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

NHTSA

DOT HS 812 385 March 2017

Urgency Coding Validations

Final Report

DISCLAIMER

Suggested APA Format Citation:

Lewis, B. A., Eisert, J. L., Baldwin, C. L., Singer, J., & Lerner, N. (2017, March). *Urgency coding validations* (Final report. Report No. DOT HS 812 385). Washington, DC: National Highway Traffic Safety Administration.

Technical Report Documentation

1. Report No. DOT HS 812 385	**2. Government Accession No.**	**3. Recipient's Catalog No.**
4. Title and Subtitle Urgency Coding Validations		**5. Report Date** March 2017 **6. Performing Organization Code**
7. Author(s) Bridget A. Lewis, Jesse L. Eisert, Carryl L. Baldwin, Jeremiah Singer, & Neil Lerner		**8. Performing Organization Report**
9. Performing Organization Name and Address Westat, Inc. 1600 Research Boulevard Rockville, Maryland 20850-3129		**10. Work Unit No. (TRAIS)** **11. Contract or Grant No.** DTNH22-11-D-00237
12. Sponsoring Agency Name and Address National Highway Traffic Safety Administration 1200 New Jersey Avenue SE. Washington, DC 20590		**13. Type of Report and Period Covered** Draft final report / December, 2013 – December, 2015 **14. Sponsoring Agency Code**
15. Supplementary Notes Christian Jerome (COR)		

16. Abstract

This report describes a series of experiments conducted to validate the auditory parameters associated with urgent forward collision warnings (FCWs) and to expand previous research examining vibrotactile driver assistance alerts. Work conducted in NHTSA's CWIM research program found five auditory characteristics were primarily responsible for perceived urgency – frequency, peak-to-total-time ratio, harmonics, pulse duration, and interburst interval. Driving simulator studies were conducted to validate the relative importance of these characteristics. The first simulator experiment was conducted using a desktop simulator, and exposed participants to an unexpected crash avoidance scenario. Of the five key auditory parameters examined, variations in frequency appeared to be the most detrimental to first exposure perception. The second experiment, conducted in a moving-base simulator was intended to further validate the previous experiment's findings, as well as explore the potential for negative transfer as a result of exposure to varying FCWs across three crash avoidance events. Results show that the warning that met all five criteria was more effective than the warning out of the recommended frequency range. When comparing participants who received the same warning across all three drives relative to those who received two different warnings, there was some evidence of negative transfer when the warning presented to participants was switched, but it was limited to the non-recommended warning. A sorting task conducted in this simulator experiment also found signal categorization results that were consistent with those found in earlier studies. The third experiment investigated the characteristics of vibrotactile signals that lead to categorization as urgent warnings in a sort task. Three key criteria were identified, though even signals that met all criteria were categorized as alarms by fewer than 80 percent of participants. The final experiment in this project was an on-road experiment designed to validate lab and simulator findings in on-road driving at 60 mph. Results further validated previous findings, with auditory signals that met four or five key criteria being perceived as urgent warnings more often, and being categorized faster, than other signals. Signals were presented at 70 dB (A-weighted) and were well detected both with music off and with music playing at 75 dBA, with the exception a few signals of short duration which were detected less than 50 percent of the time in music. Vibrotactile seat pan vibration signals were also investigated in this experiment. Results show little differentiation between vibrotactile signals and all vibrotactile signals used in this experiment were categorized as social notifications by a plurality of participants, providing evidence that tactile signals are not often thought of as warnings when presented alone. This experiment also included a surprise event in which participants experienced an unexpected signal – either an urgent warning, status notification, or social notification. Although there was no real threat present, results show that the urgent warning signal led to a nominally more rapid return of attention to the road, more scanning of the roadway, speed decreases of larger magnitude, and more warning interpretations than the other signals.

17. Key Words human factors, driver vehicle interface (DVI), connected vehicle (CV), distraction, forward collision warning (FCW), ambient noise		**18. Distribution Statement** This document is available to the public from the National Technical Information Service www.ntis.gov.	
19. Security Classif. (of this report) Unclassified	**20. Security Classif. (of this page)** Unclassified	**21. No. of Pages** 88	**22. Price** N/A

Form DOT F 1700.7 Reproduction of completed page authorized.

Table of Contents

List of Figures

List of Tables

Executive Summary

Introduction and objectives

Modern vehicles have increased capabilities to alert drivers to impending hazards and to provide various other sorts of information. Forward collision warnings (FCWs) and other imminent crash warnings have the potential to help drivers avoid or mitigate serious crashes and are therefore of particular priority. If urgent crash warnings are to be effective they must be quickly and reliably detected, correctly understood, and responded to appropriately. However, there is no standard or consensus for the driver-vehicle interface (DVI) for crash warnings. The particular sounds, visual displays, or vibrotactile signals used vary considerably among vehicles. Also, the array of specific warning functions varies among vehicles and these may be mapped to the interface in various ways. Furthermore, such safety-critical displays occur in the broader context of communications within the vehicle and the range and frequency of such messages may be expected to expand with the development of connected-vehicle technology, smartphones, and other communications technologies. Drivers may receive driving-related alerts and messages of lower priority, or messages unrelated to driving altogether. Signals may come from portable aftermarket devices that are independent of the vehicle or driving. For example, a smartphone might present an auditory signal that indicates that the user has received a text message.

This is the context in which crash warnings will occur and in which they must be effective. Therefore it is important to determine the physical parameter boundaries within which warning signals reliably convey the intended message. It is also important to consider the boundaries within which non-critical signals can be designed so as to minimize potential confusion with urgent crash warnings. The research described in this report addresses the attributes of effective crash warnings, particularly auditory FCWs, and the considerations for appropriate categorical perception of in-vehicle alerts in general. Within the context of this report, an "effective" crash warning signal is one that is quickly and reliably categorized as an urgent warning.

This project extends and validates research findings from previous NHTSA-funded projects: Crash Warning Interface Metrics (CWIM) and Human Factors for Connected Vehicles (HFCV). Reports on these projects may be found in Lerner et al. (in press) and Jenness et al. (in press). A key product of this work was the derivation of five key design criteria for auditory signals that results in a sound being quickly and reliably perceived as an urgent crash warning. These criteria were derived primarily from perceptual laboratory and driving simulator experiments conducted by George Mason University. The criteria were:

- A base frequency of 1000 Hz or higher but less than 2500 Hz;
- A peak-to-total time ratio (ratio of time that the signal is at peak intensity) of 0.7 or higher;
- At least three harmonic components (more harmonic components contribute to harsher quality of sound);
- A perceived interburst interval (IBI) of 125 ms or less but greater than 15 ms between sound components; and

- A burst or pulse duration of 200 ms or less.

IBI is defined as the time gap between bursts of sound, which contributes to tempo. The “real” IBI for a sound is the gap of silence between bursts. When bursts have onset and offset ramps, however, the “perceived” IBI includes both the “real” gap and the portion of the onset and offset ramps in which the burst is at less than 90 percent of its full intensity. This is demonstrated in the plots shown in Figure E-1, where time is plotted on the X-axis and sound signal intensity is plotted on the Y-axis. As shown in the top portion of Figure E-1, a sound with a real IBI of zero (no gap between bursts) still has a perceived IBI due to the reduced intensity of the sound during its offset and onset ramps. IBI, as used in this report, refers to perceived IBI.

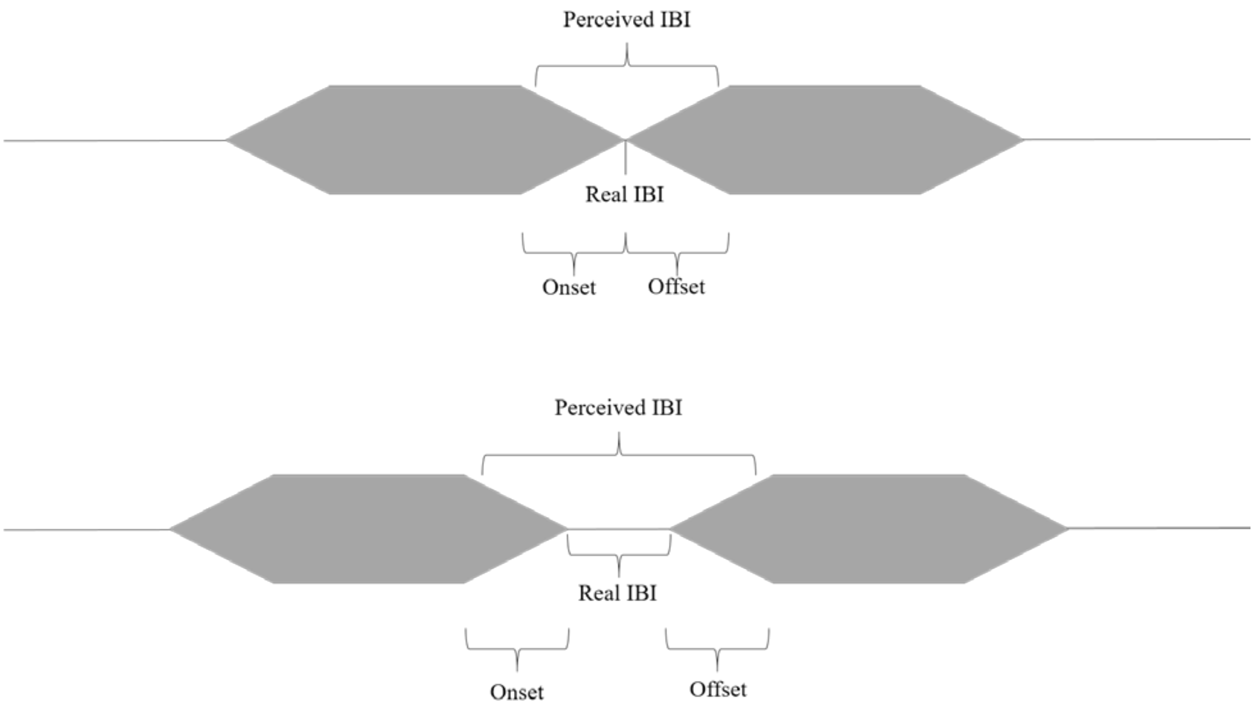


Figure E-1. Time-intensity plots showing the relationship between real and perceived IBI

The objectives of the research described in this report included the following:

- Replicate and validate laboratory and low-fidelity simulator research findings regarding warning design parameters. Validation includes overt driver response to alerts, particularly following first exposure, in imminent crash situations in both low and higher-fidelity driving simulations and driver perception of alerts when presented during actual on-road driving.
- Refine design recommendations for urgent crash warnings, particularly for FCW.
- Refine design recommendations for promoting appropriate categorical perception of auditory and vibrotactile alerts for various categories of meaning and urgency.

Four experiments were conducted as part of this research. Each is described briefly below.

Desktop simulator validation

Participants engaged in simulated driving in a desktop simulator. The primary driving task was to follow a lead vehicle at a safe following distance (at least 1.8 second headway) at 50 mph. The secondary task was a visual perception task involving determining whether Gabor patches on roadside billboards were slanted to the left or to the right. After several minutes of completing these two tasks simultaneously, the lead vehicle abruptly slowed down to 10 mph, creating the potential for a collision to occur. At the moment that the lead vehicle began slowing, participants received one of seven auditory signals or a no-warning control.

The warning sounds are described in Table E-1 below. The base warning met all four of the criteria developed in previous research. The GMU Prime warning met all four of the criteria plus a fifth criterion (having a pulse rate of 200 ms or less) that was identified as important in an early phase of this research. The other warnings met three of the original four criteria. By comparing the base warning to warnings missing one of the four criteria, researchers could determine the relative importance of each criterion.

When all warning conditions were compared against each other, there was a significant difference in accelerator release time after controlling for the covariate of headway time at warning. While none of the warnings differed in accelerator release times from each other, the low frequency, long IBI, high frequency, and short pulse duration warnings all had significantly faster response times than no warning. These data are shown in Figure E-2. There were no significant differences between warnings when comparing minimum distance to lead car or emergency maneuver response time (brake or evasion).

Table E-1. Varied physical warning parameters

FCW	Description	Frequency (Hz)	Harmonics	IBI (ms)	Peak-to-Total-Time Ratio	Offset Time (ms)	Pulse Duration (ms)
1	Base Warning	1576	5	18	.95	10	400
2	Low Frequency	700	5	18	.95	10	400
3	No Harmonics	1576	1	18	.95	10	400
4	Long IBI	1576	5	344	.95	10	400
5	Low Ratio	1576	5	18	.6	150	400
6	High Frequency	3000	5	18	.95	10	400
7	GMU Prime	1576	5	18	.95	10	200

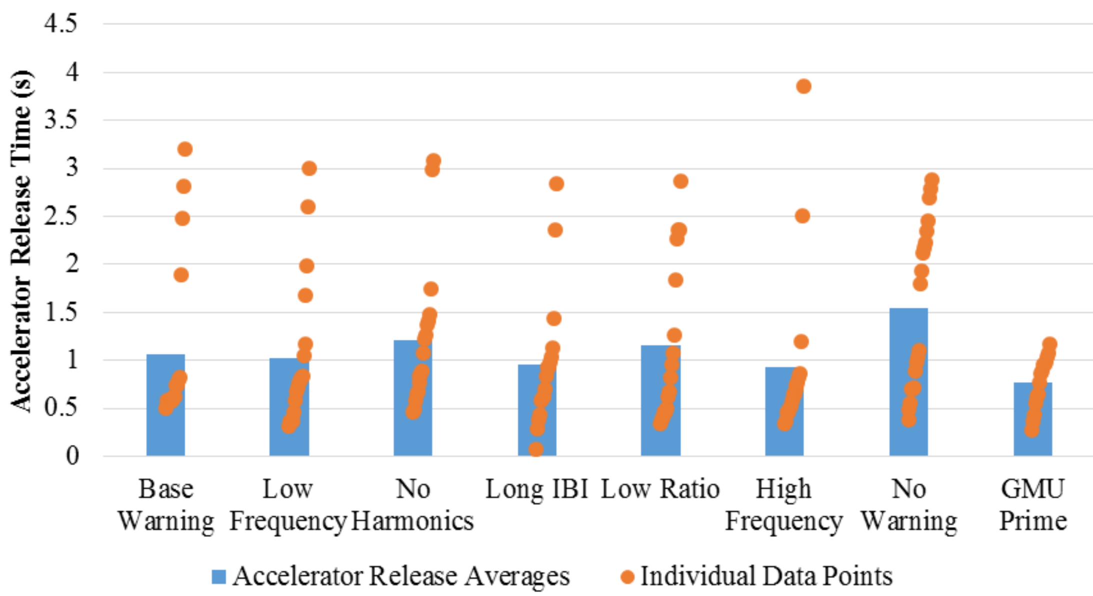


Figure E-2. Accelerator release data from desktop simulator validation

When comparing all eight warning conditions, there was no significant effect of warning condition on likelihood of collision with the lead vehicle. When the various warning conditions were collapsed into groups (GMU Prime, meets most parameters, doesn't meet frequency, and no warning), however, rank collision scores showed that the GMU Prime warning had the fewest collisions, the high/low frequency group lead to the most collisions, and the difference between these two groups was significant.

One key result from this experiment was the finding that the GMU Prime warning, which met all five criteria of an effective warning, was indeed better than the other warning groups in terms of collision avoidance, and particularly in reducing the likelihood of long response times, as seen in the individual data points in the accelerator release data figure above. Another key result was the finding that the warnings that had a frequency either higher or lower than the criterion range performed particularly poorly, which suggests that it is especially important for auditory warnings to have base frequencies between 1000 and 2500 Hz.

Motion-base simulator validation

The next experiment examined the validity of the five recommended criteria for eliciting appropriate crash avoidance response in a higher fidelity simulator. Additionally, this second experiment investigated the potential effects of learning and negative transfer due to inconsistent warnings on driver performance in a surprise crash avoidance scenario. This experiment also included a sound sorting task intended to validate laboratory and desktop simulator findings in a higher-fidelity, motion-base simulator. Participants engaged in three separate drives in a motion-base simulator. While driving at 65 mph, they performed a visual-manual distraction task and, in the second and third drive, occasionally heard sounds that they then categorized as either an alarm, status notification, or social notification. Definitions provided to participants described an alarm as a time-critical highly urgent sound like a collision warning or a lane deviation warning, a status notification as a sound indicating something to do with the status of the car, like low tire pressure or a door is open, and a social notification would be something informing of a call or an email or a Facebook update. Within each drive, participants experienced a collision warning event. Participants were randomly assigned to one of five groups. Two groups received consistent warnings, either all GMU Prime or all a warning that meets all but some key criteria, here referred to as an edge warning (the high frequency warning from the desktop simulator validation) for each of the three experimental drives. Importantly, the edge warning was not intended to be a bad warning, but a warning that does not conform to all criteria that have been defined for a good warning. Two groups included “switch” conditions, where the warning they received was inconsistent throughout the course of three experimental drive events. In the switch conditions, the first and third drive always had the same warning, but the second drive had the other warning. A control group received no warning during any of the experimental drives.

For the first drive, in which participants experienced the collision event for the first time, results show no significant differences between the two warning and no-warning groups in likelihood of collisions, though the GMU Prime warning was associated with a lower mean speed at collision (31 mph) than the no-warning condition (45 mph), as well as a faster brake response time, which is evidence of reduced collision severity. Figure E-3 shows the data for mean speed at time of collision.

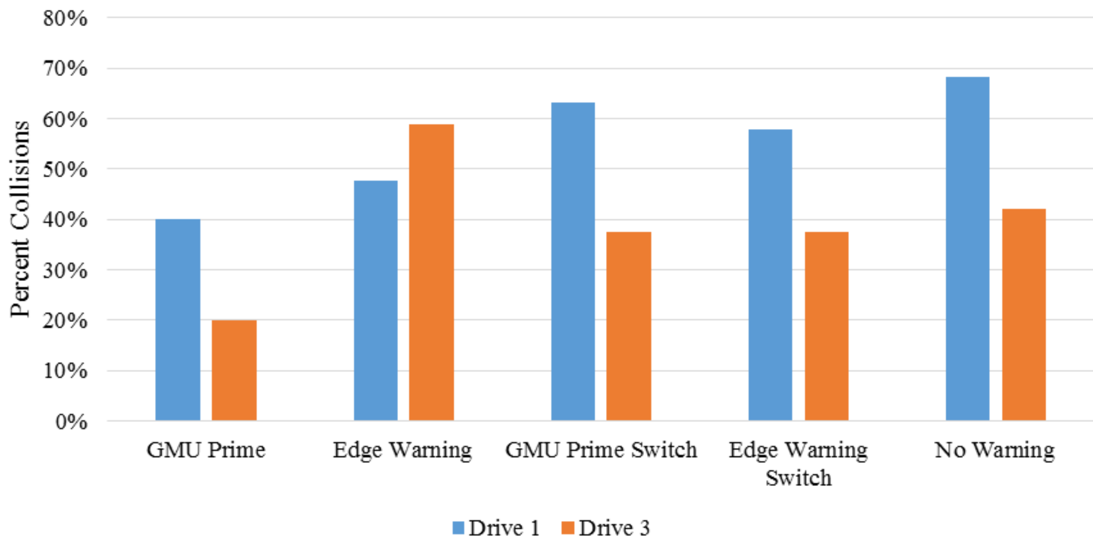


Figure E-3. Speed at time of collision by condition and drive

The third experimental drive allowed researchers to investigate the possibility of negative transfer by comparing groups that received consistent warnings in all three drives versus groups that switched warnings. Results for drive 3 collision data indicate overall decreases in collision rate, relative to drive 1. Collision rates between groups were not statistically significant. Drivers in the consistent warning 2 condition were somewhat more likely to collide than all other conditions, and even, strangely, than they were in drive 1.

Results of the sort task, which used a broader set of signals including many that met few of the key warning criteria, indicate that participants categorize signals very similarly in the higher-fidelity, moving-base simulator and in a similar paradigm using the desktop simulator. Sounds that meet more criteria are more likely to be categorized as urgent alarms, as shown in Figure E-

4. Sort task findings also confirmed previous findings that participants typically respond more quickly to sounds that meet most or no criteria than they do to sounds which meet some criteria.

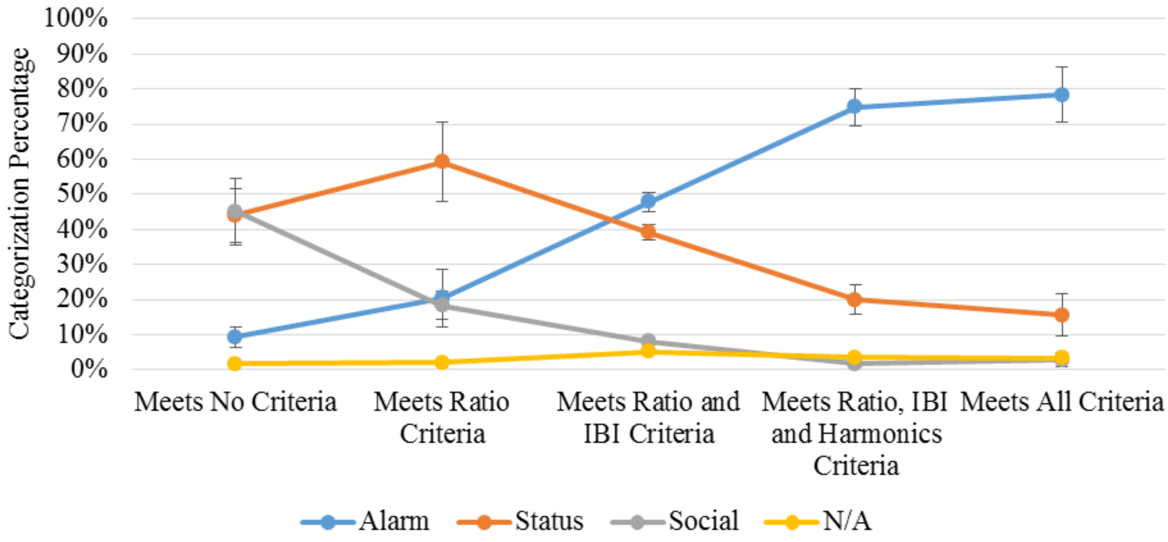


Figure E-4. Percentage categorization by criteria met for drive 3 sort task

Tactile Parameters and Urgency Category

The purpose of this experiment was to develop preliminary definitions for the perceptual space that make up the category of urgent vibrotactile collision warnings. This was done in two steps. First, a set of 27 existing vibrotactile signals was compiled based on a literature review. Then, the signals were included in a perceptual sort to discern criteria and cutoffs for highly urgent, vibrotactile alarm signals. Stimuli varied on pulse duration, interpulse interval, pulses per burst, burst duration, IBI, total number of bursts and total alert time.

A backwards, stepwise logistic regression model showed that IBI (≤ 150 ms), burst duration (≤ 200 ms), number of bursts (≤ 10), and total signal time (≤ 1000 ms) predicted about 56 percent of the variance in alarm categorization for a significant overall prediction model. Signals that met all criteria were classified as alarm signals by nearly 80 percent of participants.

On-road experiment

Previous research conducted within NHTSA's CWIM research program found that ambient noise conditions inside a vehicle can have a substantial effect on the detection and interpretation of in-vehicle signals (Singer, Lerner, Kellman, & Robinson, in press). Therefore, the intent of this experiment was to confirm that laboratory and simulator findings could be validated in an actual driving situation under both baseline and louder ambient noise conditions. For the present experiment, the behavioral measures were perceptual (detection, response time, perceived urgency/meaning) but were collected under actual on-road driving conditions and with two levels of in-vehicle ambient noise. This experiment also investigated participants' reactions to an unexpected auditory signal; this part of the procedure is described later in this section.

Participants drove an instrumented vehicle on a limited access tollway (Maryland Route 200) with a 60 mph speed limit. During the Baseline ambient noise condition, all windows were closed and music was off. During the Music On condition, an instrumental smooth jazz song was played in a continuous loop at an average loudness of 75 dBA.

There were 17 auditory signals and 9 vibrotactile seat pan vibration signals in this experiment. Six of the auditory signals were alarms either replicated from actual in-vehicle warning systems or created by GMU and used in the previously described experiments. The remaining signals were expected to be classified as status or social notifications. Participants experienced each signal once under baseline ambient noise and once with music playing. When participants experienced a signal, they verbally spoke the category in which they felt it belonged: alarm, status notification, or social notification. Detection rate, categorization, and verbal response time to categorize were then calculated.

All auditory signals were detected at high rates (greater than 95%) in the baseline ambient noise condition, but three signals were detected at relatively low rates (less than 50%) in the music condition, which showed that music interfered with detection of certain signals, particularly those with short durations. Auditory alarm signals were generally rated higher in urgency than all other signals, with the exception of Auditory Alarm 6, which had a lower frequency than the 1000 Hz criterion cutoff. Signals meeting the four key criteria in Lerner et al. (in press) were perceived as alarms by 84 to 98 percent of participants in the baseline ambient noise condition. Figure E-5 shows how each signal was categorized.

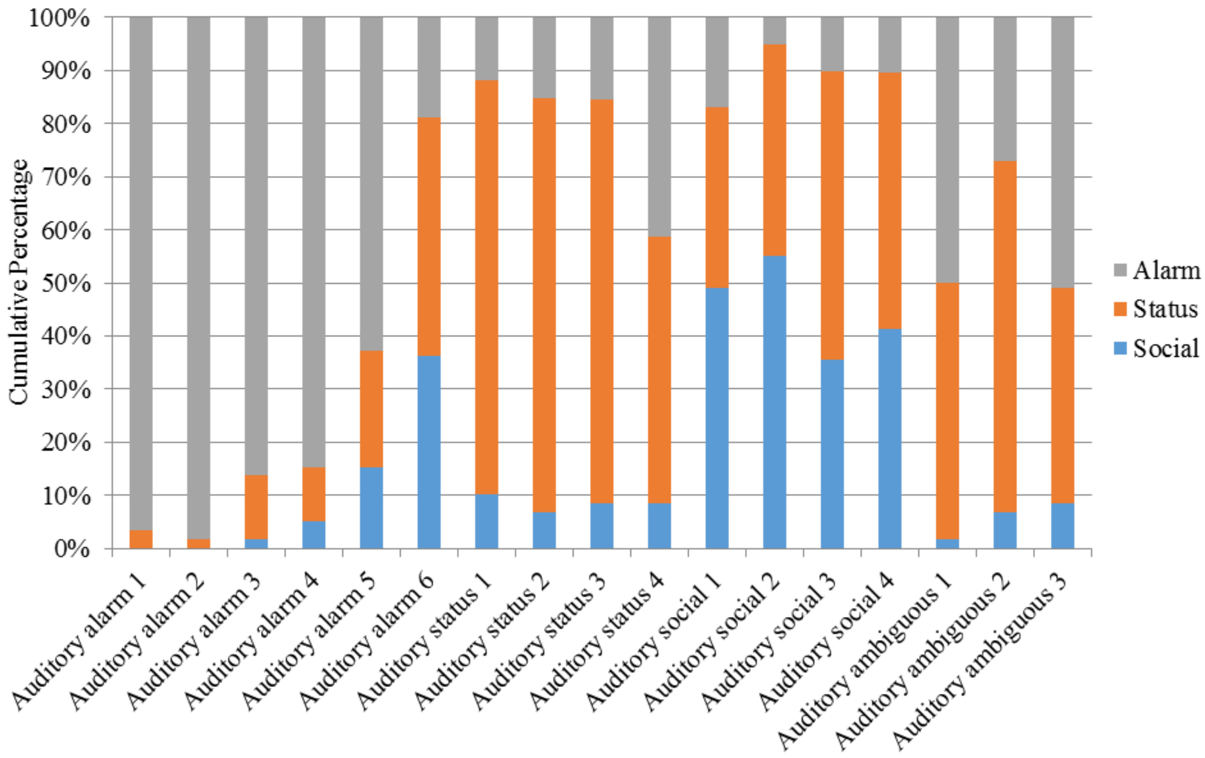


Figure E-5. Percentage of participants assigning each meaning category to an auditory signal in the baseline condition

For the dependent measure of verbal response time to categorize the signal, there were significant effects of the auditory signal, the ambient noise condition, and their interaction. Auditory alarms, which showed the best agreement among participants in perceived meaning, were responded to more rapidly than other sounds. This finding is consistent with results of the previous laboratory and simulator experiments.

Results for vibrotactile signals showed fewer differences between signals and ambient noise conditions than were observed for the auditory signals. Vibrotactile signals were detected at high rates in both baseline and music conditions, indicating that vibrotactile signals are resistant to interference from elevated ambient noise from music. However, there was little differentiation in perceived urgency between signals, with all signals being perceived as social notifications by a plurality of participants. This suggests that, as implemented in this experiment, seat pan vibrotactile signals alone are not likely to be interpreted as warnings by drivers.

In addition to the perceptual methods, in which drivers responded to auditory and vibrotactile signals that they knew might occur, this experiment also investigated driver reaction to an unanticipated auditory signal. This was done in an initial portion of the drive, before drivers were exposed to the listening procedures of the perceptual ratings. The intent in this portion of the experiment was to see if drivers responded in a different manner to unexpected signals that varied in terms of their subjective category of meaning. Participants were instructed to read navigation-related messages presented on a visual display in the vehicle's center stack while driving on a straight section of road with no surrounding traffic. During the third instance of this task, while the participant's gaze was on the center stack display, the experimenter triggered an auditory signal – either a nominal alarm, status notification, or social notification. Although there was no real threat and no driver response was needed, researchers expected that drivers might display differences in their responses to the various signals.

Results show benefits of the alarm in terms of more rapid return of attention to the road, increased scanning of the road following the signal, greater decreases in speed following the signal, and verbal responses indicating that the signal was interpreted as an alarm. Most of these effects were not statistically significant, but given the small sample size per group (n = 17) and the lack of a real threat scenario, these trends are noteworthy.

Key Conclusions

The experiments conducted in this project were intended to validate and expand the findings of previous experiments through the use of driving simulator and on-road research methods. These experiments confirmed the validity of the four criteria for auditory alarms established in previous work, and established that a fifth criterion (short pulse duration) further contributes to perception of auditory signals as urgent warnings. Evidence also suggests that the frequency criterion (1000-2500 Hz) is particularly important, because driving simulator research showed significantly more crashes and more severe crashes when this criterion was not met than when other criteria were not met.

Sorting data from the motion base simulator experiment and the on-road driving experiment validate results obtained from previous work. Results indicated that as sounds meet more of the recommended criteria they are more likely to be identified as alarms. Further, response time results indicate that participants responded fastest to sounds that met none or almost all criteria, indicating that sounds that met none of the criteria were unambiguously non-urgent and sounds that met nearly all criteria were unambiguously urgent. Sounds that met a few criteria were more ambiguous and therefore participants categorized them more slowly.

Consistent with previous work (Singer et al., in press), auditory signals presented at about 70 dBA were generally well-detected, even with music playing. The exceptions to this were the social notification signals that were presented for a briefer period than the other signals (as they likely would be in actual applications). In contrast to the auditory signals, there was not a strong effect of signal characteristics for the vibrotactile seat pan signals. The nine vibrotactile signals were all reliably detected (90% or greater), but there was not a great difference in perceived urgency among them and all were perceived by a plurality of participants as social notifications.

The on-road experiment observed that, relative to less urgent message categories, an unexpected auditory signal perceived as an urgent warning led to nonsignificant trends in the direction of faster return of gaze to the forward roadway, more time scanning the forward roadway, a larger decrease in vehicle speed, more emotional responses, and greater likelihood of describing the sound as a warning. Factors contributing to the absence of statistical significance may have included the relatively small n for each group, drivers' unwillingness to commit long glances to a distracting task while driving on real roads, and the fact that participants had no expectation of an imminent threat and could quickly determine that there was no threat.

Implications for design of alerting systems

In addition to providing validation of previous research results, this project's findings also provide the basis for supporting the design of auditory FCWs and other in-vehicle alerting displays and systems. Design goals, specific recommendations, and discussion are provided for the design of urgent crash warnings, design for non-crash message types, and use of laboratory perceptual methods for development and evaluation of in-vehicle signals. It is important to note that this project did not explore all possible warning modes, and explored a limited set of options within each mode, so the implications provided here should be interpreted within this context.

1. Introduction

1.1 Background

Modern vehicles have increased capabilities to alert drivers to impending hazards and to provide various other sorts of information. Forward collision warnings (FCWs) and other imminent crash warnings have the potential to help drivers avoid or mitigate serious crashes and so are of particular priority. If urgent crash warnings are to be effective they must be quickly and reliably detected, correctly understood, and responded to appropriately. However, there is no standard or consensus for the driver-vehicle interface (DVI) for crash warnings. The particular sounds, visual displays, or vibrotactile signals used vary considerably among vehicles. Also, the array of specific warning functions varies among vehicles and these may be mapped to the interface in various ways. For example, in one vehicle an auditory chime may be designed to inform the user of low fuel, but in another car the same chime means that the user has received a new email or message from the infotainment system. Furthermore, such safety-critical displays occur in the broader context of communications within the vehicle and the range and frequency of such messages may be expected to expand with the development of Connected Vehicle technology, smartphones, and other communications technology. Drivers may receive driving-related alerts and messages of lower priority, or messages unrelated to driving altogether. Signals may come from portable aftermarket devices that are independent of the vehicle or driving. For example, a smartphone might present an auditory signal that indicates that the user has received a text message.

This is the context in which crash warnings will occur and in which they must be effective. Therefore it is important to determine the physical parameter boundaries within which warning signals reliably convey the intended message. It is also important to consider the boundaries within which non-critical signals can be designed so as to minimize potential confusion with urgent crash warnings. The research described in this report addresses the attributes of effective crash warnings, particularly FCWs, and the considerations for appropriate categorical perception of in-vehicle alerts in general. Within the context of this report, an “effective” crash waring signal is one that is quickly and reliably categorized as an urgent warning.

“Categorical perception” refers to the categorization of stimuli into distinct, clearly defined categories. One everyday example of categorical perception is in how we perceive a rainbow. Even though the wavelength of the light changes smoothly and continuously from the top to the bottom of the rainbow, we see discrete bands of distinct colors. The analogy for in-vehicle signals is that while various physical parameters (e.g., sound frequency) and perceptual outcomes (e.g., perceived urgency) may vary in a continuous manner, there may be boundaries within which signals tend to convey a particular meaning which is distinct from the meaning of signals beyond that boundary. For complex, man-made signals such as in-vehicle alerts, the underlying perceptual basis is complex and a result of both basic perceptual processes and learned associations. Category boundaries may not be sharply defined. However, from a vehicle warning system design perspective, the important point is that we may define physical parameters and bounds that are predictive of how people are likely to categorize meaning in a reasonably consistent manner.

While there has been a considerable history of research on warnings, there has not been much work to define acceptable bounds or parameters based on the category or urgency of the warning, and particularly in keeping highly urgent warnings distinct from other messages within the in-vehicle environment. NHTSA has funded important recent work in this area under its Crash Warning Interface Metrics (CWIM) and Human Factors for Connected Vehicles (HFCV) programs. Reports on these projects may be found in Lerner et al. (in press) and Jenness et al. (in review). A key product of this work was the derivation of four key design criteria for auditory signals that cause a sound to be quickly and reliably perceived as an urgent crash warning. These criteria were derived primarily from perceptual laboratory and driving simulator experiments conducted by George Mason University (GMU) (Lewis, Eisert, Roberts, & Baldwin, 2014).

The criteria were:

- A base frequency of 1000 Hz or higher but less than 2500 Hz;
- A peak-to-total time ratio (ratio of time that the signal is at peak intensity) of 0.7 or higher;
- At least three harmonic components (more harmonic components contribute to harsher quality of sound); and
- A perceived interburst interval (IBI) of 125 ms or less but greater than 15 ms between sound components.

IBI is defined as the time gap between bursts of sound, which contributes to tempo. The "real" IBI for a sound is the gap of silence between bursts. When bursts have onset and offset ramps, however, the "perceived" IBI includes both the "real" gap and the portion of the onset and offset ramps in which the burst is at less than 90 percent of its full intensity. This is demonstrated in the plots shown in Figure E-1, where time is plotted on the X-axis and sound signal intensity is plotted on the Y-axis. As shown in the top portion of Figure E-1, a sound with a real IBI of zero (no gap between bursts) still has a perceived IBI due to the reduced intensity of the sound during its offset and onset ramps. Within this report, IBI data refers to perceived IBI.

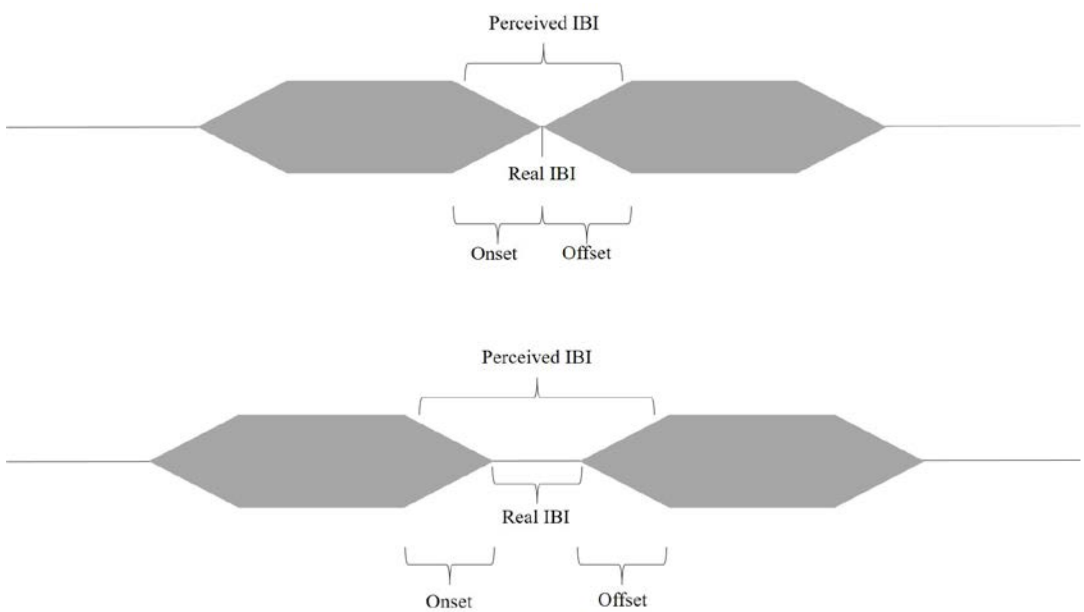


Figure 1. Time-intensity plots showing the relationship between real and perceived IBI

Auditory signals possessing all four of the criteria described above were categorized as urgent crash warnings more than 90 percent of the time. They were also categorized more rapidly than other sounds in a driving simulator procedure. Signals possessing none of these four criteria were rarely (<10%) interpreted as crash warnings. Auditory signals related to vehicle status, but not requiring an urgent response, were most distinguishable from urgent warnings and social notifications if they met only the ratio criterion.

These findings provide a basis for the design of in-vehicle auditory signals that elicit appropriate interpretation of the intended message category in untrained listeners. However, this research would benefit from extension and validation in two respects. First, the perceptual findings should be extended to include dependent measures of driver vehicle control in response to a perceived crash threat. Second, the perception of signals should be verified and refined under actual on-road driving conditions. These two aspects of validation would strengthen the empirical basis underlying the modeling of driver perception of warnings and also refine the modeling of responses so that the final recommendations fully reflect both driver vehicle control response and actual on-road perceptual processes.

1.2 Objectives

The objectives of this research included the following:

- Replicate and validate laboratory and low-fidelity simulator research findings regarding warning design parameters. Validation includes overt driver response to alerts in imminent crash situations in a higher-fidelity driving simulator and driver perception of alerts when presented during actual on-road driving.
- Refine design support for urgent crash warnings, particularly for FCW.
- Refine design support for promoting appropriate categorical perception of auditory and vibrotactile alerts for various categories of meaning and urgency.

1.3 Overview of research activities

Four experiments were conducted as part of this research. Laboratory and driving simulator experiments were carried out by GMU in parallel with an on-road experiment conducted by Westat. The first simulator experiment used a desktop simulator to replicate and expand earlier findings regarding the critical physical parameters for effective warning sounds. A second simulator experiment then evaluated various auditory alerts in a motion-base driving simulator validation, which included both driver crash avoidance behaviors and perceptual categorization. The vibrotactile signal experiment was a laboratory perceptual study that manipulated the parameters of vibrotactile signals to assess how people perceive the urgency and meaning of such signals. The on-road experiment presented drivers with both auditory and vibrotactile signals and obtained perceptual judgments of those signals under baseline and louder (music) background noise conditions. The on-road experiment also examined driver reactions to unexpected auditory signals of different types. This set of experiments was complementary in extending and validating earlier findings on auditory and vibrotactile signal parameters for in-vehicle alerts. The validation aspect included consideration of driver crash avoidance responses in realistic imminent crash situations in the moving-base driving simulator and consideration of driver perceptual and orienting responses in actual on-road (but not imminent crash) driving conditions.

1.4 Structure of this report

The methods and findings of the experiments are reported in Sections 2, 3, and 4. Section 2 covers the driving simulator experiments. Section 3 presents the research on vibrotactile signal parameters. Section 4 describes the on-road experiment. Section 5 then discusses the key findings and implications of the set of experiments, including recommendations to support the effective design of crash warnings and appropriate categorical perception of signals to convey messages of differing urgency.

2. Driving Simulator Experiments

2.1 Introduction

This set of experiments had the primary aim of providing further validation of the major auditory parameters determined to be key components of a highly urgent warning in the project team's previous research. The work described here includes two experiments: one conducted using a desktop driving simulator and the other using a high-fidelity motion base simulator, respectively. The first simulator experiment examined driver response to the first exposure to a warning in conjunction with a potential collision event in a desktop driving simulator. Forward collision warnings (FCW) with realistic timing or no warning (control) preceded the hazard event. FCWs in this experiment consisted of seven sounds specifically designed to: (1) validate the effectiveness of the criteria (faster and more appropriate collision avoidance maneuvers), (2) further examine the range of previously established criteria, and (3) determine which of five recommended characteristics is most important.

A systematic set of sounds that met all, or all but one of the suggested criteria was presented. Sound 1 (the Base Warning) met the original four criteria and Sound 7 (GMU Prime) met one further criterion identified in this work, that of pulse duration. Long pulse durations were undersampled in the previous work as the stimulus set started with actual OEM warnings almost always consisting of pulses of 200 ms or less. Preliminary investigations in this series indicated that pulse duration should be less than or equal to 200 ms. Sounds 2 and 6 examine the lower and upper bounds of the frequency parameter (Low and High Frequency). Sound 3 (No Harmonics) examines the harmonics criteria, Sound 4 (Long IBI) meets all but the tempo or IBI criteria and Sound 5 (Low Ratio) meets all but the peak-to-total time ratio criteria. It was hypothesized that the best response (in terms of the fastest and most appropriate responses as described in further detail below) would be observed for participants receiving a FCW meeting all previously defined criteria (Base Warning and GMU Prime).

The second experiment described here was designed to further validate the results of the previous research and the results of the first experiment in a higher fidelity simulator. In the second simulator experiment, the primary focus was the driver's response to the first occurrence of the auditory FCW preceding the first hazardous event. Participants were asked to complete three experimental drives all ending in a difficult, hard to avoid, potential collision scenario. Participants received either the warning that met all five key criteria or one that met three criteria. Participants were further divided into five groups in order to determine the effectiveness of warning consistency on participant response. It was hypothesized that any warning would produce better participant responses to collisions than no warning, but that a warning meeting all key criteria would produce the best collision avoidance response. It was further hypothesized that participants receiving consistent warnings (i.e., same warning for subsequent exposure) would show greater learning effects, in terms of better responses in the third collision scenario, than would participants receiving inconsistent warnings. In addition to investigating warning response, Experiment 2 included a sorting component in which participants categorized a wide range of auditory signals into one of three categories while engaged in simulated driving.

2.2 Experiment 1- Desktop simulator validation

2.2.1 Method

2.2.1.1 Participants

174 participants (56 male, 118 female, mean age= 22.3 years), recruited from the student research participant pool and via flyers posted on campus. Participants received either a small amount of research participation credit that could be applied to their classes or were compensated $10 for completing the study. All participants self-reported normal or corrected-to-normal vision and hearing and were licensed drivers.

2.2.1.2 Stimuli

Stimuli consisted of seven auditory warnings, presented between groups, and one group received no collision warning. Participants were randomly assigned to each group in equal numbers. Every effort was made to maintain equal gender and age distributions by group. Each warning deviated from a "base" warning sound on one auditory parameter. Descriptions of all seven warnings are presented in Table 1. Each warning consisted of a single burst with four pulses per burst. Pulses had a 10 ms onset time (the amount of time the sound increases in intensity from silence to its full intensity). All warnings were presented at 75 dBA and ambient noise (engine, wind, road, etc.) was presented at 60 dBA.

Table 1. Varied physical warning parameters (desktop simulator)

Description	Frequency (Hz)	Harmonics	IBI (ms)	Peak-to-Total Time Ratio	Offset Time (ms)	Pulse Duration (ms)
Base Warning	1576	5	18	.95	10	400
Low Frequency	700	5	18	.95	10	400
No Harmonics	1576	1	18	.95	10	400
Long IBI	1576	5	344	.95	10	400
Low Ratio	1576	5	18	.6	150	400
High Frequency	3000	5	18	.95	10	400
GMU Prime (Short Pulse Duration)	1576	5	18	.95	10	200

2.2.1.3 Apparatus and procedure

The experiment was run in a sound attenuated room using a Realtime Technologies, Inc. (RTI), desktop driving simulator. The Simulator uses a 23" monitor and a Logitech G23 steering wheel attached to a Playseat Racing seat to create a cab-like atmosphere. All sounds were presented via a pair of computer speakers placed directly below the monitor. Figure 2 shows the experimental setup.

Figure 2. Desktop simulator setup

Throughout the drive participants also completed a perceptual secondary task, in which they were asked to determine if Gabor patches (see Appendix A: Desktop Simulator Experiment Additional Stimuli) presented on billboards were shifted towards the left or the right. All Gabor patches were shifted five degrees either to the left or right. Participants responded using two paddle shifters located on the back of the steering wheel. To prevent participants from seeing multiple billboards at once, a building was placed after each billboard. In addition to responding to the billboards, participants were also required to maintain a safe headway while following a lead vehicle. If a participant had a safe headway (1.8 seconds or greater) they would see a green square on the screen, if they were at a close but not necessarily unsafe distance (1.5-1.8 seconds) they would see a yellow square, and if they were at an unsafe headway (less than 1.5 seconds) they would see a red square. Figure 3 shows an example of the Gabor patches and the headway maintenance system.

Figure 3. Left-shifted Gabor patch and safe headway display

Upon arrival, participants presented their driver's license for the researcher to inspect and verify it was valid and they were at least 18 years of age. After license verification, participants read and signed a consent form, then adjusted the seat until they were comfortable reaching the pedals. The first block of the experiment consisted of three training sessions. First participants completed a training drive to familiarize themselves with simulator handling. They were instructed to maintain a speed of 50 mph while keeping the vehicle in the right hand lane. Participants were also instructed to come to a complete stop twice during this training to familiarize them with the brake pedal. Once participants felt comfortable maintaining their speed and lane they began the second training portion. The second training portion was to familiarize the participant with the secondary task while continuing to maintain a speed of 50 mph, staying in the right lane. Once they felt comfortable with this they moved to the final training portion. In the final training portion participants practiced doing the secondary task while also maintaining a safe following distance from a lead vehicle, maintaining a speed of 50 mph, and staying in the right lane. Once participants felt comfortable with this final training session they moved on to the experimental portion of the study.

Participants were instructed that the purpose of the experiment was to test out a new connected vehicle technology designed to teach drivers what a safe headway was and that they were to maintain a safe headway throughout the drive while completing the secondary task. Participants were never explicitly told that the vehicle had a collision warning system in it. During the experimental drive participants drove for about 10 minutes completing two blocks of the secondary task, each block consisting of 20 billboards. Midway through the second block of Gabor patches the lead vehicle abruptly slowed down to 10 mph at a deceleration rate of 6 m/s^2, causing a potential collision. If a participant was to receive a collision warning, it was played concurrently with the presentation of the brake lights. After the event occurred participants were instructed to remain stopped and then completed a short demographic questionnaire. After the survey, participants were informed about the true purpose of the experiment and received either research credit or $10 for completing the study. The entire study lasted approximately 15 minutes on average.

2.2.1.4 Design and data analysis

The key independent variable was warning played, with main dependent variables of interest being minimum distance to lead car, accelerator release time, and evasive maneuver response time (EMRT) during the collision event. Since headway time when the event started was variable depending on the participant it was controlled for as a covariate. Collision data were examined in a separate analysis.

2.2.2 Results

2.2.2.1 Gender

Due to the disparity in the number of male and female participants in this experiment a simple t-test was conducted to ensure no statistically significant difference existed between performances due to gender. The t-test found no statistically significant difference between genders for minimum distance to lead car, accelerator release time, and EMRT.

2.2.2.2 ANCOVA results

The analysis of covariance (ANCOVA) for minimum distance to lead car and EMRT were not statistically significant. The ANCOVA for accelerator release time found that the covariate, headway time at warning was significantly related to accelerator release time, $F(1,130) = 4.45$, $p = .007$. There was also a significant effect of warning played on accelerator release time after controlling for the effect of headway time at warning, $F(7,130) = 1.23$, $p = .05$. Post hoc analysis of accelerator release time shows that none of the warnings differed in accelerator release times from each other, but low frequency ($M = .97$, $p = .014$), long IBI ($M = .88$, $p = .007$), high frequency ($M = 1.03$, $p = .030$), and GMU Prime (short pulse duration) ($M = .68$, $p = .001$) all had significantly faster response times than no warning. Figure 4 shows mean and each unique data point for accelerator release time. In looking at the individual results it is evident that GMU Prime (short pulse duration; the only warning that meets all five warning criteria) results in the least amount of variance for accelerator release time.

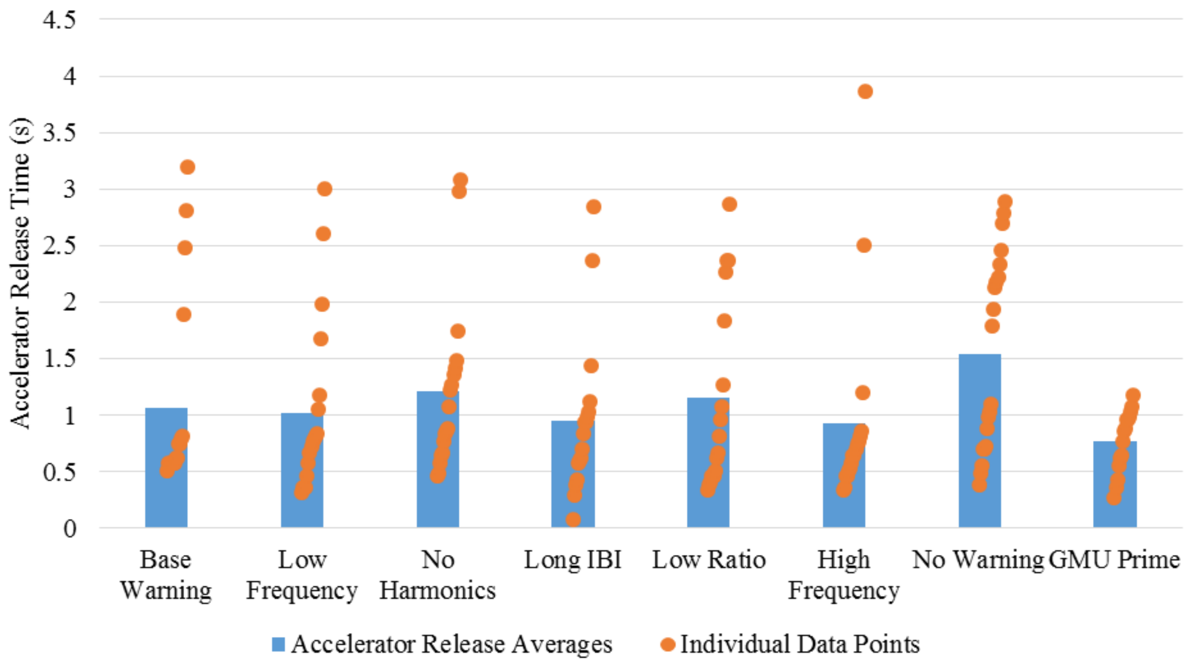


Figure 4. Accelerator release data

2.2.2.3 Collision data

Figure 5 shows the percentage of participants in each condition who collided with the lead vehicle. Collision data were examined using a Krushkal Wallis H Test, which revealed that there was no statistically significant difference in number of collisions between the eight warnings played, X^2 (7) = 9.34, p > .05.

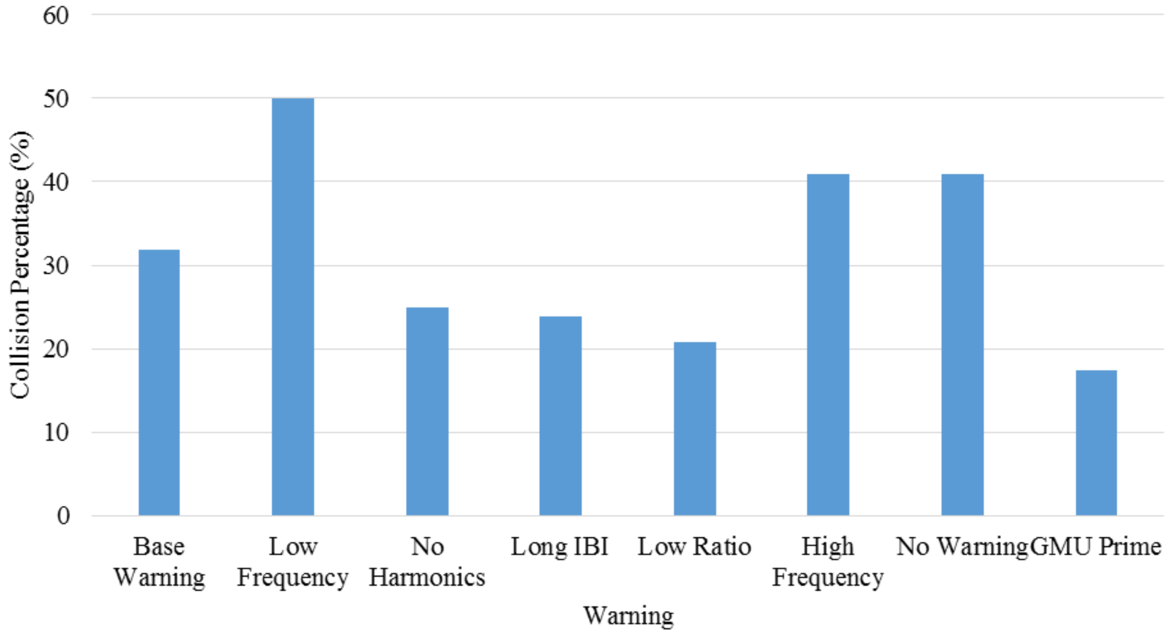


Figure 5. Percentage of participants who collided with lead vehicle

2.2.2.4 Collapsed groups

Taking these two analyses into account it was evident that the frequency adjustments resulted in the highest collision rates, among the warning groups, even though they also had some of the quickest accelerator release times. In order to look at this finding more closely warnings were split into four groups; GMU Prime, meets most parameters (including Base Warning, No Harmonics, Long IBI and Low Ratio warnings), doesn't meet frequency (including both Low and High Frequency warnings) , and no warning. Using these new groups another ANCOVA was run using headway time at warning as a covariate again. No statistically significant difference was observed for minimum distance or EMRT. ANCOVA results for accelerator release time revealed that the covariate, headway time at warning was significantly related to participants' accelerator release time, $F(1,134) = 7.03$, $p = .009$. There was also a significant effect of warning group on accelerator release time after controlling for the effect of headway time at warning, $F(3,134) = 4.29$, $p = .006$. Post hoc tests revealed that all warning groups had significantly faster accelerator release times compared to no warning.

A Kruskal Wallis H Test was used to statistically examine the ranking of all the data points for the current sample. Using the collapsed groups a Kruskal Wallis H Test showed that there was a statistically significant difference between the warning groups for number of collisions, X^2 (3) = 8.26, $p = .04$, with a mean rank collision score of 75.63 for GMU Prime, 82.50 for meets most parameters, 99.86 for doesn't meet frequency, and 96.09 for no warning. This result demonstrates that the GMU Prime warning group resulted in the fewest collisions, while the group that didn't meet the frequency criterion resulted in the most collisions. Additionally, chi-square difference test revealed that the GMU Prime group had significantly fewer collisions than the doesn't meet frequency group, $X^2 = 5.04$, $p = .025$ and while it didn't result in significantly fewer crashes than no warning, it was trending in that direction, $X^2 = 3.03$, $p = .08$.

2.2.3 Discussion

A primary purpose of this study was to begin to validate the signal criteria recommendations established in antecedent research. Using a base warning signal that met all four of the recommended criteria, one parameter at a time was systematically manipulated to determine if any one specific parameter contributed more to the perception of a warning than the others. Findings suggest that deviating from the recommended frequency range (either higher or lower), or the recommended pulse duration resulted in significantly more collisions than GMU Prime, even though they had similar accelerator release times. When looking at participant responses when asked what they thought the sound played represented, only 42 percent of participants in the "doesn't meet frequency" group thought the sound represented a collision warning, whereas 71 percent of participants in the GMU Prime group reported they thought the sound played represented a collision warning. This may have been because the wording of the question was open-ended. Other answers included thinking the study was over (~13%), thinking the sound was related to errors in the secondary task (~8%), thinking the sound came from the other vehicle (~4%) or unsure (~4%). This, in addition to the collision data, suggests that of all the recommendations, frequency may be the most important for warning perception/interpretation. Our previous recommendations only provided a low cut off point for frequency suggesting a warning should at minimum have a frequency of 1000 Hz, however these results suggest that the upper cutoff for frequency be somewhere below 3000 Hz.

The secondary purpose of this experiment was to select the two warnings that would be used in the next experiment, one to serve as the "best" warning and one to serve as an "edge" warning. For the best warning, GMU Prime was selected and for the edge warning the high-frequency adjusted warning was selected from the previous experiment. High frequency was selected because changes to this dimension resulted in a high degree of variability in responses in Experiment 1. While the low frequency warning resulted in more crashes than the high frequency warning, it was hypothesized that it may be more easily masked by the ambient noise from the motion based simulator.

2.3 Experiment 2- Motion-base simulator validation

2.3.1 Methods

2.3.1.1 Participants

101 undergraduate and graduate students (29 male, 72 female, mean age= 20.2 years), were recruited from the student research participant pool and via flyers posted on campus. Participants received a small amount of research participation credit that could be applied to their classes. All participants self-reported normal or corrected-to-normal vision and hearing and were licensed drivers.

2.3.1.2 Stimuli

Stimuli consisted of two auditory warnings, both designed to have relatively high urgency and probability of being classified as a time-critical alarm, though one warning (Warning 1, GMU Prime consisting of the five recommended criteria) was expected to result in somewhat better collision avoidance response than the other (Edge Warning, meeting four of the criteria but having a frequency higher than our recommended). Both warnings were played at the same intensity level, approximately 10 dBA above ambient background engine, and environmental road and wind noise while the simulator was running. An additional 29 other sounds, including alarms, status notification sounds, and social notification sounds previously examined were also presented. Specifically varied parameters for the two warnings are included in Table 2. Warnings played for a duration of 1600 (Edge Warning) to 2200 ms (GMU Prime), had onset and offset times of 10 ms, and multiple harmonics. The warnings had perceived IBI of 18 ms due to the onset and offset times. This is based on parameters established in medical alert standards defining perceived alert time, where “downtime” consists of any part of the sound below 90 percentage intensity (International Electrotechnical Commission, 2006).

Table 2. Varied physical warning parameters (motion-base simulator)

Warning	Pulse Duration	Base Frequency	Peak to Total Time Ratio
GMU Prime	200	1576 Hz	0.9
Edge Warning	400	3000 Hz	0.95

Participants were randomly assigned to one of five groups. Two groups received consistent warnings, either all GMU Prime or all Edge Warning for each of the three experimental drives. Two groups included “switch” conditions, where the warning they received was inconsistent throughout the course of three experimental drive events, and one group, the control group, received no warning during any of the experimental drive events. Groups and warnings are elaborated in Table 3.

Table 3. Group warning characteristics by drive

Group	Drive 1 Event	Drive 2 Event	Drive 3 Event
GMU Prime	GMU Prime	GMU Prime	GMU Prime
Edge Warning	Edge Warning	Edge Warning	Edge Warning
GMU Prime Switch	GMU Prime	Edge Warning	GMU Prime
Edge Warning Switch	Edge Warning	GMU Prime	Edge Warning
No Warning	No Warning	No Warning	No Warning

Twenty-nine additional sounds not included for use as warnings varied on all parameters, and were similar to those used in previous studies (see Lewis, Eisert, Roberts, & Baldwin, 2014). Additional sounds included currently in-use vehicle sounds (and some sounds designed by our labs) such as forward collision warnings, lane deviation warnings, curve speed warnings, fatigue alerts, backup and park assist sounds, seatbelt reminders, door open reminders, and various types of infotainment and social notifications.

2.3.1.3 Apparatus and procedure

The experiment was run in a Realtime Technologies, Inc., open-cab driving simulator on a motion-base. The motion-base allowed for 90 degrees of yaw motion to simulate turns, and one degree of pitch motion to simulate acceleration and braking. The visual component of the simulator included three 42-inch plasma displays, allowing for a 180-degree field of view (Figure 6). An RTI program called SimVista was used to create two simulated driving worlds and all scenarios. Data were collected at 30 Hz. Prior to the experiment participants gave written informed consent and verbally completed a motion sickness history screener to assess susceptibility to simulator adaptation syndrome (see Mollenhauer, 2004, for a review).

Participants scoring over a 7 on the questionnaire were given the option to opt out of the experiment. Only two female participants were unable to participate due to susceptibility.

Figure 6. Motion-base advanced driving simulator

After completion of the screener and informed consent, participants were introduced to the simulator. All participants were given basic safety instructions and were required to buckle the seat belt. Participants completed two practice drives prior to the first experimental drive. First, participants practiced driving. Participants were instructed to drive, following a lead car in front of them at a speed of 65 mph (though, due to the tightness of curves in the first driving world, participants were instructed to slow down when turning), remaining in the right-hand lane at all times. Participants drove until the experimenter felt comfortable with their driving performance (i.e., no skidding around turns, proper lane and speed maintenance). After the first practice drive, participants were introduced to the subsidiary task, a visual-manual 1-back task. The task required participants to monitor a small touchscreen to the right of the steering wheel which constantly presented numbers from 0-9 along with the words “YES” and “NO”. Participants were required to respond by pressing the corresponding affirmative or negative button based on whether the number presented matched or did not match the number presented directly preceding the currently presented number (see Figure 6). For drive 1, numbers would appear for 2 seconds during which time participants could respond. Responses, whether correct or incorrect were immediately followed by another stimulus or else the screen timed out after 2 seconds.

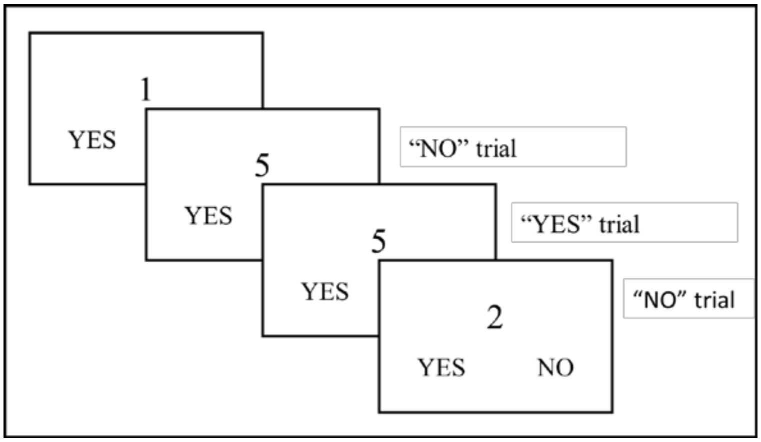


Figure 7. 1-back design and correct responses, overlapping screens indicate time progression

Once participants were comfortable completing the subsidiary task by itself, they completed the same drive as the first practice, this time while completing the subsidiary task. Participants were instructed that driving safely was to be their priority, meaning that if they were uncomfortable with the subsidiary task during complex maneuvers (such as rounding turns) they should stop doing the task, and return to it when they felt in control of the vehicle. Participants again completed the dual-task practice until the experimenter felt comfortable with their control of the vehicle during dual-task phases.

After completion of both practice drives, participants completed the first experimental drive. This drive was seemingly the exact same as the preceding practice drive, but about 3 minutes into the drive (after the first two turns) while driving on a straight section of roadway the lead car changed lanes suddenly, revealing a vehicle traveling at approximately 10 mph in the participant’s lane of travel. At this point, participants either received one of the two warnings or received no warning based on their group. Participants could either employ a hard brake or a swerve to successfully avoid collision. However, this event was designed to be difficult to avoid.

After the first experimental drive, participants were told that their vehicle was now a “connected vehicle” and given instructions for responses to sounds that they would hear. Participants were told that sounds could fall into one of three categories: alarms, status notifications, and social notifications. Refer to Appendix B for the specific instructions and definitions provided. Definitions provided to participants described an alarm as a time-critical highly urgent sound like a collision warning or a lane deviation warning, a status notification as a sound indicating something to do with the status of the car, like low tire pressure or a door is open, and a social notification would be something informing of a call or an email or a Facebook update. Each category was matched to appropriate responses where alarms should be responded to with a brake press, status notifications should be responded to by pressing a triangle indicator button that would appear in place of the secondary task and social notifications should be responded to by pressing a telephone button that would appear along with the triangle indicator on the touchscreen (see Figure 8).

Figure 8. Touchscreen images for "Status" and "Social" categorization

After practicing these responses participants completed two more experimental drives lasting 15 and 20 minutes where they responded to alarm, status and social notifications followed by an event. In drive 2 the event was a lead vehicle braking event and in the third experimental drive the event included a reveal event, identical to the event in the first experimental drive. For drive 2, an interstimulus interval (ISI) of between 4 and 7 seconds was added between trials on the 1-back task and there were between 9 and 13 1-back trials between sounds. The response window was still 2s. For drive 3 the ISI was adjusted to between 3 and 7 seconds and there were only between 4 and 7 1-back trials between sounds to decrease the amount of total time needed for the drive.

Due to different possible responses and appropriate actions for the second drive, analysis of the second drive event has been excluded from this report. The main comparison of interest is response to hazards in Drive 1 and 3, as these were the same type of event.

2.3.1.4 Design and data analysis

Independent variables included warning played (GMU Prime, Edge Warning, and no warning) and group (1-5) with main dependent variables of interest being collisions, evasive maneuver response time (EMRT) and speed at collision for those participants who did collide as an index of collision severity.

2.3.2 Results

Table 4 gives a breakdown of participant demographics by group. Every effort was made to ensure equal gender and age distributions across groups. Analysis of demographic data indicates that there were no significant differences in age or gender distribution by group.

Table 4. Breakdown of participant demographics by group

Warning Condition	Male	Female	Age	SD	Total
GMU Prime	5	15	19.6	1.67	20
Edge Warning	6	15	22.1	6.21	21
GMU Prime Switch	7	12	19.5	1.93	19
Edge Warning Switch	5	14	19.5	2.59	19
No Warning	6	16	20.3	2.64	22
Total	29	72	20.2	3.55	101

2.3.2.1 Drive 1

Results for drive 1 were analyzed in terms of warning played rather than by group, as groups 1 and 3 and groups 2 and 4 received identical alerts up until drive 2. Analysis of collisions by warning type (Figure 9) indicate that participants who received a warning collided somewhat (though not significantly) less often than did participants who received no warning, $F(1,99) = 1.84$, $p = .178$.

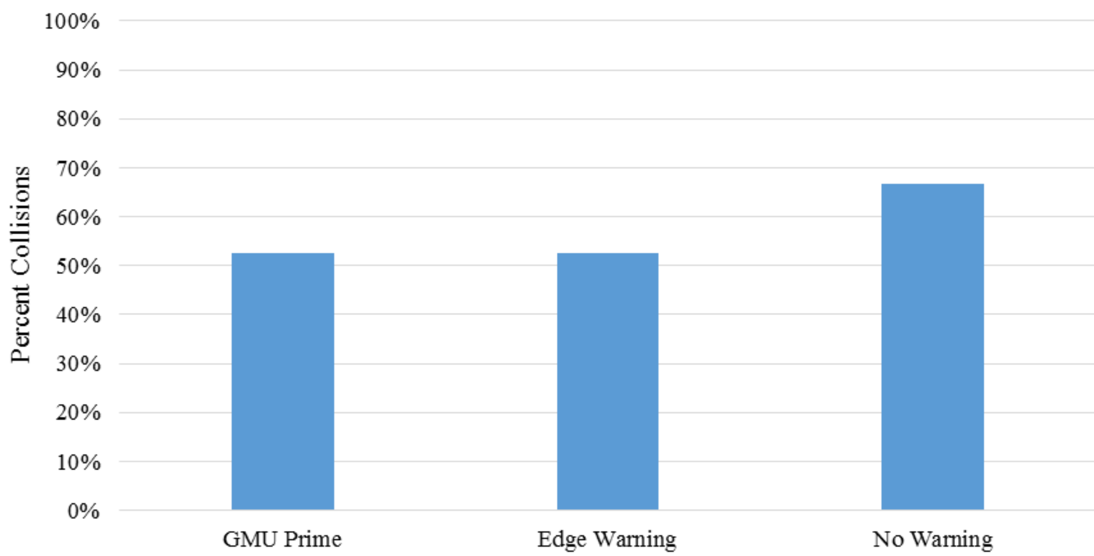


Figure 9. Drive 1 collisions by warning type

Further analysis of data for participants who collided indicated that, although there were not significant differences in number of collisions between warning types, there were significant differences in speed at time of collision, $F(2,53) = 4.01$, $p = .024$. This variable represents the speed of the participant's vehicle at the time that it collided with the stopped reveal car and can be considered as a metric of collision severity. Figure 10 shows results of speed at collision data for those participants who collided (a measure of collision severity). Post hoc comparisons using the Tukey HSD test indicate that only the difference between GMU Prime and no warning was significant ($p=.019$), where Edge Warning did not vary significantly from either group.

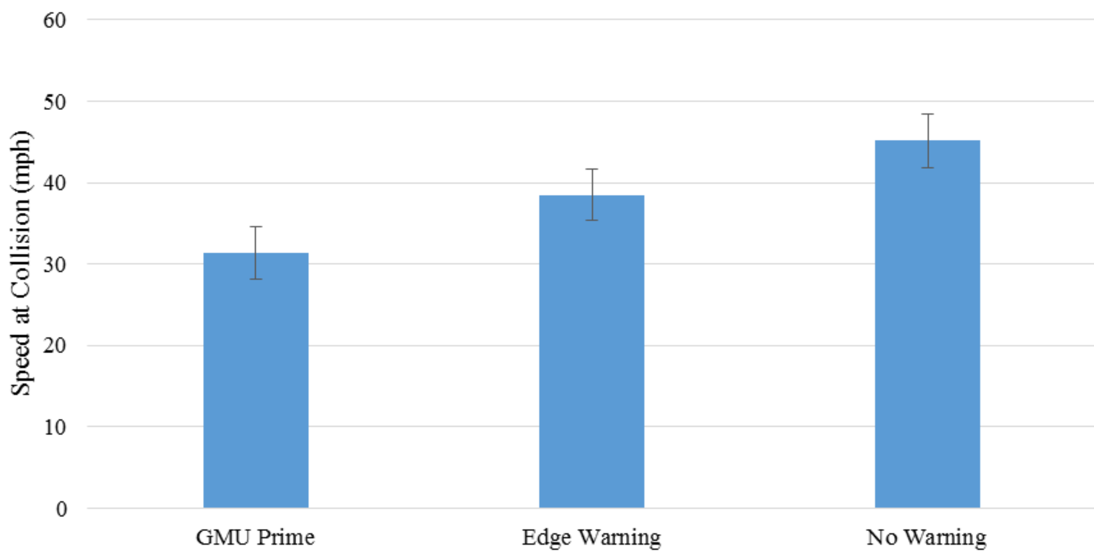


Figure 10. Drive 1 speed at time of collisions (if participant collided) by warning type

Further results indicate that brake response time (BRT), though not EMRT, varied significantly by group, $F(2,66) = 3.28$, $p = .044$. Figure 11 shows differences in BRT by warning. Tukey HSD post hoc comparisons indicate that again only the difference between GMU Prime and no warning was significant ($p =. 050$), a difference of about 350 ms.

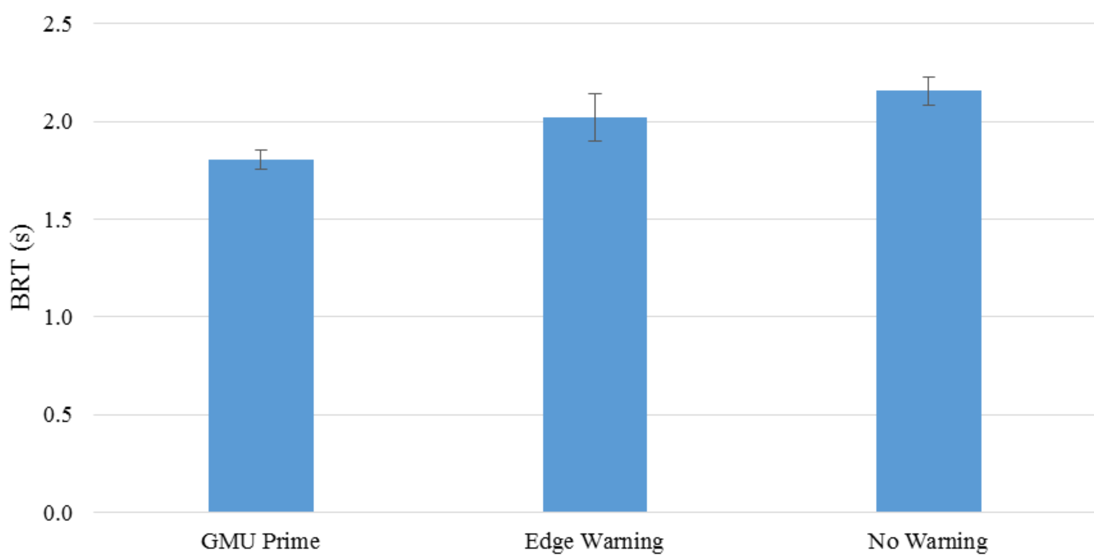


Figure 11. Brake response time by warning

It is important to point out that this event happened very early into the experiment, after adequate but short practice with the simulator. All participants included in this experiment indicated that they had no previous experience using a motion based driving simulator. Additionally, the event itself was designed to be representative of extremely difficult collision events and therefore the very high collision rate overall is not unexpected. Importantly, despite the high collision rate for all groups, results indicate that participants receiving GMU Prime collided at about 15 mph less than participants receiving no warning.

2.3.2.2 Drive 3

Data from 18 participants are not available for drive 3 due to various factors including simulator sickness, equipment failures, and corrupted data. Results therefore reflect data from 89 participants (24 male, 65 female, mean age = 20.4 years). There were no significant differences in age or gender by group. Results for drive 3 collision data indicate overall decreases in collision rate, such that participants in the group that consistently received GMU Prime decreased their collision rates by almost half (Figure 12).

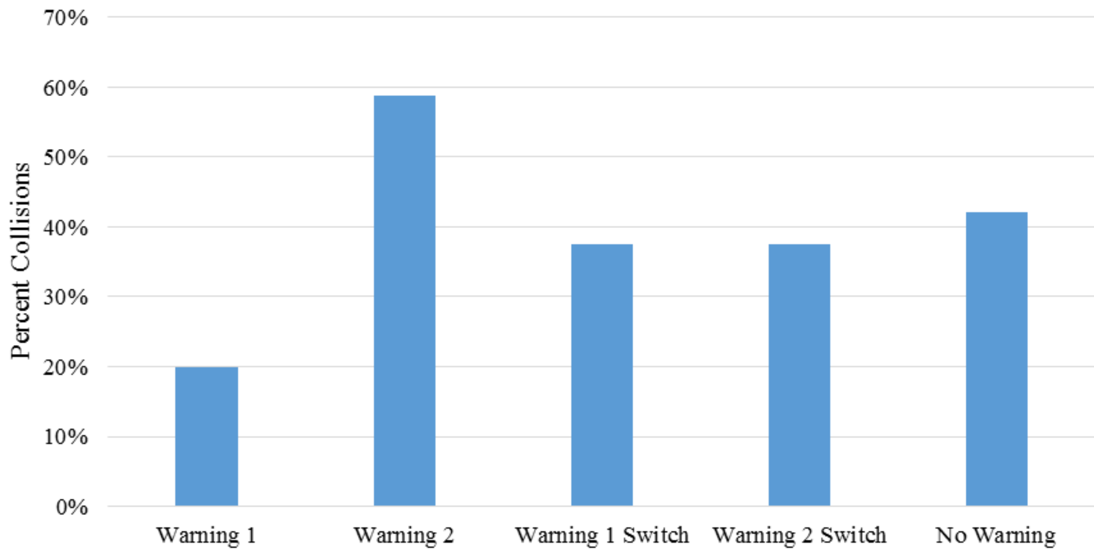


Figure 12. Percentage collisions by group

Collision rates between groups are not statistically significant. However, they may be practically significant in that only 20 percent of drivers in the consistent GMU Prime condition collided compared to over 40 percentof drivers in the no warning condition. It also would appear that drivers in the consistent Edge Warning condition were somewhat more likely to collide than all other conditions, and even, strangely, than they were in drive 1. Figure 13 shows a comparison between collision rates by group for drive 1 versus drive 3. Interestingly, all groups show some level of learning for the reveal event except for the consistent Edge Warning group. Participants consistently receiving Edge Warning actually seemed to collide slightly more often, contrary to expectations. This finding may point to Edge Warning being particularly unreliable, or may reflect anomalous results.

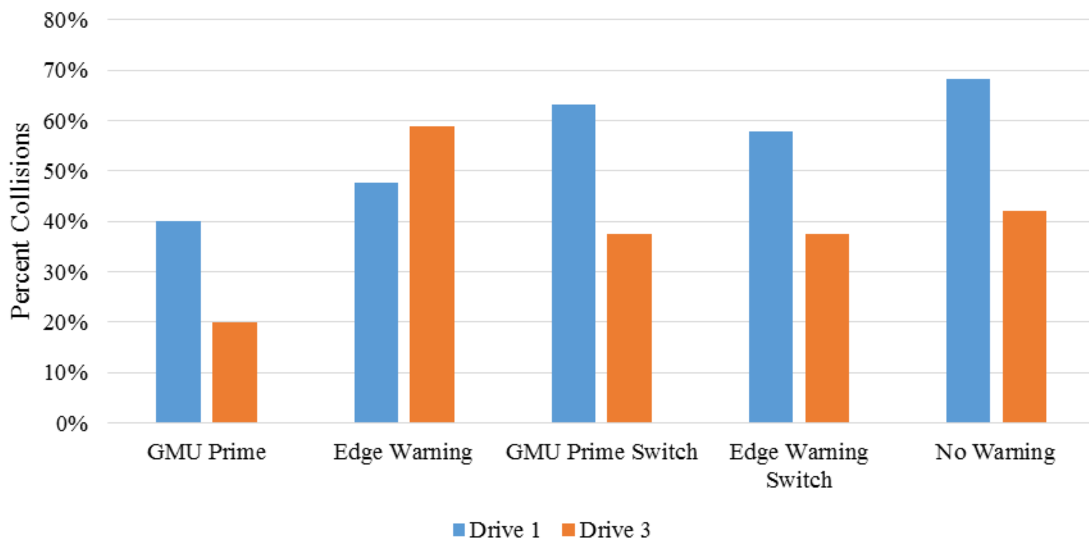


Figure 13. Comparison of collisions between drives 1 and 3

Further analysis of drive 3 collision data indicates no significant differences in speed at collision for participants who collided. The few participants in the consistent GMU Prime condition who did collide, however, did so at a lower speed than did participants in all other conditions (Figure 14). However, it was found that there was a homogeneity of variance (as assessed by Levene's Equality of Variances Test) therefore an independent samples t-test was conducted. Results indicate that there were significant differences between the GMU Prime group (M = 14.7 mph, SD = 8.3 mph) and the warning two group (M = 42.3 mph, SD = 18.5 mph); $t(11) = -2.46$, $p = .032$.

It is also important to note that participants in the consistent GMU Prime and no warning condition showed some effect of learning, where their speed at the time of collision if they collided in drive 3 was, in general, lower than their speed at time of collision if they collided in drive 1. However, for the consistent Edge Warning and both switch groups, there was not a large decrease in speed at the time of collision for participants who collided (Figure 15).

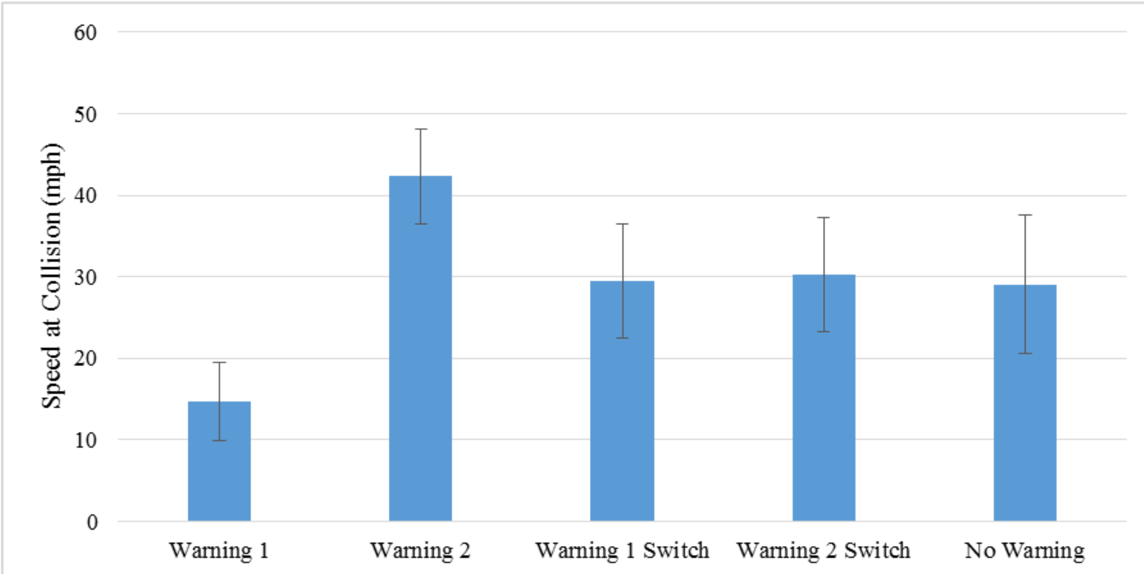


Figure 14. Speed at time of collision for participants who collided by group

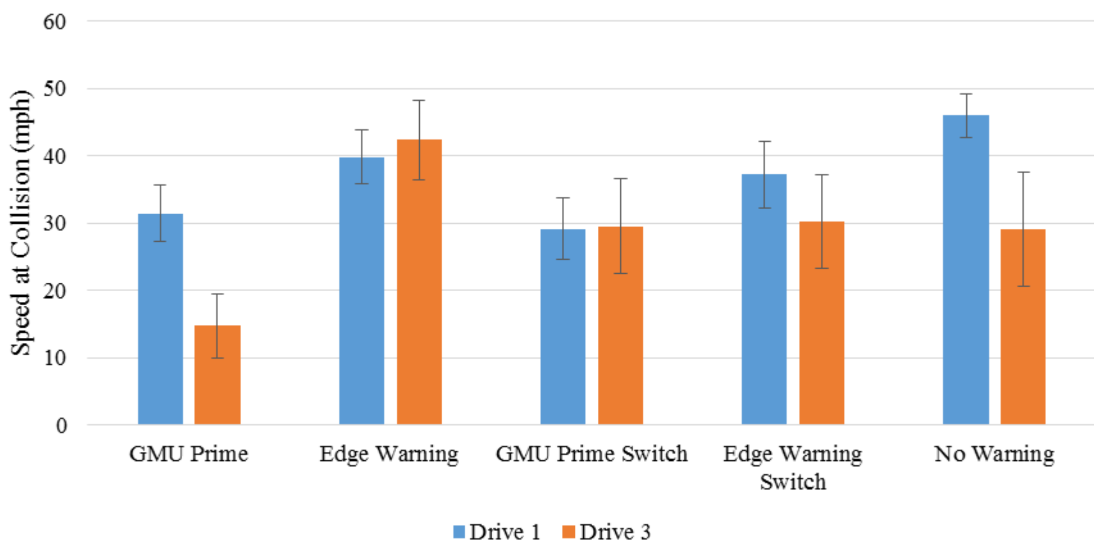


Figure 15. Speed at time of collision by condition and drive

2.3.2.3 Video data

Video data was further analyzed in terms of both subjective reactions to alerts and types of responses. Video data included one color and sound-enabled webcam, trained onto participants' faces and three black and white surveillance cameras (soundless) trained onto participants' faces, feet, and the forward roadway.

For drives 1 and 3 surveillance video data were obtained from 60 and 53 subjects, respectively. Data were lost due to a variety of factors including video failures and inability to adequately make out the face of the participant due to poor lighting. Data were coded by three independent coders who were blind to the participant's assigned experimental condition. Data were coded as the level of subjective danger of the participant's response. For example, if a participant came to a complete stop, before hitting the revealed vehicle, that response had little or no danger and was indicated as a 0. Participants who still tapped the revealed vehicle, or swerved with incomplete control and without checking their blind spot would be marked between 1 and 4. Participants who hit the lead car at speed or swerved into oncoming traffic or off the road were given a 5. Participant eye position at the time of alert (up or down), response (brake or swerve) and outcome (avoid, collision or secondary collision where participants avoided the revealed vehicle but hit the lead or another vehicle) were also analyzed. No significant results or notable trends were found.

For drive 1, further analyses of visual scanning patterns at the time of the warning were completed using the color and sound webcam video. Data from 62 participants were available. Possible responses included good, non-confused scanning, indicated by an immediate orientation to the forward roadway with a quick reaction and poor, confused scanning, indicated by looking down to the dashboard or the touchscreen, back at the researcher, or around in the car cab before orienting to the forward roadway. Results indicated that scanning was slightly better for GMU Prime than Edge Warning, with 12 out of 30 subjects (67%) exhibiting non-confused, good responses, for GMU Prime and only 10 out of 32 subjects (45%) exhibiting non-confused, good responses for Edge Warning.

2.3.2.4 Drive 3 categorization

Despite the loss of participant data for drive 3, all but two participants were still able to complete the sorting task during drive 3. Therefore, included in this analysis are 98 participants (30 male, 68 female, mean age = 20.4 years).

Results of sound categorization data indicate that participants respond very similarly in the higher-fidelity, moving-base simulator as they do in a similar paradigm using the desktop simulator. Analysis of results by way of criteria met again indicates that sounds that meet more criteria are more likely to be categorized as alarms in a connected vehicle context (see Figure 16). Further, this study confirms previous findings that participants typically respond more quickly to sounds that meet most or no criteria than they do to sounds which meet some criteria, $F(4,26) = 3.50$, $p = .02$ (see Figure 17).

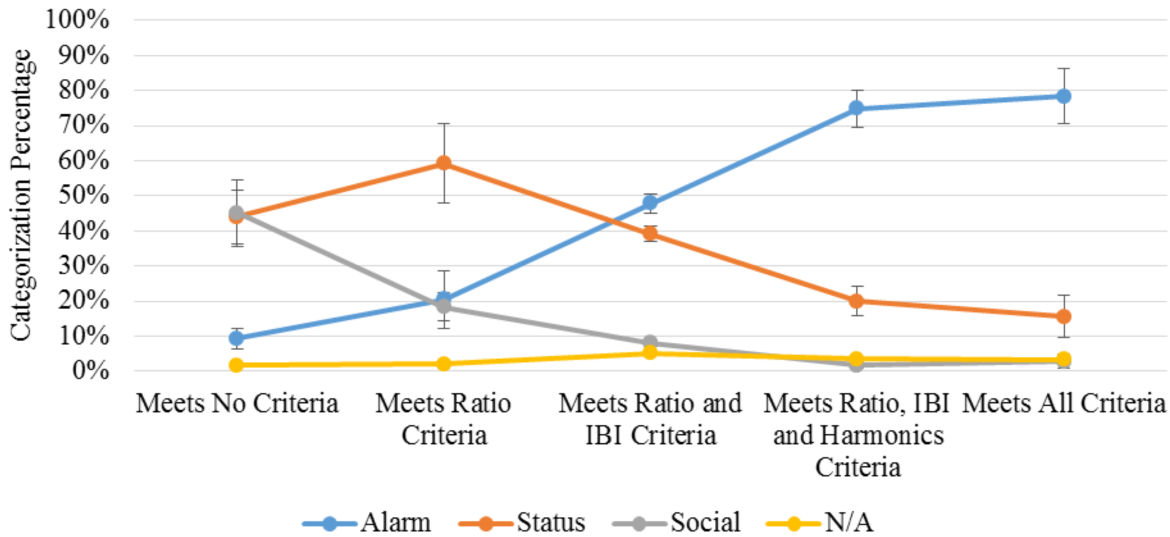


Figure 16. Percentage categorization by criteria met for drive 3 sort task (N/A indicates no response and "meets all criteria" refers to all five recommended criteria)

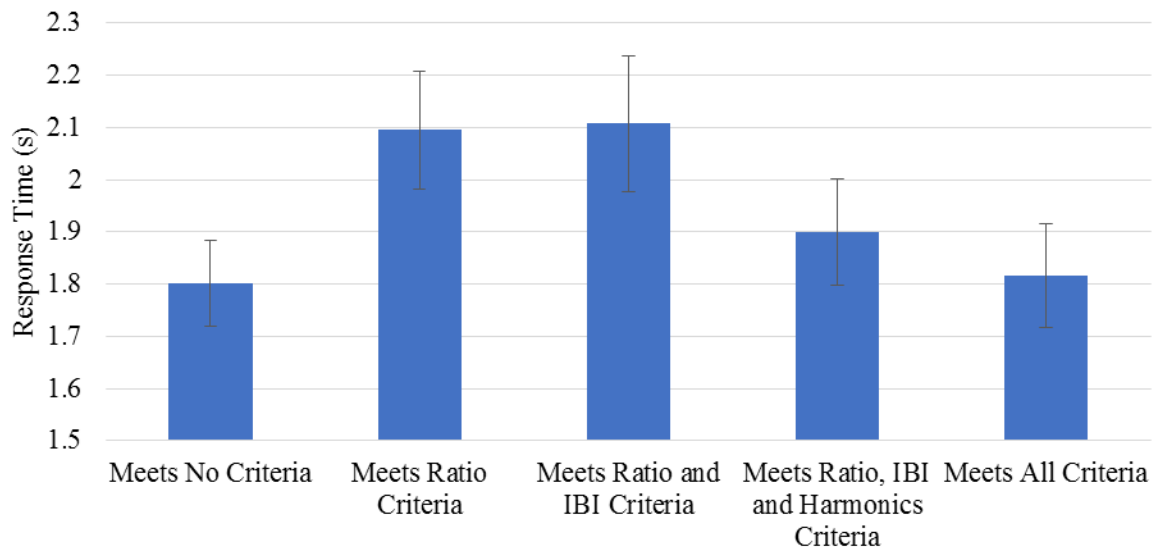


Figure 17. Mean response times by criteria met

2.3.3 Discussion

Results from Experiment 2 confirm findings in Experiment 1 that the warning meeting all key criteria performs better than the warning only meeting some criteria, as reflected by decreased collisions and decreased collision severity for participants who do collide. For the participant's first exposure, there is evidence than any warning is better than no warning, in terms of faster response times, decreased collision rates and decreased speed at the time of collision for those participants who did collide compared to participants receiving no warning. However, this effect does not hold through to the last drive. Participants receiving only GMU Prime, the warning meeting all five criteria, do perform the best overall compared to other groups in terms of collisions and speed at time of collision. Participants consistently receiving Edge Warning, the warning missing two key components (frequency too high and pulse duration too long), performed the worst of all groups. Participants consistently receiving an edge warning (one that does not meet all criteria) or inconsistent warnings did not show effects of learning, as reflected by no changes in speed at the time of collision for participants who collided, where participants in the GMU Prime and no warning condition did show decreases in collision severity by drive 3. It is important to point out that there was a difference in collision rates and that the data available for collision severity is limited, particularly for the GMU Prime group. However, as the goal of collision warnings is to decrease collisions among those who are susceptible, as demonstrated here by a distracting technique, it may be less important to help the people who weren't going to crash anyhow. The current data show that people overall crash less with the good warning and even those who still crash despite receiving the best warning available, at least do so at a lower rate than no warning or poorer warnings.

Analysis of sorting data from the drive 3 categorization task further validate results obtained from previous work. Results indicated, as previously seen, as sounds meet more of the recommended criteria they are more likely to be identified as alarms. Further, response time results indicate that participants responded fastest to sounds that met none or almost all criteria, indicating that sounds that met none of the criteria were unambiguously non-urgent and sounds that met nearly all criteria were unambiguously urgent. Sounds that met a few criteria were more ambiguous and therefore participants categorized them more slowly.. This result is also consistent with previous work.

The two validation experiments discussed here provide support for the need to carefully consider the parameters necessary to ensure that an auditory FCW will be recognized as a highly urgent signal. The present validation of the previously established parameters indicates that variations in frequency outside a range of approximately 1000 to 2500 Hz can be expected to induce a level of perceived uncertainty thus reducing driver response time and increasing inappropriate collision avoidance responses. Temporal factors work interactively to impact perceived urgency levels and FCW signals should consist of several pulses that have fast perceived temporal rates, which can be achieved by pulse durations of 200 ms or less and IBIs of less than 20 ms. Finally, ensuring that the sound quickly reaches its maximum sound level and maintains that maximum for at least 90 percent of its pulse duration is associated with a high perceived urgency level. Utilizing all five of these parameters simultaneously leads to the greatest crash reduction capability, the quickest driver response and the most reduction in crash severity relative to utilizing only some of these criteria. The current simulator validation studies support previous research establishing these auditory criteria as critical to effective FCW design and warrant further validation in on-road investigations that are currently being conducted in association with this project.

3. Tactile Parameters and Urgency Category

3.1 Introduction

The purpose of this experiment was to assess how people perceive and categorize signals presented in the vibrotacile modality. Specifically, it was intended to give preliminary definitions for the perceptual space that make up the category of highly urgent vibrotactile collision warnings. This was done in two steps. First, a set of existing vibrotactile signals was compiled, based on a literature review. Then, the signals were included in a perceptual sort (identical to the methods used for previous auditory sorts, see Lewis, Eisert, Roberts and Baldwin, 2014, for more detail), in order to discern criteria and cutoffs for highly urgent, vibrotactile alert signals.

3.2 Methods

3.2.1 Participants

Five male and 20 female undergraduate and graduate students were recruited from the student research participant pool. Participants received a small amount of research participation credit that could be applied to their classes.

3.2.2 Stimuli and apparatus

Stimuli used in the tactile sort task were created based on in-vehicle tactile alerting systems and tactile navigation systems investigated in published research. Stimuli varied on pulse duration, interpulse interval (time gap between each pulse within a burst), pulses per burst, burst duration, IBI, total number of bursts and total alert time as indicated in Table 5. There were 27 total stimuli.

Table 5. Varied tactile parameters and their ranges

Parameter	Range Investigated
Pulse Duration	10 ms – 3000 ms
Interpulse Interval	10 ms – 2000 ms
Pulses per Burst	1 or 2
Burst Duration	10 ms – 3000 ms
IBI	10 ms – 2000 ms
Total Number of Bursts	1-75
Total Alert Time	150 ms – 5000 ms

Stimuli were created as sound files in Adobe Audition CS6 and were “played” similarly to sounds through a C2 tactor connected to an amplifier. The C2 tactor was placed on participants’ left wrists over a small piece of plastic wrap and secured with a wrist sweatband.

Stimuli were presented using Microsoft PowerPoint as shown in Figure 18. Embedded in the slide were numbers which, when clicked, initiated the tactile sequence similarly to past auditory work. After playing stimuli, participants could move initiator buttons into one of three categories: “Alarms,” “Status Notifications,” and “Social Notifications.” Each category also had space such that participants could justify their inclusion characteristics and give an urgency level for each category. This design is identical to those used in previous work (Lewis et al., 2014).

3.2.3 Procedure

At the beginning of the experimental session, participants provided written informed consent. After providing consent, participants completed a short demographic questionnaire then the experimenter placed the tactor on the participant's wrist. Participants were then instructed that they should click on numbers to play stimuli and then would be able to move numbers into whichever category they deemed appropriate. Categories were defined both in writing and verbally for participants. After all signals had been sorted, participants were asked to describe the characteristics of each category and to assign a number representing the urgency level of each category (between 1 and 100). Participants were given as much time as was needed to complete the paradigm, typically less than 30 minutes.

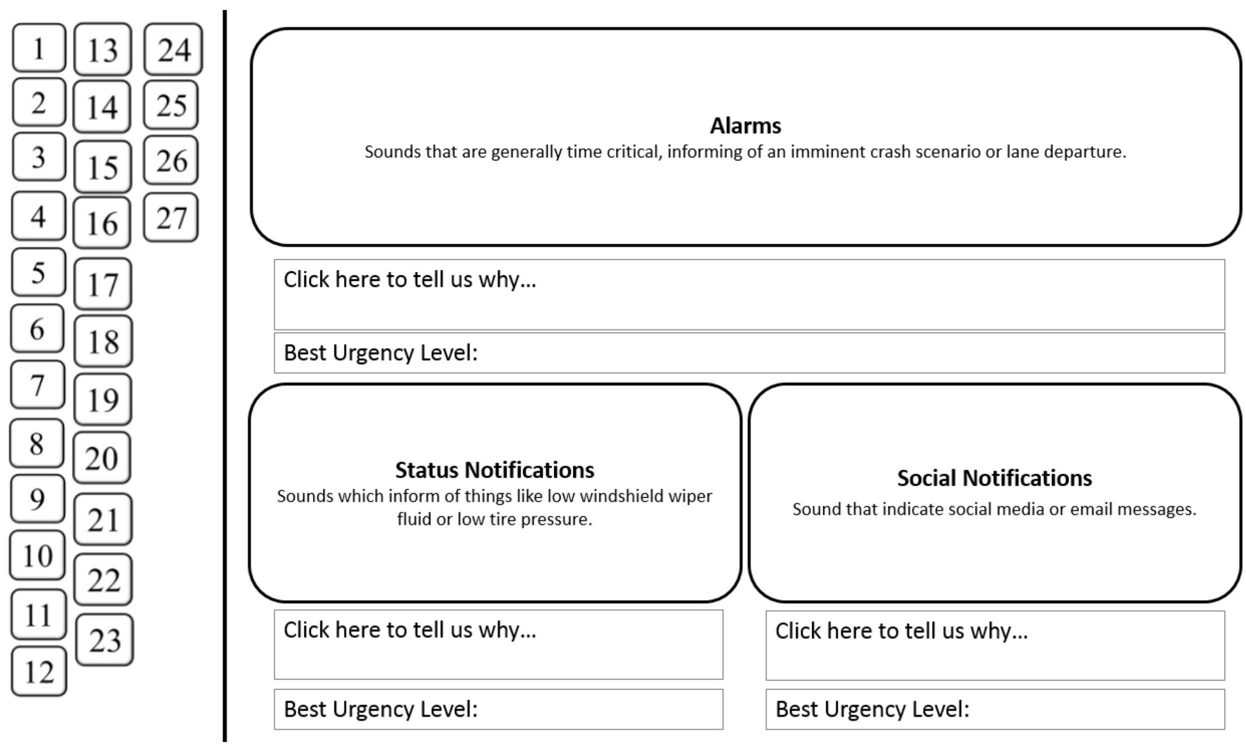


Figure 18. Tactile sorting space

3.3 Results

Results were analyzed identically to previous auditory work. The properties that explained the most variance in alarm categorization were determined using a backwards, stepwise logistic regression to predict "Alarm" category membership from all varied parameters. Specifically, IBI, burst duration, number of bursts and total signal time predicted about 56 percent of the variance in alarm categorization for a significant overall prediction model, $F(4,22) = 7.06$, $p = .001$. Using these results, cutoff criteria were determined in order of decreasing importance for each property of interest (as in previous work). Criteria and their cutoffs are reported in Table 6.

Table 6. Criteria and cutoffs for alarm categorization in tactile sort task

Criteria (in Order of Decreasing Importance)	Cutoff
IBI	≤ 150 ms
Burst Duration	≤ 200 ms
Number of Bursts	≤ 10
Total Time	≤ 1000 ms

Figure 19 shows percentage categorization by criteria met. Signals that met all criteria were classified as alarm signals nearly 80 percent of the time. It should also be noted in this preliminary study that there was not a large number of signals per category, with only one signal meeting IBI, Burst Duration, and Number of Bursts (this lack of variance is reflected in the lack of error bars for this signal).

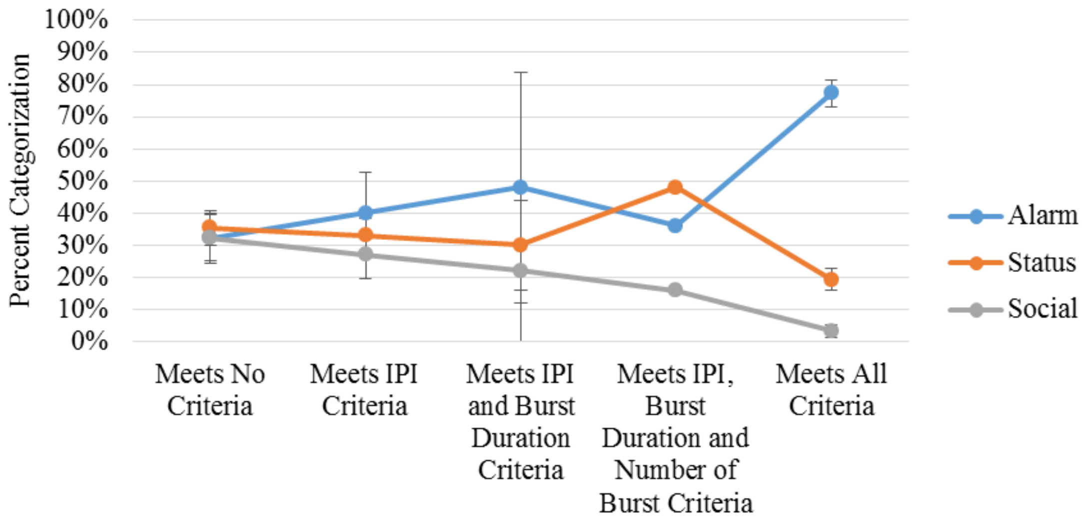


Figure 19. Percentage categorization by criteria met

3.4 Discussion

The results of the tactile perceptual sort task provide very preliminary data regarding parameters that contribute the most to perceptions of time-criticality and urgency in the tactile domain. As these criteria are verified based on the same dataset from which they were created, these results should be interpreted cautiously. These preliminary data suggest that there may be systematic changes which can increase the likelihood of a tactile alert being identified as an alarm, however, this dataset is limited in that the tactile signals used were only signals previously evaluated in the literature. It is possible that new stimulus combinations or those being used in current vehicles may shed more light on appropriate parameters contributing to alarm perception.

Unlike the auditory modality, the tactile modality affords fewer parameters that may be varied systematically and realistically in a vehicle contributing to previous findings that it is unlikely that drivers will be able to differentiate tactile signals as readily as auditory signals in an in-vehicle context (Fitch, Hankey, Kleiner, & Dingus, 2011; Meng & Spence, 2015). The context in which these tactile signals were presented to participants also differed in notable ways from the vehicle context. The signals used in this experiment were presented using a tactor attached to the wrist, which is different than current in-vehicle applications of tactile signals. Also, a moving vehicle inherently provides a vibrotactile environment due to vibrations from the road and engine, and this background vibration could potentially affect perception of vibrotactile signals. These limitations are addressed in the subsequent on-road research described in the next section of this report.

4. On-Road Experiment

4.1 Introduction

The on-road experiment described in this section was intended to validate earlier GMU findings and design criteria developed within the CWIM research program, and also to complement the concurrent research described in Section 3 of this report. For the present experiment, the behavioral measures were perceptual (detection, response time, perceived urgency/meaning) but were collected under actual on-road driving conditions and with two levels of in-vehicle ambient noise. Thus the intent was to confirm that laboratory and simulator findings could be confirmed in an actual driving situation and under both baseline and louder ambient noise conditions.

In addition to the perceptual methods, in which drivers responded to auditory and vibrotactile signals that they knew might occur, this experiment also investigated driver reaction to an unanticipated auditory signal. This was done in an initial portion of the drive, before drivers were exposed to the listening procedures of the perceptual ratings. The intent in this portion of the experiment was to see if drivers responded in a different manner to unexpected signals that varied in terms of their subjective category of meaning. A prototypical example of an auditory signal from each of three different categories was selected: an urgent alarm, a non-urgent status notification, and a social notification. The primary question is whether the perceptually more urgent signal resulted in more rapid or otherwise different visual search behavior than less urgent signals. These drivers knew they were in an experimental setting and the occurrence of an alert may not be "unexpected" in the same sense as a driver operating their own vehicle in normal driving. Thus while the absolute aspects of reaction to the signal may not be directly meaningful, this portion of the experiment was directed at comparing the relative response of participants to different categories of auditory signals.

4.2 Method

4.2.1 Design

This section provides an overview of the study design and methods. Further detail is in the sections that follow. The experiment included 60 participants who engaged in normal driving on a limited access highway under free-flow conditions. There were two parts to data collection. Descriptively these may be considered as two distinct experimental designs and methods. In the first phase of the experiment, participants were presented with an unexpected auditory signal while they were visually distracted by a text-reading task. Driver responses to three different signals were compared in a between-groups design (n=20 per group). The three auditory signals were prototypical examples of an urgent alarm, a driving-related non-urgent status notification, and a social notification. Formally, this is a single factor (message type), between-groups design. Data were collected on the timing and direction of eye glances (derived from video recordings), vehicle dynamics, and other overt behaviors (e.g., spontaneous verbalizations, shifting of posture or hand position, etc.). The analyses were directed at determining whether auditory signals that differed in subjective urgency and meaning (based on previous research) resulted in observable differences in driver behavior when presented to an unfamiliar and unsuspecting driver.

The second phase of the session presented a set of 17 auditory signals and 9 vibrotactile (seat pan vibration) signals to participants while they were driving. These 26 unique signals were presented twice: once under baseline driving conditions and once while music was playing at 75 dBA as measured at the driver's ear position (order of noise conditions was counterbalanced). Participants responded to each sound or vibrotactile signal they detected with a verbal response that indicated the category of meaning/urgency to which they assigned the signal. These categories, defined for the participant exactly as they were in the previously-described driving simulator experiments, were:

> 1 = Alarm (time-critical, highly urgent signal like a collision warning)
>
> 2 = Status notification (something to do with the status of the car, like low tire pressure or a door is open)
>
> 3 = Social notification (something telling you that you have a call or an email or a Facebook update)

This procedure defines a two-factor, within-subjects (N=60) design, with the factors of ambient noise condition (2 levels) and alerting signal (26 unique auditory and vibrotactile stimuli). The 26 signals varied on a variety of dimensions, such as modality (auditory, vibrotactile), nominal message category, and various signal parameters.

Several dimensions of driver perceptual response to the signals were recorded for analysis. These were: detection of the signal (detected or not detected), verbal response time to indicate the type of message, and perceived message category. The analyses were directed at determining driver perception of various signals while engaged in actual on-road driving, under benign and more demanding ambient background noise conditions. While the results serve as stand-alone findings, they are relatable to previous laboratory and driving simulator findings and the set of criteria recommended to define design spaces for in-vehicle signals of different urgency.

4.2.2 Participants

Participants included 60 drivers aged 21 to 50 (mean age = 35.7), with 30 males and 30 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. Anyone with a history of serious moving violations or suspensions was excluded from the study. No participants dropped out or were removed from the study.

Participants were recruited through the Volunteers section of Craigslist and through a news item posted on Westat's intranet homepage. Participants received $75 for completing the session. Prospective participants completed a screener questionnaire. The screener questions concerned age, gender, license status, and familiarity with various types of vehicles. It also included a set of questions related to hearing impairment. A recruitment ad and the telephone screener are shown in Appendix C and Appendix D, respectively.

4.2.3 Apparatus

All participants drove Westat's WesDRIVE instrumented vehicle. This vehicle is a 2011 Subaru Outback instrumented with sensors and video cameras to capture data from the vehicle and the environment.

Data collection took place on a limited access toll highway (Maryland Route 200) running east to west in Montgomery County, with a 60 mph speed limit. Participants traversed this route between Shady Grove Road and Briggs Chaney Road in both directions until data collection was complete. This span of roadway was about 13 miles in length (one way). This is a relatively new highway with smooth and uniform asphalt over most of its length. It is also generally free-flowing, with low traffic volumes. 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. Experiment sessions were conducted between 9 am and 4 pm to avoid rush hour traffic and sun glare while driving.

During the experiment session, participants wore a head-mounted noise cancelling microphone that was connected to the experimenter's laptop. This allowed the researchers to record participants' verbal response time to categorize a stimulus. The microphone was used only for response time detection and no recordings were made using this microphone. The microphone model used in this experiment was theBoom E (see Figure 20).

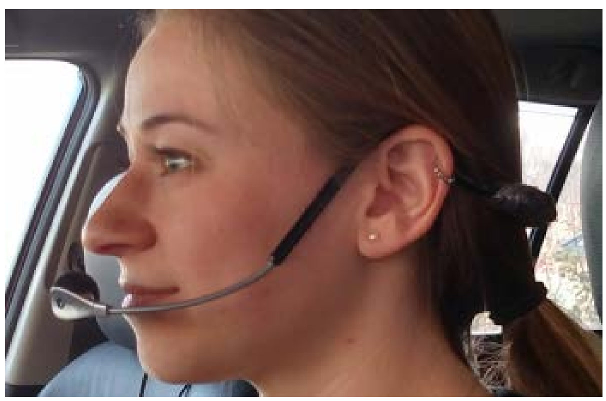


Figure 20. Verbal response microphone

4.2.4 Ambient 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 Music On condition, the song "Café Amore" by Spyro Gyra played in a continuous loop. The song could be categorized as instrumental smooth jazz. It was selected because it had been used in previous research (Brodsky, 2001; Lerner et al., in press) and has a medium tempo and relatively constant loudness through the duration of the track. The song has a dynamic range of 14 dB, where dynamic range refers to the difference between a song's maximum sound pressure level (SPL) and its average SPL. The music was played through the WesDRIVE vehicle's sound system using the AUX input on the vehicle's aftermarket Pioneer sound system head unit. The bass, treble, balance, and fade settings were set to neutral "0" values. The volume of the music was set so that the average loudness was 75 dBA, as measured in the WesDRIVE vehicle at the driver's head position with the engine turned off. Sounds were presented through four of the vehicle's speakers simultaneously: left and right A-pillar and left and right front door speakers.

Ambient SPL fluctuated somewhat within each ambient noise condition based on a variety of factors (e.g., vehicle speed, wind, dynamics of the music), but typical levels for the baseline condition were approximately 64 dBA and typical levels for the music condition were approximately 76 dBA.

4.2.5 Signals and stimulus presentation

There were 17 auditory signals and 9 vibrotactile seat pan vibration signals in this experiment, for a total of 26 stimuli. Twenty-four of these stimuli were used in the previous validation study conducted by GMU. The other two stimuli were forward collision warning (FCW) auditory alerts recreated from production vehicles, which were used in previous research by the project team (Lerner et al., in press). In addition to the experiment signals, additional distinct signals were used for practice trials. Each auditory signal is briefly described in Table 7 and each vibrotactile signal is described in Table 8.

Table 7. Auditory signal descriptions

Name	Description	Duration (s)
Auditory alarm 1	Signal created by GMU that meets all five criteria associated with identification as an urgent warning (GMU Prime)	2.2
Auditory alarm 2	Signal created by GMU that meets three of the five criteria associated with identification as an urgent warning (Edge Warning)	1.6
Auditory alarm 3	FCW recreated from a production vehicle	1.4
Auditory alarm 4	FCW recreated from a production vehicle	1.5
Auditory alarm 5	FCW recreated from a production vehicle	2.0
Auditory alarm 6	FCW recreated from a production vehicle	2.0
Auditory status 1	Seat belt reminder recreated from a production vehicle	1.3
Auditory status 2	Seat belt reminder recreated from a production vehicle	1.9
Auditory status 3	Backing alert recreated from a production vehicle	1.4
Auditory status 4	Park assist alert recreated from a production vehicle	1.6
Auditory social 1	Door open alert recreated from a production vehicle	1.5
Auditory social 2	Infotainment sound recreated from a production vehicle	1.1
Auditory social 3	Infotainment sound recreated from a production vehicle	0.6
Auditory social 4	Infotainment sound recreated from a production vehicle	0.2
Aud. ambiguous 1	FCW recreated from a production vehicle	1.5
Aud. ambiguous 2	Park assist alert recreated from a production vehicle	1.6
Aud. ambiguous 3	FCW recreated from a production vehicle	1.5

Table 8. Vibrotactile signal descriptions

Name	Duration per pulse (ms)	Duration of gap between pulses (ms)	Total number of pulses	Total signal duration (s)
Vibrotactile alarm 1	100	100	8	1.50
Vibrotactile alarm 2	30	30	25	1.47
Vibrotactile status 1	250	250	3	1.25
Vibrotactile status 2	60	270	5	1.38
Vibrotactile social 1	1500	0	1	1.50
Vibrotactile social 2	200	800	2	1.20
Vibrotactile ambiguous 1	326	261	3	1.50
Vibrotactile ambiguous 2	990	0	1	0.99
Vibrotactile ambiguous 3	1000	0	1	1.00

The name given to each signal refers to how it was categorized by participants in the sort task conducted in the motion base simulator sort task or the tactile parameters experiment rather than its intended function in the source vehicle. For example, while some FCW alerts from production vehicles are categorized as alarms, others are categorized as ambiguous signals (i.e., signals for which fewer than half of participants agreed on a single categorization of the sound). (The exceptions to this rule were auditory alarms 5 and 6, which were FCW signals recreated from production vehicles, but not used in the previous sort task.) All signals, as well as a pink noise signal (i.e., a signal of random noise containing equal amounts of energy per octave), were adjusted for equal perceived loudness using the "Perceived Loudness" adjustment in Adobe Audition CS6. This adjustment increased or decreased the loudness of each signal so that all signals would be of equal perceived loudness when accounting for the human ear's differential sensitivity to different frequencies. Signals were presented in the vehicle at a nominal level of 70 dBA at the driver's head position, as measured using the loudness-matched pink noise signal.

Auditory signals were WAV files played from a laptop computer through the vehicle's four front speakers, just as the music was played. Vibrotactile signals were presented via four C-2 tactors located below the surface of the driver's seat (see Figure 21). All vibrotactile stimuli were presented at the maximum intensity of the tactors, with the maximum rate of vibration (250 Hz). All four tactors were synchronized to present the same vibrotactile signal at the same time.

Figure 21. Experiment vehicle driver's seat viewed from above (tactor locations indicated by arrows)

It is important to note that the signals that were reproduced from in-vehicle systems were 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. Similarly, the vibrotactile signals used in this experiment recreated the pulse patterns of the source signals, but the intensity and seat pan location might differ from use in previous research.

Within each noise condition block, the experimental control software generated a random presentation order for the 26 signals. The software provided a random time gap between signals that ranged from 20 to 40 seconds. 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 auditory circumstances (e.g., a large truck passing or a patch of noisier roadway surface).

4.2.6 Procedure

4.2.6.1 Greeting and initial instruction

Upon arrival, the participant's driver's license was checked to confirm identity and status and the participant read and signed an informed consent form. The participant then sat in the driver's seat of the test vehicle and adjusted the seat position and mirrors. The experimenter was seated in the rear right seat with a laptop computer for experiment control and a live video monitor showing the participant's face.

The complete set of instructions to the participants is attached in Appendix E. The general purpose and procedure were first explained to the participant as an overview. Safety priorities were made clear and participants were asked to silence their cell phones. The participant was told that the WesDRIVE vehicle had intelligent warning systems that would be active during the session, and could provide warnings about potential safety threats. This instruction was included so that participants might think of the warning system when they receive the surprise alert during the experiment.

This was followed by a period of vehicle familiarization, during which the participant drove the vehicle around the parking lot. Next, the participant drove from Westat to the experiment location on Route 200. This drive took less than 10 minutes. Once on Route 200, the experimenter instructed the participant to maintain a speed close to 60 mph and stay in the right lane unless there is a need to pass a slow vehicle.

4.2.6.2 Surprise auditory alert

Next, the experimenter provided instructions for the surprise event. The participant was told that he or she would be provided an opportunity to practice reading messages on an in-vehicle display to get comfortable doing so before the experiment trials began. The experimenter then presented a series of three text messages on a display screen mounted low in the vehicle's center stack (see Figure 22 for display screen location, see Figure 23 for the three messages). Each message gave a navigation-related instruction. The first message had one line of text, the second message had three lines of text, and the third message had four lines of text. For each message, participants were instructed to read the message silently as quickly as possible, and then say "done." They were also instructed that the messages were for reading practice only, so they should not obey the instruction. Each message was displayed for six seconds, then disappeared.

Figure 22. Display screen with text message

Figure 23. Text messages presented on center stack display

The experimenter only triggered a text message on a straight section of roadway when no surrounding traffic was present. When the third text message was presented, the experimenter waited until the participant was visually committed to the display, then triggered an auditory signal. The experimenter had a live video monitor feed of the driver's face to help determine when the driver's gaze was directed to the display. The participant received either an urgent alarm (Auditory alarm 1), a status notification (Auditory status 1), or a social notification (Auditory social 1), according to a counterbalanced assignment. About five to ten seconds after the participant received the surprise alert, the experimenter asked the participant what he or she thought the signal meant, how they reacted to it, and why they reacted that way. The participant's responses were documented for later analysis.

4.2.6.3 Rapid categorization of alerts

Next, the experimenter directed the participant to exit Route 200 and park in a commuter parking lot for instruction for the next part of the experiment. The experimenter explained that the participant would receive auditory and seat pan vibration messages, and rapidly categorize each one as an urgent alarm, a vehicle status notification, or a social notification. These categories were verbally described as follows:

- "An alarm would be some kind of time-critical highly urgent signal like a collision warning or a lane deviation warning."
- "A status notification would be a signal indicating something to do with the status of the car, like low tire pressure or a door is open."
- "A social notification would be something telling you that you have a call or an email or a Facebook update."

The participant was instructed to categorize each message as quickly as possible into one of the three categories by saying "one" for an alert, "two" for a vehicle status notification, and "three" for a social notification. The participant put on the verbal response microphone practiced the response procedure. As an initial practice, the experimenter spoke the words "alert," "status," and "social" and the participant responded with the matching numerical response. The experimenter told the participant that he or she could begin the practice by referring to a "cheat sheet" located on the dashboard that showed the correct numerical response for each category, but that he or she should respond without looking at the codes when comfortable doing so. Thirty trials of this practice were conducted in quick succession. This process also gave the experimenter the opportunity to verify that the microphone was detecting the participant's verbal responses. Next, two practice trials were conducted. The first trial was an auditory signal and the second signal was a vibrotactile signal. Once this practice was completed, the participant was instructed to drive back to Route 200, where two additional on-road practice trials were conducted.

Following practice, participants completed all experiment trials while driving on Route 200. When a participant experienced a signal, he or she verbally responded with the perceived numerical category of that signal. The verbal response time was automatically captured by the experiment software and the numerical category was entered by the experimenter. If the participant did not make a verbal response within six seconds of signal initiation, the trial was marked as an undetected signal. The experimenter did not provide any feedback to the participant if he or she missed a signal. If the microphone did not accurately capture response time (e.g., participant response not detected by the microphone or participant made a vocalization before providing their response) the experimenter made a note to manually calculate the response time using the video data from the session.

Half of participants completed the baseline ambient noise condition first, and half completed the music condition first. The experiment was paused to turn around at the ends of the driving route (Shady Grove Metro Station and Briggs Chaney Road). It took approximately 35 minutes to complete the 52 experiment trials. The entire session took approximately 100 minutes, on average.

4.3 Results

4.3.1 Perception of signal meaning

This section describes how participants classified the signals and how quickly they did this. In order to present perceived urgency data in a more easily portrayed manner, the signal category ratings were transformed. While procedurally, participants rated an “alarm” as a 1 and a “social notification” as a 3, these numbers were reversed in the quantitative analysis, so that an alarm response was treated as 3 and a social notification as 1. In this manner, higher ratings correspond to higher perceived urgency. Group mean ratings of signal meaning are based on this scale.

4.3.1.1 Auditory signals

Figure 24 shows the percentage of participants who responded to each auditory signal under both baseline and music ambient noise conditions. All auditory signals were reliably detected under baseline driving conditions, exceeding 98 percent in every case. Based on Singer et al. (in press), it was assumed that most of the 70 dBA signals would also be well-detected under the music ambient noise condition. This was the case except for three social notification signals (Auditory Social 2, 3, and 4), which had shorter durations than other auditory signals and were detected 32 percent to 42 percent of the time. Of the remaining auditory signals, all were detected more than 90 percent of the time during music, with the exception of Auditory Status 1 and Auditory Social 1, which were at about 85 percent. Since the data that follow on perception of meaning and verbal response time are based only on participants that heard the signal, it should be recognized that the group data for Auditory Social signals 2, 3, and 4 during music are based on fewer than half of the participants. However, there is little missing data for all other signals and noise conditions.

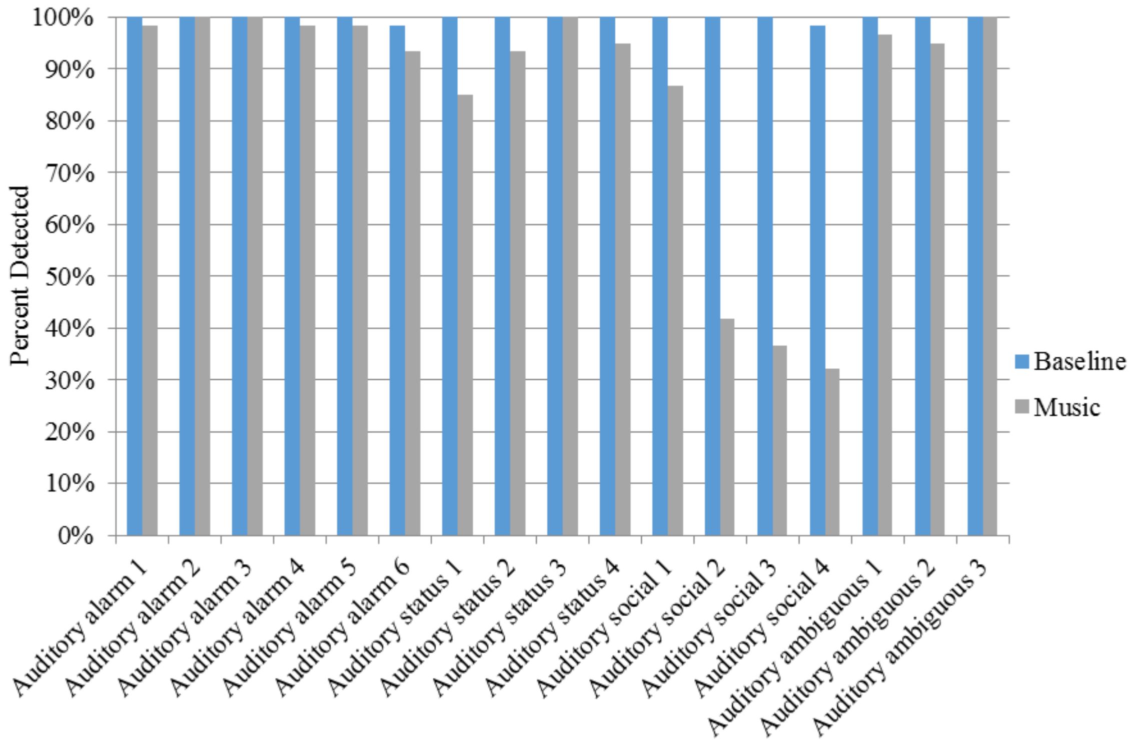


Figure 24. Percentage of participants who detected each auditory signal under baseline and music ambient noise conditions

The category of rated signal meaning varied substantially among the auditory signals, even though all were presented at the same nominal 70 dBA level. Figure 25 shows the group mean rating of message urgency level for all auditory signals under both ambient noise conditions. The 17 auditory signals are grouped along the X-axis based on the intended category of meaning, according to either suggested design criteria or the signal designer's intended message. The figure indicates that the on-road classification of messages is closely related to the intended message.

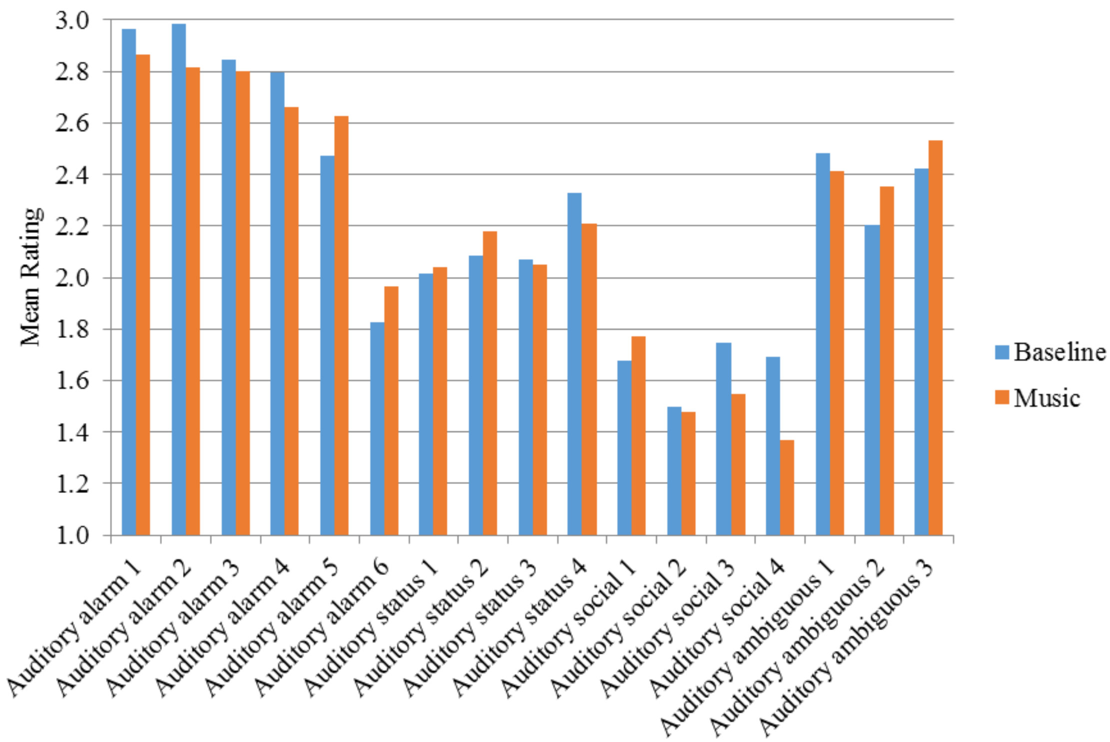


Figure 25. Mean urgency category ratings of auditory signals under baseline and music ambient noise conditions

Multinomial logistic regression was used to analyze signal meaning. Multinomial logistic regression is used to predict the probability of category membership on a dependent variable based on multiple independent variables. This approach is an extension of binary logistic regression that allows for k>2 categories of a dependent variable. Maximum likelihood estimation is used to evaluate the probability of category membership. It is an attractive approach because it does not assume normality, linearity, or homoscedasticity. In addition, it assumes non-perfect separation of the outcome variables by the predictor variables. The current model analysis was performed in SAS and used a cumulative logit model with Fisher's scoring as an optimization technique. Differences of least square means are reported with Sidak adjusted p-values. The Wald Chi-Square statistics are presented in Table 9.

Table 9. Analysis of auditory meaning responses

Effect	DF	Wald Chi Square	P value
Ambient noise	1	4.15	0.041
Auditory signal	16	586.10	<.0001
Interaction	16	25.51	0.061
Subject	59	304.22	<.0001

The analysis indicates statistically significant effects of both the auditory signal and the ambient noise condition. The interaction term fell just short of the 0.05 significance level. Subsequent to this analysis, pairwise comparisons were conducted for data grouped by the nominal category of meaning of the signal. All pairwise comparisons were statistically significant beyond the p =

0.0001 level. Thus as a group, the alarm signals were perceived as more urgent than the ambiguous signals, which in turn were perceived as more urgent than the status signals, which in turn were perceived as more urgent than social notification signals. The effect of the ambient noise condition, though small, was significant. Overall, the baseline and music ratings were very highly correlated (R=0.958).

Auditory Alarm signals were generally rated higher than all other signals, with the exception of Auditory Alarm 6. Auditory Alarm 6 was an OEM FCW alert that has also been found to be low in urgency in previous work within the CWIM project. Those signals meeting the four key design elements recommended in Lerner et al. (in press) had a mean rating near the ceiling of 3.0, which means nearly every participant perceived these as urgent alarms. Pairwise comparisons indicated that auditory alarms 1-4 were not significantly different from one another but were rated higher than all the other auditory signals, including warning signals 5 and 6. Auditory warning signal 6 was rated significantly lower than the other five warnings. The “ambiguous” signals (right side of graph) met only some criteria, and were intermediate in ratings. Auditory Status signals were generally rated around a mean of 2.0, which corresponds to the rating scale value for this message category. Auditory Social signals were rated lowest of all.

Figure 26 presents the group rating data for the baseline noise condition in the form of stacked bar charts. This illustrates the degree of agreement among participants in assigning meaning categories to the sounds. The modal category for all of the Auditory Alarm sounds was “alarm,” with the best having near-perfect agreement. The modal category for the Auditory Status sounds was “status” in all cases, with about three-quarters of participants agreeing in the best cases. For the Auditory Social sounds, there were roughly similar splits between the status and social categories. Very few of the participants interpreted these as warnings (4-17%). The Auditory Ambiguous sounds were indeed ambiguous to participants, with splits between the “alarm” and “status” categories.

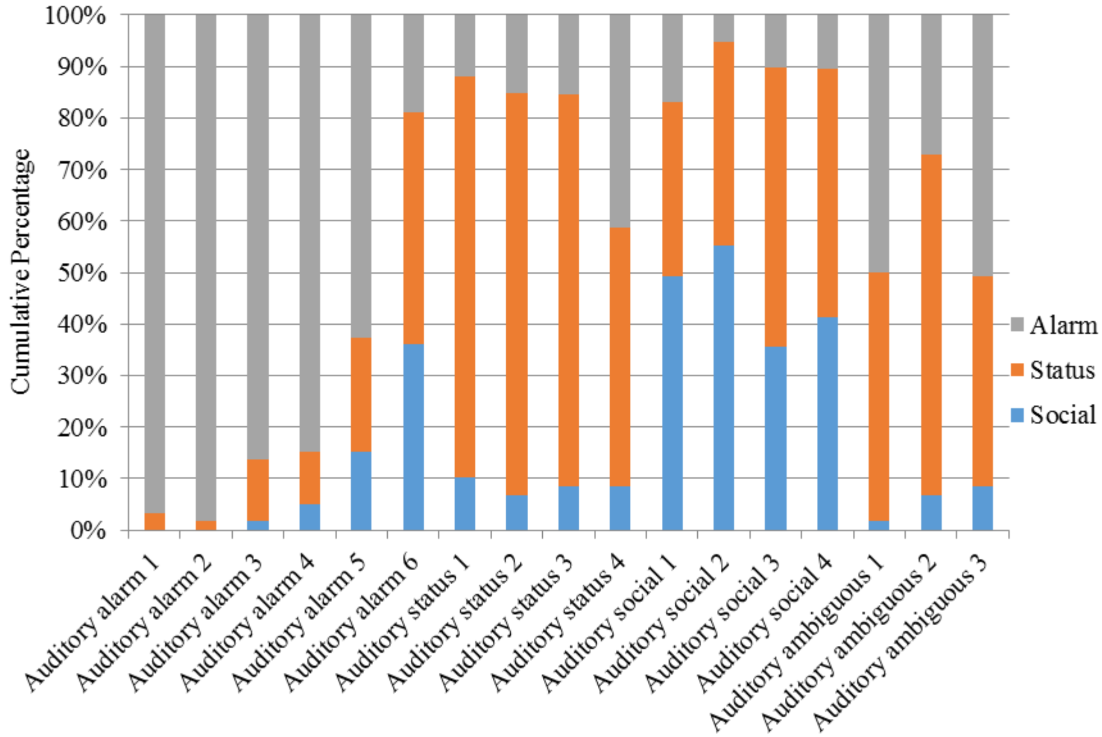


Figure 26. Percentage of participants assigning each meaning category to each auditory signal

Figure 27 shows how rapidly participants were able to verbally classify the meaning of auditory signals. Participants who did not detect the signal were excluded from analysis. Table 10 presents the results of an analysis of variance (ANOVA), which indicates statistically significant effects of the auditory signal, the ambient noise condition, and their interaction. The results parallel those of the perceived meaning findings. The response times to auditory alarms 1 through 4 did not differ significantly from one another but were significant faster than the responses to auditory alarms 5 and 6 (with the exception of the comparison of alarm 4 with alarm 5). An additional analysis was conducted collapsing signal conditions into the four categories (alarm, status, social, ambiguous). Signal, noise, and their interaction were all significant. Pairwise comparisons indicated that the verbal response time for each group differed significantly from each of the other groups. Auditory alarms, which showed the best agreement among participants in perceived meaning, were responded to more rapidly than other sounds. In particular, auditory alarm sounds had mean response times of about 1.6 to 1.8 seconds (with the exception of auditory alarm 4 during music, just over 2.0 seconds). Other message types tended to have longer response times generally above 2 seconds. The differences in verbal response times are substantial. The auditory alarm sounds (excluding auditory alarm 6) were responded to about 400 ms faster than the ambiguous and social sounds, and about 600 ms faster than the typical status sounds.

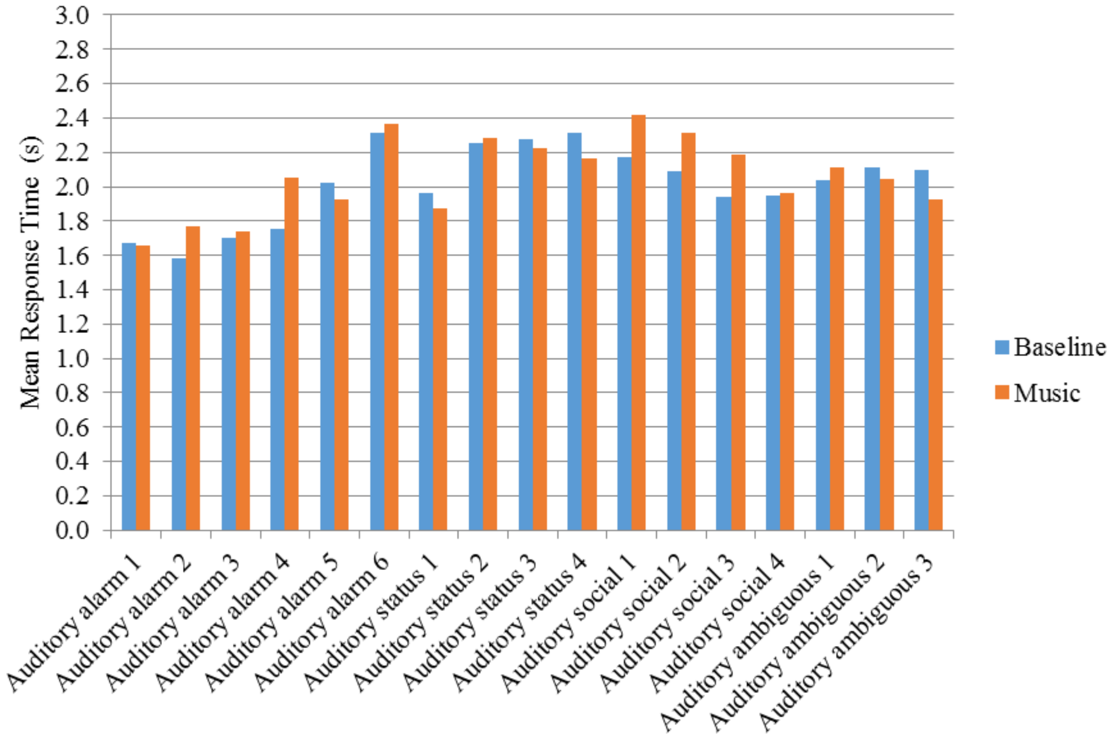


Figure 27. Mean verbal response time to categorize auditory signal meaning

Table 10. ANOVA for verbal response time to auditory signals

Effect	DF	F value	P value
Ambient noise	1	7.24	0.0072
Auditory signal	16	15.07	<.0001
Interaction	16	1.83	0.0229

The verbal response time was substantially correlated with both the mean category rating (R=0.651), suggesting a relationship with signal urgency, and the percentage of ratings in the modal category (R=0.683), suggesting a relationship with meaning consensus. However, since the mean rating and the modal percentage are themselves correlated (R=0.703), this confound makes it difficult to determine the relative importance of urgency and consensus in relation to response time. Figure 28 shows the relationship of mean verbal response time to the mean category rating. The figure indicates a curvilinear relationship. The category rating has little relationship to response time through most of the rating scale, until those sounds with the highest ratings (>2.5 on the 3-point scale). Response time then drops substantially, on the order of about 400 ms. A similar relationship to response time exists for consensus, where the drop in response time occurs when the percentage of cases in the modal category is at about 85 percent. These observations suggest that drivers interpret highly urgent, unambiguous signals more quickly than other signals, but that there is no evident gradient of response speed leading up to these most rapid cases.

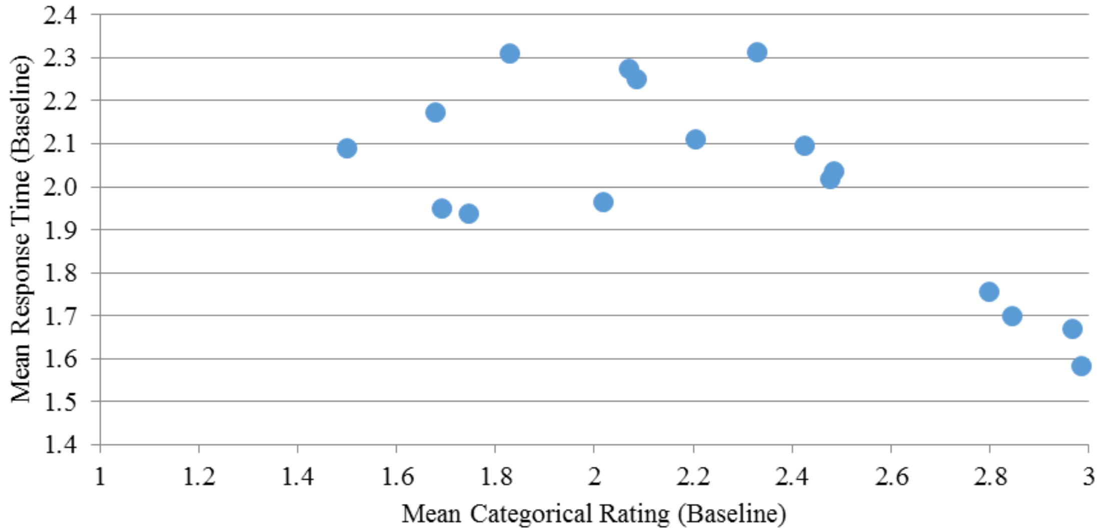


Figure 28. Relationship of categorical rating and response time for auditory signals

4.3.1.2 Vibrotactile signals

Figure 29 shows that all nine vibrotactile seat vibration signals were reliably detected in both baseline and music ambient noise conditions, with detection rates ranging from 90 percent to 100 percent. As might be expected for a vibrotactile signal, there was essentially no effect of the ambient noise condition. Despite this consistent detection, however, the vibrotactile signals were not perceived as very urgent. As seen in Figure 30, the 18 group mean ratings (nine signals, two ambient noise conditions) were all in a narrow range of 1.4 to 1.8 on the three-point scale.

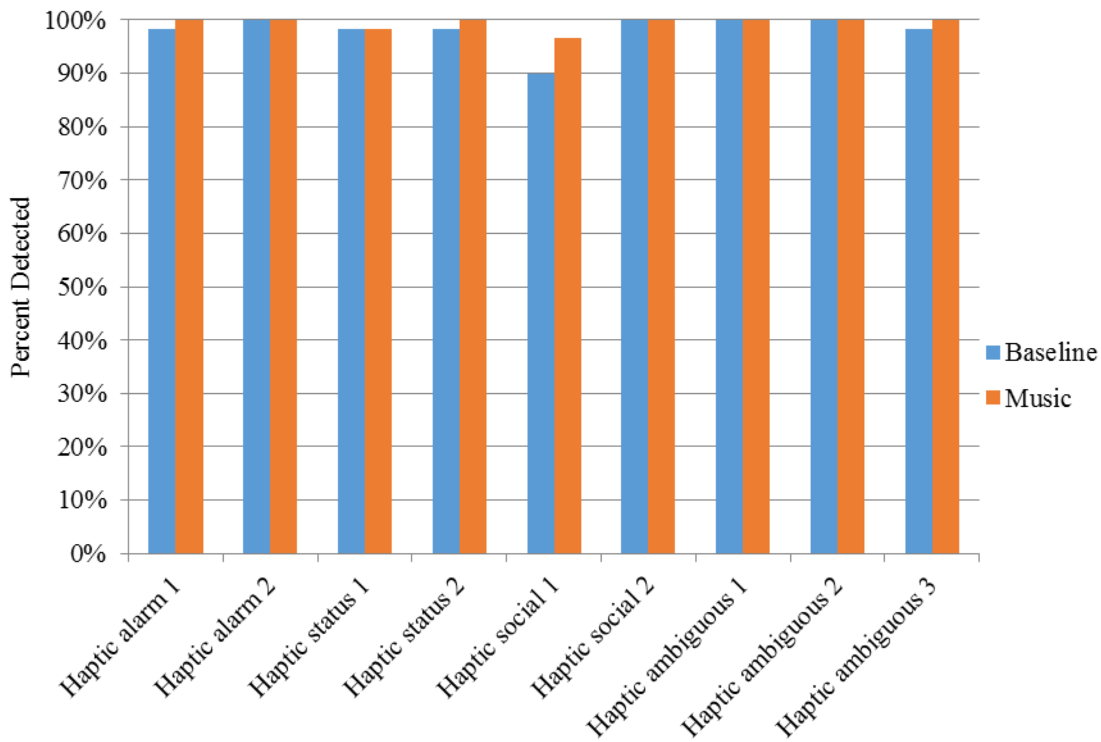


Figure 29. Percentage of participants who detected each vibrotactile signal under each ambient noise condition

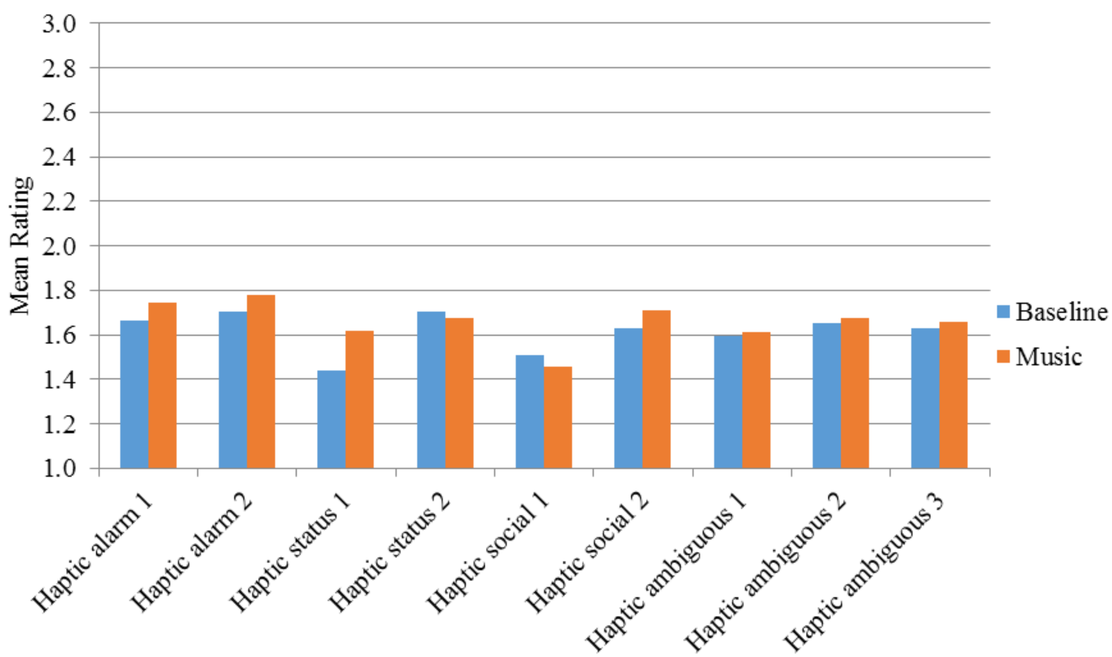


Figure 30. Mean urgency category ratings of vibrotactile signals under baseline and music ambient noise conditions

The findings of the multinomial logistic regression analysis on these data are shown in Table 11. The main effect of the vibrotactile signal was statistically significant but there was no effect of ambient noise or the interaction term. Pairwise comparisons indicated that the primary basis of the significant effect of the vibrotactile signal was due primarily to the low ratings of haptic social 1. It was rated significantly differently from all other signals except for haptic status 1. The only other significant comparison was of haptic alarm 2 with haptic status 1 (the comparison of haptic alarm 1 with haptic status 1 was close to significance (p=0.063)). Thus although the two vibrotactile alarm signals had the highest mean ratings, these were only minimally higher than the other signals.

Table 11. Analysis of vibrotactile meaning responses

Effect	DF	Wald Chi Square	P value
Ambient noise	1	0.73	0.3933
Vibrotactile signal	8	47.02	<.0001
Interaction	8	5.54	0.6985
Subject	59	326.97	<.0001

Figure 31 presents the group rating data for the baseline noise condition in the form of stacked bar charts. In every case, the modal category was "social," ranging from 43 percent to 67 percent of the ratings. The "alarm" category never accounted for more than 20 percent of the ratings for any vibrotactile signal.

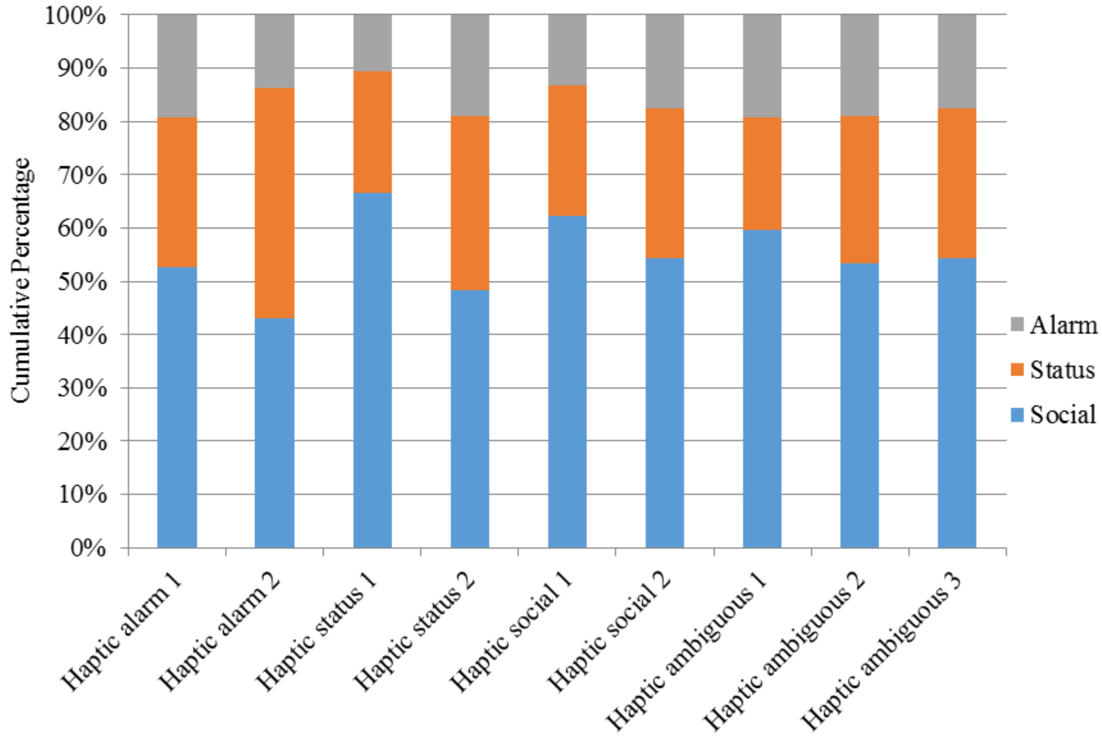


Figure 31. Percentage of participants assigning a vibrotactile signal to each meaning category

Figure 32 shows how rapidly participants were able to classify the meaning of vibrotactile signals, based on verbal response times. There was relatively little effect of the particular vibrotactile signal, with response times ranging from about 2.1 to 2.3 s in the baseline noise condition. Table 12 presents the findings of the ANOVA on these data; only the main effect of vibrotactile signal was statistically significant. An additional analysis was conducted collapsing signal conditions into the four categories (alarm, status, social, ambiguous). Only the main effect of signal was significant. Pairwise comparisons indicated that verbal response times to the alarm sounds were significantly slower than the response times to status or ambiguous sounds. Overall, the response times to vibrotactile signals are slow compared to the auditory signal response times. Verbal response times to vibrotactile signals showed no correlation with the mean category rating (R=0.066) or the percentage of ratings in the modal category (R=0.203), which is to be expected given that there was little range or systematic relationship among signals for these measures.

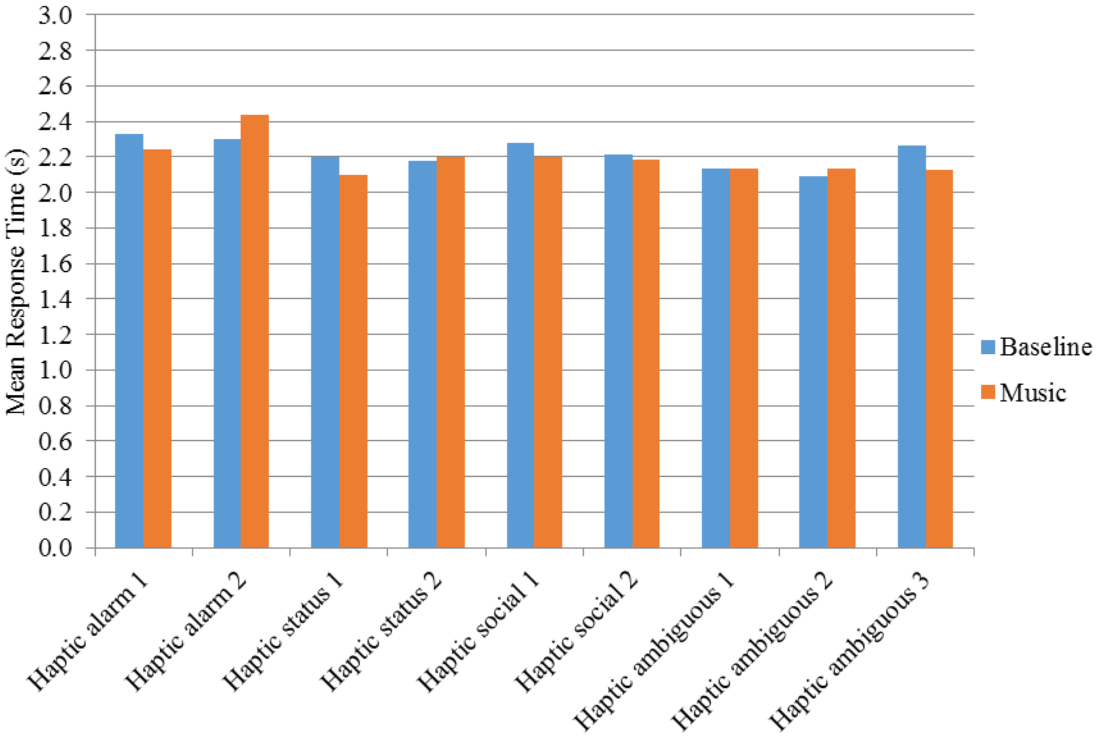


Figure 32. Mean verbal response time to categorize vibrotactile signal meaning

Table 12. ANOVA for verbal response time to vibrotactile signals

Effect	DF	F Value	P value
Background Noise	1	2.06	0.1513
Vibrotactile signal	8	2.47	0.0120
Interaction	8	0.66	0.7275

4.3.2 Driver response to unexpected signals

4.3.2.1 Data reduction

In order to assess the effects of the warning signal on visual response, it was necessary that the participant be looking down at the visual display at the moment that the signal began. In nine of the sixty cases, the participant looked up prior to signal onset. These cases were removed from the analysis of the visual data but were included in analysis of other behaviors. After eliminating the nine cases, there were 17 participants in each signal condition for the visual response analysis. No participants had to be dropped from other analyses for any reason.

4.3.2.2 Visual scanning in response to the signal

Participants in the Alarm condition directed their gaze to the forward roadway more rapidly than did participants in the other two conditions.

Figure 33 shows a cumulative relative frequency plot of visual response times to look ahead. Video analysis indicated that three participants in each group were looking forward at the moment the signal was initiated, so the figure excludes these individuals. Many participants responded quite rapidly, suggesting a reflexive orienting response as opposed to a volitional decision to scan the roadway. The 85th percentile time for the Alarm group was 0.7 s, versus 1.1 s for the Status group and 1.2 s for the Social group. The time to look forward exceeded 0.9 s for only one participant in the Alarm group, and all were under 1.5 s. In contrast, eight of the 34 participants in the other two groups took longer than 1.0 s to respond, with four exceeding 1.75 s. The distributions for the Status and Social conditions were similar to one another.

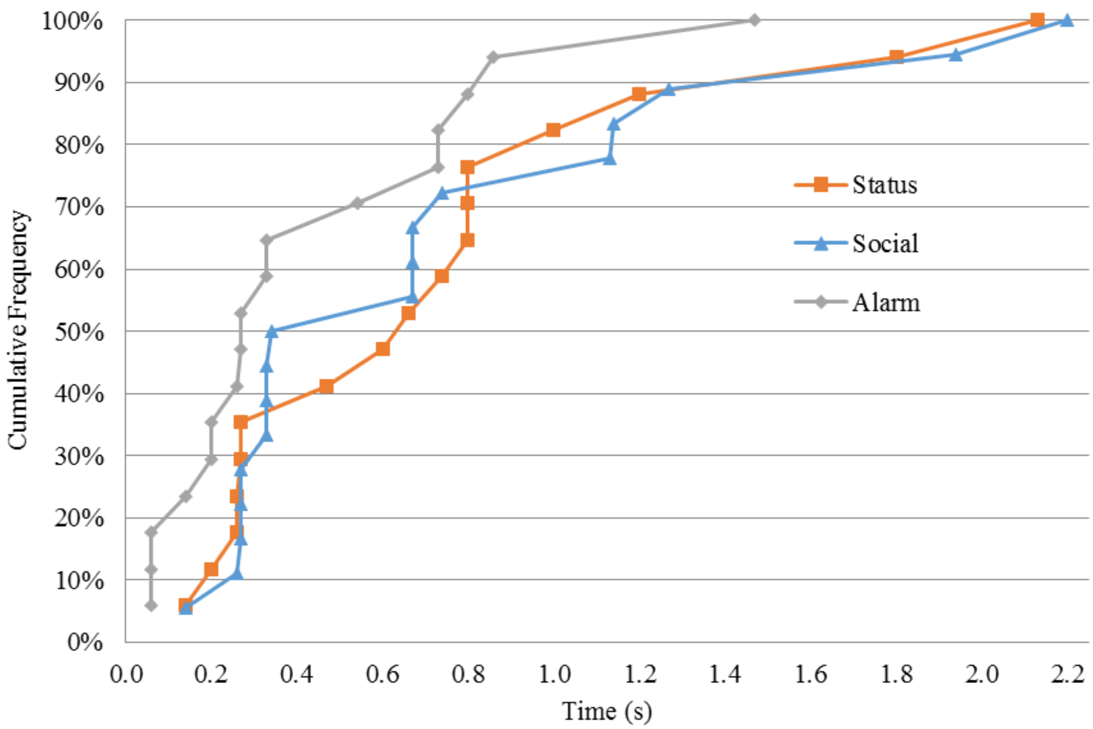


Figure 33. Cumulative relative frequency plot of time to initiate glance to forward roadway, as a function of category of auditory signal

Group mean times to initiate a forward glance were compared in an ANOVA. Response times were log transformed to accommodate the skew in the response time distributions, with the assumption of homogeneity of variance being met. Table 13 summarizes the analysis. The p value of 0.051 was just short of the 0.05 criterion. Given the substantial variance in response times as evident in Figure 33, the sample size may not have been adequate for detecting the differences in means. While Figure 33 suggests that essentially everyone looks up within a short time following the warning sound, whereas as substantial proportion of drivers took over 1 second to look forward following the other two sounds, a larger number of observations would be required to unambiguously discriminate among the response time distributions.

Table 13. ANOVA for time to initiate forward glance

Factor	Sum of Squares	df	Mean Square	F	p
Auditory signal	4.573	2	2.287	3.158	0.051
Within groups	35.476	49	0.724		
Total	40.049	51			

Participants in the Alarm condition also spent a somewhat greater portion of the post-alarm period looking at the forward roadway. Figure 34 shows the group mean percentage of time in the 5 s following signal onset that the participant's gaze was directed at the roadway. It includes all 20 participants in each group. For all three groups, nearly all of the 5 s was directed at either the forward roadway or the message display (4.84 to 4.89 s), with minimal glancing toward the dashboard, mirrors, or other locations. The Alarm group spent about 2.45 s looking forward, which is about 300 ms more than the Status group. The mean time for the Status group was inflated by a single participant who spent the entire 5 s interval looking at the roadway; without this participant, the Status group mean falls from 2.14 to 1.99 s. An ANOVA on the amount of time spent looking at the forward roadway during the 5 s interval is summarized in Table 14. Although differences are in the expected direction, the magnitude of the effect (about 15%) was not statistically discriminable.

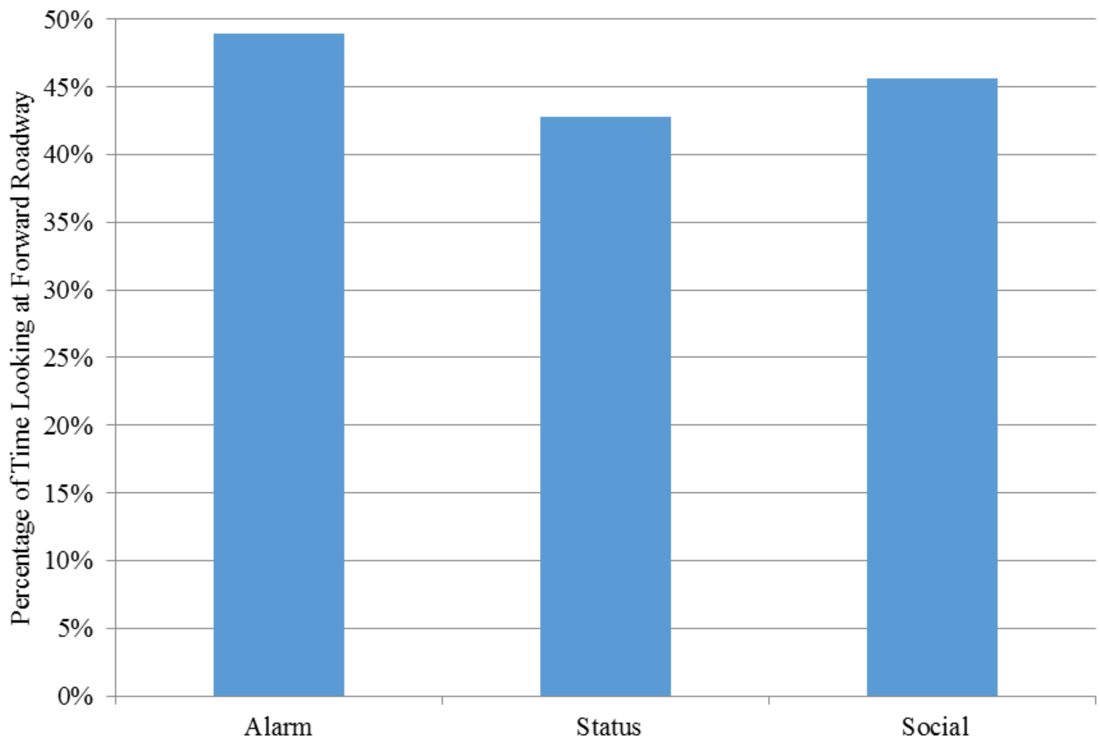


Figure 34. Percentage of 5 s post-signal interval with gaze directed at forward roadway

Table 14. ANOVA for time looking at forward roadway

Factor	Sum of Squares	df	Mean Square	F	p
Auditory signal	0.950	2	.475	0.66	0.519
Within groups	40.819	57	.716		
Total	41.769	59			

4.3.2.3 Changes in vehicle speed in response to the signal

For each participant, travel speed over the 3-second interval from 2 to 5 seconds following the unexpected signal was compared to the speed 0.5 seconds before the signal. The group mean vehicle speed dropped following the unexpected signal for the Alarm and Status groups. As Figure 35 shows, the group mean drop in speed was about 0.71 mph for the Alarm group, 0.45 mph for the Status group, and 0.09 mph for the Social group. Since there was no actual threat, or even nearby traffic, when the signal was presented, no large drop in speed was anticipated and this was a rather conservative test of driver reaction to the signal.

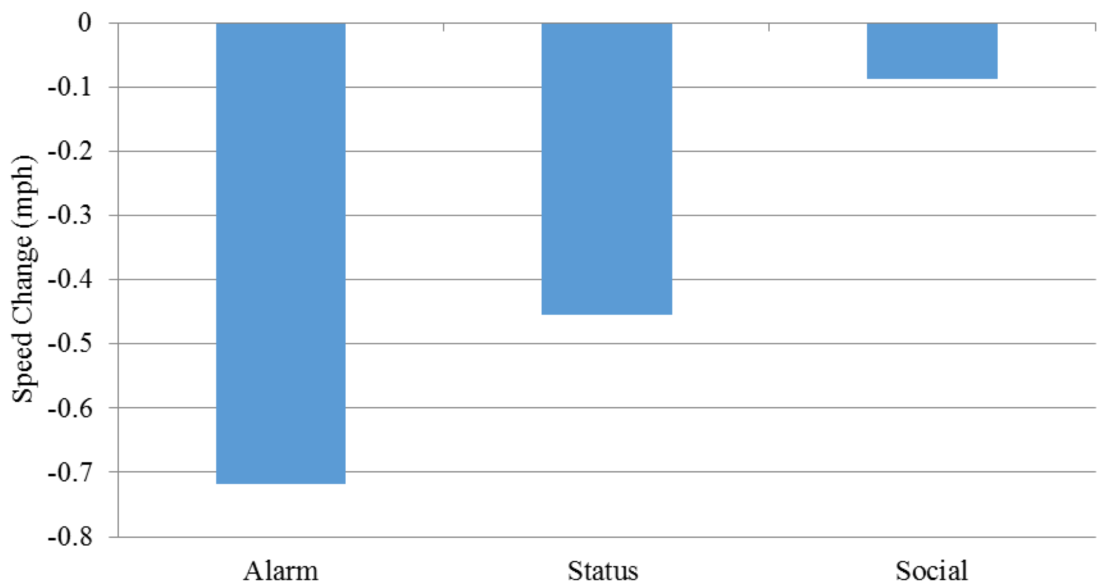


Figure 35. Mean change in speed following the unexpected auditory signal

An ANOVA was conducted on the group mean change in speed following the signal and the results are shown in Table 15. The effect of signal type on change in speed, while ordered in the anticipated direction, did not achieve statistical significance ($p=0.127$). The standard error of speed change was about 0.3 mph, so changes of the observed magnitude would be difficult to statistically discriminate with the sample size used here.

Table 15. ANOVA for change in vehicle speed

Factor	Sum of Squares	df	Mean Square	F	p
Auditory signal	3.922	2	1.961	2.14	.127
Within groups	50.400	55	9.160		
Total	54.322	57			

4.3.2.4 Other behaviors

Video recordings, brake status data, and experimenter notes were reviewed to identify any other clear overt actions to the signal, such as braking, shifting hand position, or vocalizing. However, such responses were rarely observed, with only four instances (two hand position shifts, two verbalizations, no braking) across the 60 participants.

After encountering the unanticipated sound, participants were asked "What did you think when you heard that sound? What did you think it meant?" Responses were coded for the general meaning of the sound that the participant indicated and for any self-reported emotional reactions that the participant may have raised. Table 16 summarizes the "meaning" responses, which were grouped under several categories.

- Related to an alarm/urgent warning
- Related to vehicle status
- Related to social notifications
- Related to the center stack display screen (location of the distraction task)
- Related to navigation
- Not interpretable (could not determine what participant meant)
- No response given regarding meaning.

Table 16 does not include uninterpretable responses. The percentages for each sound sum to somewhat over 100 percent because a few participants offered more than one meaning. Participants who heard the alarm sound were more likely to interpret the sound as a warning than were participants who heard the status indication or social notification sounds. About 73 percent (11 of 15) of the Alarm group participants indicated a "warning" meaning, compared to 47 percent in the other groups. Participants in the Social group appeared more likely to attribute the sound as having something to do with the center console display. Note that the navigation interpretations are likely the result of the use of navigation-type messages being used for the visual distraction task.

Table 16. Reported meaning of the unexpected sound

Group	n	Meaning of the Sound					
		Warning	Status	Social	Display	Navigation	No meaning given
Alarm	11	73%	7%	0%	7%	13%	13%
Status	19	47%	26%	0%	5%	21%	5%
Social	19	47%	16%	5%	26%	11%	5%

Six of the 20 participants (30%) who experienced the alarm sound mentioned some sort of emotional reaction to the sound (using terms such as “startled,” “scared,” “disturbed,” and “alarmed”). Participants were not asked directly about emotional reactions, so these were spontaneous reports in response to the general question of what they thought about the sound. Of the 40 participants hearing the other two sounds, only one (in the Social Notification group) reported an emotional response.

4.4 Discussion of on-road experiment

This experiment provided on-road validation of auditory signal design assistance properties derived from previous laboratory and driving simulator research. Sounds that met the criteria for an urgent warning as suggested within the CWIM program (Lerner et al., in press) were reliably recognized as urgent and were identified (by verbal response) quickly. Sounds meeting only some criteria were ambiguous in meaning and slower to be interpreted.

Sounds differed substantially in the time it took participants to classify the meaning of the signal. The range from fastest to slowest was about 0.7 s. While the fastest response times were for signals that were almost universally categorized by participants as urgent alarms, the slowest response was observed for a message (auditory alert 6) recreated from an actual OEM FCW signal. This signal, however, did not meet all criteria for an urgent warning. In particular, this signal had a base frequency of approximately 400 Hz, well below the criterion minimum of 1000 Hz. While it is not known how this response time difference might translate into overt behavioral effects, differences of this magnitude could have a substantial impact on crash avoidance and crash severity.

Consistent with previous work (Singer et al., in press), auditory signals presented at about 70 dBA at the driver ear position were generally well-detected even with music playing. The exceptions to this were the social notification signals that were presented for a briefer period than the other signals (as they likely would be in actual applications). Background music at 75 dBA did not have substantial or consistent effects on the time it took drivers to categorize the message or on the interpretation of the message. Perception of auditory signals at lower (e.g., 65 dBA) levels would be expected to be more disrupted, based on Singer et al. (in press). Other louder and more disruptive transient ambient noise conditions (e.g., heavy rain, windows fully lowered) might differentially affect driver perception of some signals more than others, even for 70 dBA signals (Singer et al., in press). Such effects are most easily studied under laboratory listening conditions, as opposed to the on-road methods used here.

In contrast to the auditory signals, there was not a strong effect of signal characteristics for the vibrotactile seat pan signals. The nine vibrotactile signals were all quite reliably detected (90% or greater). However, there was not a great difference in perceived urgency among vibrotactile signals hypothesized to differ in urgency based on laboratory experiments. The modal perceived meaning category in every case was social notification, and the signals were rated as urgent warnings in only between 10 percent and 20 percent of the cases. Verbal response times to categorize the vibrotactile signals were slow relative to auditory signals. These findings do not imply that different types and locations of vibrotactile signals (e.g., brake pulse, seat belt tensioning) might not differ in these aspects or that some other seat vibration stimuli might not differ. However, the range of vibrotactile seat pan signals employed here did not provide any reliably-perceived range of urgency.

Given the highly reliable detection of the seat pan vibrotactile signals, their general lack of interpretation as being urgent warnings, and the manner in which participants interpreted signal meaning, such vibrotactile cues may be useful for indicating low urgency events or messages, such as social media notifications.

If findings regarding driver interpretation of signal meaning are to be safety-relevant, signals with different perceived meaning should result in differences in driver response to the signals. Ideally, such differences in behavior should occur even though the driver is untrained regarding the signal and not expecting any warnings or alerts. Phase 1 of the experiment addressed this issue, presenting drivers with one of three message types while the driver was visually distracted.

This experiment observed that, relative to less urgent message categories, an unexpected auditory signal perceived as an urgent warning led to trends in the direction of faster scanning of the forward roadway, more glance time to the forward roadway, a larger decrease in vehicle speed, more emotional responses, and greater likelihood of describing the sound as a warning. While consistent in direction, inferential tests on any particular measure did not reach conventional significance levels. Factors contributing to this may have included the relatively small n for each group, drivers' unwillingness to commit long glances to a distracting task while driving on real roads and the fact that the unexpected signals were presented only on low-demand, straight sections of freeway when there was no surrounding traffic, and therefore participants had no expectation of an imminent threat and could quickly determine that there was no threat.

It should be recognized that while the participants in this experiment were not anticipating any alerts, they were aware that they were in an experimental situation and study vehicle. The signals also only were presented during safe periods where there was no other traffic in proximity. Thus participant alertness and responsiveness in general may have been different from that more typical of a normal driver in their own vehicle. Therefore, the absolute aspects of driver response (e.g., glance timing, speed changes) in this experiment may be different from that seen in truly naturalistic driving. Quite likely differences among signal types may have been constrained by a more general alertness in the experiment and the objective absence of risk factors in the period before the signals sounded.

This experiment indicates that on-road perception of in-vehicle auditory signals closely parallels the findings from laboratory and simulator work and that design assistance for categorical perception of auditory signals derived from lab and simulator methods is valid. Findings regarding vibrotactile signal perception are less consistent. Despite reliable detection, drivers were generally insensitive to differences among vibratory signals as presented through the seat pan and their urgency was not perceived as very high. While vibrotactile signals of various sorts may ultimately prove useful for various in-vehicle applications, the generalizability of laboratory and simulator findings to on-road performance should be viewed cautiously.

5. General Discussion

5.1 Replication and validation of previous findings

5.1.1 Auditory alerts

The current series of investigations provides additional converging evidence for the importance of specifying the main auditory parameters that should be used for highly urgent FCW signals. Specifically, use of the five criteria identified in this report lead to high rates of appropriate perceptual recognition both within the context of a driving simulator and on-road. In other words, classification results (e.g., participants classifying a sound as an urgent warning when it met the five criteria associated with an urgent FCW and not when it did not) obtained while engaged in simulated driving as well as in the on-road study matched closely the results of previous investigations examining the sounds in lower fidelity lab settings.

Further, the previous series of investigations (Lerner et al., in press) examined classification responses outside the context of an actual collision event. In the two driving simulator studies in this series, behavioral response was examined for first exposure to the sound presented in conjunction with a hazard event. In both the low and high fidelity driving simulations, use of all five recommended auditory parameters led to improved collision avoidance responses. Specifically, in the low fidelity simulator use of all five parameters, relative to only four, resulted in a more consistent pattern of brake responses that also tended to be faster. This can be taken as further validation that participants are recognizing the sound to be in the category it is designed to be, that of a highly urgent FCW. Notably, this consistent pattern of brake responses is obtained following first exposure to the sound with no prior expectation that collision events or warnings would be issued.

In the high fidelity driving simulation, presenting a sound meeting all five parameters demonstrated more effective collision avoidance maneuvers and evidence for a decrease in collision severity when collisions did occur, relative to a warning that met only four of the five criteria. Again, these results were obtained upon first exposure to the concurrent warning and collision event. Subsequent exposure and classification results obtained while engaged in simulated driving matched closely the classification results obtained in previous studies (Lerner et al., in press).

For obvious reasons drivers in the on-road study were not presented a highly hazardous collision situation. However, sounds meeting all five recommended parameters were presented along with sounds not meeting all five. Those sounds that met all criteria were classified as being a warning more frequently and were identified and recognizable in the presence of background noise and varying intensity levels. The on-road experiment also confirmed that drive responses (e.g., visual behavior, vehicle slowing, and emotional reaction) tended to differ for an alarm sound as compared to status or social notification sounds. Together the current series of investigations provides strong support for the need to ensure that FCWs meet the five recommended criteria and that sounds not intended to be urgent warnings reframe from using more than two of the recommended parameters.

5.1.2 Vibrotactile alerts

Vibrotactile signals were given limited consideration in this project. Past research conducted in the GMU laboratory under the Connected Vehicles program (Lerner et al., 2014) provided some initial data showing an increase in perceived urgency as a function of pulse rate for vibrotactile stimuli but only limited effects of these signals on crash rate or crash severity in a driving simulator. The present project expanded on this work by broadening the set of vibrotactile stimuli and manipulating parameters including IBI, burst duration, and number of bursts. This work was done as a laboratory assessment using a wrist-worn tactor set. The findings did not show dramatic changes in perceptual response except in the case where all three of the key criteria were met, in which case the likelihood that the signal was perceived as an alarm was increased.

When a subset of these vibrotactile signals was included as seat pan vibration stimuli in the on-road experiment, there was little evidence that any signals generated an interpretation of being an urgent warning. While there was a statistically significant effect of vibrotactile signal on perceived meaning, this was due primarily to one of the social notification signals being perceived as somewhat less urgent than some other signals. All nine signals, however, were classified by a plurality of participants as social notifications.

Thus, while some degree of change in perceived urgency could be observed in some cases in the lab settings, the on-road experiment was unable to confirm any meaningful warning effectiveness when delivered through the seat pan while driving. This was the case even though the vibrotactile signals were very reliably detected by the drivers. It is possible that participants were unable to discriminate subtle differences in vibrotactile signal characteristics through the seat pan in the moving vehicle to the same extent that they could when experiencing the signals through a tactor on the wrist in a lab setting. It is recognized that the vibrotactile signal aspects of this project were quite limited, both in terms of the types of vibrotactile signals and in how they may be implemented in a vehicle. For example, this experiment did not investigate multimodal stimuli (e.g., paired vibrotactile and auditory stimuli). It may be that some vibrotactile signal designs could serve as effective warnings. However, for the set of signals used here, and as implemented on-road, the relatively weak lab findings could not be replicated in the on-road setting.

5.2 Warning signals and response distribution

Crash warnings are intended to assist drivers by alerting them to impending threats. Therefore, a benefit of the warning may be observed only if the driver would not have readily detected the threat in the absence of the warning. This is one factor that contributes to the difficulty of obtaining statistically significant differences in driver response with relatively small samples; only a subset of the participants may have the potential to be affected. Figure 3, which shows results from the initial driving simulator experiment, illustrates this. The data points for the “No Warning” condition show a bi-modal distribution. One cluster of participants was quickly aware of the hazard (as indicated by accelerator release) while another cluster responded slowly (longer than 1.5 seconds). For the various warning groups, the number of “slow response” participants was fewer for the best warning (GMU prime); there were essentially no slow responses. The benefits of the warnings are not so much in shifting the response distribution to faster times as in eliminating slow responses. Interestingly, a similar effect may be seen in the on-road data for the time in which it takes drivers to orient to the forward roadway following an unexpected signal (see Figure 32). Most drivers look forward within one second, regardless of the signal type. But signal type did appear to reduce the number of slow times. Only one of seventeen participants took longer than 1.0 s to look up following the urgent alarm and all were faster than 1.5 s. In contrast, about one-fourth of drivers hearing the status or social notifications took longer than 1.0 s, with half of these cases exceeding 1.75 s. These observations suggest that the benefit of better in-vehicle warnings is that fewer drivers fail to orient to the roadway or hazard quickly. Potential safety benefits may be difficult to detect statistically because participants in general do not necessarily respond more quickly. Rather, some proportion of the subset of slow responders is eliminated. This may have very substantial safety benefits despite being difficult to statistically discriminate experimentally without a large sample.

5.3 Limitations

The research described in this report has several limitations that should be recognized. These include the following:

- Range of auditory signals and noise conditions: The experiments described here used a reasonably broad array of auditory signals, but of course many other examples could be included. It is possible that some novel types of sounds might identify additional features of interest or present exceptions to the recommendations derived from these experiments. Also, work conducted under the CWIM project (Singer et al., in press) indicated that the perceptual response to an auditory signal may be influenced by the ambient noise condition in the vehicle. Other than "baseline" vehicle noise, only one other noise condition was included here, that of one piece of music being played at about 75 dBA.
- Range of vibrotactile signals: This research used a very limited range of vibrotactile signals. They were presented on the participant's wrist in the laboratory study and through the vehicle seat pan in the on-road and driving simulator experiments. The number of signal parameters that varied was limited. Therefore conclusions regarding vibrotactile signals are limited in their generalizability. Vibrotactile display is a currently active area of research and findings regarding efficacy for vehicle application are varied. Vibrotactile signals of the sort used in the present study may be presented at different locations, at different amplitudes, and with "dynamic" aspects that provide a sense of movement. Other sorts of haptic interface may be used, such as steering wheel vibration or movement, accelerator pedal counterforce, seat belt tensioning, or momentary deceleration. The ability to provide guidance regarding vibrotactile displays for categorical perception is therefore quite limited, pending additional research.
- Multimodal displays: The research in this project only addressed signals occurring in a single modality, either auditory or vibrotactile. In practice, alerts, and particularly critical warnings, are likely to be presented in more than one modality. Visual displays are likely to be included. The effect of multimodal aspects on perceived meaning and urgency have not been investigated here and may be of practical significance. Previous work in this area (Lerner et al., in press) indicated that multimodal signals are not responded to any faster than auditory signals alone, nor are they perceived as more urgent than auditory only signals and that visual only signals are the most likely to be missed, relative to multimodal combinations.

- Participant population: The experiments reported here had somewhat limited sample sizes. While the sample sizes were adequate to discriminate differences in perception of signal meaning and other factors, they were not sufficiently large to obtain statistical significance for certain behavioral indices, even though trends were generally in the predicted direction. Additional observations would be helpful to establish effects at traditional levels of statistical significance. The participant sample was also intentionally screened to exclude those with hearing impairments. It may be useful to consider the design of interfaces for those with varying degrees of hearing impairment. The participant sample was also restricted in terms of age and other demographics. The lab and simulator work conducted at GMU used a student population and was predominantly female. The on-road experiment encompassed a broader age range (21 to 50), but still excluded older drivers.

- Changes in response over time and experience: The driving simulator and on-road experiments used procedures in which the participants were unfamiliar with the warning system. For research purposes, this is an important consideration, since it is desirable to have a signal evoke the appropriate meaning and response even when unfamiliar. However, over time and experience, actual driver response may vary with dynamic processes, such as learning or habituation. There may be improved discrimination, changes in perceived validity, annoyance, and so forth. The findings of the experiments described here represent initial driver reactions. Changes with exposure in the course of actual use would be of interest as well.

- Relationship to on-road crash avoidance actions and crash reduction: This project compared laboratory perceptual findings to crash avoidance behavior in a moving base simulator and to judgments and driver reactions on the road in non-crash situations. The ultimate evaluation of a system of in-vehicle alerts would be to directly measure driver crash avoidance responding during actual imminent crash events on the road and further, to measure actual effects on crash rate reduction and crash severity reduction. These are very difficult and expensive outcomes to properly control and assess and may not be practically feasible. While we can assert that the findings of this project show a correspondence with safety-relevant measures, the actual safety benefits of the better performing signals cannot be quantified.

References

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. http://doi.org/http://dx.doi.org/10.1016/j.trf.2011.02.001

International Electrotechnical Commission. (2006). International Standard IEC 60601: Medical electrical equipment. *Part 1-8: General Requirements for Safety. Collateral Standard: General Requirements, Tests and Guidance for Alarm Systems in Medical Electrical Equipment and Medical Electrical Systems*. Geneva: Author.

Jenness, J.W., Baldwin, C., Chrysler, S., Lee, J.D., Manser, M., Cooper, J., Salvucci, D., Benedick, A., Huey, R., Robinson, E., Gonzales, C., Lewis, B., Eisert, J., Marshall, D., Lee, J., Morris, N., & Becic, E. (in press). Connected Vehicle DVI Design Research and Distraction Assessment. Washington, DC: National Highway Traffic Safety Administration.

Lerner, N., Robinson, E., Singer, J., Jenness, J., Huey, R., Baldwin, C., & Fitch, G. (2014). Human factors for connected vehicles: Effective warning interface research findings. (Report No. DOT HS 812 068). Washington, DC: National Highway Traffic Safety Administration.

Lerner, N., Singer, J., Huey, R., Brown, T., Marshall, D., Chrysler, S., Schmitt, R., Baldwin, C., Eisert, J., Lewis, B., Bakker, A.I., & Chiang, D.P. (in press). Driver-Vehicle Interfaces for Advanced Crash Warning Systems: Research on Evaluation Methods and Warning Signals. Washington, DC: National Highway Traffic Safety Administration.

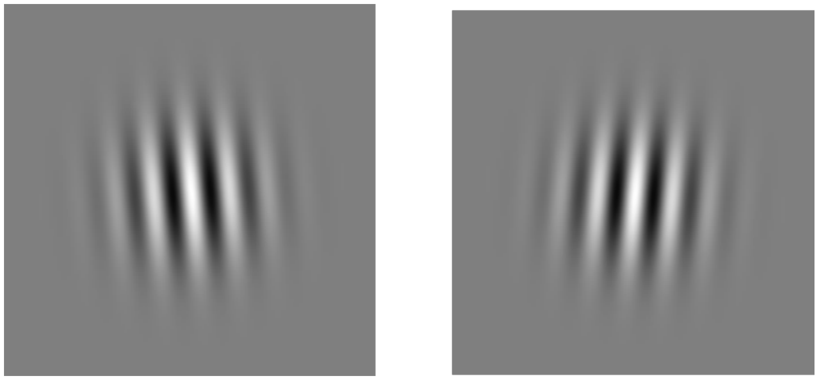
Lewis, B. A., Eisert, J. L., Roberts, D. M., & Baldwin, C. L. (2014). Designing Unambiguous Auditory Crash Warning Systems. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *58*(1), 2078–2082. http://doi.org/10.1177/1541931214581437

Meng, F., & Spence, C. (2015). Tactile warning signals for in-vehicle systems. *Accident Analysis & Prevention*, *75*, 333–346. http://doi.org/10.1016/j.aap.2014.12.013

Mollenhauer, M. A. (2004, January 16). *Simulator adaptation syndrome literature review*. Royal Oak, MI: Realtime Technologies Inc. Available at www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA513655

Singer, J., Lerner, N., Kellman, D., & Robinson, R. (in press). Crash Warning Interface Metrics: Warning and Message Perception under Ambient Noise Conditions Laboratory Experiments. Washington, DC: National Highway Traffic Safety Administration.

Appendix A: Desktop Simulator Experiment Additional Stimuli



Gabor patches used in Experiment 1, secondary task. Patches were tilted to the left or to the right at a 5° angle

Images used for social notification and status notification categorization, respectively

Appendix B: Protocols and Instructions to Participants for Simulator Experiments and Tactile Parameters Experiment

Desktop Driving Simulator Protocol

The following includes the protocol used for Experiment 1 including all experimenter instructions but not including procedural items (ex. Load Drive 1.in, press R to run):

- When participant arrives:
  - Check license
  - Give and explain informed consent
  - Give demographic survey
- Practice 1: Driving Only
  - Instructions

    We will now begin the training portion of the experiment so you can familiarize yourself with the simulator. I would like you to practice maintaining a speed of 40 mph and staying in the right lane. After a short period of time I will have you come to a complete stop so that you can get accustomed to how much force is needed to bring the simulated vehicle to a complete stop.
  - Practice outcomes: participant must be capable of maintaining a speed of ~40 mph behind the lead car. Participants will come to at least one full stop, may be asked to stop multiple times if participant is unable to stop the vehicle in a timely manner).
- Practice 2: Driving and Secondary Task
  - Instructions

    We will now begin the secondary task training portion of the experiment to familiarize yourself with the secondary task. While you are completing this portion we ask that you also maintain a speed of 40 mph and stay in the right lane. To your right you will see a bunch of billboards separated by buildings with stimuli on them called gabor stimuli on them. Your task is to identify which direction the stimuli are facing and press the corresponding paddle for whether they are leaning to the left or to the right. If they are leaning to the right, press the right paddle (behind the steering wheel), if they are leaning to the left, press the left paddle. In between each billboard will be a blockbuster building to prevent you from seeing multiple billboards at one time. So remember to maintain a speed of 40 mph, stay in the right lane, and press the paddle based on whether the lines on the billboard are oriented to the left or right. Please note that driving safely is the most *important part of the task, please prioritize your speed maintenance above all else.*

  - Practice outcomes: participants must maintain ~40 mph behind the lead car and respond correctly to billboards (practice block includes 20)

- Experimental Drive
  - Instructions

    We will now begin the actual experiment. The vehicle that you will be driving is a connected vehicle, which is equipped with multiple sensors that communicate and send information with other vehicles on the roadway. These sensors can send information about how close someone is to another vehicle, about if someone is about to hit another vehicle, if inclement weather is detected and other types of information. This experiment in particular is testing out a new connected vehicle technology that is designed to teach drivers what a two second following distance is. In this experiment you will be following a lead vehicle that you must stay behind at all times. Your task is to keep a two second following distance from this lead vehicle, if you get too close a red triangle will appear telling you that you are following at an unsafe distance, if you are following closely but not necessarily unsafely, you will see a yellow triangle and if you are following at a safe and appropriate distance a green triangle will appear. In addition to maintaining the two second headway you will also be completing the billboard task and are again asked to press the corresponding paddle, press left if the billboard stimulus is leaning to the left and press right if the stimulus is leaning to the right. The posted speed is 40 mph for this drive. And please again note that driving safely is the most important part of this task. Please prioritize your speed and headway maintenance above all else. Any questions?

  - Regardless of whether or not the participant crashes, they will drive until the end of the block of billboards.

- Complete post drive survey

High-Fidelity Simulator Validation Protocol

The following includes the protocol used for Experiment 2 including all experimenter instructions but not including procedural items:

- When participant arrives:
  - Check license
  - Give and explain informed consent
  - Orally administer sim sickness questionnaire
  - If participant passes questionnaire, introduce to sim
- Practice 1: Driving Only
  - Instructions

    The first thing we're going to do today is practice driving. The sim operates just like a regular car, but the handling might be a little different from your car. I'd like you to drive, following the car in front of you, at a speed of around 65 mph (you can see your speed in the meter cluster in the dashboard, just like on a normal car). You should be aware that the simulator will move to the left or right when you go around turns, similarly to the feeling of a normal car. There will be multiple curves, although the speed limit is 65 your vehicle is relatively heavy so you should get your speed below 50, around 40, when you take turns to avoid skidding then get back up to 65 on the straightaways. After I feel comfortable that you are capable of controlling the vehicle I will ask you to come to a few complete stops so that you can get used to the braking system. The brakes are pretty tight but they work just fine. Go ahead and start driving.
  - Practice outcomes: Participant will take at least 3 curves, must be able to maintain speed (around 65 mph) and lane position (right lane). Participants will come to at least one full stop, may be asked to stop multiple times if participant is unable to stop the vehicle in a timely manner).
- Practice 2: Secondary Task Practice
  - Instructions

    Now we're going to practice driving while doing an n-back task, presented on the touchscreen to your right. An n-back refers to a task that asks you to respond yes or no to whether the number you are currently being presented matches a number which was presented "n" numbers back. In this case you will be performing a 1-back. This means we would like you to respond yes or no by pressing the corresponding buttons on the touchpad to whether or not each number presented matches the number which was presented right before it. For example, if you see a 1 and then a 2 the answer to the 2 would be no, because it *does not match the 1, if you then see another 2 the answer would be yes, because it does match the previous 2. You don't ever need to respond to the first stimulus you see because there is no answer, there was no number for it to match but then after that you do need to respond to every number, just press yes if the number you're seeing DOES match the one you just saw and no if the number you're seeing does NOT match the one you just saw. The numbers will go pretty fast, basically about 1 a second, so you'll need to respond pretty quickly. Also, the screen will beep to register your response has been recorded.*

  - Practice outcomes: Participants must get at least ten n-back answers correct in a row to complete practice.

- Practice 3: Driving and Secondary Task
  - Instructions

    Okay, now I'd like you to start driving, once you are up to speed, I will start the n-back back up and I'd like you to do that at the same time as you drive. Remember that you should be following the lead car, the speed limit is 65 mph and that driving safely is the most important task here. I will tell you when to stop.

  - Practice outcomes: Participants must maintain control of the simulated vehicle and get at least ten n-back answers correct in a row to complete practice.

- Drive 1
  - Instructions

    Alright, now all you're going to do is exactly what you just did in the last practice. Drive, following the car in front of you, while doing the n-back to the best of your ability. Remember that you should be following the lead car, the speed limit is 65 mph and that driving safely is the most important task here. Go ahead and start driving, I will start the n-back when you get up to speed.

  - The drive ends when either the participant crashes or ~30 seconds after a successful avoid procedure.

- Practice 4: Driving with Secondary Task and Sound Responses
  - Instructions

    Alright, so now we're going to add another component. This time, your vehicle will be what we call a connected vehicle, meaning it can receive and send information about things like the weather, traffic, social media and vehicle status. This drive is going to be just like the n-back and driving you were just doing, in that you will drive and complete the n-back task, but this time, sounds are going to be played randomly from speakers. When you hear a sound, we would like you to identify it as either an alarm, a status notification or a social notification. An alarm *would be some kind of time-critical highly urgent sound like a collision warning or a lane deviation warning, in which case you should hit the brakes, but you don't need to come to a stop just hit the brakes however you think the alarm is telling you and then catch back up with the lead car, a status notification would be a sound indicating something to do with the status of the car, like low tire pressure or a door is open, in which case you should press the button on the touch screen that looks like a triangle indicator, and a social notification would be something telling you that you have a call or an email or a facebook update, in which case you should press the button on the touchscreen that looks like a telephone. In this practice, you won't have to decide what the sound is, it's going to tell you what it is, for example if it says "Alarm" hit the brakes, if it says "Status" hit the triangle and if it says "Social" hit the phone button. Additionally, the n-back is going to be a little slower this time so you'll be able to rest your hand on the steering wheel when you aren't responding and the road will be more of a highway drive where the turns aren't quite as sharp. Remember that you should be following the lead car, the speed limit is 65 mph and that driving safely is the most important task here. Go ahead and start driving, I will start the n-back when you get up to speed.*

  - Practice outcomes: participant must have responded correctly to at least the second of each category example while completing the secondary task and maintaining lane position and speed.

- Drive 2
  - Instructions

    Alright, for this drive, you will be doing just what you've been practicing except this time it won't tell you what the sounds are you have to decide for yourself. So if you hear a sound and you think it's an alarm sound, you press the brakes, if you think it seems like a status notification you press the triangle and if you think it seems like a social notification you press the telephone. Drive, following the car in front of you, while doing the n-back and responding promptly to any sounds you might hear using either the brakes or one of the buttons on the touchscreen. Remember that you should be following the lead car, the speed limit is 65 mph and that driving safely is the most important task here. Go ahead and start driving, I will start the n-back when you get up to speed.

  - The drive ends when either the participant crashes or ~30 seconds after a successful avoid procedure.

- Drive 3
  - Instructions

    Alright, this last drive will be a little bit different. This time what we want you to do is drive and categorize the alerts into alarms, status *notifications and social notifications. This will be similar to before but a little longer. This time, when you hear any sound, even if nothing is happening we want you to respond promptly by pressing the phone button if you think the alert should be a social notification, the triangle if you think the alert should be a status notification and the brakes if you think the alert should be an alarm. If possible, we'd also like you to press the brake as hard as you think the alarm is urgent, so if you think it's not a very urgent alarm, just tap the brakes, as if you were disengaging cruise control, but if you think the alarm sounds very urgent press the brakes harder and slow down or come to a stop before catching back up with the lead car. Remember that you should be following the lead car, the speed limit is 65 mph and that driving safely is the most important task here. Go ahead and start driving, I will start the n-back when you get up to speed.*

  - The drive ends when either the participant crashes or ~30 seconds after a successful avoid procedure.

- Post Drive
  - Participants complete demographic survey

Tactile Parameters Experiment Protocol

The following includes the protocol used for Experiment 3 including all experimenter instructions but not including procedural items:

- When participant arrives:
    - Check license
    - Give and explain informed consent
    - Give demographic survey
    - Attach tactor to participant's left wrist
- Task
    - Instructions

        Vehicles are currently being designed using tactile, or haptic, systems which use vibrations to send messages to the driver. This task is designed to allow you to sort vibrations into different categories of signal type you might receive in a vehicle. The categories here are "Alarms", "Status Notifications" and "Social Notifications". Alarms should include vibrations that you believe to be time critical, collision warning vibrations. Status notifications should include vibrations that indicate something about the status of your car, for example, low windshield wiper fluid or low tire pressure. Social notifications should include vibrations used by a car's social media system to indicate a social media (like facebook or an email) notification. We would also like you to indicate why you put vibrations in each category (for example did you group based on their speed or something else about how they felt to you). Finally, we would like you to indicate the urgency level that you think would best represent vibrations that should be in each category on a scale of 1-100. You can double click on numbers to feel the signal and you can play them as many times as you like and move them as many times as you like until you are satisfied with your groupings. You must place at least one number in each category. Please let me know when you are done.

Appendix C: On-Road Experiment Recruitment Ad Text

Volunteers Needed for Driving Safety Study - Receive $75 (Rockville, MD)

Volunteers 21 to 50 with valid driver's licenses are needed for a Federally funded safety research study. Participants will drive a vehicle on public roads while occasionally experiencing different kinds of messages and giving feedback about them. Sessions will be held on weekday mornings and afternoons. Each session will last up to 2 hours and will take place in Rockville, Maryland.

To participate, you must not have had your driver's license suspended or revoked, or received a citation for driving under the influence of alcohol, drugs, or other controlled substances, within the past five years. A motor vehicle record check is required for participation. You must also have normal or corrected-to-normal vision and hearing. For more information or to sign up, please call [redacted].

More information about Westat can be found at www.westat.com

Appendix D: On-Road Experiment Telephone Screener

Thank you for your interest in the In-Vehicle Message Study. If you participate in this study, you will drive a vehicle provided by Westat on local roads and on the Inter County Connector while providing feedback about messages that will be played in the vehicle. You will receive $75 for completing the study.

I have a few questions I need to ask to verify your eligibility. Your ability to participate will depend on your eligibility and our need for participants with a variety of characteristics. If you are invited to participate, we will first need to verify your driving records to ensure that you have not had any major driving violations in the past few years.

1. What is your age? ___________
2. For how many years have you had a valid U.S. driver's license?
3. Has your license been suspended or revoked within the past five years __Yes __No
4. How many days per week do you typically drive? ______
5. Have you ever been diagnosed with a hearing impairment? __Yes __No
6. Do you have any reason to believe you have a hearing impairment? __Yes __No
7. Do you use a hearing aid? __Yes __No
8. Do you have any issues with your vision, including colorblindness? __Yes __No
9. *If Eligible*: What times can you be available for a 2-hour session in Rockville?
   a. ___weekday mornings
   b. ___weekday afternoons
10. May I have your name? ______________________________
11. May I have your daytime phone number? _______________________
12. Is there an email address where I can send you information about the study? ________________ ____ ___

Before you can participate in the study, we will need to check your driving records to ensure that you haven't had any recent major violations. We will mail a form to you to fill out and return to us. You will need to provide your name, address, date of birth, and driver's license number. All information you provide will be kept confidential and will only be used to determine your eligibility.

13. What address would you like us to send the form to? ____________________________________ ____

Thank you for your interest in this study. We will mail the driving records release form to you shortly. Please sign and return it to us at your earliest convenience. Once we verify your driving records we can schedule you for a session.

Appendix E: On-Road Experiment Protocol and Instructions to Participants

Adjustments and calibration: Before we get started, please silence your cell phone. You can also adjust the seat and mirrors to get comfortable in the car. [*wait for participant to make adjustments*] Are you comfortable with your seat and mirror positions? During this session, please do not adjust the heat or fan settings – we need to keep the fan low so it doesn't make much noise. But please let me know if you get too warm or cold.

Purpose and Procedure: The purpose of this study is to see how people understand different sorts of messages that might be used in cars. Cars in the near future will be able to communicate many kinds of information to the driver using sounds, visual displays, or even seat vibration.

This research vehicle has special capabilities. It has intelligent warning systems that can warn you if there is a forward crash threat or if you are drifting out of your lane. It also has the ability to present messages and warnings to you in many different ways, including sounds, visual displays, and seat vibration. This vehicle's communication and warning systems are always active, so it is possible that a message or warning could occur while we are driving.

This vehicle also has cameras that will record video of your face and upper body, as well as audio inside the car. These recordings will be used to analyze data from this session. At the end of the session, I will give you a video release form that will give you the option to approve use of video from this session for other scientific and educational purposes.

Now let's talk about what you will be doing today. After you get some practice driving this vehicle, you will drive us to Route 200, which is also known as the Intercounty Connector where the actual experiment will begin. While you drive normally, I will occasionally ask you to give me your interpretation of signals that you hear, see, or feel as you are driving. Your job will be to tell us what the signals mean to you.

Safety precautions and vehicle familiarization. During today's session, safety is the top priority. You will be required to wear your seat belt at all times while driving and obey posted speed limits and other traffic laws. I will be giving you navigation directions while you drive, but please only make driving maneuvers, such as a lane change, when it is safe to do so. I would prefer you to miss a turn rather than do something risky. Remember that it is your responsibility to drive safely. If at any point during this session you feel that you cannot drive safely, like if you start to get drowsy, please let me know. Please do not use cruise control in this vehicle.

Now you can practice driving. We will take a minute to drive around the parking lot. Please pull out of the parking space when it is safe to do so. I'll give you directions around the parking lot. *[Drive one lap around parking lot.] While driving, ask:* Are you comfortable driving this car? Would you like to make any more adjustments before we go out on real roads? Now let's start driving toward I-270, which will take us to Route 200. I'll give you step by step directions. I'll give you more instructions about this study once we get onto the ICC [*give directions toward I-270*]

Once on I-370. We're on I-370 now which will eventually become Route 200. While on the ICC, please try to maintain your speed close to the speed limit, which is 60 miles per hour. Be aware that the police frequently pull over speeders on this road. Stay in the right lane unless you need to pass a slower vehicle. If you need to pass, please let me know before you change lanes, use your turn signals, and always look carefully to make sure it is safe to change lanes. When we get close to the end of the road, I'll give you directions to exit and get back on in the other direction. Do you have any questions?

Part 1

(begin reading after passing Shady Grove Road Metro exit)

Before we get into the main part of the experiment, I want you to get comfortable reading messages on the display screen. While you drive, you will occasionally see a message appear on the display to your lower right *[point to display to make sure participant knows which display you're referring to]*. Each message will show a driving instruction, but you don't have to actually do anything the instruction says; you only have to read the message. Does that make sense? I'll let you know just before a display appears. Once a display appears, please read the entire message silently as quickly as you can and then say "Done" when you have finished. Any questions?

[Before presenting each display, ensure that there is no close surrounding traffic. If there is surrounding traffic, continue to drive until it clears out. Make sure the participant is on a straightaway. Avoid concrete when triggering signals. Do not trigger signals in the tunnel.]

The first message will appear…now. <Activate text display at the <u>exact</u> time you say "now">

The next message will appear…now. <Activate text display at the <u>exact</u> time you say "now">

The next message will appear…now. <Activate text display at the <u>exact</u> time you say "now" When subject begins reading, trigger the unexpected alert> (note what kind of traffic is around when you trigger this signal)

If participant doesn't react to sound after about 5 seconds: Did you notice a sound a few seconds ago?

What did you think when you heard that sound? What did you think it meant? [*Record responses*]

Before we continue the experiment, I need to give you some additional instructions. We'll exit the ICC at Georgia Avenue southbound and park in a commuter parking lot so I can explain more. When we pull into the lot, please put the car in Park, but DO NOT turn off the engine. If the engine turns off, I will need to restart the data collection systems. *<Exit and park in the commuter lot. If you miss the Georgia Ave exit, take Layhill Rd exit northbound and make your first right into Layhill Park, then park on your left facing the baseball field..>*

That sound you heard while you were looking at the display on the road was an example of the kinds of signals you will experience as you continue this session. Some signals will be sounds and some will be seat vibration. As I mentioned earlier, cars are beginning to be capable of presenting all kinds of messages to drivers. These can include crash warnings, other driving-related notifications, and notifications about incoming messages such as phone calls, texts, or emails.

<As the participant exits the ICC, give detailed instructions to the parking lot. Have the participant park in an open space and **remind them not to turn the car off when they park**>

<u>Microphone setup</u>. For the next part of the driving session, we will get back on the ICC and while you drive you will occasionally receive a signal – either a sound or a seat vibration. Whenever you detect a signal, you will tell me what it means to you as soon as you can after

experiencing it. You will wear a microphone that will record how quickly you respond to the signal.

- Please put on the mic so that the curved elbow joints sit on top of your ears and the microphone is on the left side of your mouth.
- The microphone needs to be positioned carefully so that it picks up your voice when you speak, but doesn't get activated by your breathing. Please bend the tip of the microphone so that it is about an inch in front of the left corner of your mouth. *<verify that the mic is well positioned>*
- <Ask participant to remove a coat with collar that will interfere with Boom back band.>
- <Ask participants with long hair to place the mic band underneath their hair and directly over their ears >

<VERY IMPORTANT. Tell participant this is the good stuff. Need to pay attention>

When you hear a sound or feel a vibration, you will identify it as either an alarm, a status notification or a social notification. An alarm would be some kind of time-critical highly urgent signal like a collision warning or a lane deviation warning. A status notification would be a signal indicating something to do with the status of the car, like low tire pressure or a door is open. A social notification would be something telling you that you have a call or an email or a Facebook update.

So once again, your job will be to say out loud which of those three categories best represents the signal you just experienced. Was it an alarm indicating a time-critical highly urgent warning? Was it a status notification indicating something to do with the status of the car? Or was it a social notification indicating a phone call, email, or other social update? You can also think of these three categories as three levels of urgency where an alarm is high urgency, a status notification is medium urgency, and a social notification is low urgency. Any questions so far?

So when you think of these three categories, you can think of them by number, where an alarm is a one, a status notification is a two, and a social notification is a three. There is a cheat sheet located on the dashboard in front of you to help you remember.

Let's practice this before we go out on the road so you can get used to it. For now, I'd like you to respond to what I say rather than an actual signal. When you hear me say "alarm," you respond by saying "one." When you hear me say "status," you respond by saying "two." When you hear me say "social," you respond by saying "three." You can start by using the cheat sheet to help you, but try to get used to responding without looking at the sheet. Please be sure to speak loud and clear so the microphone can hear you. *<Go through the following set of practice trials, waiting about 2 seconds after a response to say the next word. Check Boom threshold to ensure that participant's voice is peaking the meter (red bar). If not, either lower the threshold, instruct participant to speak louder, or move mic closer to mouth.>*

Please say your answers quickly and clearly because we want your first impression. There are no right or wrong answers – we want to know what these signals mean to you. Also please try hard not to say anything or make sounds like "umm" or "uhh" before you give your answer. The microphone uses the sound of your voice to record how quickly you respond to a signal. If you make any noise before you say one, two, or three, we will get an incorrect response time. Any questions?

Let’s do some practice before we get back on the road. Remember that signals can be either sound or seat vibration. <*conduct first two practice trials*>

Let’s get back on the ICC and start the experiment <direct participant to ICC eastbound. Give a participant a refresher on the three categories while waiting at the light to exit the lot or while you are entering the ICC.>

<*When you are up to speed on ICC…*> Let’s do a little more practice before get started. I won’t give you any advance notice before you receive a message. Any questions before we start? <*Conduct practice trials.*>

Good. Now we will start the actual experiment. Do you have any questions before we begin? […] For half of this session we will drive with music playing. For the other half, we will have the music off.

Before Block 1: For this next set of messages, we will have the music off.

Before Block 2: For this next set of messages, we will have the music on. *[Start music playing on loop]*

Experimenter Notes:

- Watch the experimenter console after triggering a message. The response options should appear on screen at the exact moment a participant begins to say a number. If it appears significantly before participant decided on answer (> 1 second), type “bust rt” in the comments box, or “close rt” if it appears < 1 second before participant decided on answer.
- Listen carefully for “umm” and “uhh” before participant speaks response.
- Give participant feedback if they are making any noises that trigger the mic before their numerical response
- If quiet nonverbal noises trigger the mic (like the sound of participant’s mouth opening) raise the mic threshold to 15 or 20, and/or recheck mic position relative to participant’s mouth.
- Look for upcoming concrete sections/overpasses before triggering
- Click button quietly and avoid giving any subtle triggering cues
- If participant fails to hear a sound, you can trigger the next one without waiting for the countdown
- Try to be silent in back seat at all times
- Keep an eye on participant speed
- Do not allow cruise control use
- Watch for signs to exit onto Briggs Chaney Rd (shortly after Route 29); and then Shady Grove Rd.
- Watch carefully for transition between Blocks 1 and 2 (aka 3)

- During final block, choose turnaround spot to minimize drive back to Westat at end of session.

Feedback & Debrief

- <*While driving back to Westat*> Now that we're done with the alerts, I'd like to ask you a few questions on the way back to Westat. <*read questions and write answers on session info sheet*>
- What was it that made a message seem like an urgent alarm to you? <*Clarify if necessary*> In other words, what aspects of a sound or vibration made it seem like an alert?
- What was it that made a message seem like a vehicle status notification to you?
- What was it that made a message seem like a social notification to you?
- <*If not already addressed*> What aspects of the seat vibrations affected how you categorized them?
- <*Back at Westat visitor lot, give participant copy of video release form*> Now that we've finished, please read this video release form and sign your name if you approve of our use of video from this session for scientific and educational purposes. You are not required to sign.
- Pay participant

DOT HS 812 385
March 2017

U.S. Department
of Transportation

National Highway
Traffic Safety
Administration

NHTSA

12380-032217-v2a