42 |
| Adjusted Minimum TTC | 0.42/0.79 | 0.55/.076 | | | 0.49/0.78 |
| Brake Reaction Time | 0.22/1.00 | 0.75/.1.01 | | | 0.49/1.00 |
| Area of Lane Exceedance | | | 0.63/0.79 | 0.34/0.38 | 0.48/0.59 |
| Duration of Lane Exceedance | | | 0.66/0.90 | 0.33/0.39 | 0.50/0.65 |
| Maximum Lane Exceedance | | | 0.48/0.69 | 0.30/0.33 | 0.39/0.51 |
| Steering Reaction Time | | | 0.92/1.23 | 0.33/0.53 | 0.63/0.88 |

## 3.2 Crash Warning Evaluation in a Connected Vehicle (CV) Environment

There is a concern that drivers may soon be visually or cognitively distracted by greater levels of in-vehicle information than they are now. This may be related to V2V/V2I communications, henceforth referred to as CV, or come from other information sources. This experiment modified the protocol as developed in the previous set of experiments described in this section to include such messages as the distraction task that preceded a critical event trial. The goal was to access

### 3.2.1 Experiment 6: Evaluation of Protocol with Information Search Task

Introduction

The aim of this study was to provide an understanding of how distraction task type affects evaluation of crash warning DVIs. This was done by replicating the test protocol elements executed in the Alert Timing and Protocol Sensitivity to Systems study using distraction tasks that did not interfere with assessing visual FCW presented in the driver's forward view. This study examined alert modality for both the FCW and LDW systems through three levels of alert modality presented to participants. However, only the FCW events are included in the current analysis. The incentive structure and level of familiarity used during this data collection were those used in Alert Timing and Protocol Sensitivity to Systems study. After the surprise events, false alarm events were presented to the driver to examine their response to false positive or nuisance events.

There was one main hypothesis for this study: The protocol and warning set developed to date is robust enough to detect differences in driver response to a visual and/or haptic crash warning compared to a no-warning baseline condition while the driver is engaged in a cognitively and visually demanding information search task emulating a CV system.

Method

This study had a three-level between-subject experimental design with 24 participants aged 35-55 completing all experimental procedures. Three levels of alert modalities were used: no alert, visual, and haptic (eight participants, balanced by gender, in each modality condition) for both the FCW and LDW systems. Although the LDW events remained in the study drive, they are not included in this experimental design and are not part of the data reduction or analysis. The late alert timing implemented in previous studies, Figure 2, was used for the FCW events in this study.

The platform, apparatus and driving scenarios were the same as those used in the System Training and Exposure study with the exception of the in-vehicle tasks. A new task was added, a CV information search task. This task was modeled after the information search task used in Connected Vehicle DVI Design Research and Distraction Assessment (DTNH22-11-D-00237/0001, final report not published at time of this submission) and was modified to include additional faux applications and presented search questions visually (see Figure 38). Additionally, the new screens were updated for display on a larger 7-inch center stack touchscreen display. These screens are based on the Windows 8 icon design and layout. The font size and contrast were adjusted to be in compliance with ISO 15008: 2009 standards for in-vehicle visual presentations. Both the CV information search task and the message task were used.


**Figure 38. Information search screen modeled after the Windows 8 tile layout**

The CV task and the message reading task from previous studies were both moved to a location near the top of the center stack to more closely mimic tasks that may be performed through an in-vehicle touch screen. This location was the same as for the message reading and menu selection task in the experiment on Secondary Task and Incentive Considerations described above. The information search task was always associated with the FCW and LDW events. As in previous studies, participants were instructed to maintain position in the green zone of the headway display provided.

Compensation for this study was selected based on the results of the prior study. Base compensation was set at $45. The base pay was prorated. If participation lasted less than one hour, participants earned $10. If participation lasted for over one hour, participants earned the full base pay.

<u>Key Results</u>

When looking at the two primary endpoints for response to the alert for forward crash events, several challenges in analyzing the data were encountered. As each participant had the opportunity for two braking events, there were a total of 48 possible braking events to analyze. In examination of the data, there were a total of eleven braking events (22.9%) where the end of visual commitment could be attributed to the alert (seven for the first event, and four for the second event); additionally, there were a total of seventeen braking events (35.4%) where the braking came after the alert and could have been affected by the alert. Overall, 25% of the FCW events did not trigger because the driver was not appropriately engaged in the task. The biggest challenge with the scarcity of the data is that, despite the fact that each driver was attending to the display as the event began for one of their two events, there are no usable data for the no-warning condition due to participants looking back too soon, making it impossible to assess the benefit of the system relative to the no-warning condition.

With this in mind, the data from the seventeen brake applications can still be examined. In analyzing these data, there was no statistical difference between the visual warning and seat belt tensioner ($p=0.4813$), with the visual warning having a mean brake reaction time of 2.32 s versus the seat belt tensioner having a mean brake reaction time of 2.26 s.

Engagement with the secondary task is considered when trying to better understand engagement with the CV task. Due to the nature of the task, overall duration of the task was longer for the CV task than for the numbers task by design; however, examination of the initial visual commitment does shed some light on the differences between experiments. For the CV task, there was no statistical difference in initial visual commitment duration by warning type ($p=0.6375$), but there was by gender (0.0341) with females have shorter initial visual commitments than males. Figure 39 shows the differences by gender for both the CV task and for the numbers task. As was stated in the discussion of the Alert Timing and Protocol Sensitivity Study, there was also a statistical difference in initial visual commitment duration between females and males with the numbers task (p=0.0335); although the magnitude of the differences differed with a 250 ms difference for the numbers task and a 720 ms difference for the CV task. An even bigger difference exists between the two tasks than was observed for gender with the initial visual commitment duration for the CV task being 1140 ms shorter than for the numbers task.

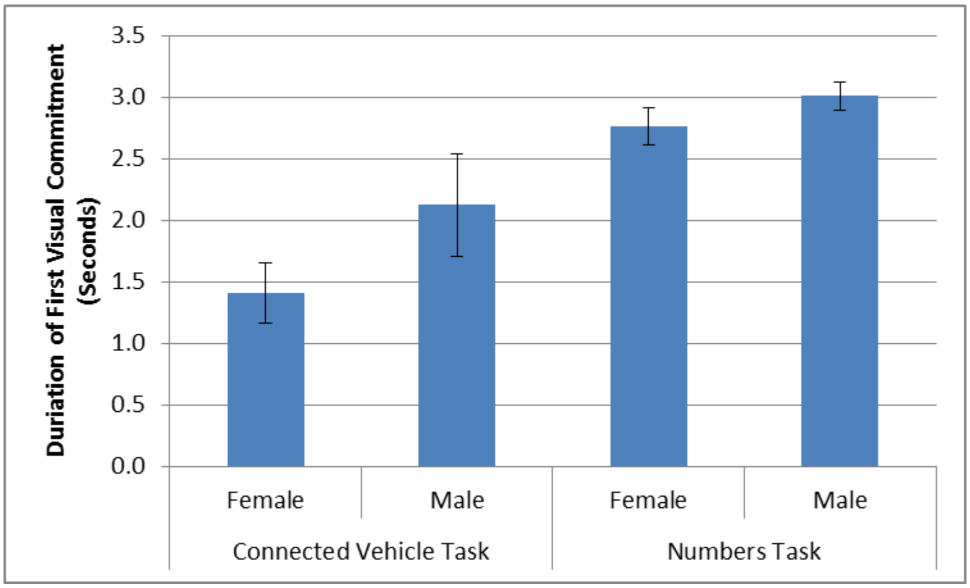


**Figure 39. Duration of first initial visual commitment by gender and task**

Discussion

One of the premises of the CWIM effort was that drivers sometimes, if rarely, engage in tasks that require unusually long glances or choose to employ long glances away from the forward view. Whether this is due to the task design or the driver simply finds the task compelling in that moment, the protocol developed in this effort allows the evaluation of FCW and LDW DVIs within that context. During the CV portion of this effort, a task that did not require long glances and was positioned near the forward view was shown to be unsuitable for this alert evaluation protocol. The CV task was designed to meet NHTSA's distraction guidelines for in-vehicle tasks, which require short glance times and tasks to be interruptible. The position of the CV task near the forward view also allowed changes in the driving environment to be detected in the driver's peripheral vision. Since this protocol was designed to evaluate FCW and LDW alerts when drivers employ longer glances away from the forward view to engage in a task that is not interruptible, it is not surprising that a task that met NHTSA distraction guidelines did not produce appropriate glance behavior. However, it should be noted that the goal of this effort was not to evaluate the distraction levels associated with specific tasks, but to evaluate DVIs within the context of tasks that require significant levels of visual commitment. The main recommendation is to employ in-vehicle tasks that require long glances away from the forward view, although it is possible that with large sample sizes, subtle differences in DVIs could be detected with tasks located near the forward view that require short glances.

### 3.2.2 General Discussion

When taken in conjunction with the preceding crash warning evaluation experiments in this effort, the most important consideration is the placement of apparatus for and type of secondary task. The preceding set of experiments showed that the choice of tasks can have significant impact on the evaluation of the DVI. This was in the context of short focused tasks in a forward configuration and in a rightward and slightly rear configuration. When considering the context of the task, a fuller understanding may be gleaned from the data. In general, forward-facing tasks do not provide the opportunity to reliably distract the participant long enough to effectively evaluate crash warning DVIs that would generally alert the driver in rare situations. Additionally, tasks that are easily divided into smaller chunks for completion, such as the CV task, result in reduced effectiveness of the evaluation due to frequent opportunities to glance to the forward roadway.

## 3.3 Exploration of an On-Road Methodology for LDW Assessment

### 3.3.1 Experiment 7: On-Road LDW

Introduction

The objective of the work described in this section was to develop a test track or on-road methodology for use in a protocol to evaluate LDW DVIs. As the research described in Section 3.1 indicated, protocols were developed for both FCW and LDW interfaces based on a driving simulator approach. There was also a parallel on-road procedure for the FCW application. However, there was no available method for providing an LDW test track procedure that parallels the LDW driving simulator method. The driving simulator approach was used because a lane deviation could be unobtrusively initiated in a controlled manner in the simulator without awareness by a distracted driver. The simulator "world" can simply be modified while the driver's gaze is distracted. However, one major problem with a simulator approach is that the OEM system must be "reverse engineered" so that it can be programmed into the simulator. Perfect replication can be very difficult and the protocol outcome may always be questioned on the basis that the simulation was not correct. On-road methods are also preferred over simulator methods by some and would not require specialized driving simulator facilities. Therefore, it would be helpful to have a method by which an actual OEM vehicle could be systematically evaluated for its LDW DVI on the road.

It is possible to devise procedures whereby distracted drivers are apt to make occasional "natural" lane deviations (e.g., Finley, Funkhouser, & Brewer, 2009). Such methods might be useful for experimental research or product development. However, for purposes of a systematic and replicable evaluation protocol, it is essential to be able to induce reliable, controlled lane drifts. While this can be done rapidly and without perceptible motion cues in a driving simulator, it is much more challenging to do on an actual roadway.

The intent here was to explore an on-road (or test track) methodology with the following characteristics:

- A production vehicle could be used to evaluate its native (OEM) system, without modification or damage to the vehicle

- Lane drifts could be introduced unobtrusively, without driver awareness (assuming a distraction task)
- Sufficient lateral drift could be made to occur while the driver was distracted such that LDW alerts would be triggered
- These lane drifts would not provide motion or auditory cues that make the driver aware of the vehicle action
- The lane drifts appear to the driver to be self-induced
- Successful lane drifts adequate to trigger a LDW without driver awareness can be made to occur reliably and in a controlled manner

The section below describes this effort and the outcome. The method was successful in many respects, but could not achieve a level of reliability suitable for a standardized protocol. Ultimately, it was determined that while the lane drift system worked in principle, and can be activated without driver awareness, the results are simply too inconsistent to provide a basis for an LDW evaluation protocol. Further piloting or formal data collection was not undertaken within this project.

Method

The initial concept used hydraulic means to slow the right front wheel relative to the others. This system was implemented in Westat's instrumented vehicle (WesDRIVE). Although functionality was achieved, it presented several issues that ultimately made it infeasible. Namely, it was quite invasive to the vehicle, required modification of a key safety system, and would require considerable plumbing and bleeding to allow it to work in both directions. In the long run, it would not have been an ideal solution for studying a large number of vehicles and their characteristics under LDW conditions. As such, the researchers returned the brake system on the research vehicle back to its original condition and devised a more flexible, less invasive instigation system.

Westat subsequently developed a method for inducing lane drift in unmodified vehicles, using a towed trailer as a means of inducing yaw without driver awareness of the movement. The concept would be to have the driver's eyes distracted from the roadway through use of a secondary task (using procedures parallel to the methods in Section 3.1) and induce a lane drift in a particular direction that forces the vehicle toward the lane edge, triggering the LDW alert. The challenge is to be able to reliably induce a lane drift of sufficient magnitude so that the LDW is triggered while the driver's attention is still diverted (i.e., within about 1.5 s), yet to have the vehicle motion be sufficiently subtle that the driver is unaware of the lateral motion, slowing, vibration, noise, or other cues. Initial piloting with project staff was promising, although lane departures were not consistent and frequently not within desired time boundaries. There then followed an extensive and iterative process of hardware modifications and procedural modifications to improve response.

The resulting instigation system used the concept of individually-controllable electronic trailer brakes, similar to those often used on heavy recreational vehicles and utility trailers. The concept was that a trailer with sufficient lateral braking yaw force would exert a force in that direction on the tongue of the trailer and would effectively cause the towing vehicle to yaw in the opposite direction. This seemed feasible from a design and engineering perspective, but would require the construction of a custom trailer with which to attempt it, since no such trailer existed on the market.

First, it was desirable to have a trailer small enough so that it would not be visible behind the vehicle to the driver in most passenger vehicles. A small (40"x48") trailer was acquired on which a box to hold ballast could be built. Ballast was required because the trailer was quite light and without it, the trailer's motion would be unable to cause a lateral movement in the towing vehicle. The box was built and provided an inconspicuous attachment for normal drives. Drivers knew a trailer was hitched but could not see it when seated in the driver's seat. Figure 40 shows the trailer hitched to the WesDRIVE vehicle.


**Figure 40. LDW instigation trailer behind WesDRIVE vehicle**

The next step in the configuration of the trailer was to add brakes. Although the small trailer size was desirable for its low visual profile, it was incompatible with trailer brakes because the wheels were too small and were not available with brakes (especially not electronically-controlled ones). One could find replacement wheels that were compatible and equipped with electronically-controlled brakes, but they required a larger hub than the one that had come on the trailer. One could buy larger hubs, but they were incompatible with the axle that was on the trailer. And, one could buy a bigger axle, but it was too long to fit the small trailer footprint. These incompatibilities led the research team to shorten a longer axle, have new hubs welded to it, and to install the larger wheels with electronic brakes on it.

The control mechanism was then re-engineered to control each wheel individually and provide the proper level of braking force to allow braking from either side. The brakes on the trailer were re-wired to allow individual control and then a typical electronic brake control box was reverse engineered to allow individual and fine control of the brakes.

The controller for the system was constructed from a typical auxiliary brake control box, like those used on travel and utility trailers. The purpose in those cases is to apply supplemental braking force at the trailer in harmony with the application of the towing vehicle's brakes. The braking force is typically adjusted by a rheostat on the control box mounted near the driver's brake pedal. The idea here is that when the brakes are applied in the towing vehicle, a consistent level of current is supplied to the solenoids on the trailer's drum brakes, causing the trailer to assist with the braking of the trailer (rather than leaving that entire responsibility to the towing vehicle). The adjustment on the controller is intended to adjust the braking force from the trailer brakes to create a smooth, straight braking force. The control is normally adjusted once by the driver, based on the size of the trailer load, to obtain a predictable and controllable automatic braking force each time the towing vehicle brakes. Alternatively, many of these controllers also possess the ability for the driver to manually engage the brakes prior to or without engaging the towing vehicle's brakes. For this application, the team used the electronics from such a controller, added a more sensitive, single-turn potentiometer, a toggle switch to select which wheel would be activated, and calibrated supplemental resistors in line with each brake solenoid to set and maximize the control/sensitivity provided by the potentiometer. The controller that was initially used included a digital display indicating the level of actuation (as a percentage) being provided to the selected brake during each trial. This provided visual feedback to the experimenter controlling a given brake control event about how fast and to what level the actuation was being supplied.

The only additional instrumentation designed and installed in the towing vehicle was a camera and display for use by the experimenter. The experimenter (seated in the rear seat) had a number of responsibilities, among them monitoring surrounding and following traffic in order to confirm when it was safe to initiate a test trial. In order to allow monitoring behind the vehicle without requiring the experimenter to continually turn around, a rear facing camera was installed and the camera image was displayed on a monitor in front of the experimenter. This required the use of a special monitor in conjunction with the camera. This monitor, unlike many other conventional monitors, allowed mirroring of the camera output so that objects seen on the left of the car would appear on the left of the display. Conventional rear-facing cameras would display the left on the right and vice versa, which was perceptually confusing and difficult for the experimenter to use.

Given the number of interacting factors, the development of the specific details of the system required trial and error, with a large number of adjustments in ballast, braking force, and onset timing to create the subtle control desired for this effort. Beyond the obvious factors in the braking performance, other factors also made a difference in the handling yaw instigation effectiveness, including road surface, weather, vehicle speed, electronic solenoid strength, the current to brake application profile, the best controller sensitivity profile for subtle application, brake pad adjustments, and distribution of the ballast (front to back and side to side). In fact, because of the number and interplay of these factors, finding the perfect balance of settings for every situation to create a reliably controllable, subtle, but quick vehicle yaw was quite difficult to achieve and then to maintain. This meant that reliable control was impossible to guarantee from day to day without re-adjustment via various means each time one of the factors shifted, making its use as a repeatable and efficient tool for studying the LDW performance more demanding and less desirable.

A test protocol was developed and revised throughout the piloting effort. Participants first received instructions and practice to familiarize them with the research task. They were told that the purpose of the study was to observe how participants drive when distracted. Participants were informed of the instrumentation on the WesDRIVE vehicle and were told that the trailer was carrying research equipment. Safety instructions were provided to participants, informing them in general terms that they would drive while occasionally performing a distracting task, and that safety must be their first priority. Participants then received practice driving the vehicle, first in the Westat parking lot and then on local roads. Next, participants practiced the three tasks that they would be asked to do while driving. Two – changing lanes and changing air conditioning settings – were included to add diversity to the driving task. The third task – number recall – was the key task during which lane departures would occasionally be induced. The number recall task was used in previous simulator research to induce distraction. In this task, when verbally prompted, the driver must look over his or her right shoulder at a display screen mounted to the side of the passenger's headrest. Five numbers would then be presented in rapid succession. The participant's task was to watch all five numbers, then repeat them back to the researcher. Each number was presented for 0.4 s, for a total display duration of 2 s. Participants were asked to visually commit to this task for the full duration of presentation. Participants practiced this task at least four times while stationary, or more if necessary to achieve good number recall performance. Participants then drove to the pilot testing road segment and performed the study procedures. Following testing, participants were asked if they knew what the purpose of the study was and if they noticed the LDW warnings or any induced braking or lane shifting.

Key Results

The system was first tested on local roads with project staff as drivers and experimenters. Extensive testing and revision was conducted to be able to achieve a braking force that could initiate a lane departure quickly and without providing any obvious cues to the driver. Variables adjusted during this task included weights and distributions of weights in the trailer, resistors of various measures on the electronic brake lines to fine tune braking force, and rheostat force and application speed used by the experimenter. While this round of testing showed that lane departures could be achieved by the system, it also revealed that the lane departure force that could be applied and the speed of lane departures was dependent upon road smoothness, wetness (even slightly damp roads would cause the trailer brakes to lock up), and vehicle speed (different speeds resulted in different levels of maximum available brake force before shudders or vibrations could be detected).

Following initial system calibration, a series of pilot test sessions were conducted with naïve participants at a closed course test track. Participants went through instruction and practice as described above, and engaged in multiple number recall task events. Results showed that lane departures occurred with some, but not all participants. When lane departure events did occur, they tended to occur very close to the end of the number recall task's visual demand. These results suggested that the system was working, but not reliably enough to achieve usable LDW event data from a sufficient number of participants or trials per participant. It became clear in piloting that while the trailer braking system could induce a bias, that bias could be overcome by a participant who held the wheel firmly in position while distracted. Another finding of interest is that even when shudder, brake noise, or vehicle slowing was detectable by the experimenter, none of the participants reported noticing it. The conclusion from this piloting was that further refinements were necessary to achieve an adequate frequency of lane departures.

As a next step, additional refinements were tested. Trailer weight and weight distribution were again adjusted to try to increase maximum available brake force. Alternative roads and driving speeds were evaluated to determine how they might affect the speed of lane departures. As adjustments were made, pilot tests were periodically run with naïve subjects to determine whether adequate lane departure rates could be achieved. During these sessions, the experimenter and an additional safety observer were present in the vehicle and ensured that distraction events were only triggered on straight sections of road when no surrounding traffic was present. As this testing was conducted on actual roads, additional challenges were identified. While on active roads, participants appeared to be less willing to commit to the distraction task. Participants often stole glances at the road or otherwise diverted their attention from the distraction task. The instructions were revised to emphasize the importance of committing to the task, but participants still had trouble committing to the task. Participants also tended to drive relatively slowly during test sessions, and allowed the vehicle to slow further during number recall tasks. Ultimately, the influence of real world driver behavior, workload and distraction factors, and subjective sense of control or safety were not able to be sufficiently overcome by reasonable procedural modifications in order to be sure of achieving an adequate frequency of lane departures.

Discussion

This task was an attempt to implement in actual driving, with a production vehicle, an induced lane drift similar to scenarios that may be used in a driving simulator. If successful, an on-road or test track LDW evaluation protocol might have advantages over simulator-based methods. In the formal sense of developing a low-cost system that could induce lane drift without driver awareness, this task was successful. But in the more important sense of providing a practical LDW evaluation methodology, the effort was not successful. Primary difficulties included:

- The number and interaction of factors influencing system response limited the repeatability of the lane drift induction and imposed requirements for frequent re-adjustment.
- It is difficult to get reliable driver commitment to a secondary task for the period required for subtle lateral motion to achieve warning triggering criteria.
- The behavior of experimentally naïve drivers is such that they provide resistance to the lane drift, counteracting the system's induced lateral movement.
- Ultimately, a "good" trial (triggering an LDW without the driver looking up or sensing the lateral motion) occurs in only a minority of cases. The methodology simply does not reliably induce the necessary conditions.

# 4 Research on Variability among Warning Signals

In the absence of standardized DVIs for ACWS functions, displays may vary substantially from vehicle to vehicle. This could result in driver confusion, unclear messages, and less than optimal DVIs in some cases. Research under a previous CWIM project (Robinson et al., 2011) began to address some of these concerns, including experiments on negative transfer of learning and on the comprehension of status displays. The experiments included in this section continue research to address variability among warning signals. Section 4.1 continues work on negative transfer. Section 4.2 addresses the bounds within which DVI variability might occur while still conveying a consistent message. Section 4.3 concerns how different audible messages can remain effective under different in-vehicle ambient noise conditions.

## 4.1 Negative Transfer with an Auditory FCW

Introduction

This experiment continued the research begun under a previous CWIM contract. The objective of this research project was to determine whether there are benefits to having consistency in FCW characteristics across different vehicle models, or, conversely, if there are adverse effects (i.e., negative transfer) that could limit the overall effectiveness of the technologies in a broadly deployed marketplace scenario if FCW characteristics are not consistent. If there are benefits to consistency, a third objective is to specify the necessary consistent elements.

At the heart of the CWIM effort is the recognition of the need for an objective, repeatable evaluation protocol to assess the adequacy of a particular DVI for a particular ACWS function. FCW and LDW are the two functions directly addressed in this program. However, it is also recognized that even if various DVIs from different vehicles test adequately individually, there may be driver performance decrements associated with the variability of DVIs among vehicles. Previous work (Lerner et al., 2011) noted considerable variability among vehicles in terms of the warning systems present, the means of providing the warning, and the displays used to indicate the presence and operational status of various systems. Lerner et al. (2011) also found evidence for concern over the existence of negative transfer in a simulator study.

The previous negative transfer study (Robinson et al., 2011) had employed two FCW sounds taken from Green et al. (2008). During simulated driving, participants were exposed to one of two FCW sounds (A or B) over two separate testing sessions separated by at least a day. During a third testing session they experienced either the same FCW or received the other sound. Brake response time results showed "negative transfer" (slower response) after distracted drivers responded to auditory FCW A and then encountered an unfamiliar auditory FCW B in a "different" vehicle. This effect, however, was only observed in one of the two directions of change (i.e., A to B versus B to A). One aim of the current negative transfer study was to re-examine this effect and specifically to determine whether the previous finding may have been a feature of the specific warnings used in that experiment or potentially an artifact of the environmental sounds used in the scenario.

The aim of the present research was to determine if the negative transfer effects observed in the previous phase (Robinson et al., 2011) were replicable and to determine the extent to which they generalize across auditory warnings and acoustic environmental variables. Specifically, the research team manipulated the presence and absence of environmental sounds that were similar to the warnings used in the investigation. The researchers also examined the potential for a different but acoustically similar "Distinctive" FCW sound to reduce the presence of negative transfer.

Method

The current negative transfer study is modeled closely after the previous one (which found that switching from warning A1 to A2 resulted in a significant increase in brake response time after participants had been repeatedly exposed to the A1 sound), though with several changes as discussed in more detail later in this section. The impact of warning variability was examined following a period of Warning familiarization and followed by a simulated vehicle switch. Participants were engaged in a visual distraction task throughout. The visual distraction task required participants to watch for the onset of a red LED light and then monitor the light for a period of 1-2 s to determine if the light flashed. This secondary task was used to see if results would be replicable if the distraction task's location was changed from a center stack out into the periphery. 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 first exposed to an alternate auditory FCW in a nominally "different" vehicle.

The experimental design (see Table 10) was a between-subjects manipulation of the Test Phase: Control group (no alert change) versus Experimental group (alert change with background alerts) versus Comparison Experimental group (alert change with no background alerts) versus Distinctive alarm group (using distinct alerts).

**Table 10. Experimental conditions in the negative transfer study**

| **Condition 1** | **Condition 2** | **Condition 3** | **Condition 4** | **Condition 5** | **Condition 6** |
|---|---|---|---|---|---|
| Control: A1 | Control: A2 | Control: No warning | Experimental | Comparison: No background alerts | Distinct |
| A1 → A1<br><br>background alerts | A2 → A2<br><br>background alerts | No Warning<br><br>background alerts | A1 → A2<br><br>background alerts | A1 → A2<br><br>no background alerts | A1 → A3<br><br>background alerts |
| **Participants** | **Participants** | **Participants** | **Participants** | **Participants** | **Participants** |
| 11 | 12 | 10 | 12 | 12 | 14 |

Participants engaged in nine drives (in each drive participants were exposed to one collision event) that were approximately 5-9 minutes long over three sessions (three drives per session). Sessions took place on different days with 1-3 days between sessions. With training and post-experiment questionnaires, participants spent approximately 4 hours total per person in the study. The first 6 drives constituted the Learning Phase and drives 7-9 constituted the Test Phase (see Table 11).

**Table 11. Experimental sessions breakdown**

| | Learning Phase | | Test Phase |
|---|---|---|---|
| Session/Day: | **1** | **2** | **3** |
| Activity: | Training + $D_1$+D2+D3 | $D_4$+$D_5$+$D_6$ | $D_7$+$D_8$+$D_9$ +Post |
| Duration: | 1.5 hours | 1 hour | 1.5 hours |

Participants were given a short break between each drive (D) and each drive included at least one FCW event. FCW events consisted of the lead car suddenly braking and rapidly decelerating. This event could be triggered automatically at a predetermined location in the scenario, or manually by the experimenter. For key Test drives (D1, D6, and D7) the experimenter monitored the glance behavior of the driver through a video feed and manually triggered a FCW event when the driver's eyes were averted from the forward field of view. Each Drive was a different distance/duration and included different surface visual scenery in an attempt to reduce the buildup of expectation and anticipation of events in each drive.

On Day 3 the Vehicle "change" occurred. Participants were told that they were driving a simulated vehicle that would have different handling characteristics and a different DVI than the simulated vehicle they had been driving on the previous days. Superficial visual stimuli (i.e., different section of the roadway) were altered to increase the realism of this switch.

The FCW event in Drive 7 contained the main experimental manipulation, which consisted of a different auditory FCW for each of the three Switch conditions: 1) Experimental switch from A1 to A2 – the condition that had resulted in negative transfer in the previous negative transfer study; 2) Comparison – switch from A1 to A2 in the no background alert condition; and 3) Distinctive – a switch from A1 to A3 with A3 being a FCW with similar temporal characteristics as A1, but with altered spectral components.

Drives 8 and 9 also had FCW events, both with the changed alert in the Experiment condition. For two of the Control conditions, no change in auditory alert occurred despite all other vehicle "changes." One control condition experienced no warning during any of the drives. For the Comparison condition, no extra environmental noise was presented throughout drive, but as in the Experimental condition a switch in FCW sounds from A1 to A2 was presented on each drive of the third day while everything else was same as the Experimental condition. Complete, valid data sets were acquired from 71 participants.

Training focused on getting participants accustomed to the motion-based driving simulator and the various distraction tasks they were asked to complete. The training drive lasted approximately 10 minutes. The simulator features 180° of yaw motion and 1° of pitch motion; additionally it has three 42" screens allowing for 180° field of view. Participants heard all auditory signals for their condition via a laptop computer prior to participation to familiarize them with each sound. Greater detail regarding the methods of this study can be found in a separate appendix (Appendix C).

Key Results

Response time was measured from the time the collision warning was issued to the time when either a steering or brake response was initiated. A steering response was defined as a derivative value (instantaneous rate of change in steering inputs) of steering input greater than 1.5 standard deviations of the participants' mean steering input during the collision event. A brake response was defined as a brake force input value greater than 20 Newtons during the collision event. The fastest response between the two was taken as their evasive maneuver response time (EMRT). Collisions were determined whenever a participant's vehicle occupied the same space as the lead vehicle. If a participant did not collide with the lead vehicle, their minimum TTC was calculated to determine how close they came to colliding. If participants did not make a response they were given an EMRT of 1.8 s, which was the initial TTC when the warning was issued. If participants were missing data from drives 1, 6, and 7, or had an EMRT faster than 0.2 s for drives 1, 6, and 7, they were not included in data analysis. A few participants' data were lost due to experimenter error (e.g., they received the wrong warning sound during one of the critical drives). Since to the maximum extent possible, experimenters were blind to the condition that each participant was in and were following a code sheet, this procedure (which was implemented in an attempt to reduce experimenter bias) did leave greater room for experimenter error. Some participants' data were lost because they were already braking before the event began or because they made some type of anticipatory maneuver before the onset of the warning (e.g., initiated braking and/or steering less than 0.2 s after the onset of the warning). Only participants who had valid data for drives 1, 6, and 7 were analyzed, resulting in a total of 71 participants included in the main analyses.

The EMRT data were submitted to a repeated measures ANOVA with condition as a between-subjects factor and exposure as a within-subjects factor. A main effect of exposure was found, $F(2,130)=40.83, p<.01$. Post hoc comparisons show that this effect was driven by the differences between the first and sixth exposure, demonstrating that participants exhibited a learning effect for the warning, the potential collision scenario, or both. This relationship is illustrated graphically in Figure 41. However, there was no significant interaction between exposure and condition $F(10,130)=.95, p>.05$, when analyzed across all conditions. When specifically looking at conditions 1, 4, 5, and 6, there still is no significant interaction between exposure and condition; however, it was trending in the expected direction, $F(6,90)=1.53, p=.18$. This trend appears to be primarily driven by an increase in EMRT in condition 5 (the switch with no background alerts experimental condition).

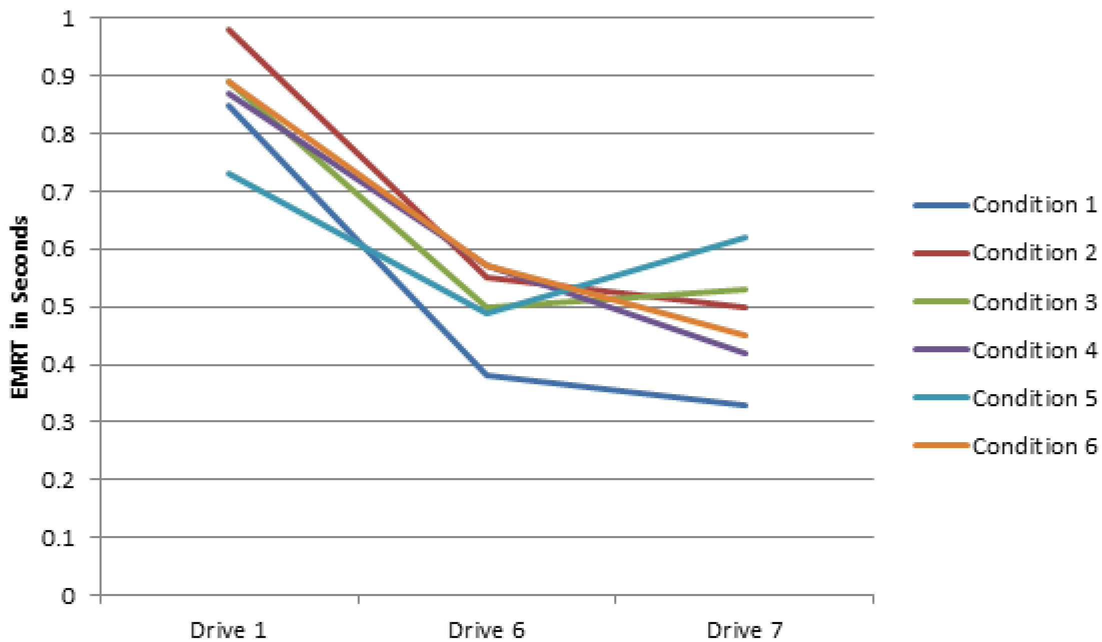


**Figure 41. EMRT as a function of exposure and condition**

Analysis of minimum TTC was examined as another means of determining the potential impact of a consistent versus novel FCW sound. Examination of minimum TTC reveals how close to colliding a participant came regardless of when they responded. A mixed design ANOVA with exposure as a repeated-subjects variable and warning condition as a between-subjects variable revealed a main effect for exposure, $F(2,130)=42.48$, $p<.001$, indicating that regardless of warning condition, participants who were able to avoid colliding came the closest to colliding during their first exposure relative to the 6th and 7th, which did not differ significantly from each other. The interaction between exposure and condition was not significant, $F(10,130)=.91$, $p>.05$. However, there was a trend pattern similar to that observed in the dependent measure EMRT in the post-switch exposure 7.

Collision data were examined using the likelihood ratio statistic. On the participants' first and seventh exposure to the event there were no differences in associations between conditions for whether they crashed or not, $L\,X^2(5)=1.97$, $p>.05$ and $L\,X^2(5)=9.17$, $p=.1$, respectively. Further examination of this non-significant trend for the seventh exposure indicated that participants in both Conditions 4 and 5 had significantly more crashes post FCW switch relative to participants in Condition 1 who experienced a consistent FCW throughout, $L\,X^2(1)=4.32$, $p<.05$ and $L\,X^2(1)=5.98$, $p<.05$. More detailed analyses of these and additional dependent metrics can be found in a separate appendix (Appendix C).

Discussion

This experiment was conducted to investigate the potential for negative transfer of auditory warnings. It did not investigate the potential for negative transfer in other modalities (e.g., haptic, visual). A primary purpose of the current investigation was to further examine a finding from the previous CWIM negative transfer study (Robinson et al., 2011) that had revealed negative transfer (slower response) among distracted drivers when they became familiar with one FCW "A1" and then encountered an unfamiliar auditory FCW "A2" in a "different" simulated vehicle. Specifically, the research team sought to determine if the previous results might have been a feature of the specific warnings used in that experiment or potentially an artifact of the environmental sounds used in the scenario. A further objective was to examine whether or not a FCW sound change to another sound that retained certain distinct characteristics with the initially learned sound would be resistant to negative transfer.

Results of the current investigation are mixed. Some evidence for negative transfer was revealed. However, this evidence was primarily observed in only one of the two main switch conditions. The research team had expected that participants might experience negative transfer in both Condition 4 and 5 when participants encountered a new FCW sound (A2) for the first time on their 7th exposure drive (on Day 3). However, only Condition 5 (no background alerts) demonstrated slowed response time on the 7th drive. There are several possible reasons why negative transfer might not have been observed in the Switch Condition 4. One is the possibility that there was no effect because changing the auditory sound does not result in negative transfer. This alternative is highly questionable, however, since evidence to the contrary was observed in the previous CWIM negative transfer study (Robinson et al., 2011), as well as in the current study in Condition 5. An alternative explanation is that the precise nature of the warning sound is less important in very environmentally noisy situations. It is possible that in all conditions *except* Condition 5, participants grew accustomed to reorienting their vision forward any time they heard an audible alert due to the presence of the nonverbal speed notifications.

Another potential reason for not observing more robust negative transfer effects may be limitations of the current methodological protocol. Participants in the control condition (Condition 3) who received no warning were only slightly more likely to crash than participants who received warnings. In Drive 6, which occurred before the potential switch in warning sound, 4 of 10 or 40% of participants in the control condition crashed. With the exception of Condition 1 whose crash rate in Drive 6 was only 9% relative to a crash rate of 81.8% in Drive 1, the crash rates of the other experimental conditions that received a warning were only slightly lower. All conditions had high crash rates in the first drive but crash rates dropped sharply in all conditions by Drive 6. Specifically, for Drive 6, Conditions 2 and 4 had crash rates of 33.3% (or 4 of 12), Condition 5 had a crash rate of 25%, and Condition 6 had a crash rate of 35.7%. Evasive response times were also similar between the control condition and the other conditions receiving a warning. It appears that with the exception of Condition 1 where a specific (A1) consistent warning was presented across all drives, even participants receiving a warning had relatively high crash rates.

The observation that participants in the control condition exhibited relatively similar collision avoidance response times and behaviors as those participants receiving the FCW is a limitation of the current study. This observation indicates that participants may not have had ample time to respond to the warnings or perhaps they found the secondary peripheral discrimination task too challenging. The sudden brake events in Drives 1, 6, and 7 occurred when the participant's eyes were diverted from the forward roadway. The warnings were presented when the TTC between the participant's car and the lead vehicle decreased to 1.8 s. This lead time is well within the existing guidelines (Campbell, Richard, James L. Brown, & McCallum, 2007) and is ecologically valid. Including a longer lead time under actual driving conditions would not be realistic because most people drive with less than the 2 s lead time required by Virginia State law. However, a survey of existing crash warning literature indicates that many researchers in fact use significantly longer lead times for triggering warnings. For example, Kramer, Cassavaugh, Horrey, Becic, & Mayhuh (2007) used a lead time of 2.12 s and Gray (2011) used a lead time of 3 s. Providing the warning sooner would give participants more time to avoid the collision situation, though the tradeoff is that it might not be as ecologically valid. It is also worth noting that the scenarios in this experiment may have been particularly difficult to avoid relative to some other simulator studies due to the deceleration rate of the lead vehicle. While not all studies report this information, some prior studies have reported using a lead vehicle deceleration rate of 6 $m/s^2$ (Ho, Reed, & Spence, 2007; Gray, 2011) and the rate used in this experiment was much more rapid (13.41 $m/s^2$). This likely contributed to the high crash probabilities observed.

One finding the current investigation does make clear is that use of a consistent warning sound similar to the A1 warning resulted in the best collision avoidance response by all measures examined (e.g., EMRT, TTC, crashes). The current investigation also provides support for the idea that minor changes in alert characteristics (from A1 to A3) are unlikely to have dramatic negative performance consequences. No evidence was observed for a negative impact on a FCW sound switch in Condition 6 where the sound remained similar. However, since there was also little evidence of a negative effect in Condition 4 this observation must be interpreted with caution.

In conclusion, some additional evidence was obtained in the current investigation that there is reason to be concerned about the potential for negative transfer in auditory warnings. However, the pattern of conditions that might yield it is complex and not entirely clear at present. Further work should examine the range of key sound parameters that can be changed without disrupting people's general perception of a sound as a highly urgent signal. In other words, it may be most important that people recognize a sound as an urgent warning, regardless of exactly what the sound is. Better understanding of the range within these parameters that can be used to convey high urgency will ensure that drivers can respond appropriately to sounds intended to convey warnings rather than confuse them for less important types of signals. That goal was the aim of the next set of experiments designed to gain better understanding of the perceptual space that defines highly urgent auditory warnings.

## 4.2 Categorical Perception of Warnings and Alerts

### 4.2.1 Perceptual Space Sort Task

Introduction

The primary aim of this series of investigations was to establish criteria regarding the key characteristics of a warning sound that enable the sound to be effectively perceived as representing a highly urgent collision warning. Toward this aim, three methodologies were implemented to specifically examine the range of key parameters that result in a majority (85%) of listeners perceiving a given sound as an urgent collision warning. Methods for each of the experiments are discussed in this section.

First, a survey of existing FCW sounds currently in use in automobiles was conducted. An inventory of the sounds was compiled from this survey. A focus was placed on obtaining representative OEM FCW sounds (i.e., those with the greatest vehicle fleet penetration). Once obtained, the sounds were analyzed in terms of their acoustic properties (e.g., spectral frequency components, pulse duration, onset and offset, harmonics, inter-pulse interval, etc.). From this inventory, as well as existing published guidelines (e.g., Campbell, Richard, James L. Brown, & McCallum, 2007), a list of key parameters was developed and examined in the following three experiments. The goal of this series was to define the perceptual space associated with a highly urgent FCW sound. Specifically, the research team sought to define the key parameters that would be associated with at least 85% of listeners perceiving the sounds as a highly urgent, time-critical collision warning. Additional details about this experiment are provided in a separate appendix (Appendix D).

Method

Participants were 21 subjects recruited through the George Mason University (GMU) research pool or by word of mouth. Stimuli initially included in the Perceptual Sort Task were 29 varying DVI-related sounds. Sounds were created in Adobe Audition CS5.5 and Audacity and were based closely on real, in-vehicle systems. Stimuli included sounds from many categories of in-vehicle system including those similar to FCW, LDW, backup warning, fatigue alerts, door open, low fuel, seatbelt, and park assist sounds. Stimuli were equated for loudness using Adobe Audition and were presented at 84 decibels (dB).

Stimuli were presented using Microsoft PowerPoint as shown in Figure 42**.** Sounds were embedded in a large slide such that they could be played by clicking on the sound and a pressing a small play triangle. After playing sounds, they could be moved using the cursor into one of four categories. Categories included "Warnings," "Alerts," "Status Notifications," and "Social Notifications." Each category also held a small section labeled "Prototype." Under each category were two text boxes. The first text box held the text "Click here to tell us why…" and was intended to allow the subject to give justification for how or why they grouped sounds together. The second box held the text "Best Urgency Level:" and was intended to allow participants to give a numerical rating of the desired urgency level for each category between 1 and 100.

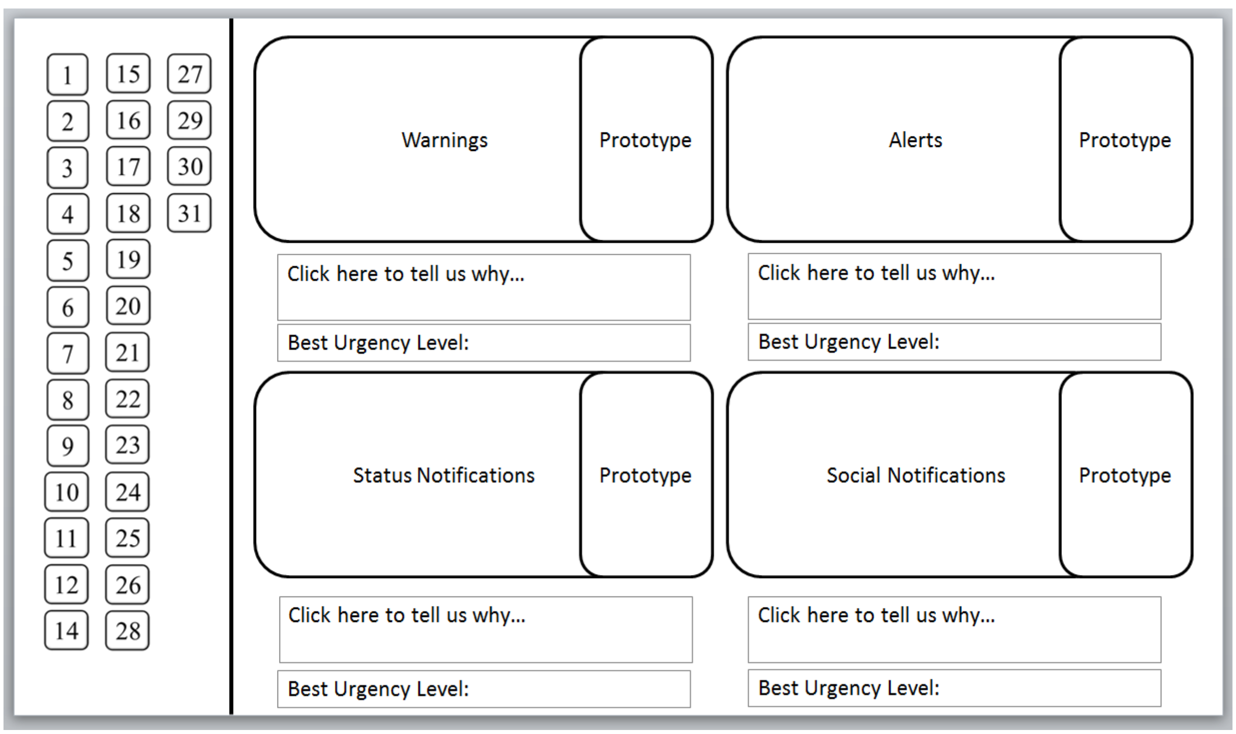


**Figure 42. Sort task design and setup**

Before the experiment began, subjects gave written informed consent and completed a short demographic survey. Subjects were then given a short practice Sort Task using sounds not used in the experimental task. Subjects were instructed to play each sound and then categorize it into one of the four available categories. “Warning” sounds were suggested to “include sounds that you believe to be time-critical, collision warning sounds.” “Alert” sounds were suggested to “include non-critical alerting sounds, for example, a sound that might indicate a lane departure or a parking assist sound.” “Status Notifications” were suggested to “include sounds that indicate something about the status of your car, for example, low windshield wiper fluid or low tire pressure.” “Social Notifications” were suggested to “include sounds used by a car’s social media system to indicate a social media (like Facebook or email) notification.” Subjects were instructed that they could have as many sounds as they desired in each category but must have at least one sound in every category. Subjects were further instructed that after categorizing each sound, they should play every sound again and choose one sound which was the most prototypical of each category and move it into the “Prototype” section. Subjects were finally instructed that they should then attempt to explain their grouping choices and choose an urgency level for each category between 1 and 100.

After the practice task, subjects were given the experimental task which was identical to the practice but with more sounds. The experiment took under half an hour.

Key Results

Results were first analyzed only in terms of the percentage of times they were placed in a given category. It was immediately obvious from the result that, while subjects distinguished between warnings and alerts as opposed to status or social notifications, there was little agreement on what constituted a warning or an alert. This was further made obvious based on explanations for categorization where it was possible for two subjects to categorize based on the same criteria but to choose opposite categories for which those criteria apply. Therefore, all “warning” and “alert” categorizations were reclassified into a new category called “Alarms” after data collection.

Using a backward stepwise logistic regression to predict membership in the category “warning”, and disregarding redundant properties, the four most important (i.e., explaining the most variance) properties that related most to alarm-like warnings were determined. These properties were interburst interval (IBI), base frequency (spectral), number of harmonics (which contributed to the harshness of the tone) and peak to total time ratio. These properties explained 58% of the variance in Alarm classification ($R = .800$, adj. $R^2 = .581$). Peak to total time ratio defines a property to encompass the perceptions created by longer or shorter onset or offset values. See Figure 43 for a visual representation of peak to total time ratio. Figure 43 also shows the interpulse interval (IPI), which is the gap between pulses during a burst of sound. IBI is the gap between multiple bursts of sound. Both IPI and IBI contribute to the perceived tempo of a sound.

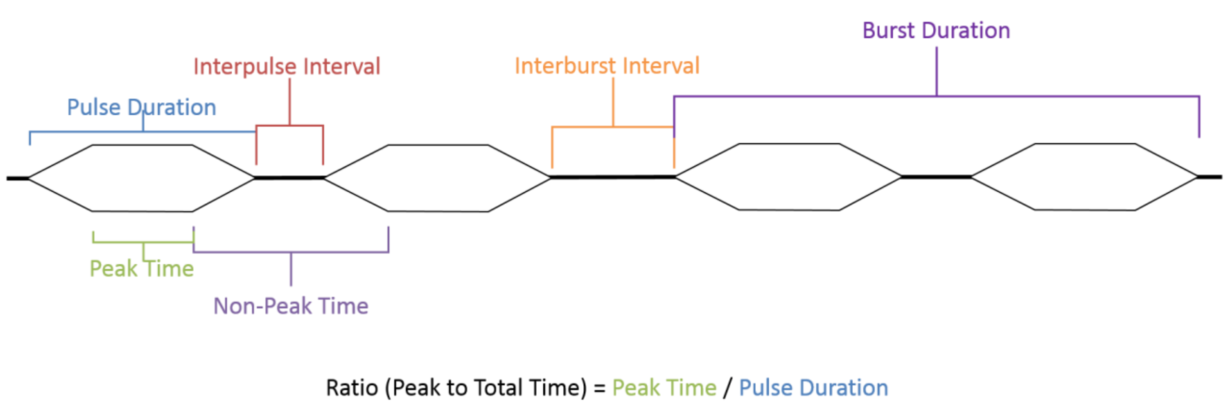

**Figure 43. Visual representation of peak to total time ratio**

Using the results of the backwards stepwise regression, criteria were defined in order of decreasing importance for each property of interest and cutoffs were determined. Cutoffs allowed the researchers to organize the data in terms of whether or not it met each criterion sequentially, and eliminating any sound that did not pass at each level. Cutoffs in this case were based partially on results from the subsequent Method of Adjustment Experiments (described in Sections 4.2.2 through 4.2.4 below) and partially arbitrarily, based on the research team's experiences from previous studies and examples from the literature. Cutoffs are reported in Table 12.

**Table 12. Criteria and cutoffs for alarm categorization in the perceptual sort task**

| Criteria (in order of decreasing importance) | Cutoff |
|---|---|
| Peak to Total Time Ratio | ≥ .7 |
| Interburst Interval (IBI) | ≤ 125 ms |
| Number of Harmonics | ≥ 3 |
| Base Frequency | ≥ 1000 Hertz (Hz) |

As shown in Figure 44, these criteria do a good job of predicting classification for alarms and show an inverse relationship with non-alarms. Sounds that met all four of the above criteria were classified as alarms by 90% of the listeners. However, it should be noted that, due to the relatively small number of sounds classified, many sounds are present at the left end of the graph (many sounds meet few or no criteria) and only one sound meets all but the base frequency criteria and only one sound meets all criteria.

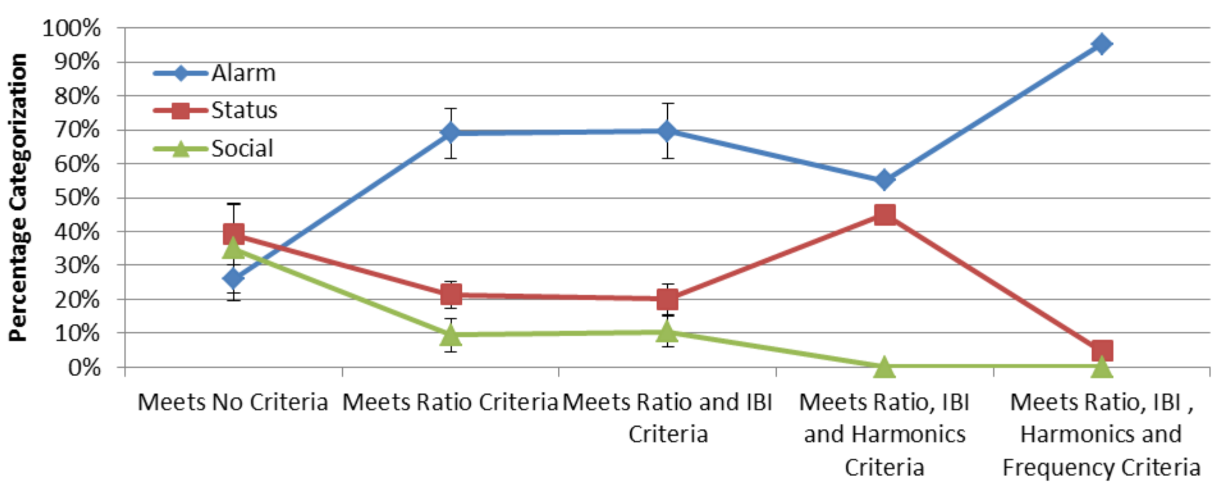


**Figure 44. Percentage categorization by criteria met**

Discussion

The results of the Perceptual Space Sort Task provided preliminary cutoffs for key parameters that contribute to sounds being classified as urgent time-critical collision warnings. However, since the criteria were established by and verified on the same data set, further validation of the criteria was warranted. Additionally, since only one sound in the set actually met all criteria and only one other sound met three of the four criteria, their interpretation requires caution. Furthermore, not all sounds had the same presentation length.

Concurrent with the conduct of this first Sort experiment, the researchers collected data aimed at further examining the parameter levels associated with a sound being perceived as an urgent collision warning using a psychophysical technique known as the method of adjustment. The method of adjustment experiments are described in the following sections.

### 4.2.2 Method of Adjustment - Single Parameter

Introduction

This experiment was run in parallel with the perceptual space Sort Task to further examine perception of the individual acoustic parameters associated with an urgent collision warning. The primary purpose of this experiment was to see at what point participants perceived an auditory tone to either be a highly urgent time-critical collision warning sound or to no longer be a highly urgent time-critical collision warning sound. A psychophysical method of adjustment procedure was implemented using ascending and descending thresholds for the key parameters. Participants would begin with a high or low parameter value and then adjust it up or down, respectively, until the parameter fell within what they perceived to be a level consistent with a critical warning. For this experiment, frequency, tempo (IBI), pulse duration, and pulses per burst were examined one at a time. Averages of the crossover points using an ascending and descending series of parameter adjustments could then be used to further define the perceptual space. Further, these results could be compared to the parameters obtained from the previously described Sort Task. Obtaining similar parameter values for the cutoffs would validate the Sort Task results.

Method

Participants were 20 students recruited through the GMU research pool or via word of mouth. The base sound used for this experiment was a single 300 Hz tone lasting 200 ms with a 20 ms onset and offset. The tone was repeated eight times with 118 ms between each burst, lasting for a total time of 2426 ms. The sound was played through Sennheiser headphones via a Matlab program. The Matlab program allowed participants to adjust the sound by frequency, pulse duration, tempo (IBI), or pulses per burst. Sounds were presented at 84 dB. All parameters included a minimum and maximum possible value with a set increment by which the parameter could be adjusted, shown in Table 13.

**Table 13. Parameter adjustment values**

| Parameter | Minimum Possible Value | Maximum Possible Value | Adjustment Increment |
|---|---|---|---|
| Frequency | 200 Hz | 2500 Hz | 100 Hz |
| Pulse Duration | 160 ms | 800 ms | 40 ms |
| Tempo | 10 ms | 800 ms | 40 ms |
| Pulses Per Burst | 1 | 5 | 1 |

Upon arrival participants signed a consent form then filled out a basic demographic survey. After completing the survey, participants were informed that they would be presented with a single sound (the base sound previously described) and would be asked to adjust the sound on frequency, pulse duration, tempo, and pulses per burst, separately. For this experiment participants only adjusted one parameter at a time. Each screen indicated which parameter would be adjusted and whether the parameter should be adjusted to sound like a highly urgent, time-critical, collision warning or no longer like a highly urgent, time-critical, collision warning. Participants adjusted each parameter six times. In a randomized order participants adjusted each parameter three times in ascending order to make it sound like a highly urgent, time-critical collision warning sound and three times in descending order to make it no longer sound like a highly urgent, time-critical collision warning sound. Participants received real time feedback, such that each time they made an adjustment they would hear the new sound based upon their adjustment. Participants could continue to make adjustments until they were satisfied that the sound they had created either sounded like or no longer sounded like a highly urgent, time-critical, collision warning.

Key Results

Table 14 shows the mean value for each parameter as well as at what value at least 85% of participants found the sound to be a highly urgent, time-critical collision warning.

**Table 14. Mean values for parameters and value in which at least 85% of all participants found the sound to be a highly urgent, time-critical warning**

| Parameter | Mean Value | 85% Value |
|---|---|---|
| Frequency | 931.71 Hz | 1200 Hz |
| Tempo (IBI) | 330 ms | 240 ms |
| Pulse Duration | 460 ms | 360 ms |
| Pulse Per Burst | 2.73 | 4 |

Discussion

Averages of each participant's ascending and descending parameter values were obtained. These averages were then averaged across participants to determine the overall average values for each parameter and the values associated with 85% of participants perceiving the parameter to be associated with a highly urgent warning. Results provided a base minimum value for when a sound switches from either being or not being a highly critical time urgent collision warning sound. Based upon this experiment, a highly critical time urgent collision warning sound should be at least roughly

900 Hz, 460 ms in length, have a tempo of 330 ms and have at least 3 pulses per burst. However, as can be seen by the 85% value, for the majority of people to perceive the sound as a highly critical time urgent collision warning the sound's frequency and pulses per burst would need to be higher and tempo and pulse duration would need to be decreased. With the exception of tempo, the parameter values obtained using this procedure reflected the cutoff criteria obtained in the Sort Task quite closely. That is, initial validation of the Sort Task parameters cutoffs was obtained for the base frequency and number of pulses per burst.

Next, the research team implemented the same method of adjustment procedure but with the addition of the ability to manipulate multiple parameters of a sound at the same time.

### 4.2.3 Method of Adjustment - Multiple Parameters

Introduction

This experiment was conducted to explore the possibility that some acoustic parameters may play a stronger role than others in shaping the perceptual space. This experiment allowed individuals to adjust all four parameters simultaneously, allowing them to change the parameter they felt was more important and thus dictating their perception. Furthermore, this experiment makes it possible for the participant to always make a sound seem like or unlike a highly critical time urgent collision warning sound, where in the single parameter adjustment it was possible that someone may not have been able to effectively create a highly urgent sound by changing levels of only one parameter. For example, it is possible that participants could not adjust the sound enough to perceive it as being either a highly critical time urgent collision warning sound or no longer being a highly critical time urgent collision warning sound.

Method

Participants were 20 students recruited through the GMU research pool or via word of mouth. For this experiment the same base sound from the Single Adjustment Task was used in addition to two DVI-related FCW sounds. Sounds were played through Sennheiser headphones via a Matlab program. The Matlab program allowed participants to adjust the sound on frequency, pulse duration, tempo (IBI), and pulses per burst simultaneously (by the same values as presented for the Single Adjustment Experiment). Sounds were presented at approximately 78 dB.

The procedure was exactly the same as the Single Adjustment Task with the exception that participants in this experiment were asked to adjust each of the four parameters for a single sound. Participants were allowed to adjust all four parameters at the same time, allowing them to freely choose which parameter to adjust first and allowed them to continually switch between parameters until they were satisfied with the resulting sound. As in the previous adjustment study, participants were instructed that their task was to make the sound presented seem either like a highly urgent, time-critical collision warning (in the ascending trials) or to no longer sound like a highly urgent, time-critical collision warning (in the descending trials).

Key Results

Table 15 shows the average value for each parameter as well the value at which at least 85% of participants found the sound to be a highly urgent, time-critical collision warning.

**Table 15. Average values for parameters and value in which at least 85% of all participants found the sound to be a highly urgent, time-critical collision warning**

| Parameter | Average Value | 85% Value |
|---|---|---|
| Frequency | 1094.17 Hz | 1300 Hz |
| Tempo (IBI) | 500 ms | 360 ms |
| Pulse Duration | 510 ms | 440 ms |
| Pulse Per Burst | 3.03 | 4 |

Discussion

Similar to the previous adjustment study and in line with the results obtained from the Sort Task, the current study indicated that the base frequency of the sound must be over 1000 Hz in order to be perceived as an urgent, time-critical warning. Further, the results of this study indicate that when participants were allowed to adjust the sound on multiple parameters, frequency and pulses per burst were the ones dictating their perceptions. One caveat is that similar to the previous adjustment study, while there was no true IPI, there still was a perceived IPI, due to the 20 ms offset and 20 ms onset ramp. Therefore, though not specifically manipulated it is likely that the IPI of 40 ms derived from the onset and offset times was driving the perceived tempo of the sound. Again, the 85% values show that for there to be a high level of agreement the cutoffs need to be slightly higher.

### 4.2.4 Method of Adjustment - Prototype Development

Introduction

The primary purpose of this experiment was to gain information regarding what values of each of the key parameters would result in individual perceptions of that sound being their "ideal or prototypical" collision warning sound.

Method

Participants were 25 students recruited through the GMU research pool or via word of mouth. For this experiment the same base sound from the Single Adjustment Task was used. The sound was played through Sennheiser headphones via a Matlab program. The Matlab program allowed participants to adjust the sound on frequency, pulse duration, tempo (IBI), and pulses per burst simultaneously. Sounds were presented at approximately 78 dB.

The procedure was the same as the Multiple Adjustment Task with the exception that participants were now asked to adjust the sound until it matched their prototype of the ideal highly urgent, time-critical collision warning sound.

Key Results

Table 16 shows the average value for each parameter as well as the value at which at least 85% of participants found the sound to be a highly urgent, time-critical collision warning.

**Table 16. Average values for parameters and value in which at least 85% of all participants found the sound to be a highly urgent, time-critical collision warning**

| Parameter | Average Value | 85% Value |
|---|---|---|
| Frequency | 1576 Hz | 1900 Hz |
| Tempo (IBI) | 254 ms | 40 ms |
| Pulse Duration | 398 ms | 200 ms |
| Pulse Per Burst | 4.16 | 5 |

Discussion

As can be seen from the table, the average values of the prototype sound indicate a slightly higher base frequency and a faster tempo, relative to the previous method of adjustment experiments. The tempo of the prototype is faster and the prototype sound has more pulses per burst.

The next experiment involved replicating the Sort Task with new participants and with additional sounds created from the results of the Method of Adjustment experiments. If the criteria established in the first Sort Task are valid, then they should be replicable with a new sample of listeners. Further, the researchers reasoned that if the Method of Adjustment procedure results in reliable parameter values, then a sound created with the obtained values should have a high probability of being classified as a warning. A validation Sort Task was conducted next to examine these issues.

### 4.2.5 Validation Sort Task

Introduction

The primary purpose of the Validation Sort Task was to validate the results of the previous Sort Task and the Method of Adjustment tasks. Specifically, for the Validation Sort Task, the researchers chose to investigate parameters that were deemed to be important by the previous tasks. Recall that only one sound met all four of the criteria established in the first Sort Task. Therefore, for this Validation Sort Task the researchers directly manipulated key acoustic aspects of several sounds to create new sounds designed to examine the robustness of criteria cutoff values. Specifically, the number of harmonics present in several sounds was manipulated by adding harmonic components to sounds that previously had only one harmonic. The presence of multiple harmonics was found in the previous tasks to influence the "harshness" of the sound, and indirectly, its perceived urgency or importance. Further, it was found in the first Sort Task that some sounds were inadvertently disproportionately long. Therefore in this validation task, sounds were recreated to be of more similar lengths, around 1500 ms.

Method

Participants were 15 subjects recruited from the GMU research pool or through word of mouth. Stimuli for the Validation Sort Task were similar to the Perceptual Sort Task but included additional stimuli that were identical to current stimuli but included added harmonics such that 52 total stimuli were presented. All apparatus and presentation methods were identical to the Perceptual Sort Task.

The procedure for the Validation Sort Task was identical to the Perceptual Sort Task except that the procedure did not include a practice as it was concluded from the Perceptual Sort Task that the practice was unnecessary.

Key Results

Results were analyzed identically to the Perceptual Sort Task, using the same criteria and cutoffs. A backwards stepwise regression was able to account for 61% of the variance in Alarm classification (R=.802, adjusted $R^2$=.612). Figure 45 shows that sounds that meet all criteria (or even criteria after the first cutoff) are much more likely to be considered "alarm" sounds than status or social sounds.

Discussion

As has been shown in the results, the manipulations of specific parameters and previously defined criteria and cutoffs not only hold in these new data, but fit slightly better with more sounds. This indicates that the criteria and cutoffs do a relatively good job of predicting classification, especially when two or more criteria are met. However, these sounds were still classified in a relatively low fidelity, low context experiment with no distraction. It was therefore important to retest these criteria when participants were driving and when they needed to make more realistic decisions under divided attention.

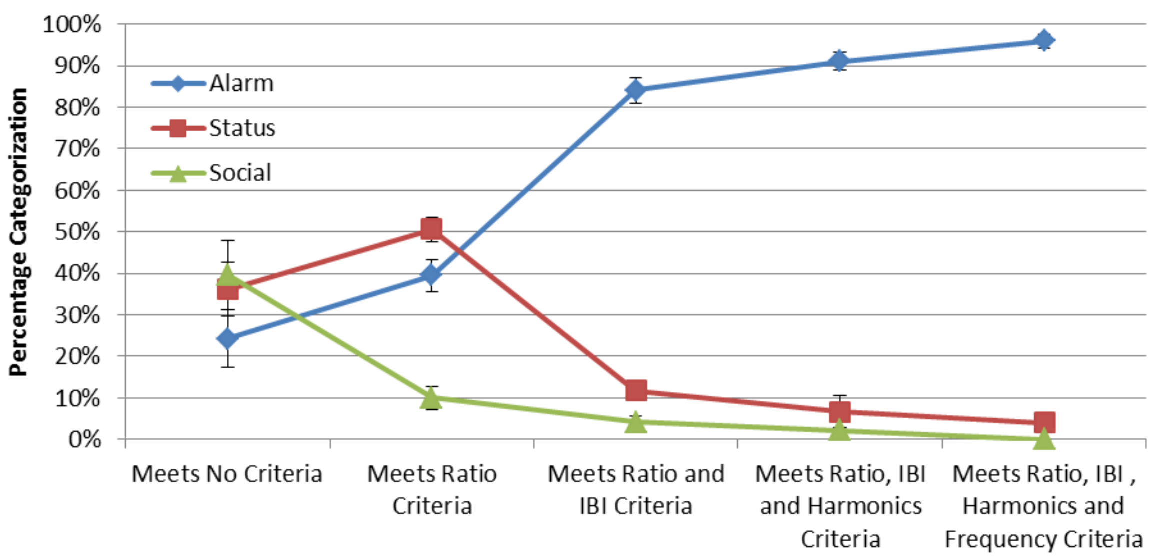


**Figure 45. Percentage categorization by criteria met**

### 4.2.6 Rapid Categorization under Divided Attention

Introduction

The Rapid Categorization under Divided Attention task was designed to extend and provide additional validation of the results from the earlier studies in this series. It has previously been shown that perceptual cognitive load can degrade participants' responses to warnings while driving

(Lewis, Penaranda, Roberts, & Baldwin, 2013; Santangelo & Spence, 2007). Therefore, this experiment was designed to examine whether or not people's rapid categorization of sounds into critical and non-critical categories differed when they were experiencing the multiple demands of simulated driving and being engaged in a distracting secondary task relative to when they could devote their full attention to the categorization task. In this experiment participants first completed a Sort Task similar to the two reported earlier in this series. Next they were asked to categorize sounds by providing a behavioral response (either a brake or button press) while they were concurrently engaged in both a simulated driving task and an n-back number recall task.

Method

Participants were 22 students recruited through the GMU research pool or via word of mouth. For this Sort Task, 26 sounds were used, with specific emphasis placed on those sounds that were shown to be consistently categorized the same across the Sort Tasks.

A RealTime@ technologies desktop driving simulator was used, with a Logitech Driving Force GT steering wheel for simulated driving. Ambient sound was presented via two monaural computer speakers at 60 dB. While participants were driving they also completed a visual n-back task that was shown on a touch screen interface, using the "x" and "o" buttons to respond positively or negatively to the n-back task. The experimental apparatus are shown in Figure 46. At periodic intervals the n-back task would pause and be immediately followed by the presentation of an experimental sound. Participants were instructed to categorize the sound as an alarm, a status notification, or a social notification using the brake pedal to indicate an alarm sound, the left button on the steering wheel to indicate a status notification, and the plus button on the steering wheel to indicate a social notification. For the experimental task the same sounds were used as in the Sort Task. All sounds were repeated twice except for some of the social sounds that were repeated four times. Sounds were presented via a pair of Bose computer speakers at 75 dB. This set of sounds included several sounds that, based on criteria established earlier in this line of research, were expected to consistently be classified as alarms. Additionally, several sounds shared some attributes of an alarm but not all. The research team expected these sounds to be more difficult to classify and therefore result in not being classified as an alarm and/or requiring more time to make a classification response.

**Figure 46. Experimental task setup**

When participants arrived they signed an informed consent document then completed a brief demographic survey. Upon completing the survey participants began the Sort Task. The Sort Task was similar to the two previous Sort Task experiments with some small changes. In this case the separate categories of "Warning" and "Alert" were merged into one single "Alarm" category, as previous experiments indicated that there is little consistent differentiation between "Warnings" and "Alerts" by sound parameters. Additionally, the "Alarm" category allowed subjects to rank order sounds from "most warning-like" to "most alert-like." The prototype section of the experiment was also removed; otherwise, all aspects of the Sort Task were identical to the previous sort task experiments. The sort task design and setup are shown in Figure 47.

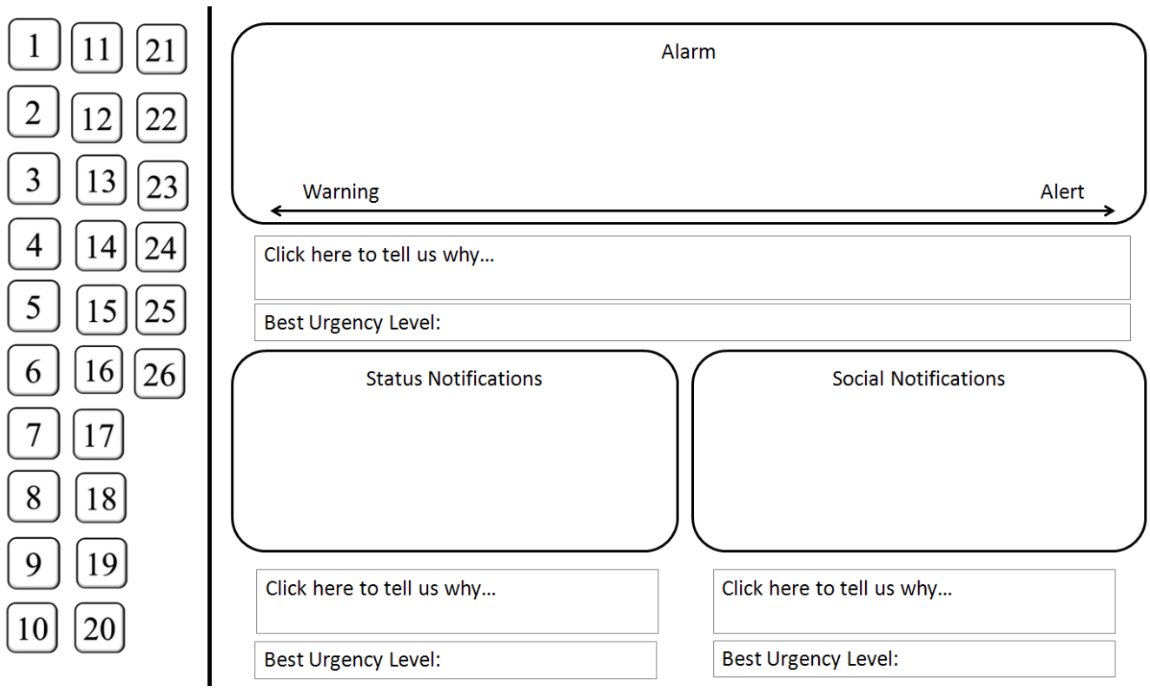


**Figure 47. Sort task design and setup**

Upon completion of the Sort Task, participants were seated in the driving simulator. Participants practiced driving while maintaining a speed of 45 mph and holding a steady lane position. Once participants were comfortable with driving using the simulator they practiced the n-back task alone and then with categorization. Participants were instructed that they would be doing a visual 1-back task, such that numbers between 1 and 10 would be presented via a small screen to the right of the steering wheel and that they would be asked to respond via the steering wheel positively or negatively depending on whether or not each number matched the number presented directly before it. Participants were instructed that after a variable amount of time on the 1-back task, a sound would play and that they would be required to indicate via the brake, left button, or plus button whether they considered the sound to be an alarm, a status notification, or a social notification. During the practice task, the sounds included were spoken "Alarm", "Status" or "Social," in order to familiarize participants with the response actions without including a decision portion. Midway through the practice, participants were asked to begin driving and to complete all three tasks simultaneously. After completing the practice task, participants completed the experimental task. The experimental task was similar to the practice task with the only main difference being that sounds were now abstract tones (drawn from sounds used in the Sort Task experiments). Participants were instructed to make their classifications as quickly as possible while maintaining control of the vehicle.

Key Results

Results from the Sort Task portion of the Rapid Categorization Under Divided Attention, using the same cutoffs and criteria as in the Perceptual Space Sort Task and Validation Sort Task, indicate that the sounds from Rapid Categorization Under Divided Attention are categorized similarly to previous experiments and that the criteria and cutoffs result are highly predictive of alarm categorization ($R$=.865, adj. $R^2$=.701) as shown in Figure 48.

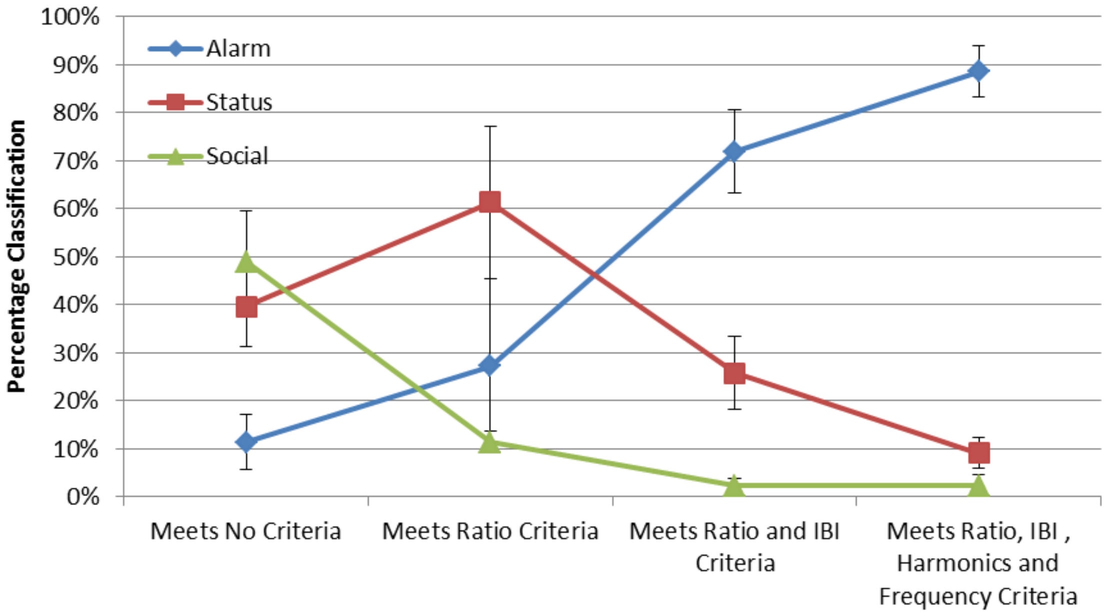


**Figure 48. Percentage categorization by criteria met for Sort Task**

As can be seen in Figure 48, if the abstract sound met the first two of the previously established criteria it was classified as an alarm at least 70% of the time and only misclassified as a status or social sound the remaining percentage of the time. Sounds that met all of the previously established criteria were classified as alarms nearly 90% of the time. Note that these results are highly similar to the results of the previous two Sort Tasks. This provides additional support for the reliability and validity of the previous results. The next aim of the current study was to determine if the previously established alarm criteria were also predictive of alarm perception under divided attention rapid categorization conditions.

Figure 49 shows the results of the rapid categorization task while participants were engaged in the driving and n-back distraction task. As can be seen, results are remarkably similar to those obtained from the Sort Task without concurrent load. Alarm categorization in the Sort Task obtained while driving was highly correlated with alarm categorization during the rapid categorization task (Pearson Correlation=.962, $p<.001$). Additionally, status and social categorization during the Sort Task were highly correlated with status and social categorization during the rapid categorization task (Pearson correlation for status categorization =.934, $p<.001$ and for social categorization =.982, $p<.001$).

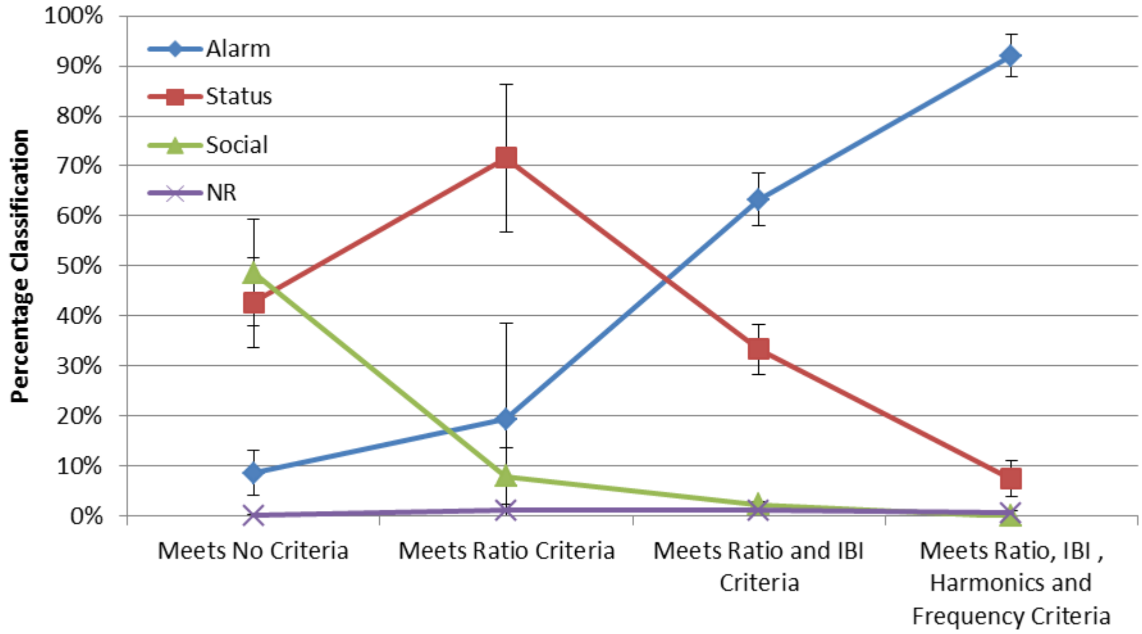


**Figure 49. Percent categorization while driving by criteria met during driving and distraction task performance**

Of further note, results indicate that the number of criteria met was indicative of response times where fastest response times were made for sounds that met all the previously established alarm criteria, and slowest response times were for sounds that met some of criteria as shown in Figure 50. These results indicate that when a sound matches with the criteria, people are able to decide more quickly whether or not the sound is a critical alarm relative to status or social sound.

Additionally, the more likely a sound was to be classified as an alarm, the faster the response time for that alert, with alarm classification being significantly predictive of lower response times, $B = -.398$, $p=.004$. Further, alarm classification even in the Sort Task was significantly predictive of faster response times during the driving task, $B = -.296$, $p=.031$. These results indicate that the more homogenous the classification (the more subjects classify a sound as an alarm), the faster subjects are able to respond to that sound.

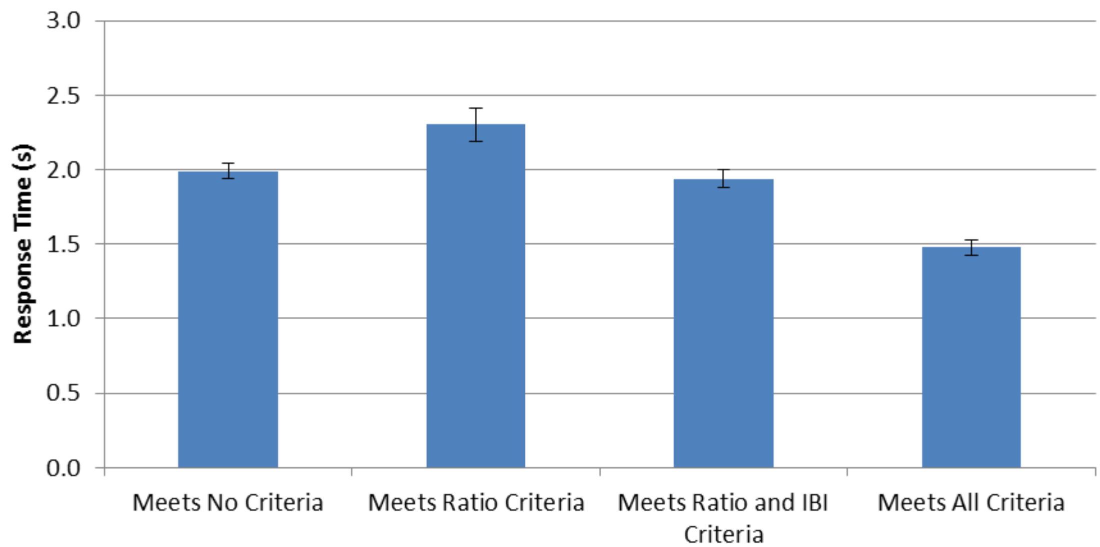


**Figure 50. Response time by criteria met**

Discussion

Results indicate that responses during rapid categorization while driving were remarkably similar to those found in the sorting only tasks. More interestingly, results show that sounds that meet more criteria have faster response times than those that only meet some criteria, indicating that when a sound meets all of the previously defined cutoff criteria participants find it easier to decide how to classify the sound and respond faster, almost always via the brake pedal. As illustrated in Figure 50, when a sound met all of the established criteria people responded on average within 1.5 s. However, when a sound met only the ratio criteria it took participants over 2 s to make a response. These results are quite dramatic and may suggest that non-urgent sounds should not be allowed to have ratio numbers that are above the criteria cutoff of .7 in order to prevent them from potentially being confused with a warning.

They are also faster when a sound meets none of the cutoff criteria, indicating that they find it equally easy then to decide that the sound is not an alarm. Sounds that meet a middle number of criteria showed longer response times, possibly indicating that those sounds were more ambiguous to participants, as is reflected in the percent categorization data.

### 4.2.7 General Discussion

The primary aim of this series of perceptual space investigations was to determine the key acoustic characteristics and the parameter values associated with of a sound being perceived as belonging to the category of a highly urgent collision warning sound. Toward this aim, the research team employed three methodologies in multiple experiments. The first experiment employed regression techniques to identify key characteristics that accounted for a substantial amount of the variance in category classifications. Cutoff scores for each of these key parameters were then established by examining the values of each characteristic that resulted in a high probability (defined as 85% in the research plan) of classification as an alarm (warning plus alert). Specifically, a sound that had a peak to total time ratio (Ratio) of greater than or equal to .7, an IBI of less than or equal to 125 ms, at least three harmonics, and a base frequency of greater than or equal to 1000 Hz was classified as an alarm over 90% of the time. Meeting only the first of these criteria – Ratio greater than or equal to .7 – resulted in an alarm classification rate of 70%. These resulting criteria were able to classify participants' perceptions of sounds. However, only one sound actually met these criteria and since the criteria were established by the same data set that was being used to examine classification performance, further validation was warranted.

Using alternative methodologies (method of adjustment with single and multiple parameters) yielded results that confirmed several of the parameters (base frequency, presence of at least three harmonics). Though not specifically manipulated, the Ratio parameter was also met by the resulting sound created in each of the method of adjustment procedures. Though the method of adjustment yielded a longer IBI than the criteria had established, it could reasonably be determined to be an artifact of the fact that when adjusting IBI participants were actually listening to bursts that contained three pulses that each contained onset and offset times of 20 ms each. Using criteria established in IEC 60601-1-8, a medical alarm standard, a sound with onset and offsets create the perception of tempo when they alternate in and out of 90% of their peak amplitude. Therefore, the sounds being adjusted would have inadvertently yielded a tempo within the criteria established in the first Sort Task.

A final validation in the series examined the predictive capabilities of these criteria while participants were engaged in a simulated driving task and were concurrently performing a distraction task consisting of a visual version of an n-back task. The criteria had strong predictive capabilities under these conditions. Sounds that met all four of the criteria were classified as alarms over 90% of the time. Further, when exposed to sounds meeting all four criteria participants were able to make their classifications much faster than when they met only one to three of the criteria. This suggests that sounds meeting these criteria have not only a high probability of being understood to be a warning but also that participants come to this understanding much faster than sounds that meet only a few of the criteria.

Together, the results of this series of investigations indicate that it is possible to define key characteristics and their associated parameter values that should reliably yield effective alarms for use in designing collision avoidance system DVIs. Further research examining the effectiveness of alarms matching these criteria in more ecologically valid contexts is warranted. The actual behavioral responses of drivers in potential collision situations after exposure to alarms designed in accordance with these criteria must be examined. However, the cross validation of these criteria across multiple experimental methodologies with multiple samples, combined with the observed speeded behavioral responses to these alarms in a simulated driving context when all criteria were met provide strong support for their potential effectiveness as collision warnings.

## 4.3 Warning Perception in Ambient Noise Environments

Introduction

In order to be reasonably effective, in-vehicle crash warnings must be reliably and rapidly detected by the driver and properly interpreted. They must convey the proper degree of urgency so that driver response is quick and appropriate. They should be distinguishable from less urgent alerts and messages, so that distraction, annoyance, and false alarm mistrust effects are minimized. Considerable research has addressed these issues, both within the CWIM project and broadly in the literature. However, the vast majority of this work has been conducted under relatively benign in-vehicle ambient noise conditions. Warnings, however, must remain effective under the likely range of noise conditions that may be anticipated in vehicles. Very little information exists on perception of meaning and urgency in noise.

The objective of the present study was to measure aspects of driver perception of warnings and alerts under a range of ambient noise driving conditions on actual roads. The characteristics and sound level of in-cab ambient noise may vary due to the vehicle's physical characteristics, the road surface, surrounding traffic, travel speed, and interior noise sources. As an initial study of this topic, only a limited set of ambient noise conditions could be included. Likewise, there are a great many types of auditory displays that might be evaluated, including various sounds as well as voice messages. Only a limited set of auditory displays could be included. The goal, then, was to provide an initial assessment of the nature and magnitude of the effects of ambient noise conditions on key aspects of driver perception of warnings. The intent was to encompass a range of noise conditions and auditory signal types. More refined investigations of listening conditions, signal characteristics, and driver reactions were warranted based on these initial results. This experiment is described in greater detail in a separate stand-alone report (Singer, Lerner, Walrath, & Gill, 2014). Follow-on research described in a separate report (Singer, Lerner, & Kellman, 2015) provides a laboratory replication of these findings, an investigation of additional ambient noise conditions, and investigations of the effects of signal loudness, annoyance, and consumer acceptability.

<u>Method</u>

The experiment was a three-factor design, with one between-groups factor (vehicle type) and two within-groups factors (interior noise condition, acoustic signal). Three different vehicles were used in the experiment in order to provide a representative range of vehicle types: (1) a small car, (2) a larger sedan, and (3) an SUV. Each participant drove only one of these vehicles. During the drive, data were collected under three different interior noise conditions: (1) windows up, music off; (2) windows down, music off; and (3) windows up, music on. The order in which each noise condition block was presented to participants was counterbalanced within each vehicle condition.

A set of 15 different acoustic signals was presented under each noise condition. These included three unique voice messages and eight unique non-voice sounds. All eleven of the unique sounds and voices were presented at a sound pressure level (SPL) of approximately 65 dB (A-weighted) as measured near the driver's right ear. One of the voice messages and three of the non-voice sounds were also presented at 75 dB, with the resultant total of 15 signals. The lower 65 dB level is representative of a number of acoustic alerts as measured in actual current practice. The higher 75 dB level is more consistent with human factors guidance, assuming a moderate level of ambient vehicle cab noise.

Five different dependent measures were recorded to evaluate driver response. These included: (1) a measure of reaction time for the participant to detect the occurrence of a signal; (2) a rating of signal noticeability; (3) a rating of signal urgency; (4) a rating of speech intelligibility (for voice messages only); and (5) perceived meaning of the signal (chosen from a set of four alternatives).

*Participants*. Participants included 34 drivers aged 22 to 49, with 13 males and 21 females. No participants reported having hearing decrements or using hearing assistive devices. All drove regularly, held valid U.S. driver's licenses and passed a screener of their motor vehicle records.

*Vehicles*. Three different classes of passenger vehicles were used in order to provide a range of vehicle types. These types were small car, sedan, and SUV. The specific vehicles used were selected from among the most popular (highest sales) models in that class and with good rental availability. The specific vehicles were:

- Small car: 2013 Hyundai Accent GLS
- Sedan: 2013 Toyota Camry LE
- SUV: 2013 GMC Terrain SLT

Twelve participants drove the small car, 11 drove the sedan, and 11 drove the SUV.

*Roadway.* Data collection took place on a limited access toll highway (Maryland Route 200) with a 60 mph speed limit. This is a relatively new highway with smooth and uniform asphalt over most of its length. It is also generally free-flowing, without congestion. These attributes permitted good control over ambient road noise and speed conditions. The roadway has three travel lanes in each direction. Participants were instructed to travel in the right lane except when needing to pass slower vehicles.

*Noise conditions.* All drives were conducted during clear weather on dry roads, with a target speed of 60 mph. The fan on the climate control system was on but set to a low setting. During the baseline condition, all windows were closed and music was off. During the "windows down" condition, the front windows on both sides of the vehicle were fully opened. During the "music on" condition, the smooth jazz song "Café Amore" by Spyro Gyra played in a continuous loop. The volume of the music was adjusted by the participant to the volume they would typically use for their own music while driving alone in their own car. However, the experimenter required participants to set the volume at a level equating to at least 60 dB, as measured in an otherwise silent vehicle interior. The minimum SPL was established to ensure that the music could potentially affect participants' detection and ratings of messages.

*Auditory signals and stimulus presentation.* Fifteen auditory signals were compared. There were 11 unique alerts presented at approximately 65 dB. Four of these sounds were also presented at approximately 75 dB. All sounds were initially volume-adjusted to these levels, but were then adjusted for perceptual equivalence of loudness, as determined by six individually tested raters.

The alerts used in this experiment were adapted from examples of current in-vehicle warnings and alerts of various types, other sounds found in various sources, and synthetic speech messages created using an online text-to-speech generator.[1] It is important to note that the signals that were sourced from current in-vehicle systems were presented using a different speaker in a different vehicle interior, and are not necessarily presented at the same SPL as the original alerts. Therefore, the results of this experiment do not necessarily reflect upon the messages as used in their native vehicle environments. The alerts used in this experiment are briefly described below. Note that alerts 1 through 8 are sounds presented at the 65 dBA level; alerts 9-11 are voice messages at the 65 dBA level; and alerts 12-15 are the subset of alerts presented at the 75 dBA level.

1. FCW 1: One burst of 20 fast beeps with a relatively high frequency profile.
2. FCW 2: Four bursts of four fast beeps with a relatively low frequency profile.
3. Blind spot warning: Three bursts of four fast beeps, each with a smoothed onset and offset and a sustained low intensity sound between beeps.
4. Pedestrian warning: A single sustained beep with a duration of 2 s.
5. Seat belt alert 1: A single chime that decays to silence in the span of about 2 s, with intensity varying in a wavelike pattern.
6. Seat belt alert 2: Two chimes, each of which decays to silence in the span of about 1 s.
7. Park assist 1: One burst of eight beeps.

[1] http://www.oddcast.com/home/demos/tts/tts_example.php

8. Park assist 2: Two bursts of three beeps.
9. Female voice – not urgent: Female voice says "Attention."
10. Female voice – urgent: Female voice says "Warning, warning."
11. Male voice – urgent: Male voice says "Warning, warning."
12. FCW 1 (high): Same as FCW 1, but presented at 75 dB.
13. Blind spot (high): Same as Blind spot, but presented at 75 dB.
14. Park assist 1(high): Same as Park assist 1, but presented at 75 dB.
15. Female voice – urgent (high): Same as 13, but presented at 75 dB.

During the experimental drives, the auditory signals were presented by an X-Mini II XAM4 capsule speaker mounted on top of the dashboard immediately behind the steering wheel.

Within each noise condition block, the experimental control software generated a random presentation order for the 15 auditory signals. The software provided a random time gap that ranged from 10 to 50 s and averaged 30 s from the completion of the previous sound's ratings to the presentation of the next sound. Once the random time had passed, the software indicated to the experimenter that the next signal could be activated. The actual triggering of the trial was done by the experimenter, who first determined that there were no usual acoustic circumstances (e.g., a large truck passing or a patch of noisier roadway surface). When the participant detected the signal they pressed a microswitch button, worn on their finger or thumb, to provide a response time. The data collection system recorded the response time and then cued the experimenter, who was seated in the right rear seat, to verbally present a series of rating and choice questions. The participants rated:

- Noticeability (defined as "The sound is easily noticeable among other sounds and noises in the vehicle); 1=not very; 7=extremely
- Urgency (defined as "The sound conveys a sense of importance, motivating you to make an immediate response); 1=not very; 7=extremely
- Perceived meaning. The sound was placed in whichever of four categories the participant felt best matched the meaning conveyed by the signal
    1. Urgent crash warning
    2. Safety information
    3. Information not related to safety
    4. Incoming personal communication (e.g., call, text message, email)

If the participant did not respond to the alert by pressing the microswitch within 8 seconds of signal initiation, the participant was deemed to have failed to detect the signal and no rating questions were asked. The participant received no feedback that there was a missed signal.

<u>Results</u>

Driver perception of the auditory signals was influenced by both the nature of the signal and the ambient noise background. Figure 51 shows the percent of signals detected by the driver for each signal under each ambient noise condition. Few participants missed any signals in the baseline condition (windows up, music off). In the music condition, a few signals were detected in the 70% or fewer range, but most were detected at least 80% of the time. The open windows condition interfered with detection more dramatically. All four of the signals presented at the 75 dB level continued to be detected well even with the windows open, but for the 65 dB signals there was a wide disparity in performance, with a few detected by more than 90% of participants and three others detected by only around 10% of participants.

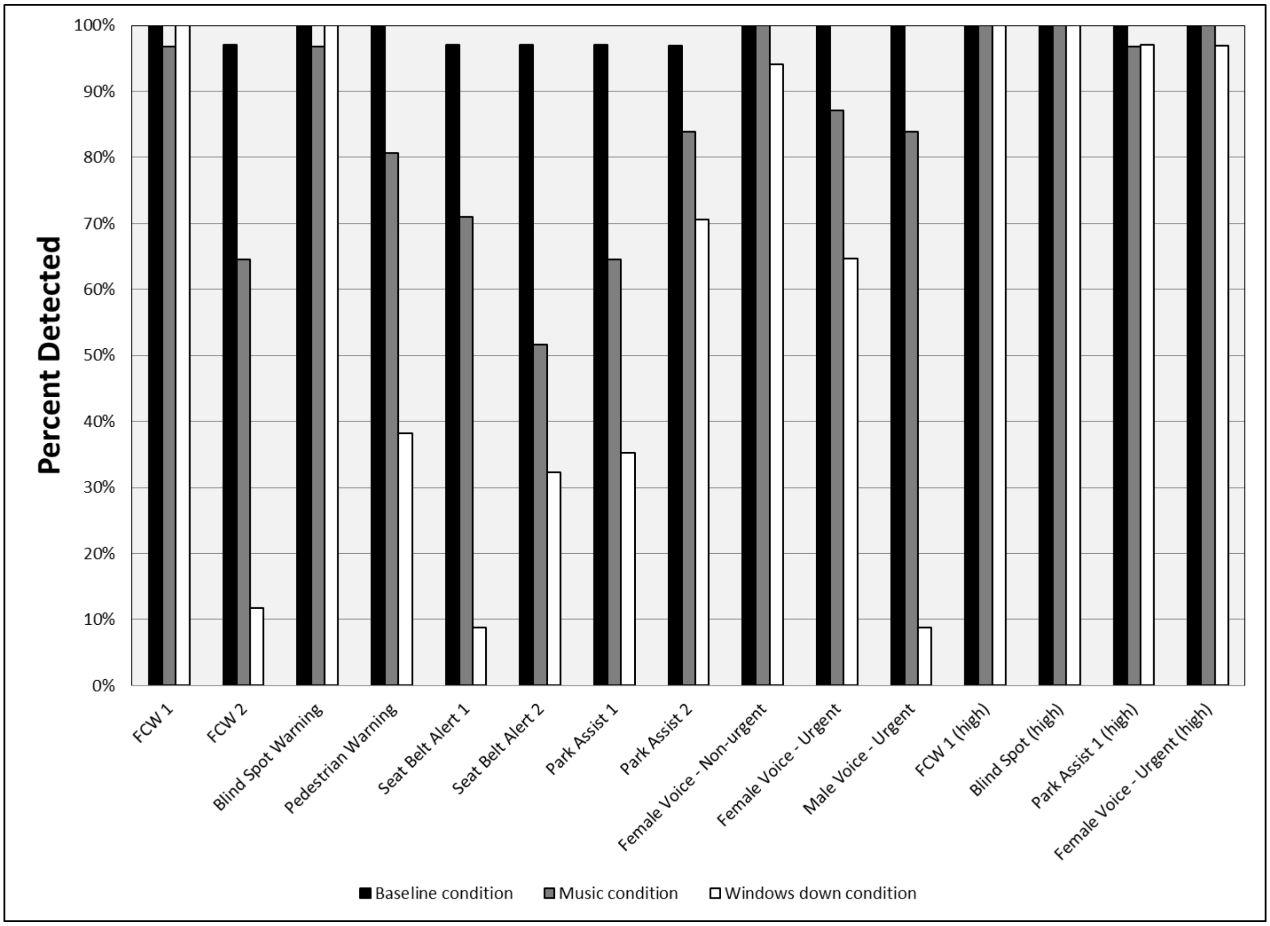


**Figure 51. Percentage of participants who detected alerts under each ambient noise condition**

Three-factor (signal, ambient noise background, vehicle type) ANOVAs were conducted for the measures of rated noticeability, rated urgency, and response time. The three parallel analyses yielded identical conclusions. Signal, noise condition, and their interaction were all statistically significant ($p<.0001$). There was no main effect of vehicle type and no interaction of vehicle type with ambient noise background. There was a significant interaction of signal with vehicle type, though there was no clear pattern of effects. It may be expected that due to the complex and varied geometry of the vehicle cabin space and the nature of the reflective and absorbing materials, the differences in the acoustic space might differentially affect some signals. There was no three-way interaction. Figure 52 shows the group mean rating of noticeability for each signal under each ambient noise condition and Figure 53 shows this for the group mean rating of urgency. Figure 54 shows this for response time. The figures show that there were dramatic differences in perceived noticeability and urgency, even among sounds that were equated for similar perceived loudness under relatively quiet conditions.

For both measures, group mean ratings for particular signal/ambient noise conditions ranged from less than 2 to slightly under 7 on the 7-point rating scales.

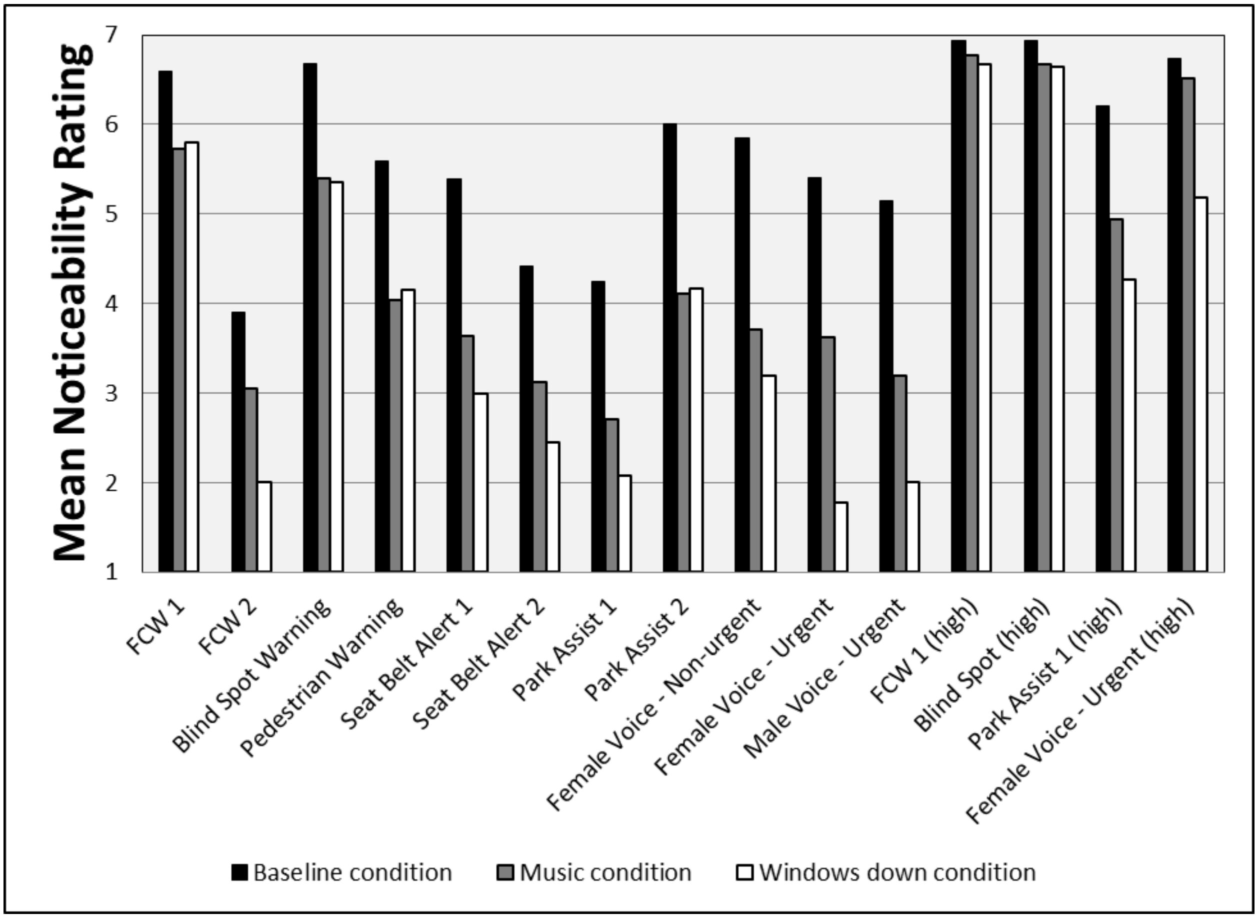


**Figure 52. Mean noticeability rating for each combination of signal and ambient noise condition**

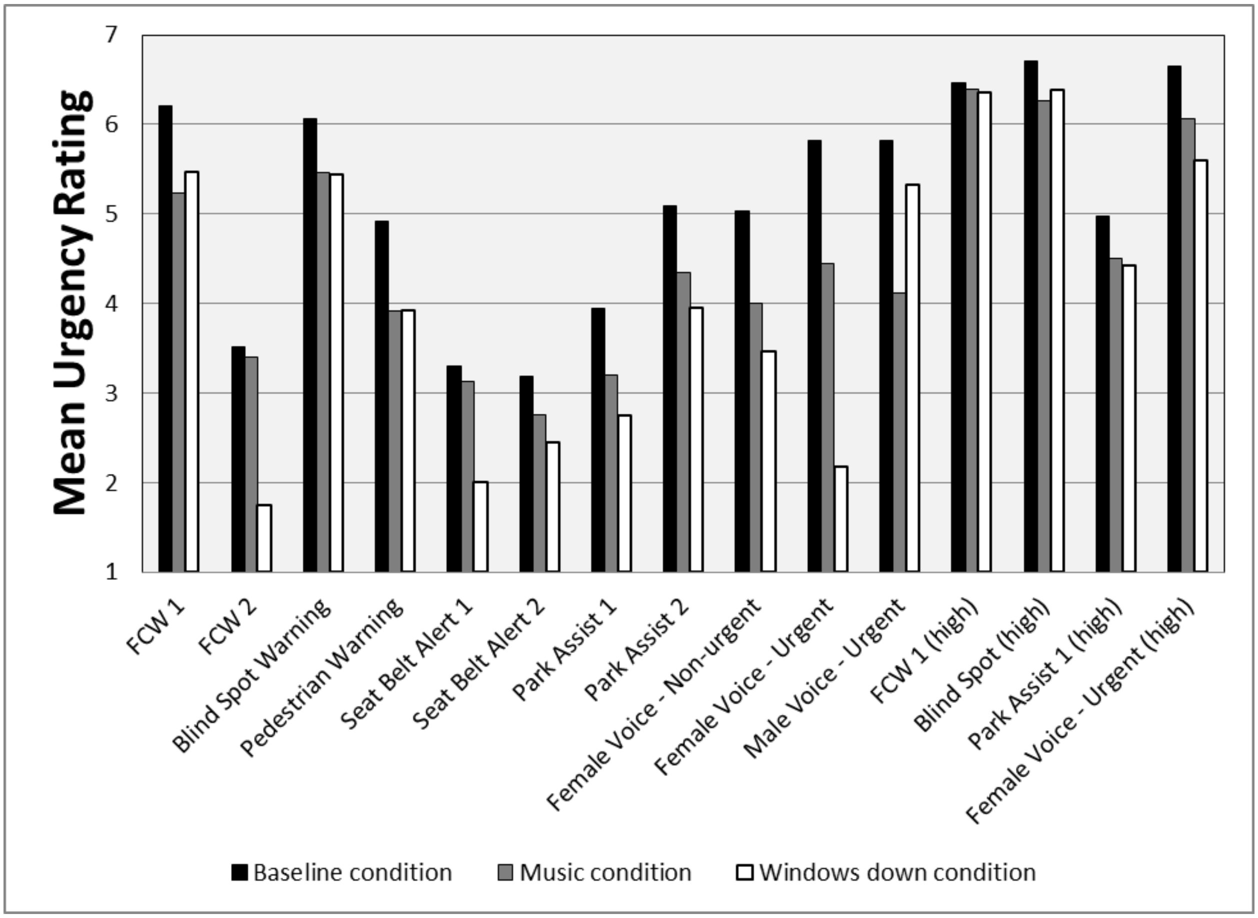


**Figure 53. Mean urgency rating for each combination of signal and ambient noise condition**

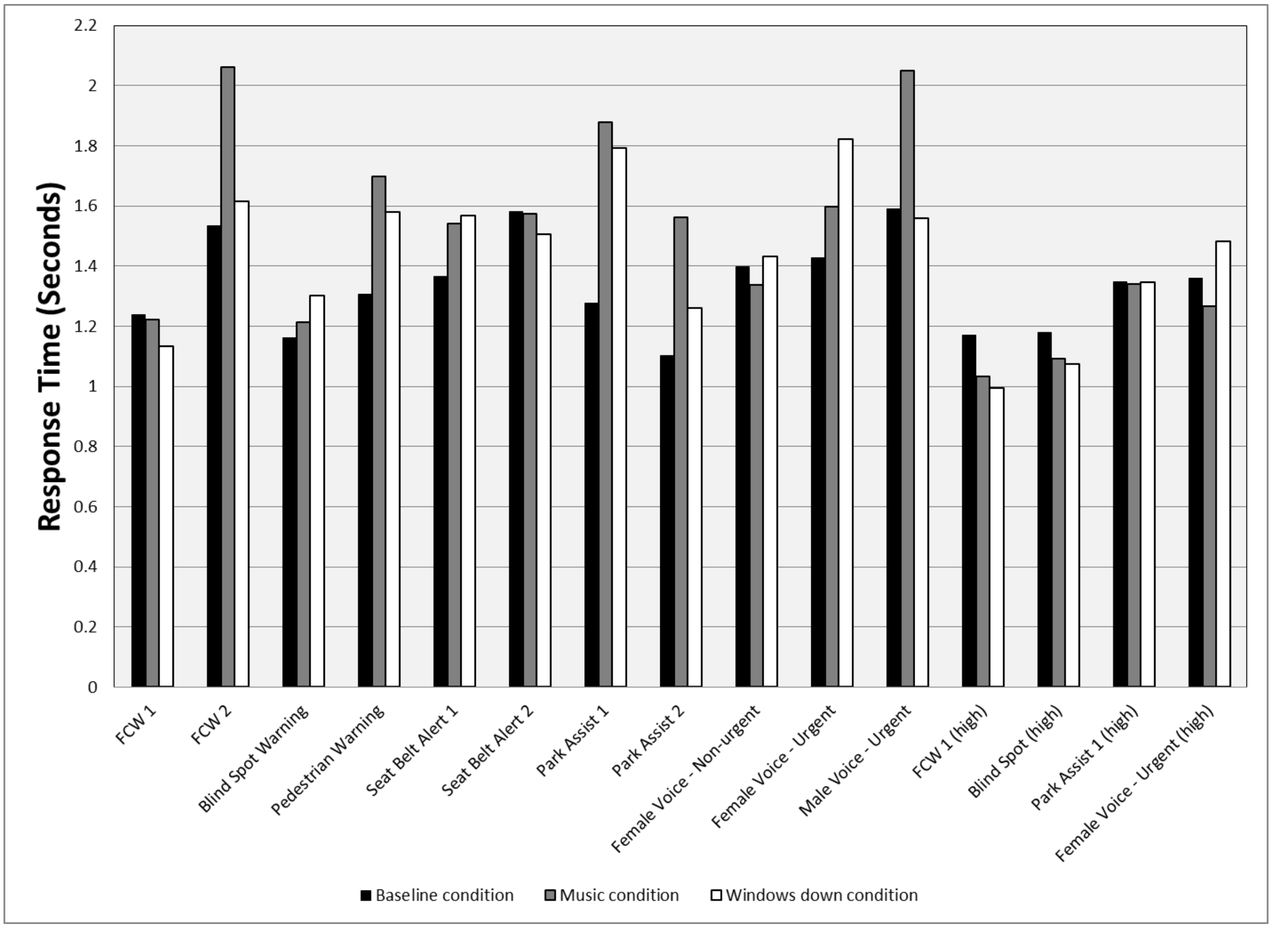


**Figure 54. Mean response time for each combination of signal and ambient noise condition**

In addition to providing ratings, participants categorized each signal as either an urgent crash warning, non-urgent safety information, information unrelated to safety, or incoming personal communication. Of particular interest here was the effect that different ambient noise conditions had on the categorization of signals that were intended to function as crash warnings. Findings suggest that signals that are most likely to be categorized as warnings under quiet conditions lose some of their effectiveness in communicating urgency under conditions of elevated ambient noise. Figure 55 shows two examples of this for the 75 dB versions of the blind spot warning and urgent female voice (“warning, warning”). For both alerts, the likelihood that a signal will be classified as an urgent crash warning decreases in higher ambient noise conditions.

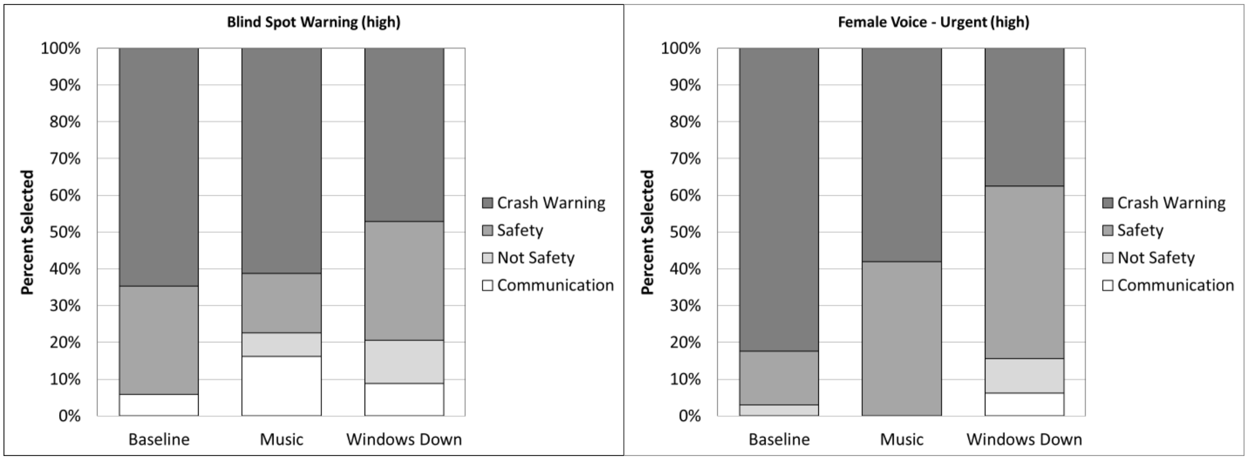


**Figure 55. Categorization of signals by ambient noise condition for blind spot warning (high) (left) and female voice - urgent (high) (right)**

<u>Discussion</u>

Background noise from music, and especially from open windows, interfered with the perception of auditory signals presented at 65 dB. Interference was not very pronounced for the set of 75 dB signals, although only four signals were included at this level. The set of sounds and voice messages equated for approximately equal loudness under relatively quiet listening conditions differed substantially in noticeability and urgency even under the baseline condition and even more under the music and open windows conditions. Under noise conditions, 65 dB signals were less likely to be detected, and when they were detected they typically lost much of their perceived urgency, which may compromise their effectiveness as urgent crash warnings. Some sounds suffered low detection rates under noise, particularly the windows down ambient condition.

This experiment was designed to provide an initial examination of the extent to which possible ambient noise conditions might interfere with signal detection and meaning. It was not intended to provide any systematic evaluation of signal features or parameters regarding their resistance to noise effects. However, based on the limited sample of sounds and conditions, it appeared that the predominant frequencies that characterize a signal may relate to perceived urgency under noise. Sounds with base frequencies below 1000 Hz generally performed worst and those with base or significant components above 1500 Hz performed best. However, this observation is based on a very limited sample of sounds that also differed in a number of other respects, and so it should be considered tentative.

As an initial study on this topic, the experiment demonstrated very sizable effects of ambient noise conditions that might reasonably be expected to occur on occasion. Following this on-road experiment, a series of three laboratory experiments was conducted to replicate and extend these findings. The first experiment provided a successful replication of the on-road results using headphone playback of recorded ambient noise conditions and alerts in a lab setting. The second experiment investigated additional ambient noise conditions, including different pavement conditions and environmental factors (e.g., rain). The third experiment investigated the effects of varying alert loudness, and added alert annoyance and consumer acceptability as dependent variables. These experiments are described in a separate report (Singer, Lerner, & Kellman, 2015).

# 5 Research on Temporal Aspects of Interference from other In-Vehicle Messages

## 5.1 Driving Simulator Research on Near-Simultaneous Safety Alerts and Crash Warnings

New vehicles feature an increasing number of warning systems, safety features, and entertainment systems. Although the design of each system is taken into consideration to ensure each performs as intended, little research has addressed the integration of these systems and how they may affect the driver response. Time-critical or urgent systems such as those designed to avoid collisions, lane departures, and other potentially critical situations may be in competition with less urgent systems such as those designed for entertainment purposes. Past research has found that if tasks share the same perceptual and response mode, then resource competition is greater (e.g., Wickens, 2008). Therefore, two auditory alerts have greater resource competition than systems that differ in perceptual mode. Just as important is the coordination and timing of alerts and how one might affect the driver's response time to another. Wiese and Lee (2004) hypothesized that a less urgent event (an auditory email message alert) would interfere with a more urgent alert (auditory collision avoidance alert). In their first study, Wiese and Lee triggered an email alert 300 ms before a collision warning alert and found that the email alert interfered with driver's response to the collision warnings. These results also support more recent findings from Hibberd et al. (2013), which found that when drivers were given a two-choice response in-vehicle task where they had to discriminate between two stimulus pairs and a critical lead vehicle braking event, the in-vehicle task had a negative effect (slower reaction times) on driver braking when presented 350 ms prior to the critical event. However, in the second study, Wiese and Lee triggered an email alert 1000 ms before the critical event and found the opposite effect; the onset of the email alert improved the braking process and induced a faster accelerator pedal release.

A study by Levy, Pashler, and Boer (2006) tested predictions of the central bottleneck hypothesis, which applies when the information processing stage of response selection limits dual-task performance by acting as a bottleneck. Levy et al. also tested predictions of hypotheses about interference from common stimulus and response modalities. Participants in the study were asked to perform two tasks simultaneously while performing a simulated driving task. A choice task was used where participants responded either manually or vocally when an auditory or visual stimulus occurred. The participants were asked to brake as soon as they saw the lead vehicle brake lights illuminate. The stimulus onset asynchrony (SOA), which is the time between the onset of the task stimuli, varied between 150, 350, and 1200 ms. Although the choice task was considered to be an easy task, it still suffered from dual-task interference. Reaction times at 0 to 350 ms SOA produced a braking delay of more than 16 ft for a vehicle traveling at 65 mph. Brake reaction times increased as the SOA was reduced, which provided evidence of interference. In contrast to the Wiese and Lee study, however, the 1200 ms SOA participants had slower brake reaction times than those with 350 ms SOA, but not slower than those with 0 or 150 ms SOA. This may be due to the fact that participants were less prepared to brake, having gotten used to shorter SOAs.

### 5.1.1 Phase 1: Research with Undistracted Drivers

Introduction

The overall goal of this Phase 1 study was to evaluate the effects of an alert that occurs shortly before an urgent crash warning on driver response to the warning. The study also sought to determine if the results may be similar to those in past studies (e.g., Wiese & Lee, 2004; Hibberd et al., 2013) where shorter time gaps between early alert and warning generally resulted in slower brake reaction times. Safety-related early alerts include non-verbal (tonal) and verbal auditory alerts. In addition, the study sought to determine if there would be similar results as those found in past research where alerts with similar perceptual modes interfered with one another more than alerts with different perceptual modes (Wickens, 1984, 2002, Haigney & Westerman, 2001). For example, pairing haptic FCW alert events with auditory early alerts (both tonal and verbal) may result in quicker braking times than when auditory FCW alert events are paired with auditory early alerts.

The test plan was to study driver performance and behavior in potentially risky scenarios in a safe, repeatable, and controlled environment, using the Dynamic Research, Inc. (DRI) motion base driving simulator. The DRI driving simulator is a dynamically realistic, moving base, "driver-in-the-loop" device.

Method

The objectives of this study were to evaluate various onset time intervals between less urgent safety-related alerts and an ACWS, and, as stated in the similar research by Wiese & Lee (2004), if "coordinating the onset of alerts to avoid temporal conflict may be a critical consideration in integrating in-vehicle information systems." The independent variables in this experiment were: Two types of lead vehicle braking events, three interval SOA times, two early alert types, two FCW types, and a baseline "no early alert" condition. The study was designed so that the SOA times and FCW type were within-subject variables, while the early alert type was a between-subject variable.

Simulated car sounds and other roadway sounds were present during each simulation. The ambient sound level inside the simulator was approximately 57 dB. The FCW alert types consisted of an auditory FCW alert and a haptic FCW alert. The auditory FCW alert consisted of a pulsing tone with a period of 400 ms and a 50% duty cycle for a total duration of 4 s. The auditory FCW alert had an approximate peak SPL between 70-75 dB(A). The haptic FCW alert consisted of a simulated deceleration (i.e., brake pulse) of approximate 0.5 g in 0.2 s. FCW alerts were issued when the lead vehicle began the critical braking.

Two non-urgent early alert types were used, a tonal early alert and a verbal early alert. Early alerts were non-urgent safety related alerts that were presented prior to an FCW alert or on their own. The early alerts chosen were curve ahead, traffic ahead, and construction ahead. The type of early alert given was determined by the group number in which the participant was assigned. In the tonal early alert a sound would play, accompanied by a visual icon. In the verbal early alert the phrases "curve ahead," "traffic ahead," and "construction ahead" were spoken and were accompanied by an icon. SOAs of 150 ms, 300 ms, and 1000 ms were used between the early alert and the FCW.

Prior to the lead vehicle critical braking event, the subject vehicle was traveling at approximately 55 mph on a straight and level road following a lead vehicle that was also traveling at 55 mph. During the lead vehicle critical braking event the lead vehicle suddenly decelerated to 25 mph at a rate of 0.5 g. A total of six critical events were used and twenty non-critical events, per road.

The headway distance (i.e., following distance) to the lead vehicle was displayed using a simulated analog headway display located in the dashboard, which gave the driver feedback on how they should adjust their speed to meet the needs of the study. Drivers were asked to maintain their speed within the green zone as much as possible by using the distance gauge to determine if they needed to speed up or slow down. If the needle was in the right red portion the subject vehicle should slow down to get back in the green area. If the needle was in the left red portion the subject vehicle should speed up to get back in the green area as described in Figure 56. The early alert icon was also displayed on the instrument panel. The icon helped drivers identify early alerts for both the verbal and tonal conditions. The same icons were used for verbal and tonal early alerts. The FCW did not have an icon associated with it.

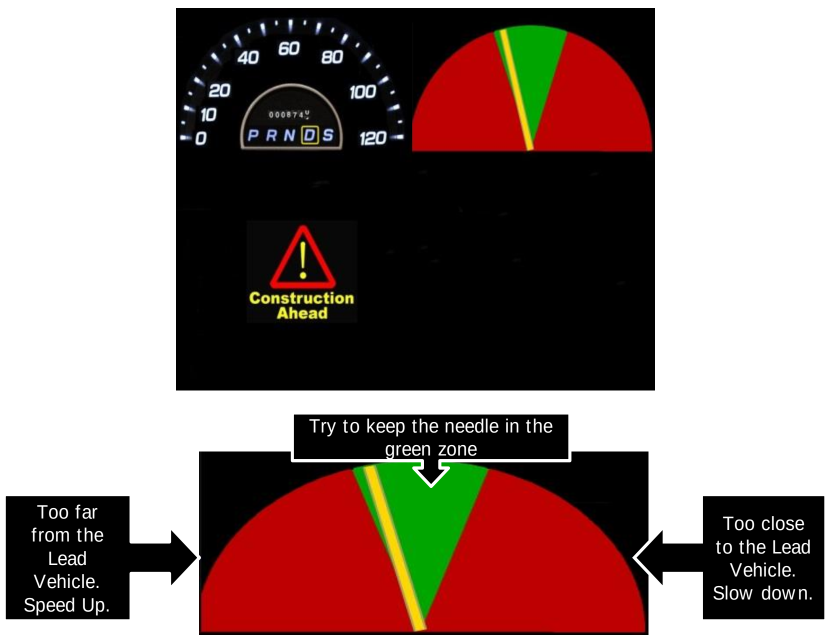


**Figure 56. Headway display gauge**

A total of 36 participants were included in the driving simulator evaluation. The order of the early alerts and FCW were counterbalanced with the braking events in order to minimize learning effects. Null braking events where alerts were presented without the presence of a critical event were included to minimize an association of braking events paired with a particular alert type. Participants were randomly assigned to one of the two main early alert type groups (verbal and tonal) with 18 participants per group. There were two FCW (auditory and haptic) conditions and three SOA intervals, resulting in six counterbalanced configurations within each early alert type group. The baseline (auditory FCW or haptic FCW without an early alert) was presented to participants as the first event. There were three participants per configuration (3 participants x 6 counterbalance configurations x 2 early alert types = 36 participants). All participants drove a 15-minute practice road followed by two 45 minute drives.

Three different types of secondary tasks were used. These tasks were used to try to mask the true purpose of the study and were presented independently of any lead vehicle braking events. The tasks included a number recall task, a CD task, and a trivia task. Both the recall task and trivia task were displayed on an LCD monitor mounted in the center console to the right of the driver.

The two measures of most interest were total response time and distance. Total response time was defined as the start of the lead vehicle critical braking event to the end of movement to the brake pedal. The data for total response time provided a more complete analysis than accelerator reaction time and accelerator-to-brake movement since it captured the missing accelerator response and movement times. Distance was defined as the minimum distance between the subject vehicle and lead vehicle after a critical lead vehicle braking event. A timeline and definition of measure are shown in Figure 57.

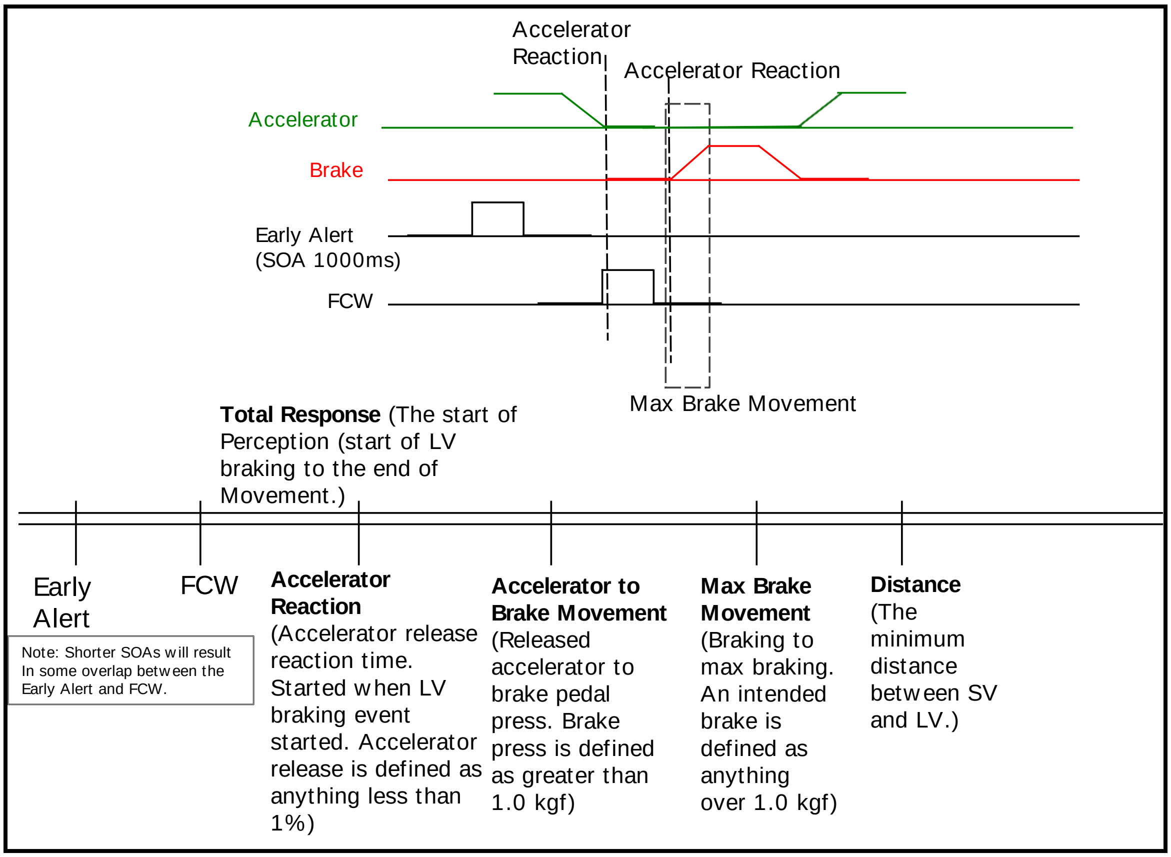


**Figure 57. Timeline and definitions of measures**

Key Results

There were significant main effects when comparing tonal and verbal groups. There was a main effect of distance ($p$=0.025) where the verbal early alerts resulted in longer minimum distances between the subject vehicle and lead vehicle. There were shorter total response times for the verbal early alert group ($p$=0.002).

There was a significant difference for FCW modality ($p$=0.006) in the verbal early alert group where the haptic FCW alert resulted in shorter total response times than the auditory FCW alert. There was also a significant difference of SOA within the verbal early alert group ($p$=0.020) where having an early alert resulted in shorter total response times than the baseline group, with 1000 ms SOA having the shortest total response time.

Similar results were found in the tonal early alert group. There was also a significant statistical difference between the various SOA times. Having an SOA of 150 ms or 300 ms resulted in longer total response times, but having an SOA of 1000 ms resulted in the shortest total response times as shown in Figure 58. There was also an interaction between SOA and early alert type ($p$=0.007) wherein total response time decreased between 300 and 1000 ms for tonal early alerts, but stayed virtually unchanged for verbal early alerts.

For all participants, there was a significant difference between auditory and haptic FCW on total response time ($p$=0.007) wherein haptic FCW alerts produced shorter total response times. There was also a significant difference between SOA times ($p$=0.001). As SOA increased, total response time decreased with a decrease in 1000 ms SOA as shown in Figure 59.

There was also a marginal interaction between FCW and SOA ($p$=0.052) where for auditory FCW total response time had a decrease at 1000 ms but for haptic FCW total response time decreased slightly with every SOA increase.

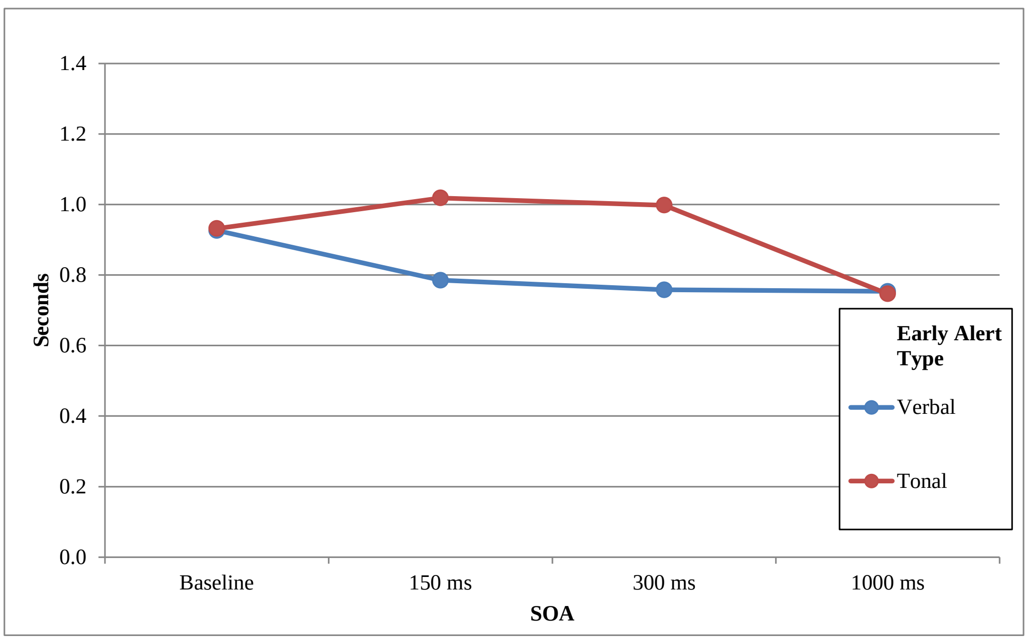


**Figure 58. Total response time as a function of early alert type and SOA (undistracted drivers)**

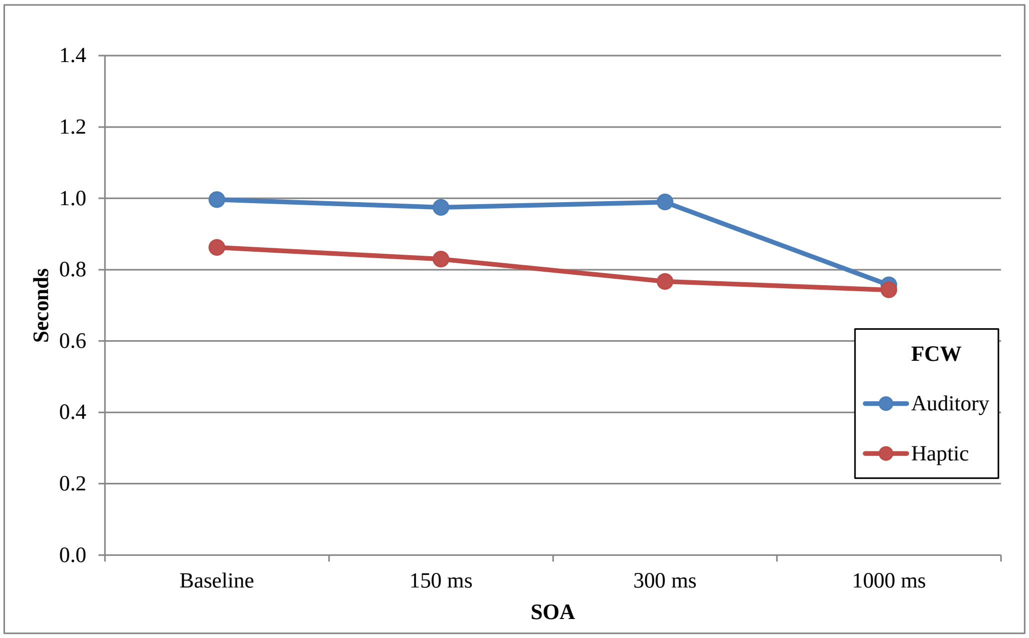


**Figure 59. Total response time as a function of FCW modality and SOA (undistracted drivers)**

Paired t-test comparisons of total response times between the baseline events and early alert events show a significant difference between the baseline events, verbal early alert events, and tonal early alert events with an SOA of 1000 ms. There were also significant differences between haptic FCW baselines and verbal early alert events with an SOA of 150 ms and 300 ms. There was also a significant difference between baseline events and tonal early events with an SOA of 1000 ms as shown in Table 17.

**Table 17. Paired comparisons of total response times between baseline events**

| **Verbal** | | | | | | | |
|---|---|---|---|---|---|---|---|
| Audio FCW | 150 ms | 300 ms | 1000 ms | Haptic FCW | 150 ms | 300 ms | 1000 ms |
| | 0.089 | 0.181 | 0.011* | | 0.037* | 0.010* | 0.091 |
| **Tonal** | | | | | | | |
| Audio FCW | 150 ms | 300 ms | 1000 ms | Haptic FCW | 150 ms | 300 ms | 1000 ms |
| | 0.458 | 0.281 | 0.014* | | 0.730 | 0.700 | 0.026* |

* result significant at p<.05

Only two collisions occurred during this study, one during an auditory FCW baseline event and one during a haptic FCW with tonal early alert and SOA of 150 ms.

There was a statistical difference between FCW alerts ($p$=0.025) in the verbal early alert group where the haptic FCW alert events resulted in longer minimum distances between the subject vehicle and the lead vehicle. In contrast, there were no statistical differences within the tonal early alert group.

A repeated measures analysis found that there was a significant difference between auditory and haptic FCW ($p$=0.002) where haptic FCW resulted in longer minimum distances between the subject vehicle and lead vehicle. As depicted in Figure 60, there was also an interaction between SOA and early alert type ($p$=0.042), where for verbal early alerts, distance decreased as SOA increased, but for tonal early alerts there was an increase in distance for alerts with SOA of 300 ms and 1000 ms.

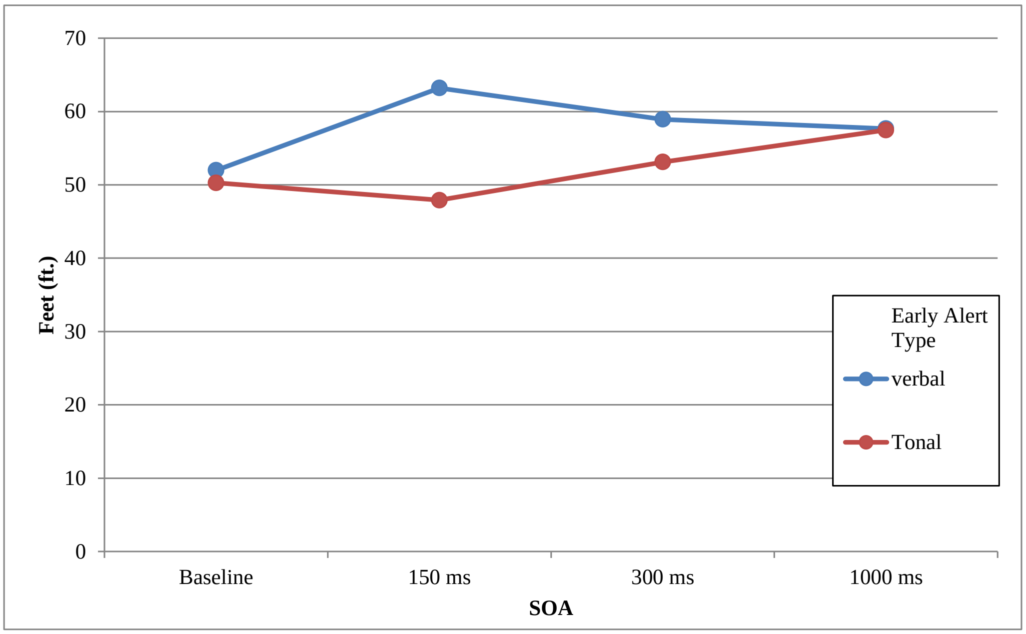


**Figure 60. Minimum distance between subject vehicle and lead vehicle as a function of early alert type and SOA (undistracted drivers)**

<u>Discussion</u>

The overall goal of this experiment was to evaluate various onset time intervals between less urgent safety-related alerts (early alerts) and an ACWS. This research also sought to determine if the results would be similar to those in past studies where shorter SOA times resulted in slower max brake movement times. In contrast, it found that longer SOA times (1000 ms) for max brake movement times were not negatively affected by early alerts. This research also sought to determine if alerts with similar perceptual modes would interfere with one another more than alerts with different perceptual modes.

One of the key questions addressed in this experiment asked if shorter SOA times slowed total response time. The data showed that an SOA of 1000 ms resulted in the shortest total response times for both the verbal and tonal early alert groups. These results are consistent with past research from Wiese and Lee where an SOA of 1000 ms enhanced response time, but shorter SOA times interfered with response time.

Interactions between early alert type and SOA showed that for the tonal early alert group, total response time decreased for 1000 ms SOA, but only gradually decreased with the verbal early alert group. In this instance, the tonal early alert events were consistent with the findings from Hibberd et al. (2013) which implied that negative effects of in-vehicle distractors could be mitigated through controlling the presentation of in-vehicle task, such that they are not presented in the period immediately before the driver is required to brake, or before 350 ms. Total response times were still better than when early alerts were presented at SOA 150 ms.

For the tonal early alert group, having an early alert aided total response time when it was presented at 1000 ms when compared to no early alert (baseline). For the verbal early alert group, having an early alert presented at any SOA resulted in shorter total response times than if not having an early alert (baseline). Early alerts, in general, may be acting as pre-cues to a critical event. This, combined with the fact that verbal alerts did not require that participants look for an icon to decipher what the alert meant, may explain why verbal alert events resulted in shorter total response times than the tonal early alert events.

The data also revealed an interaction between FCW type and SOA times. When verbal early alert events were presented at 300 ms, distance was decreased, but increased when tonal early alert were presented with the same SOA.

For the verbal early alert group, having any type of early alert resulted in larger minimum distances than having an FCW on its own (baseline). In contrast, for the tonal early alert group, having an early alert presented at an SOA of 150 ms caused shorter distances than any other SOA.

In addition to the main effects of SOA, this experiment investigated whether warnings with similar perceptual modes interfere with each other. For both the verbal and tonal early alert groups, total response time was shorter when paired with a haptic FCW alert than when paired with an auditory FCW alert. Distance results were similar to those found for total response and accelerator to brake movement, where having a haptic FCW was more of a benefit than having an auditory FCW.

Although haptic FCW resulted in shorter total response times overall, auditory FCW events had a decrease in total response time for SOA times of 1000 ms. Auditory and haptic FCW events resulted in similar 1000 ms total response times. Again, these results supported previous research that found that shorter SOA times delayed response compared to 1000 ms SOA times which enhanced response time.

### 5.1.2 Phase 2: Research with Distracted Drivers

Introduction

The key methodological difference between this Phase 2 experiment and the previously described Phase 1 experiment was that drivers in Phase 2 were engaged in a distractor task that was activated prior to the event and cued before the onset of early alerts during FCW trials, while in Phase 1 they were not. The Phase 2 experiment examined how the temporal relationship of the early alert to FCW onset can affect how distracted drivers respond to an FCW. It is unclear how drivers' reaction times will differ when their attention is divided between driving and a distractor task at the onset of a critical event. Previous research by Hibberd et al. (2010) found that the presentation of an in-vehicle task 350 ms before a braking event delayed braking 174 ms, in contrast to 146 ms without an in-vehicle task, and resulted in a 5.45 m increase in stopping distance. This study hypothesized that engagement in a distracting in-vehicle task prior to an early alert and a critical braking event will cause a delay in total reaction and braking reaction times, along with shortened distance between the subject vehicle and lead vehicle.

Method

Thirty-six drivers participated in Phase 2. A visual-verbal number task was used as a distractor task and paired with both critical and non-critical events. The number task required that participants verbally repeat a sequence of five digits displayed on an LCD screen mounted in the center console on the right of the driver seat as in the previous Phase 1study. The main goal of the distractor task was to have the driver look away from the forward roadway scene so as to not visually detect the onset of the lead vehicle critical braking event. This task was altered to ensure that participants looked away from the road during the duration of the task. The number task started 1.6 s after the number task pre-recorded auditory instruction was given. The number presentation lasted a total of 0.82 s. A previous study by Sugimoto and Sauer (2005) found that a mean reaction time for participants to respond to a warning is 0.82 s. This same reaction time was used to trigger a distraction task prior to an ACWS warning in a study by Van Auken et al. (2011) to ensure that participants had enough time to turn their attention and respond to the alert. The current study used the same timing to give participants sufficient time to turn their head and react to the number task. The number task also appeared on its own (without early alert or FCW alerts) more than half the time in order to reduce the participant's anticipation of a lead vehicle critical event.

Two different types of baseline trials were included in the experiment. The first type of Baseline condition (Baseline 1 events) were lead vehicle braking events that occurred without any alerts. The second type of Baseline condition (Baseline 2 events) were lead vehicle braking events that included an FCW, but no early alert.

Eye glance data were extracted for a set amount of time for each type of event. Eye glance reduction started at the end of the distractor task audio cue and continued for 2.85 s, which is at the early alert activation, and in some events, at the FCW alert (and lead vehicle braking activation).

Although participants were told that accuracy of the number task was important, and reminded again during the study, most participants still tended to look up at the road during the number task. It is likely that after having experienced a lead vehicle braking event in the baseline 1 event, most participants tended to look up more during the baseline 2 and critical events in anticipation of a lead vehicle braking event. Before any alert, participants spent most of their time looking down at the LCD monitor but did occasionally look up at the road while performing the task. Although participants tended to look up during the first 2.85 s after the distractor audio cue, they were still engaged in the task and did not detect the lead vehicle braking event prior to alerts.

Key Results

As SOA increased, total response time decreased, with an increase at 1000 ms SOA. There was a significant difference between auditory and haptic FCW on total response time ($p$=0.035) such that haptic FCW alerts produced shorter total response times. There was also a significant difference between SOA times ($p$<0.005).

Figure 61 shows that there was an interaction between SOA and early alert type ($p$=0.016) where for tonal early alert, as SOA increases, total response time decreases with an increase at 1000 ms. In contrast, for verbal early alert, as SOA increases, total response time decreased slightly (see Figure 61). There was also an interaction between FCW and SOA ($p$=0.004) where for haptic FCW, total response time had an increase at 1000 ms, but for auditory FCW, total response time decreased slightly with every SOA increase (see Figure 62).

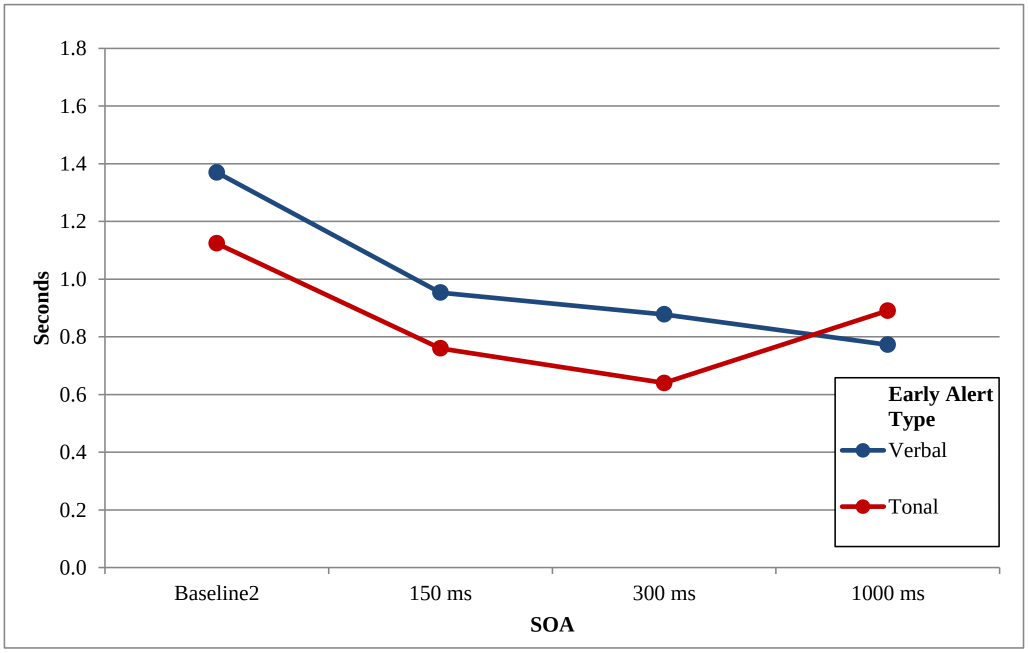


**Figure 61. Total response time as a function of early alert type and SOA (distracted drivers)**

Paired t-test comparisons for verbal early alert total response times showed a significant difference between both auditory and haptic baseline 2 events and events with an SOA of 150, 300, and 1000 ms. There were also significant differences for tonal events between auditory FCW baseline 2 events and events with an SOA of 150, 300, and 1000 ms. There was also a significant difference for haptic FCW baseline 2 and events with an SOA of 150 and 300 ms (see Table 18).

When looking at the number of collisions during baseline events, the baseline 1 events, which were critical lead vehicle braking events without any alerts, resulted in the highest likelihood of a collision (72%). That number was greatly reduced for baseline 2 events which were events with an FCW during a critical lead vehicle braking event. Three collisions (8%) were observed for the audio FCW events while five collisions (14%) were observed for the haptic FCW events. It should be noted that all baseline 1 events occurred during the warm-up drive for each participant while the baseline 2 events occurred during the actual study drives.

A repeated measures analysis found that there was a significant difference between audio and haptic FCW ($p$=0.005) where haptic FCW resulted in longer minimum distances between the subject vehicle and lead vehicle. There was also a significant difference between SOA times ($p$<0.005). When SOA increased, minimum distance also increased with a decrease in 1000 ms SOA.

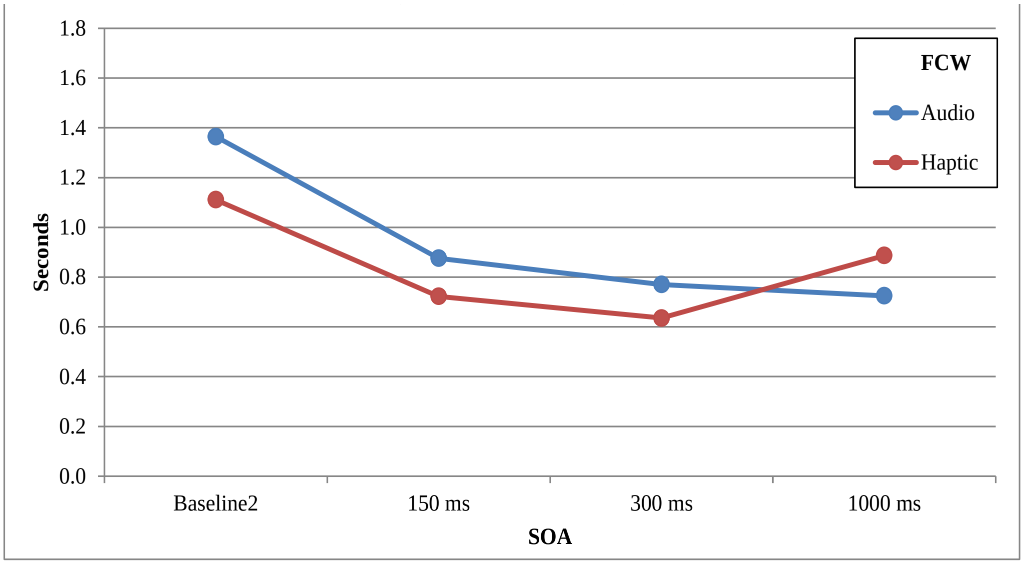


**Figure 62. Total response time as a function of FCW modality and SOA (distracted drivers)**

**Table 18. Paired comparisons of total response times between baseline events (distracted drivers)**

| Verbal | | | | | | | |
|---|---|---|---|---|---|---|---|
| Audio FCW (Baseline 2) | 150 ms | 300 ms | 1000 ms | Haptic FCW (Baseline 2) | 150 ms | 300 ms | 1000 ms |
| | 0.019* | 0.002* | <0.001* | | 0.008* | 0.002* | 0.055 |
| **Tonal** | | | | | | | |
| Audio FCW (Baseline 2) | 150 ms | 300 ms | 1000 ms | Haptic FCW (Baseline 2) | 150 ms | 300 ms | 1000 ms |
| | <0.001* | <0.001* | 0.012* | | <0.001* | <0.001* | 0.296 |

* result significant at p<.05

There was an interaction between SOA and early alert type (*p*=0.043), where for verbal early alert, distance increased as SOA increased, and leveled at 1000 ms. But for tonal early alerts distance increased as SOA increased from baseline to 150 ms, but with a decrease in distance for alerts with SOA of 1000 ms (see Figure 63). There was also an interaction between SOA and FCW type where for the audio FCW, distance increased as SOA increased, but for haptic FCW, distance increase as SOA increased, with a decrease in distance at 1000 ms (see Figure 64).

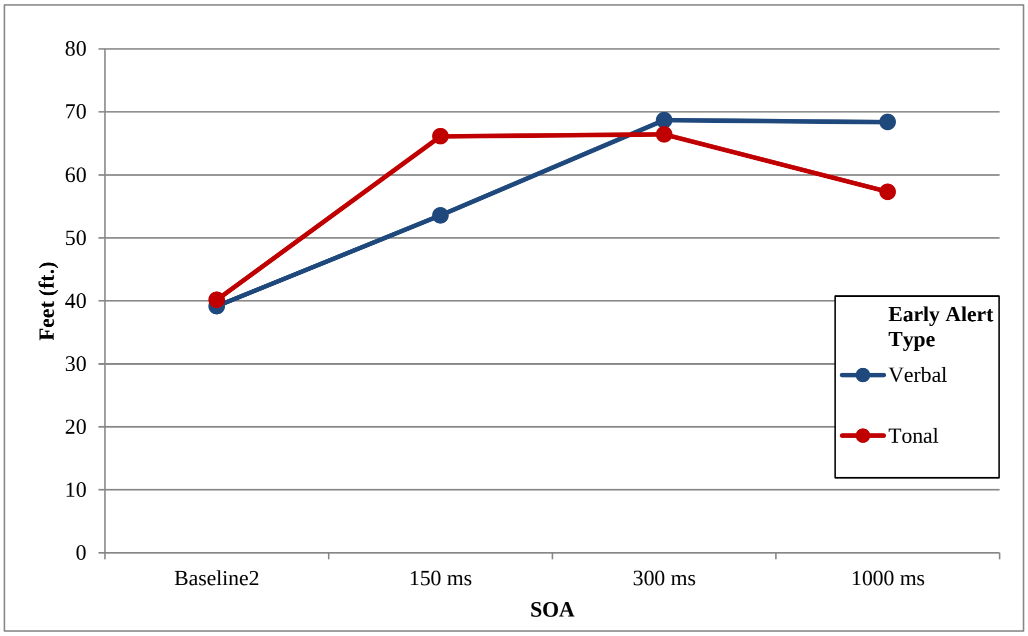


**Figure 63. Minimum distance between subject vehicle and lead vehicle as a function of early alert type and SOA (distracted drivers)**

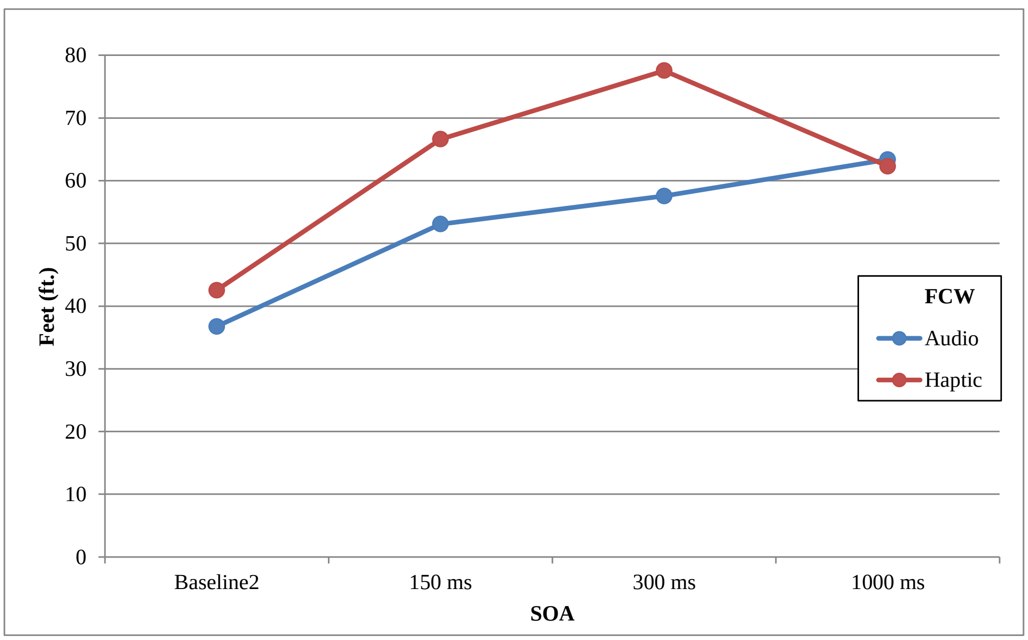


**Figure 64. Minimum distance between subject vehicle and lead vehicle as a function of FCW modality and SOA (distracted drivers)**

Paired comparisons of verbal events revealed that there were statistical differences between auditory baseline 2 events and events with an SOA of 1000 ms. There was also a significant difference between haptic baseline 2 events and events with an SOA of 300 ms for distance. When comparing tonal early alert events, there were statistical differences between auditory FCW baseline 2 events and events with SOA of 150, 300, 1000 ms. When comparing haptic FCW events there was a statistical difference between the haptic FCW baseline 2 events and events with an SOA of 150 and 300 ms (see Table 19).

**Table 19. Paired comparisons between baseline events for distance between subject vehicle and lead vehicle (distracted drivers)**

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Verbal | | | | | | | |
| Audio FCW (Baseline 2) | 150 ms | 300 ms | 1000 ms | Haptic FCW (Baseline 2) | 150 ms | 300 ms | 1000 ms |
| | 0.212 | 0.052 | 0.013* | | 0.098 | 0.004* | 0.061 |
| Tonal | | | | | | | |
| Audio FCW (Baseline 2) | 150 ms | 300 ms | 1000 ms | Haptic FCW (Baseline 2) | 150 ms | 300 ms | 1000 ms |
| | 0.001* | 0.030* | 0.006* | | <0.001* | <0.001* | 0.121 |

* result significant at $p<.05$

Discussion

This Phase 2 experiment examined how the temporal relationship of the early alerts and FCW affects how distracted drivers respond to a FCW. This study hypothesized that presenting a distracting in-vehicle task prior to a critical braking event would cause a delay in total reaction and braking reaction times, along with shortened distance between the subject vehicle and the lead vehicle.

As in the previous Phase 1 experiment the two measures of most interest were total response time and distance. Total response time was defined as the start of the lead vehicle critical braking event to the end of accelerator to brake movement. Distance was defined as the minimum distance between the subject vehicle and lead vehicle after a critical lead vehicle braking event. For both of these measures, baseline 2 (FCW with no early alert) resulted in worse performance than any of the three early alert SOA times. This finding suggests that an early alert might serve as cue or prime the driver's response in a way that improves response to an imminent subsequent warning. Similarly, the early alert might help to draw the driver's attention back to the road from the distraction. As noted in the subsequent General Discussion, however, the repeated measures design of these experiments may also have resulted in participants associating the early alerts with the possibility of a subsequent FCW.

When evaluating the total response time, the baseline 2 condition resulted in the longest total response times. Baseline 2 events resulted in slightly longer total response times in Phase 2, where participants were distracted, than in Phase 1, where participants were not distracted. Furthermore, an SOA of 150 or 1000 ms resulted in the second longest total response times followed by an SOA of 300 ms (on average for audio and haptic FCW, and tonal and verbal early alert).

Interactions between early alert type and SOA were found such that for the tonal early alert group, total response time increased at 150 and 300 ms (relative to baseline), and decreased for 1000 ms

SOA (relative to 150 and 300 ms SOA). For the verbal early alert group, total response time increased from 150 through 1000 ms SOA (relative to baseline). The tonal early alert group findings are not consistent with those from the Phase 1 experiment where the total response times slightly increased for 150 and 300 ms, followed by a decrease at 1000 ms.

There was a significant difference between FCW modes where haptic FCW events elicited shorter total response times than audio FCW events. These findings were identical to those found in Phase 1. The Phase 2 data also revealed an interaction between FCW type and SOA times where for the audio FCW events, total response time decreased as SOA increased, but for the haptic FCW event with an SOA of 1000 ms, total response time had an increase.

When distances for FCW modes were compared, there was a significant difference between audio and haptic FCW, where haptic FCW resulted in longer minimum distances. These results were identical to those found in Phase 1. There was also a significant difference between SOA times where the larger the SOA, the longer the minimum distance, except for SOA of 1000 ms. An interaction between SOA and FCW mode was also observed whereas SOA increased so did minimum distance except for events with haptic FCW and 1000 ms SOA.

The interaction between SOA times and early alert group illustrates contrasting results for each group. For the tonal early alert group, an increase of SOA resulted in an increase of minimum distance. In contrast, for the verbal early alert group, an increase of SOA resulted in a decrease of minimum distance. These findings are different from those found in Phase 1, where an increase of SOA for the verbal group resulted in a decrease of minimum distance, and an increase of minimum distance for the tonal group.

This experiment found that the total response times for baseline 2 were longer for Phase 2 (distraction and FCW only) than for Phase 1 (FCW without distraction), which suggests that the distraction was effective in shifting participants' attention away from the lead vehicle braking event. This finding was supported by eye glance data, which showed that the distractor task did as intended and kept participants engaged in a task before the critical lead vehicle braking event. Although glances to the road were recorded during this period, eye glance data showed that participants kept most of their glances to the LCD monitor. A closer look at eye glances at FCW alert onset revealed that most participants were looking at the LCD monitor when the critical event occurred.

This experiment also found that having participants distracted during lead vehicle braking events resulted in an increase in collisions. There were 11 baseline 2 collisions in Phase 2 whereas there were only two baseline 2 collisions in Phase 1. For the tonal early alert group, the total response times for the 150 and 300 ms SOA were shorter compared to the Phase 1 results. Minimum distance was also affected by distraction in Phase 2 where it was shorter in comparison to Phase 1.

### 5.1.3 General Discussion

The two experiments on near-simultaneous safety alerts and crash warnings found significant effects of SOA intervals on driver response to FCW events. The effects, however, were complex, interacted with characteristics of the early alert and the FCW modality, and differed from the findings observed in other contexts. In this pair of experiments, early safety alerts often facilitated FCW response, even for short SOA intervals.

The findings from the Phase 1 experiment (undistracted drivers) tend to suggest that having a non-urgent, safety-related early alert prior to a FCW during a lead vehicle braking event was a benefit most of the time, as compared to not having an early alert present. The only times this was not the case was when a tonal early alert was presented at 150 ms SOA. It should be noted that in doing a quick review of foot behavior for non-critical events (events which set off the early alerts without an FCW), data showed that participants tended to either release the accelerator or move their foot from accelerator to brake after the early alert more often times than not. This may suggest that the early alerts served as a type of pre-alert or pre-cue for participants in this study. Further analysis would have to be performed to determine how often this occurred for early alert-only events (non-critical events).

If considering an FCW system in a vehicle, these results tend to support a haptic FCW over an auditory FCW in producing shorter response times and longer minimum distances during lead vehicle braking events, with the caveat that only a limited set of modes and alert characteristics were tested.

When combining an FCW system with an earlier non-urgent safety related warning, these results suggest that a combination of haptic FCW and verbal early alert would enhance response time and thus improving braking response. These findings support Pashler's (1994) finding that alerts with similar perception modes would interfere.

In this experiment the FCW and early alerts were presented to the participant without distraction. The next phase of this study investigated whether results change when participants are given a distraction task before an early alert and FCW are presented.

Participants in the Phase 2 experiment (distracted drivers) had slower total response times than those in Phase 1 for baseline 2 trials, but faster total response times when an early alert preceded a FCW. Video data shows that several Phase 2 participants reacted to the distractor task audio cue in apparent anticipation of a critical lead vehicle braking event. These results suggest that participants may have associated the distractor task with the likelihood of a lead vehicle braking event as a result of this experiment's repeated measures design.

If considering an FCW system in a vehicle, the results from Phase 2 continue to support a haptic FCW over an audio FCW in producing shorter response times and longer minimum distances during lead vehicle braking events regardless of whether participants are distracted or not

In contrast with the findings of Wiese and Lee, a SOA of 300 ms caused less temporal conflict and less of an effect on the driver's performance. Overall, SOA of 300 ms resulted in shorter total response times and longer minimum distances. The total response times for an SOA of 1000 ms were also similar to those that Wiese and Lee found where participants had slower brake reaction times than those with 350 ms (300 ms for this study) but not slower than those with 0 or 150 SOA.

As a result of the distraction task, participants in Phase 2 had more collisions, longer total reaction times, and shorter minimum distances for baseline events than they did in Phase 1. In Phase 1, having an SOA of 1000 ms resulted in quicker total response times than other SOAs, but this was not the case in Phase 2.

# 6 Protocols for FCW and LDW Interfaces

A key product of this project is a set of preliminary evaluation protocols for determining the acceptability of DVIs for FCW and LDW systems. As noted earlier in this report, the purpose of the protocols is to define well-specified, repeatable procedures that will allow a common method for evaluating the DVI of a production or near-production system. The intent is not to assess a vehicle's warning system, per se, but specifically the driver interface portion of that system. Other NHTSA programs, such as the Advanced Crash Avoidance Technologies (ACAT) program, have had the specific objective of modeling and predicting the safety impact of new crash avoidance technologies (e.g., Funke, Srinivasan, Ranganathan, & Burgett, 2012). The ACAT program, however, did not include specific consideration of the DVI. The protocol efforts under the CWIM project were intended to evaluate the effectiveness of the DVI in engendering rapid and appropriate driver responses, regardless of other aspects of the system.

Although the CWIM project directed empirical research effort toward a number of methodological aspects of the protocols, it was not possible to address every aspect of the protocol through direct experimental testing. Furthermore, the findings from the experiments were sometimes complex and did not lend themselves to easily resolving procedural options. The protocols developed in this project therefore represent a combination of direct empirical evidence from project research, other literature, and expert judgment. Because the protocols have not yet been subjected to industry or other outside review, and have not yet been exercised by outside organizations, the research team characterizes these specifications as preliminary.

## 6.1 Overview of Protocols

Three separate, though related, protocols were developed: An FCW driving simulator protocol; and LDW driving simulator protocol; and a FCW test track protocol. The protocols themselves are provided as a separate deliverable product of this project, under separate cover. This section describes the general format and content of the protocols.

The three protocols share a common broad approach, all being derived from a basic procedure reported in Forkenbrock et al. (2011). The participant is not aware that the purpose of the testing is to evaluate warning systems or that potential crash events might occur. Periodically during the driving task, the participant is also instructed to engage in a secondary task. This secondary task requires the participant to monitor and remember five numbers sequentially displayed on a screen located about 90 degrees to the driver's right. The secondary task is intended to direct the driver's line of sight away from the forward roadway for a sufficient period so that the potential crash event (rear end collision for FCW, lane departure for LDW) can unfold without visual awareness, but not so long that participants will look back at the road before the task is completed. In the case of the number recall task used by Forkenbrock et al. and in the present research, this time is at least 1.32 s (time from first number appearing on the task display to FCW issuance). After a number of times of engaging in the secondary task with no incident, the potential crash event occurs during the task. Some participants experience the warning associated with the DVI being tested. Control participants receive no warning. Various measures of driver response to the warning are collected for analysis, allowing the effectiveness of the DVI warning to be compared in a statistically meaningful manner with the response of the control group.

Although this general method provides a common approach to evaluating the DVI, numerous procedural details need to be specified if the procedure is to be repeatable by a range of organizations that may engage in testing. The protocols spell out these specific details, with supporting rationale. Each of the three protocols includes the following sections:

- Participants: age, gender, sample size, inclusion/exclusion criteria
- Method: ruse, recruitment, incentive, training
- Apparatus: simulator/test track, secondary task, headway display (if applicable)
- Simulator/test track scenario: event timing and choreography, driving environment, practice and exposure to DVI
- Post data collection procedure: video coding, secondary task interaction
- Data handling: dependent measures, data reduction, data analysis

A number of the requirements for a common protocol method are complex and may not be resolvable by empirical data alone. For example, the characteristics of the participant sample could be subject to argument on a variety of grounds: inclusiveness; stability of performance; representativeness of the general driving public; representativeness of the likely consumer basis; inclusion/exclusion of high risk driving groups; etc. Likewise, the degree of familiarity that participants might have with the technology, either coming into the study or through training/exposure during the study, is debatable on a number of grounds. Since the protocols require specificity, explicit requirements for such factors are provided in the documents as the best judgments of the authors, with supporting rationale. It is recognized that there may be a variety of opinion on such matters that may require further consensus. Likewise, there are aspects to the procedures that are imperfect and also reflect the philosophy of the developers. For example, there is no straightforward choice of a secondary (distracting) task. The findings of an evaluation may well depend on the nature of the distraction as well as upon the DVI itself. The number recall task described in the protocol has a number of features the developers considered desirable. It was visually and cognitively engaging. It had the ability to retain the participant's visual attention for an adequate period of time. It represented a realistic worst case in drawing the eyes well away from the forward road, limiting peripheral vision (especially useful for the LDW application); it resulted in postural shifts that made lane drifts appear to be self-induced, which was particularly important for the LDW DVI protocol. It was highly structured to allow specific and repeatable timing of driver attention and easy replication by any researchers. The number recall task also has some limitations. It remains difficult to consistently get participants to visually commit to the task for a duration long enough to allow the potential crash event to unfold. This results in some level of broken trials and participant loss, requiring replacement. It also requires careful monitoring of participant behavior in the analysis of video records, in order to confirm compliance with the visual task. Thus there are costs in terms of additional sessions and data reduction. These examples serve to illustrate the potentially controversial nature of specific protocol requirements, no matter what the choice of options may be.

Subsequent activities would be valuable in refining these preliminary protocols and ensuring their acceptance and consensus. One step would be to receive critical feedback from a range of stakeholders, including the automobile industry, suppliers, transportation safety advocacy groups, and researchers. This should also include government agencies and offices that may make use of protocol findings, for purposes of regulation, rulemaking, compliance, or public information. Another important step would be for the protocols to be thoroughly exercised to confirm the ability of others to successfully employ the procedures and to confirm the consistency of the findings. Such an exercise should include a range of different types of testing organizations. It should also include a diverse array of DVIs, both traditional and emerging or novel, to confirm that the methods work across display modes (auditory, tactile, visual), locations, and specific display aspects. It should be noted, however, that because the number recall secondary task requires drivers to direct their gaze at least 90 degrees to their right, this method is not appropriate for the explicit testing of visual displays outside of the distracted driver's field of view. Additional work may also be done to define performance benchmarks, in addition to the basic comparison of a given DVI with a "no warning" control. If protocol outcomes are to be used for public information purposes (e.g., NCAP or similar ratings), work should also be conducted on how to best categorize and present the findings to the public.

## 6.2 Protocols (available as separate documents)

The protocols themselves are provided as separate deliverables under this project. The documents are:

- Forward Crash Warning Simulator Protocol (Timothy Brown & Dawn Marshall)
- Lane Departure Warning Simulator Protocol (Timothy Brown & Dawn Marshall)
- Forward Crash Warning Test Track Protocol (Timothy Brown & Dawn Marshall)

# 7 Implications for the Design and Implementation of Crash Warning DVIs

Much of the research conducted under the CWIM project addressed methods for *evaluating* a DVI (for FCW and LDW applications), rather than design of the DVI. However, a number of implications for ACWS DVI design were drawn from the findings of the various experiments under this project. Experiments conducted under the CWIM project generally addressed lines of inquiry that have received little direct research attention in the past.

This section presents implications for DVI design and implementation. These are cast as implications rather than explicit guidelines or design principles because they are derived specifically from the findings of the CWIM project. An objective of this project was to support the development of design principles, which should incorporate and integrate findings from the broader literature. Currently, such principles are under development in other NHTSA programs (e.g., Campbell et al., in preparation). The findings and implications of the CWIM project are presented here in support of this and other efforts at formal DVI design development.

<u>Important Note</u>: **The research implications provided here are not intended to serve as explicit recommendations, guidelines, or design principles. For much of the research conducted under the CWIM project, design implications were not a primary goal, so the research was not designed to comprehensively compare alternatives and the implications presented are therefore drawn from research that compared a limited set of alternative design options. As such, comparisons and conclusions can only be drawn from among the options investigated in this research. Also, as noted earlier in this report, the CWIM project focused specifically on the DVI rather than the ACWS as a whole. Therefore, DVI effectiveness, as defined in this report, does not directly refer to crash reduction or mitigation. Rather, effectiveness refers to the ability of the DVI to influence attention, comprehension, or behaviors in a way that is appropriate for the hazard context.**

The implications that follow are grouped under the following topic headings:

- Compatibility of auditory alert features with intended message meaning categories
- Maintenance of auditory alert meaning under in-vehicle noise conditions
- Haptic signal applicability for FCW
- LDW alert modality
- Warning signal response disruption by preceding non-urgent alerts
- Conveying information on the presence and status of ACWS

## 7.1 Compatibility of Auditory Alert Features with Intended Message Meaning Categories

<u>Design Goal</u>: Assure that auditory crash warnings are inherently perceived as urgent alarms and are not confused with other less urgent message categories that might be presented by the vehicle or other mobile devices. Urgent crash warning signals should remain distinct from other categories of information and convey high urgency without prior exposure or learning.

CWIM project findings:

- Auditory signals were most likely to be perceived as urgent crash warnings if they had:
    - a base frequency of 1000 Hz or higher but less than 2500 Hz.
    - a peak to total time ratio of .7 or higher.
    - at least 3 harmonic components.
    - a perceived tempo of 125 ms or less but greater than 15 ms between sound components (tempo is primarily determined by the duration between multiple bursts (IBI) but is also influenced by IPI).
- Auditory signals were most likely to be perceived as urgent crash warnings if all of the above criteria were met.
- Auditory signals that met none of the four criteria above were most likely to be perceived as distinct from urgent crash warnings.
- Auditory signals related to vehicle status, but not requiring urgent response, were most distinguishable from urgent warnings and from social notifications if they met only the ratio criterion.

Rationale and Discussion:

Although design principles are available for auditory warning signals for passenger vehicle applications (e.g., Campbell et al., in preparation), the research literature underlying these principles has generally not consisted of a systematic examination designed to determine the most important parameters and their value ranges that ensure that a given sound will be inherently perceived and responded to as a critically urgent warning within the context of a vehicle. Research conducted under the CWIM program examined a broad set of in-vehicle auditory signals obtained from samples currently existing OEM alarms, alerts, and notifications, as well as laboratory-derived signals based on published literature. A wide range of parameters was systematically tested, including but not limited to base frequency, highest frequency, number of harmonics, IBI, IPI, onset and offset time, total duration, nonpeak alert duration, peak duration, pulse duration, burst duration, duty cycle, and peak to total time ratio. Figure 43 (within Section 4 of this report) clarifies the terminologies used for the sound components.

A variety of methods used in a series of investigations provided converging evidence for the conclusions listed above. In a set of three investigations, participants were asked to listen to sounds and sort them into perceptual categories of “alarm-like,” “status notifications” (vehicle status, e.g., low windshield wiper fluid), or “social notifications” (e.g., email, Facebook). Each sound was characterized by its sound components and these components were entered into backward and stepwise binary logistic regression analyses to identify the sound components that accounted for a significant amount of variance in whether or not participants perceived a given sound as alarm-like. These analyses revealed the criteria identified above and sounds that met the four criteria were perceived as alarm-like by over 90% of respondents. In a second set of experiments, participants were asked to use the method of adjustment for the key parameters of base frequency, IBI, pulse duration, and number of pulses per burst either individually (in one experiment) or together (in a separate experiment). Results of these investigations provided converging evidence for the conclusions listed above.

Participant responses to a variety of auditory signals were evaluated while driving in a simulator and the findings confirmed the ability to predict message meaning based on the criteria from the preceding experiments. Sounds meeting all four of the listed criteria were classified as alarms over

90% of the time while sounds meeting 0, 1, or 3 criteria were classified as alarms less than 10, 20, and 65% of the time, respectively. Figure 49 within Section 4 shows this relationship. Further, sounds meeting all four criteria resulted in significantly faster brake press responses than did sounds not meeting all criteria.

The conclusions provided here are based on convergent findings from experiments using distinct methods. However, it should also be noted that this research comes from a single laboratory and is based on a limited set of acoustic signals. Furthermore, it would be helpful to validate the findings under actual vehicle driving conditions and with actual behavioral responses to unexpected signals while driving.

## 7.2 Maintenance of Auditory Alert Meaning Under In-Vehicle Noise Conditions

Design Goal: Assure that auditory crash warnings maintain their effectiveness under a range of anticipated in-cab ambient noise conditions. Urgent crash warning signals should remain distinct from other categories of information and convey high urgency under all listening conditions. Section 7.1 addressed the characteristics of sounds that contribute to interpretations of urgency. This section complements those findings by addressing message detectability and interpretation under elevated ambient noise conditions.

CWIM project findings:

- Urgent crash avoidance warnings for passenger vehicles presented at 65 dB(A) were often not detected in conditions of elevated noise (i.e., open windows or music playing), while warnings presented at 75 dB (A) were detected at high rates and better maintained their perceived urgency. These findings suggest that a value between 65 and 75 dB(A), such as 70 dB(A), might be an appropriate compromise. This hypothesis was tested in a follow-up laboratory experiment which is described in a separate report (Singer, Lerner, & Kellman, 2015)
- Urgent crash warnings with significant frequency components higher than 1500 Hz were most easily detected in the vehicle environment and were perceived as more urgent than messages lacking significant frequency components in this range.
- Urgent crash warnings with multiple dominant frequency components were most easily detected in the vehicle environment and were perceived as more urgent than messages of a single dominant frequency.
- Voice messages were less intelligible under noisier conditions and may need to be supplemented with tonal elements if used to present urgent warning information.

Rationale and Discussion:

Although design principles are available for auditory warning signals for passenger vehicle applications (e.g., Campbell et al., in preparation), the empirical research literature underlying these has generally not considered the range of likely in-cab noise conditions. Research conducted under the CWIM program provided unique driver perceptual data for two higher-noise conditions: listening to music and having the front windows lowered. Although this single study could include only a limited number of vehicle types (3), driving conditions (60 mph asphalt-surface freeway), alerting signals (15), and ambient noise conditions (3), several findings emerged.

Differences in findings among passenger vehicles (small car, sedan, SUV) were minor. Ambient noise levels in all vehicles averaged approximately 65 dBA in the baseline (quiet cab) condition. The various auditory signals were equated for perceived loudness (under quiet conditions) to 65 dB pink noise. Although equally loud in this sense, signals differed substantially in how noticeable they were under noise conditions and how urgent they appeared. Under higher noise conditions, 65 dB alerts typically lost perceived urgency and in many cases failed to even be detected. A subset of alerts presented at 75 dB maintained perceived urgency and meaning fairly well. This study did not examine any sound levels between 65 and 75 dB. Based on these findings, signals that meet or exceed 70 dB are likely to retain effectiveness in noise. Many OEM warning signals in recent vehicle models, however, are presented at levels below 70 dB (Lin & Green, 2013). Recommendations in the literature vary, with some sources suggesting that alerts be 10-15 dB above the masked threshold (International Organization for Standardization, 2005; Lee et al., 2004) and others suggesting that alerts be at least 20 dB above the masked threshold (e.g., Campbell, Richman, Carney, & Lee, 2004). The CWIM data do not suggest that all alerts should always be presented at 70 dBA or more. Most sounds were detected easily at 65 dBA in the baseline (no elevated noise) condition. Some sounds were more easily detected and interpreted at 65 dBA than others under noise conditions. It may be possible for vehicle systems to detect current ambient noise conditions and present signals at an appropriate SPL, or to mute sources of noise such as the vehicle's radio.

Not all auditory signals were equally capable of retaining their effectiveness (detection, meaning) under noise conditions. Based on comparison of characteristics among the set of sounds tested, it appeared that those with predominant frequencies below 1000 Hz performed worst, and those with primary or significant components at or above 1500 Hz and with multiple peaks performed best.. Similarly, the "Method of Adjustment - Prototype Development" experiment found that 85 percent of participants identified a sound as an urgent crash warning at 1900 Hz. Based on the findings from these two experiments, 1000 Hz may generally be considered the low end of the effective base frequency range, though higher base frequencies and/or harmonics, as well as other characteristics of the sound, may also influence interpretation.

Although only a few voice messages were included in the set of alerts tested, it appeared that voice warnings may be more susceptible to interference by ambient noise. Design principles from Campbell et al. (2007) indicate that caution should be used in using speech messages for in-vehicle warnings and that preceding a voice warning with an alerting tone should not be done "unless a benefit for doing so can be demonstrated."

## 7.3 Haptic Signal Applicability for FCW

Design Goal: Use appropriate haptic signals for FCW.

CWIM project findings:

- Brake pulsing or seat belt tensioning were found to be effective for returning distracted drivers' attention back to the forward roadway and eliciting desirable vehicle control responses; seat vibration similar to a virtual rumble strip (vibrating the front of the seat) was not found to rapidly and reliably return driver attention to the forward roadway within the CWIM research. It is important to note, however, that only one example of each haptic signal was used, so it is not appropriate to draw conclusions about the most effective or appropriate haptic warning signals from this finding.

Rationale and Discussion: Various CWIM project experiments related to protocol development and temporal message interference included warnings in different modalities. In both cases, haptic warnings were found to be effective, compared to auditory warnings. However, not all example haptic signals were equally effective in this research. Brake pulsing and seat belt tensioning were effective in multiple experiments, but research conducted in the NADS-1 driving simulator found that seat belt tensioning and brake pulsing were more effective than seat vibration. Therefore, this research implies that brake pulsing and seat belt pretensioning can be effective alerts. The absence of a strong benefit for seat vibration should not be taken to imply that this cannot be effective as well; only that the specific implementation used in this study found no such evidence.

These findings are preliminary, in that they are based on limited experiments and specific implementations of each signal type. No systematic parametric research was done to determine the most effective version of each signal type. While CWIM experiments found weaker observed benefits of seat vibration, it may be the case that more effective implementations of seat vibration could be designed. Previous research suggests that seat pan vibration can be effective for the presentation of warnings (e.g., Fitch, Hankey, Kleiner, & Dingus, 2011). These findings should be considered within the broader context of haptic warning design guidance in Campbell et al. (2007) and anticipated in Campbell et al. (in preparation).

## 7.4 LDW Alert Modality

Design Goal: Assure that the DVIs for LDW encourage drivers to return their attention to the forward roadway and to produce a reduction in risk associated with being out of lane.

CWIM project findings:

- A seat vibration that is directional and represents either a center line or edge line rumble strip was not found to be effective in returning the driver's attention back to the road or decreasing the duration of the lane departure within the CWIM research.
- Auditory alerts were found to be effective for decreasing the duration and extent of the lane departure.
- Steering wheel vibration similar to that used in the NADS CWIM studies was found to be effective for decreasing the duration and extent of the lane departure.
- Seat belt tensioning was found to be effective for decreasing the duration and extent of the lane departure.

Rationale and Discussion:

Research under the CWIM program considered the effectiveness of DVIs for LDW systems in the context of refining an evaluation protocol. None of the warning DVIs produced a statistically significant reduction in time to end of visual commitment (eyes back on the road).

Despite the lack of effect for returning gaze back to the road, there were differences in outcome measures based on the DVI of the alert. Auditory, seat vibration, steering wheel vibration, and steering torque were associated with a significantly smaller duration of lane exceedance and area of exceedance than when no warning was provided, but were not different from each other. When maximum lateral extent of the lane departure was considered, auditory, steering wheel vibration, and steering torque were all different from no warning, but not from each other. Seat vibration was not different from any of the other conditions.

It should be noted that these findings are based on a single, though typical, exemplar of each type of warning. It is conceivable that more, or less, effective DVIs could be designed for each. As noted previously, these findings should not be construed as generalizable to all signals of the same modality.

## 7.5 Warning Signal Response Disruption by Preceding Non-Urgent Alerts

Design Goal: Design crash warnings and potentially competing messages to minimize message interference with rapid driver response to crash warnings. [*Note: the findings below are based on limited and complex data and are viewed as highly tentative*]

CWIM project findings:

- Non-urgent messages that closely precede an urgent crash warning do not necessarily impair response to the warning, assuming the non-urgent message is terminated when an urgent crash warning is triggered.
- Intervals of at least 1 second between alerts may result in the lowest likelihood of interference.
- Interference was minimized when messages preceding an FCW were presented in different modalities than the FCW.

Rationale and Discussion: CWIM research on near-simultaneous safety alerts and crash warnings found that non-urgent messages that shortly preceded FCWs did not necessarily slow responding to the FCW. In many cases, the preceding message actually improved the speed of returning attention to the road or initiating braking in response to the FCW. However, observed effects were complex and depended on the mode of the warning, the mode of the preceding alert, and the duration of the interval between the two signals. The repeated-measures design of the experiments may also have influenced how participants responded to early alerts and FCWs. Haptic FCW signals led to faster responding than auditory signals and this relative advantage was generally more pronounced when there was only a brief (150 ms or 300 ms) interval between the preceding alert (verbal or tonal) and the FCW. When the FCW was auditory (pulsing tone), responding was generally faster when the preceding alert was verbal rather than tonal at the briefer (150 ms or 300 ms) intervals.

These relative performance results provide some support for theories of resource competition, whereby tasks that share the same perceptual and/or response mode compete more than when these modes are distinct. However, the anticipated finding of absolute decrements in performance due to preceding tasks at brief intervals (150 ms, 300 ms) was not observed. Performance was generally enhanced, relative to no preceding message or to shorter intervals, when the interval was 1 second. This is consistent with findings from previous research.

The results of the experiments on near-simultaneous safety alerts and crash warnings suggest that "competing" messages that occur shortly before an FCW do not necessarily interfere and may even serve to prime the response to the FCW. However, the findings from these experiments are complex and it is conceivable that some results could be artifacts of the experiment design (e.g., repeated measures of FCW events) or specific to certain interface aspects (e.g., use of icons to accompany verbal or tonal preceding messages). Also, the experiments employed only limited examples of message modes, content, and timing. Therefore, the findings should be treated as quite tentative. However, given the frequent assumption that non-urgent alerts that precede an urgent crash warning will necessarily be interfering, the implications of the present findings bear highlighting.

## 7.6 Conveying Information on the Presence and Status of ACWS

Design Goal: Assure that the DVI clearly conveys to drivers the presence of particular crash warning and driver assist functions in the vehicle and indicates the current operational status of such systems.

CWIM project findings:

- The driver interface should provide an intuitively meaningful indication of the presence of a warning or assist function in the vehicle and its current status.
- Providing owner's manuals, quick start guides, or similar materials to convey information about the presence of a function or to explain the meaning of interface display elements resulted in only minor increases in comprehension.
- Clearly labeled function buttons were preferable to hierarchical menu structures.
- Understandable icons or full words were more effective at identifying functions; acronyms were generally not effective, particularly if they are not commonly used.
- Descriptive terms were clearer to driver than terminology or acronyms that are unique to the manufacturer.
- Status displays in readily viewable locations that were consistent with user expectations were most easily identified and understood.
- Color coding in a manner consistent with population stereotypes (e.g., green means "on" or "functional") aided comprehension of system status.

Rationale and Discussion: In order for crash warnings and other driver assist systems to be optimally effective, the driver of a particular vehicle should be aware of the presence (or absence) of a particular function in that vehicle and the current status of that system (on or off, fully functioning, etc.). CWIM NADS driving simulator research conducted on driver familiarity (in support of protocol development) found that simply informing the driver that a function (FCW, LDW) was present in the vehicle led to improved driver response to the warning. On the other hand, if a driver mistakenly believes a function is present in a vehicle, or that it is currently operational when in fact it is not, there may be unwarranted reliance on the vehicle system. As part of the CWIM program, research was conducted on driver comprehension of several actual in-vehicle status displays (Robinson et al., 2011). This study observed substantial problems with participants' understanding of system function availability and operational status. Furthermore, having participants read relevant sections of owner's manuals or quick start guides in advance of the experimental session provided only limited benefits; significant comprehension problems still remained. Variability among vehicles was a further problem, as evidenced by the fact that familiarity with one vehicle's system (through the owner's manual/quick start guide) did not provide any measureable benefit over no familiarity when the participant was confronted with a different vehicle. Participants had difficulty interpreting acronyms and idiosyncratic terminology used by particular manufacturers. The application of color coding was sometimes inconsistent with user expectations and general human factors guidelines. Displays related to the presence or status of warning and assistance functions were sometimes located in areas where viewing and reading were difficult. Display locations were often inconsistent with user expectations and with general human factors principles regarding location and grouping of display elements. Distinctions between messages indicating whether the system was turned on or off and those indicating operational status (e.g., system capable of providing warnings or not due to malfunction or certain criteria not met) were sometimes difficult and based on subtle cues.

# References


Alliance of Automobile Manufacturers (2006). *Statement of principles, criteria and verification procedures on driver interactions with advanced in-vehicle information and communication systems*. Washington, D.C.: Alliance of Automobile Manufacturers.

Brown, T., Schwarz, C., & Marshall, D. (2011). *Evaluating forward crash warning on the NADS for CWIM* (N11-001). Iowa City, IA: National Advanced Driving Simulator, The University of Iowa.

Campbell, J., Brown, J., Graving, J., Richard, C., Lichty, M., Sanquist, T., Bacon, L., Woods, R., Li, H., & Williams, D. (in press). *Human Factors for Connected Vehicles (HFCV) Driver-Vehicle Interface (DVI) Design Principles*. Washington, DC: National Highway Traffic Safety Administration.

Campbell, J. L., Richard, C. M., Brown, J. L., & McCallum, M. (2007). Crash warning system interfaces: Human factors insights and lessons learned. (Report No. DOT HS 810 697). Washington, DC: National Highway Traffic Safety Administration.

Campbell, J. L., Richman, J. B., Carney, C., & Lee, J. D. (2004). In-vehicle display icons and other information elements, Volume I: Guidelines. (Report no. FHWA-RD-03-065). McLean, VA: Federal Highway Administration.

Dingus, T. A. & Klauer, S. G. (2008) *The relative risk of secondary task induced driver distraction*. (Paper no. 2008-21-0001). Warrendale, PA: SAE International.

Finley, M. D., Funkhouser, D. S., & Brewer, M. A. (2009). *Studies to determine the operational effects of shoulder and centerline rumble strips on two-lane undivided roadways*. Texas Transportation Institute.

Fitch, G. M., Hankey, J. M., Kleiner, B. M., & Dingus, T.A. (2011). Driver comprehension of multiple haptic seat alerts intended for use in an integrated collision avoidance system. *Transportation Research Part F: Traffic Psychology and Behaviour, 14* (4), 278–290.

Forkenbrock, G., Snyder, A., Heitz, M., Hoover, R. L., O'Harra, B., Vasko, S., & Smith, L. (2011). *A test track protocol for assessing forward collision warning driver-vehicle interface effectiveness*. (Report no. DOT HS 811 501). Washington, DC: National Highway Traffic Safety Administration.

Funke, J., Srinivasan, G., Ranganathan, R., & Burgett, A. (2012). Safety impact methodology (SIM): Application and results of the Advanced Crash Avoidance Technologies (ACAT) Program. Paper presented at 2012 SAE Government/Industry Meeting, Washington, DC.

Gray, R. (2011). Looming auditory collision warnings for driving. *Human Factors, 53(1),* 63-74.

Green, P., Sullivan, J., Tsimhoni, O., Oberholtzer, J., Buonarosa, M.L., Devonshire, J., & Sayer, J. (2008). *Integrated vehicle-based safety systems (IVBSS): Human factors and driver-vehicle interface (DVI) summary report*. (Report No. DOT HS 810 905). Washington. DC: National Highway Traffic Safety Administration.

Haigney, D., & Westerman. S. J. (2001). Mobile (cellular) phone use and driving: a critical review of research methodology, *Ergonomics, 44(2),* 132-143.

Hibberd, D. L., Jamson, S. L., & Carsten, O. M. J. (2010) Managing in-vehicle distractions: evidence from the psychological refractory period paradigm. In: *Proceedings of the 2nd International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '10).* New York, NY.

Hibberd, D. L., Jamson, S. L., & Carsten, O. M. 2013), Mitigating the effects on in-vehicle distractions through use of the psychological refractory period paradigm, *Accident Analysis and Prevention, 50,* 1096-1103.

Ho, C., Reed, N., & Spence, C. (2007). Multisensory in-car warning signals for collision avoidance. *Human Factors, 49(6),* 1107-1114.

IEC 60601-1-8 ED. 2.0 B: 2006. Medical electrical equipment - Part 1-8: General requirements for basic safety and essential performance - collateral standard: General requirements, tests and guidance for alarm systems in medical electrical equipment and medical electrical systems. American National Standards Institute.

International Organization for Standardization. (2005). Road vehicles: Ergonomic aspects of in-vehicle presentation for transport information and control systems: Warning systems (ISO/TR 16352).

Kramer, A. F., Cassavaugh, N., Horrey, W. J., Becic, E., & Mayhugh, J. L. (2007). Influence of age and proximity warning devices on collision avoidance in simulated driving. *Human Factors, 49(5),* 935-949.

Lee, S. E., Knipling, R. R., DeHart, M. C., Perez, M. A., Holbrook, G. T., Brown, S.B., Stone, S. R., & Olson, R. L. (2004). *Vehicle-based countermeasures for signal and stop sign violations: Task 1. Intersection control violation crash analyses. Task 2. Top-level system and human factors requirements.* (Report no. DOT HS 809 716). Washington, DC: National Highway Traffic Safety Administration.

Lees, M. N. & Lee, J. D. (2007) The influence of distraction and driving context on driver response to imperfect collision warning systems, *Ergonomics, 50(8),* 1264-1286.

Lerner, N. D., Jenness, J., Robinson, E., Brown, T. L., Baldwin, C. L., & Llaneras, R. (2011). *Crash Warning Interface Metrics.* (Report no. DOT HS 811 470a). Washington, DC: National Highway Traffic Safety Administration.

Levy, J., Pashler, H., & Boer, E. (2006), Central interference in driving – Is there any stopping the psychological refractory period, *Psychological Science, 17(3),* 228-235.

Lewis, B. A., Penaranda, B. N., Roberts, D. M., & Baldwin, C. L. (2013). Effectiveness of bimodal versus unimodal alerts for distracted drivers. Paper presented at the *International Conference on Driving Assessment*, Bolton Landing, NY.

Lin, B., & Green, P. (2013). *Measurements of driver-assistance warning-signal characteristics in 2013 cars.* (Report No. UMTRI-2013-03). Ann Arbor, MI: University of Michigan Transportation Research Institute.

Pashler, H. (1994), Dual-task interference in simple tasks - data and theory, *Psychological Bulletin, 116(2),* 220-244.

Robinson, E., Lerner, N., Jenness, J., Singer, J., Huey, R., Baldwin, C., Kidd, D., Roberts, D., & Monk, C. (2011a). *Crash Warning Interface Metrics task 3 report: Empirical studies of effects of DVI variability.* (Report No. DOT HS 811 470b). Washington, DC: National Highway Traffic Safety Administration.

Robinson, E., Lerner N., Jenness, J., Singer, J., Huey, R., Baldwin C., Kidd, D., Roberts D., & Monk, C. (2011b). *Crash Warning Interface Metrics: Task 3: Report appendices.* (Report No. DOT HS 811 470c). Washington, DC: National Highway Traffic Safety Administration.

Santangelo, V., & Spence, C. (2007). Multisensory cues capture spatial attention regardless of perceptual load. *Journal of Experimental Psychology: Human Perception and Performance, 33*(6), 1311-1321.

Singer, J., Lerner, N., Walrath, J., & Gill, M. (2014). *Warning and message perception under ambient noise conditions: On-road experiment.* Washington, DC: National Highway Traffic Safety Administration.

Sugimoto, Y., & Sauer, C. (2005), Effectiveness estimation method for advanced driver assistance system and its application to collision mitigation brake system. (Paper No. 05-0148-O).

In *Proceedings of the 19th International Technical Conference on the Enhanced Safety of Vehicles (ESV)*. June 6-9, 2005. Washington D.C.

Wiese, E. E., & Lee, J. D. (2004), Auditory alerts for in-vehicle information systems: The effects of temporal conflict and sound parameters on driver attitudes and performances. *Ergonomics, 47(9),* 965-986.

Wickens, C. D. (1984). *Processing resources and attention.* In R. Parasuraman & R. Davies (Eds.), *Varieties of Attention.* New York: Academic Press.

Wickens, C. D. (2002), Multiple resources and performance prediction, *Theoretical Issues in Ergonomics Science*, 3(2), 159-177.

Wickens, C. D. (2008). Multiple resources and mental workload. *Human Factors, 50(3)* 449-455.

Van Auken, R. M., Zellner, J. W., Chiang, D. P., Kelly, J., Silberling, J. Y., Dai, R., Broen, P. C., Kirsch, A. M., & Sugimoto, Y. (2011). *Advanced Crash Avoidance Technologies (ACAT) program final report of the Honda-DRI team: Volume I: Executive summary and technical report.* (Report No. DOT HS 811 454a). Washington, DC: National Highway Traffic Safety Administration.

**DOT HS 812 208**
**November 2015**

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

www.nhtsa.gov

11709a-110515-v2