

# **AUDITORY ALERTS IN VEHICLES: EFFECTS OF ALERT CHARACTERISTICS AND AMBIENT NOISE CONDITIONS ON PERCEIVED MEANING AND DETECTABILITY**

**Jeremiah Singer**

**Neil Lerner**

Westat

USA

**Carryl Baldwin**

George Mason University

USA

**Eric Traube**

National Highway Traffic Safety Administration

USA

Paper Number 15-0455

## **ABSTRACT**

Two complementary studies were conducted. In Study 1 (categorical perception of alerts), a series of experiments examined the key parameters that result in listeners perceiving a given sound as an urgent warning versus other less urgent message categories. Initial perceptual sorting experiments identified the most significant parameters and subsequent driving simulator experiments confirmed and refined the findings. Study 1 found that four auditory characteristics accounted for most of the variance in classification of auditory alerts as urgent warnings. Sounds were classified as alarms over 90 percent of the time when they had a peak-to-total-time-ratio (ratio of the time that the sound is at full intensity to the entire pulse duration including onset and offset) of greater than or equal to .7, an interburst interval of less than or equal to 125 ms, at least 3 harmonics, and a base frequency of greater than or equal to 1000 Hz. These results were observed initially in laboratory studies, and replicated during simulated driving. In Study 2 (warning perception in ambient noise environments), an experiment was conducted to investigate the effects of different in-vehicle ambient noise conditions on auditory signal detection and perception. Participants driving on a freeway experienced three ambient noise conditions (windows closed and no music, music playing, front windows open). Fifteen auditory alerts, presented at 65 or 75 decibels (A-weighted), occasionally occurred. Participants pressed a button as soon as they detected the sound, then provided ratings of the noticeability, urgency, and intended meaning of the sound. Study 2 found significant main effects for ambient noise condition and for alert sound for perceived noticeability, urgency, and response time to alert. Detection was impaired by the presence of music, and even more so with the front windows open. Even when auditory signals were heard, noise conditions modified their perceived urgency and meaning. There were also interactions between ambient noise condition and sound, indicating differences in how well sounds of similar loudness tolerated interference from noise. Results also demonstrate that the perceived urgency and meaning of auditory messages can change under noisier ambient conditions, and some features of more noise-resistant signals were suggested by the data. The findings of this research may help interface designers to create auditory signals that indicate the appropriate type and urgency of message.

## **INTRODUCTION**

Advanced technologies in vehicles and on mobile devices when used to complete the driving task such as smartphones are expanding the amount of information available to drivers. Information is increasingly available about safety-critical events, vehicle status, traffic and navigation, and so forth. In this increasingly information-dense environment, it is important that drivers can quickly and easily recognize and distinguish time-critical safety information from other, less urgent information.

This paper describes two complementary studies conducted within the National Highway Traffic Safety Administration's (NHTSA) Crash Warning Interface Metrics (CWIM) research program. The CWIM program included a series of research studies intended to address research needs related to the driver-vehicle interface for advanced crash warning systems. Study 1 (Categorical Perception of Alerts) included a series of research tasks designed to establish criteria regarding the key characteristics of a warning sound that would enable the sound to be quickly and reliably perceived as representing a highly urgent collision warning. Study 2 (Warning Perception in Ambient Noise Environments) investigated the effects of different in-vehicle ambient noise conditions on auditory signal detection and perception. Together, these two studies provide insights into the criteria required for an acoustic

signal to be unambiguously interpreted as an urgent warning with the consideration of the effects of varied in-vehicle ambient noise conditions on the detection and interpretation of signals.

## **STUDY 1: CATEGORICAL PERCEPTION OF ALERTS**

This study, which was led by George Mason University (GMU), included a series of research tasks designed to establish criteria regarding the key characteristics of a warning sound that enable the sound to be quickly and reliably categorized as an urgent warning rather than a vehicle status message or a social notification. Three methodologies were implemented across six research tasks to specifically examine the range of key parameters that result in 85 percent of listeners perceiving a given sound as an urgent collision warning. The 85-percent criterion was selected because it is a common de facto criterion to establish a boundary of typical behavior or perception.

Before beginning the experiments, the research team conducted a survey of forward collision warning (FCW) sounds currently in use in automobiles. The objective of this task was to collect an inventory of warning sounds that could be used to develop an initial list of characteristics common among acoustic warnings. A focus was placed on obtaining representative OEM FCW sounds (i.e., those with the greatest vehicle fleet penetration). Once obtained, the sounds were analyzed in terms of their acoustic properties (e.g., spectral frequency components, pulse duration, onset and offset, harmonics, tempo, etc.). From this inventory, as well as existing published guidelines (e.g., Campbell, Richard, Brown, & McCallum, 2007), a list of key parameters was developed and examined in subsequent experiments. Select non-warning sounds (e.g., vehicle status notifications, ringtones) were also surveyed and recreated for use in this research.

The first four research tasks in this study were a sort task and a series of three method-of-adjustment laboratory tasks. These studies were designed to define and refine the key acoustic characteristics that cause individuals to categorize a message as an urgent crash warning as opposed to other less urgent categories, using the characteristics identified in the inventory of existing in-vehicle warning sounds as a starting point. Specifically, the research team sought to define the key parameters that would be associated with at least 85 percent of listeners perceiving the sounds as a highly urgent, time-critical collision warning. The final two tasks validated the findings of the first four tasks using a sort task and a driving simulation task, respectively. Methods and results for each of the six research tasks are briefly discussed below.

### **Perceptual Space Sort Task**

**Introduction** The goal of this series was to define the perceptual space associated with a highly urgent FCW sound. Specifically, the research team sought to define the key parameters that would be associated with at least 85 percent of listeners perceiving the sounds as a highly urgent, time-critical collision warning.

**Method** Twenty-one GMU students participated in this research task. Stimuli initially included in the Perceptual Space Sort Task were 29 varying driver-vehicle interface (DVI)-related sounds. Sounds were created in Adobe Audition CS5.5 and Audacity and were based closely on sounds used in real, in-vehicle systems. Stimuli included sounds from many categories of in-vehicle system including those similar to FCW, LDW, backup warning, fatigue alerts, door open, low fuel, seatbelt, and park assist sounds. Stimuli were equated for loudness using Adobe Audition and were presented at 84 decibels, A-weighted (dBA).

Sounds were embedded in a large Microsoft PowerPoint slide such that they could be played by clicking on the sound and a pressing a small play triangle. After a participant played a sound, he or she could move it using the cursor into one of four categories. The available categories were “Warnings,” “Alerts,” “Status Notifications,” and “Social Notifications.” Each category also held a small section labeled “Prototype.” Under each category were two text boxes. The first text box held the text “Click here to tell us why...” and was intended to allow participants to give justification for how or why they grouped sounds together. The second box held the text “Best Urgency Level:” and was intended to allow participants to give a numerical rating of the desired urgency level for each category between 1 and 100.

Participants were instructed to play each sound and then categorize it into one of the four available categories. “Warning” sounds were suggested to “include sounds that you believe to be time-critical, collision warning sounds.” “Alert” sounds were suggested to “include non-critical alerting sounds, for example, a sound that might indicate a

lane departure or a parking assist sound.” “Status Notifications” were suggested to “include sounds that indicate something about the status of your car, for example, low windshield wiper fluid or low tire pressure.” “Social Notifications” were suggested to “include sounds used by a car’s social media system to indicate a social media (like Facebook or email) notification.” Participants were instructed that they could have as many sounds as they desired in each category but must have at least one sound in every category. Participants were further instructed that after categorizing each sound, they should play every sound again and choose one sound that was the most prototypical of each category and move it into the “Prototype” section. Participants were finally instructed that they should then attempt to explain their grouping choices and choose an urgency level for each category between 1 and 100. After the practice task, participants were given the experimental task which was identical to the practice but with more sounds.

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

The four most important (i.e., explaining the most variance) properties that related most to alarm-like warnings were determined using a backward stepwise logistic regression to predict membership in the category “alarm.” These properties were interburst interval (IBI), base spectral frequency, number of harmonics (which contributed to the harshness of the tone) and peak to total time ratio. These properties explained 58 percent of the variance in alarm classification ( $R = .800$ , adj.  $R^2 = .581$ ). Peak to total time ratio defines a property to encompass the perceptions created by longer or shorter onset or offset values. IBI is the gap between multiple bursts of sound, which contributes to the perceived tempo of a sound.

Using the results of the backward stepwise regression, criteria were defined in order of decreasing importance for each property of interest and cutoffs were determined. Cutoffs allowed the researchers to organize the data in terms of whether or not it met each criterion sequentially, and eliminate any sound that did not pass at each level. Cutoffs in this case were based partially on results from the subsequent Method of Adjustment experiments and partially based on the research team’s experiences from previous studies and examples from the literature. Sounds that met all four of the criteria below were classified as alarms by at least 90 percent of the listeners. The criteria identified were:

- Peak to total time ratio  $\geq .7$
- Interburst interval (IBI)  $\leq 125$  ms
- Number of harmonics  $\geq 3$
- Base frequency  $\geq 1000$  Hertz (Hz)

### **Method of Adjustment – Single Parameter**

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

**Method** Twenty GMU students participated in this experiment. The base sound used for this experiment was a single 300 Hz tone lasting 200 ms with a 20 ms onset and offset. The tone was repeated eight times with 118 ms between each burst, lasting for a total time of 2426 ms. The sound was played through Sennheiser headphones via a Matlab program. The Matlab program allowed participants to adjust the sound by frequency, pulse duration, tempo

(IBI), or pulses per burst. Sounds were presented at 84 dBA. All parameters included a minimum and maximum possible value with a set increment by which the parameter could be adjusted. For this experiment participants only adjusted one parameter at a time. Each screen indicated which parameter would be adjusted and whether the parameter should be adjusted to sound like a highly urgent, time-critical collision warning or no longer like a highly urgent, time-critical collision warning. Participants adjusted each parameter six times. In a randomized order participants adjusted each parameter three times in ascending order to make it sound like a highly urgent, time-critical collision warning sound and three times in descending order to make it no longer sound like a highly urgent, time-critical collision warning sound. Participants received real-time feedback, such that each time an adjustment was made, a new sound was played based upon the adjustment. Participants could continue to make adjustments until they were satisfied that the sound sounded like or no longer sounded like a highly urgent, time-critical collision warning.

**Results** Table 1 shows the mean and 85th percentile values at which at least 85 percent of participants found the sound to be a highly urgent, time-critical warning.

**Table 1.**  
**Mean values for parameters and value in which at least 85 percent of all participants found the sound to be a highly urgent, time-critical warning**

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

#### Method of Adjustment – Multiple Parameters

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

**Method** Twenty GMU students participated in this experiment. For this experiment the same base sound from the Single Adjustment Task was used in addition to two DVI-related FCW sounds. Sounds were played through Sennheiser headphones via a Matlab program. The Matlab program allowed participants to adjust the sound on frequency, pulse duration, tempo (IBI), and pulses per burst simultaneously (by the same values as presented for the Single Adjustment Experiment). Sounds were presented at approximately 78 dBA. This sound pressure level was 6 dBA lower than used in the previous Method of Adjustment study. This change was made to reduce signal loudness to a level closer to what might be found in vehicles.

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

**Results** Table 2 shows the average value for each parameter as well the value at which at least 85 percent of participants found the sound to be a highly urgent, time-critical collision warning. The table shows that the values were similar to those found in the previous method of adjustment experiment.

**Table 2.**  
Average values for parameters and value in which at least 85 percent of all participants found the sound to be a highly urgent, time-critical collision warning

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

### Method of Adjustment – Prototype Development

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

**Method** Twenty-five GMU students participated in this experiment. For this experiment the same base sound from the “Method of Adjustment – Single Parameter” task was used. The sound was played through Sennheiser headphones via a Matlab program. The Matlab program allowed participants to adjust the sound on frequency, pulse duration, tempo (IBI), and pulses per burst simultaneously. Sounds were presented at approximately 78 dBA.

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

**Results** Table 3 shows the average value for each parameter as well as the value at which at least 85 percent of participants found the sound to be a highly urgent, time-critical collision warning. Compared to the results of the previous method of adjustment task, these results show that an “ideal” collision warning sound has a higher frequency, faster tempo, and faster pace than a “threshold” collision warning.

**Table 3.**  
Average values for parameters and value in which at least 85 percent of all participants found the sound to be a highly urgent, time-critical collision warning

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

### Validation Sort Task

**Introduction** The primary purpose of the Validation Sort Task was to validate the results of the initial sort task and the method of adjustment tasks. To do so, researchers chose to investigate parameters that were found to be important in the previous tasks. Since only one of the previously-evaluated sounds met all four of the criteria established in the first Sort Task, the researchers directly manipulated key acoustic aspects of several sounds to create new sounds designed to examine the robustness of criteria cutoff values. Specifically, the number of harmonics present in several sounds was manipulated by adding harmonic components to sounds that previously had only one harmonic. The presence of multiple harmonics was found in the previous tasks to influence the “harshness”

of the sound, and indirectly, perceived urgency or importance. Further, it was found in the first sort task that some sounds were inadvertently disproportionately long. Therefore in this validation task, sounds were recreated to be of more similar lengths; approximately 1500 ms.

**Method** Fifteen GMU students participated in this experiment. Stimuli for the Validation Sort Task were similar to the Perceptual Space Sort Task but included additional stimuli using added harmonics such that 52 total stimuli were presented to participants. All apparatus and presentation methods were identical to the Perceptual Space Sort Task discussed above.

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

### **Rapid Categorization under Divided Attention**

**Introduction** The Rapid Categorization under Divided Attention task was designed to extend and provide additional validation of the results from the earlier tasks in this series. It has previously been shown that cognitive load can degrade participants’ responses to warnings while driving (Lewis, Penaranda, Roberts, & Baldwin, 2013; Santangelo & Spence, 2007). Therefore, this task was designed to examine whether or not people’s rapid categorization of sounds into critical and noncritical categories differed when they were experiencing the multiple demands of simulated driving while being engaged in a distracting secondary task. In this experiment, participants first completed a sort task similar to the two reported earlier in this study. Then they were asked to categorize sounds by providing a behavioral response (either a brake or button press) while they were concurrently engaged in both a simulated driving task and an n-back number recall task. The n-back task is described in the Method section.

**Method** Twenty-two GMU students participated in this experiment. For the sort task, 26 sounds were used, with specific emphasis placed on those sounds that were shown to be consistently categorized across the sort tasks. The sort task was similar to the two previous sort task experiments with some small changes. In this case the separate categories of “warning” and “alert” were merged into one single “alarm” category, as previous experiments indicated that there is little consistent differentiation between “warnings” and “alerts” by sound parameters. The prototype section of the experiment was removed; otherwise, all aspects of the sort task were identical to the previous sort tasks.

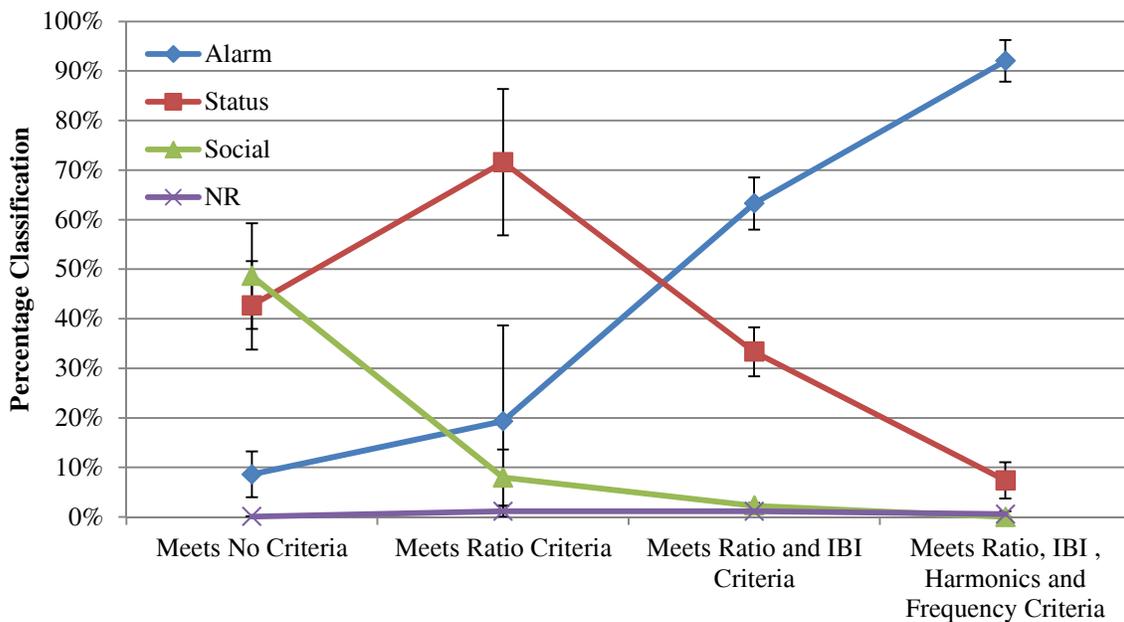
Upon completion of the sort task, participants were seated in the driving simulator. A Realtime Technologies desktop driving simulator was used, with a Logitech Driving Force GT steering wheel. Ambient sound was presented via two monaural computer speakers at 60 dBA. Sounds were presented via a pair of Bose computer speakers at 75 dBA. The same sounds were used as in the sort task. All sounds were repeated twice except for some of the social sounds, which were repeated four times to increase the temporal gap between more urgent sounds. This set of sounds included several sounds based on criteria that were earlier established and were expected to consistently be classified as alarms. Additionally, several sounds were used that shared several but not all attributes of an alarm. The research team theorized that these sounds would be more difficult to classify and potentially result in not being classified as an alarm and/or requiring more time to make a classification response.

Participants practiced driving while maintaining a speed of 45 mph and holding a steady lane position. Once participants were comfortable with the simulator they were instructed that they would be doing a visual 1-back task. For this task, numbers between 1 and 10 were presented on a small screen to the right of the steering wheel. Participants responded via the steering wheel positively or negatively depending on whether or not each number matched the number presented directly before it. After a variable amount of time on the 1-back task, a sound played and participants indicated via the brake, left button, or plus button whether they considered the sound to be an alarm, a status notification, or a social notification, respectively. For the initial practice task, instead of actual signals, the experimenter spoke the words “alarm”, “status” or “social.” Participants then repeated the category name that they heard. This allowed participants to become familiar with the response actions without exposing them to actual signals or requiring them to make a decision about the categorization of a signal. Midway through the practice, participants began driving and practiced all three tasks simultaneously. After completing the practice task, participants completed the experimental task. The experimental task was similar to the practice task with the main difference

being that sounds were now abstract tones (drawn from sounds used in the sort task experiments). Participants were instructed to make their classifications as quickly as possible while maintaining control of the vehicle.

**Results** Results from the sort task, using the same cutoffs and criteria as in the Perceptual Space Sort Task and Validation Sort Task, indicate that the sounds from Rapid Categorization Under Divided Attention are categorized similarly to previous experiments and that the criteria and cutoff results are highly predictive of alarm categorization ( $R=.865$ ,  $\text{adj. } R^2=.701$ ). If the abstract sound met the first two of the previously-established criteria (peak to total time ratio  $> .7$ ; IBI  $< 125$  ms) it was classified as an alarm at least 70 percent of the time. Sounds that met all four of the previously-established criteria (ratio and IBI, plus number of harmonics  $> 3$ ; base frequency  $> 1000$  Hz) were classified as alarms nearly 90 percent of the time. Note that these results are highly similar to the results of the previous two sort tasks. This provides additional support for the reliability and validity of the previous results. The next aim of the current study was to determine if the previously-established alarm criteria were also predictive of alarm perception under divided attention rapid categorization conditions.

Figure 1 shows the results of the rapid categorization task while participants were engaged in the driving and 1-back distraction task. Results are remarkably similar to those obtained from the sort task without concurrent load. Alarm categorization in the sort task obtained while driving was highly correlated with alarm categorization during the rapid categorization task (Pearson correlation= $0.962$ ,  $p<0.001$ ). Additionally, status and social categorization during the sort task were highly correlated with status and social categorization during the rapid categorization task (Pearson correlation for status categorization = $0.934$ ,  $p<0.001$  and for social categorization = $0.982$ ,  $p<0.001$ ).



**Figure 1. Percent categorization while driving by criteria met during driving and distraction task performance.**

Results indicate that the number of criteria met was indicative of response times, where the fastest response times were for sounds that met all four of the previously-established alarm criteria, and slowest response times were for sounds that met only the ratio criterion. These results show that when a sound matches with the criteria, people are able to decide more quickly whether or not the sound is a critical alarm relative to a “status” or “social” sound.

Additionally, the more likely a sound was to be classified as an alarm, the faster the response time for that alert, with alarm classification being significantly predictive of lower response times,  $B = -0.398$ ,  $p=0.004$ . Further, alarm classification even in the sort task was significantly predictive of faster response times during the driving task,  $B = -0.296$ ,  $p=0.031$ . These results indicate that the more homogenous the classification (the more participants classify a sound as an alarm), the faster participants are able to respond to that sound. There was a potential confound, however, because participants categorized sounds by pressing a different button on a steering wheel for

each category. Therefore, differences in button response times could have potentially influenced results, though there is no direct evidence that this was the case.

## **Study 1 Discussion**

The primary aim of this series of perceptual space investigations was to determine the key acoustic characteristics and the parameter values associated with sounds being unambiguously perceived as being or not being highly urgent collision warnings. Toward this aim, the research team employed three methodologies across a series of research tasks. The first task (Perceptual Space Sort) employed regression techniques to identify key characteristics that accounted for a substantial amount of the variance in category classifications (see Perceptual Space Sort Task Results section). Cutoff scores for each of these key parameters were then established by examining the values of each characteristic that resulted in an 85 percent probability of classification as an alarm. The cutoffs established were a peak to total time ratio of greater than or equal to .7, an IBI of less than or equal to 125 ms, at least three harmonics, and a base frequency of greater than or equal to 1000 Hz. When all four of these criteria were met, the sound was classified as an alarm over 90 percent of the time. Meeting only the first of these criteria (peak to total time ratio greater than or equal to .7) resulted in an alarm classification rate of 70 percent. However, because only one sound in the initial set actually met all four criteria and since the criteria were established by the same data set that was being used to examine classification performance, further validation was warranted.

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

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

Together, the results of this research suggest that it is possible to define key characteristics and associated parameter values that could yield effective alarms for use in collision warning systems. A limitation of this research is that the laboratory methods had participants categorize sounds rather than respond to them in a more ecologically valid context (e.g., orientation and response to a perceived threat while driving). Follow-up research funded by NHTSA, and conducted by Westat and GMU, is ongoing to investigate signal categorization and response to unexpected signals during simulated and on-road driving. This follow-up research also includes haptic signals, a wider range of acoustic signals, and a broader age range of participants.

## STUDY 2: WARNING PERCEPTION IN AMBIENT NOISE ENVIRONMENTS

### Introduction

This study, which was conducted by Westat, was an experiment to measure aspects of driver perception of warnings and alerts under a range of ambient noise driving conditions on actual roads. Considerable research has addressed perceptions of in-vehicle warnings and messages, but the vast majority of this work has been conducted under relatively benign in-vehicle ambient noise conditions. Warnings, however, must remain effective under the likely range of noise conditions that may be anticipated in vehicles. Very little information exists on perception of meaning and urgency in noise. The characteristics and sound level of in-cab ambient noise may vary due to the vehicle's physical characteristics, the road surface, surrounding traffic, travel speed, and interior noise sources.

### Method

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

A set of 15 different acoustic signals was presented under each ambient noise condition. These included three unique voice messages and eight unique non-voice sounds. All eleven of the unique sounds and voices were presented at a sound pressure level (SPL) of approximately 65 dBA as measured near the driver's right ear. One of the voice messages and three of the non-voice sounds were also presented at 75 dBA, with the resultant total of 15 signals. The lower 65 dBA level is representative of a number of acoustic alerts as measured in production vehicles (Lin & Green, 2013). The higher 75 dBA level is more consistent with human factors guidance (e.g., Campbell, Richman, Carney, & Lee, 2004), assuming a moderate level of ambient vehicle cab noise.

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

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

**Vehicles** Three different classes of passenger vehicles were used in order to provide a range of vehicle types. These types were small car (2013 Hyundai Accent GLS), sedan (2013 Toyota Camry LE), and SUV (2013 GMC Terrain SLT). The specific vehicles used were selected from among the best-selling models in their class and with good rental availability. Twelve participants drove the small car, 11 drove the sedan, and 11 drove the SUV.

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

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

interior. The minimum SPL was established to ensure that the music could potentially affect participants' detection and ratings of messages.

**Auditory signals and stimulus presentation** Fifteen auditory signals were compared. There were 11 unique alerts presented at approximately 65 dBA. Four of these sounds were also presented at approximately 75 dBA. All sounds were initially volume-adjusted to these levels, but were then adjusted for perceptual equivalence of loudness, as determined by a jury of six individually tested raters. Thus, the 65 dBA sounds were of subjectively equivalent loudness to a 65 dBA continuous pink noise signal (i.e., a signal of random noise containing equal amounts of energy per octave). The 75 dBA sounds were digitally amplified by 10 dB from the 65 dBA level.

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

1. FCW 1: One burst of 20 fast beeps with a relatively high frequency profile.
2. FCW 2: Four bursts of four fast beeps with a relatively low frequency profile.
3. Blind spot warning: Three bursts of four fast beeps, each with a smoothed onset and offset and a sustained low intensity sound between beeps.
4. Pedestrian warning: A single sustained beep with a duration of 2 s.
5. Seat belt alert 1: A single chime that decays to silence in the span of about 2 s, with intensity varying in a wavelike pattern.
6. Seat belt alert 2: Two chimes, each of which decays to silence in the span of about 1 s.
7. Park assist 1: One burst of eight beeps.
8. Park assist 2: Two bursts of three beeps.
9. Female voice – not urgent: Female voice says “Attention.”
10. Female voice – urgent: Female voice says “Warning, warning.”
11. Male voice – urgent: Male voice says “Warning, warning.”
12. FCW 1 (high): Same as FCW 1, but presented at 75 dBA.
13. Blind spot (high): Same as (3), but presented at 75 dBA.
14. Park assist 1(high): Same as (7), but presented at 75 dBA.
15. Female voice – urgent (high): Same as (10), but presented at 75 dBA.

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

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

- Noticeability (defined as “The sound is easily noticeable among other sounds and noises in the vehicle”);  
1=not very; 7=extremely

---

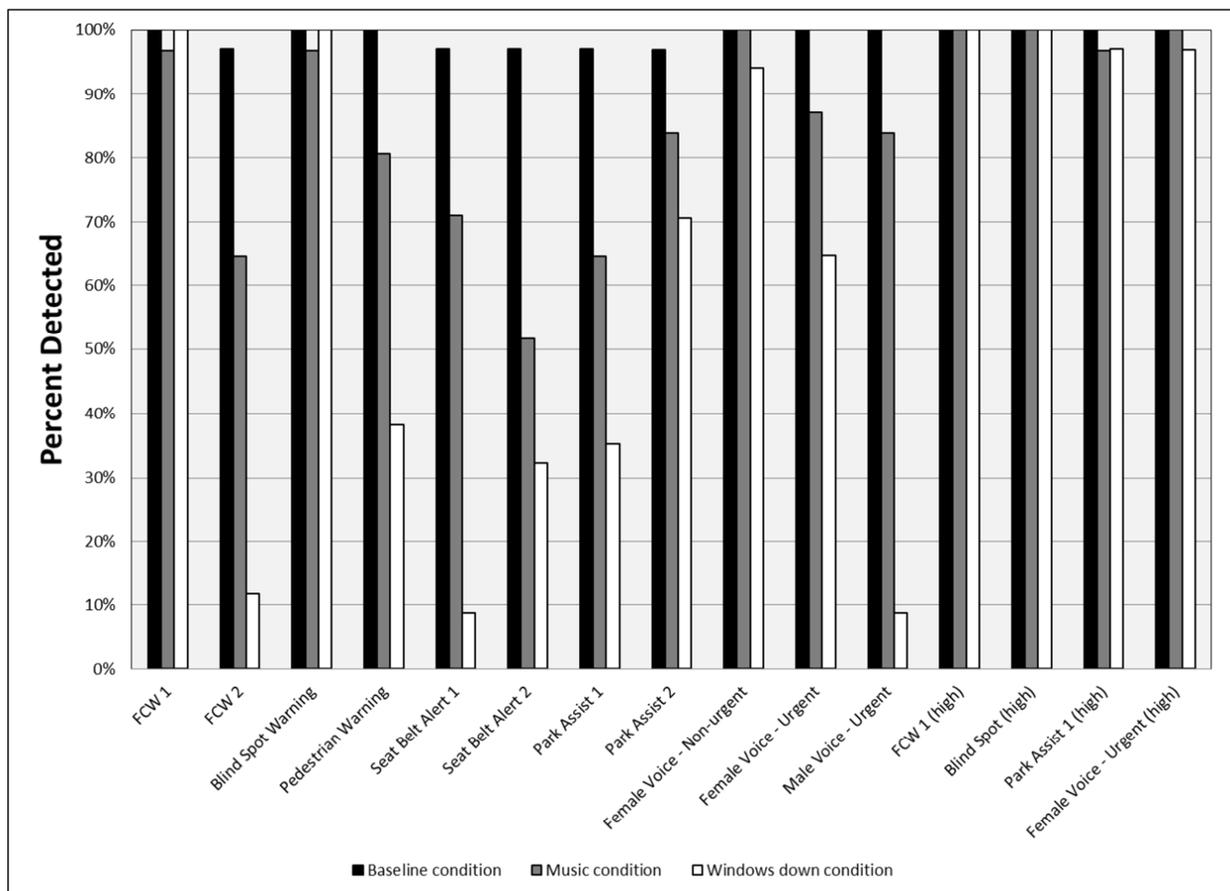
<sup>1</sup> [http://www.oddcast.com/home/demos/tts/tts\\_example.php](http://www.oddcast.com/home/demos/tts/tts_example.php)

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

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

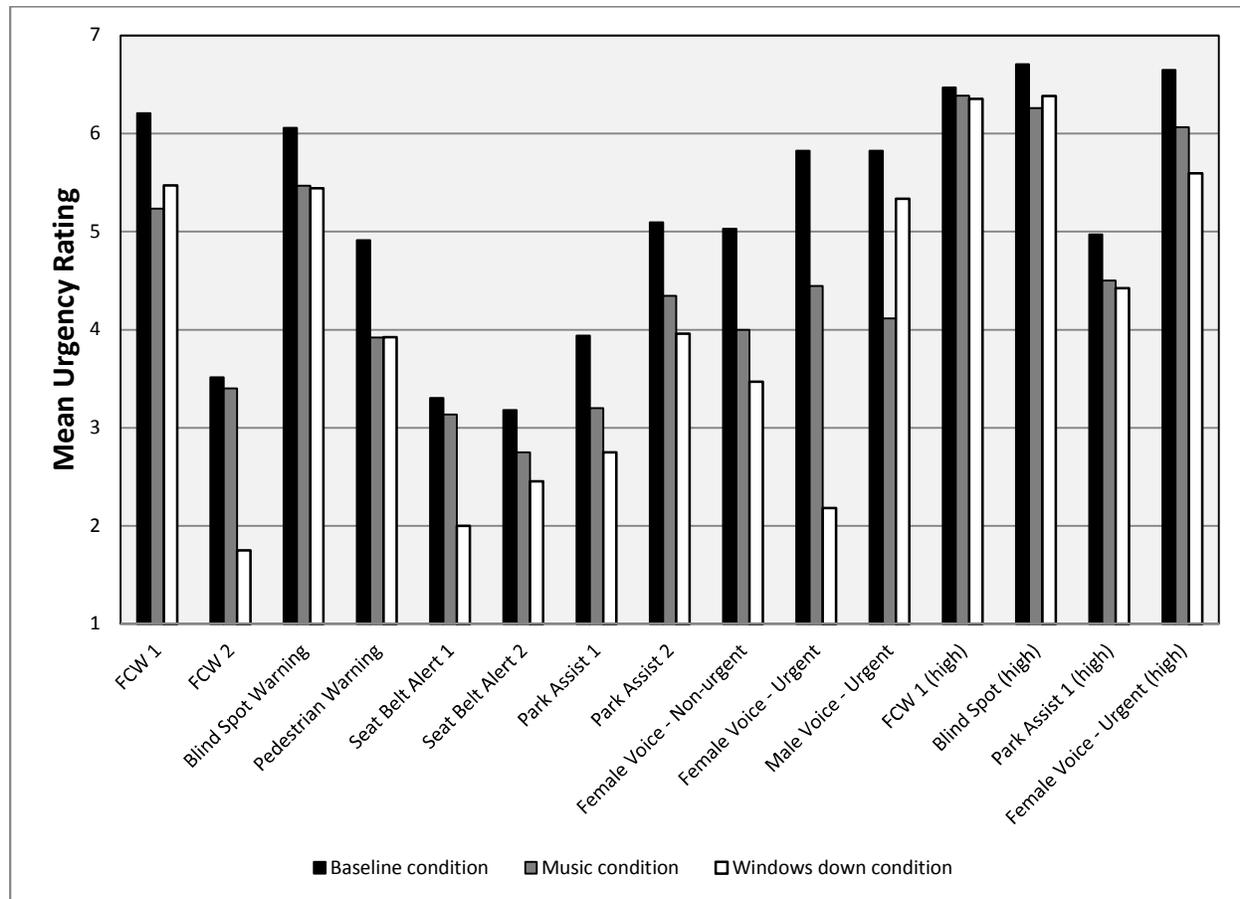
## Results

Driver perception of the auditory signals was influenced by both the nature of the signal and the ambient noise background. Figure 2 shows the percentage of signals detected by the driver for each signal under each ambient noise condition. Few participants missed any signals in the baseline condition (windows up, music off). In the music condition, a few signals were detected in the 70 percent or fewer range, but most were detected at least 80 percent of the time. The open windows condition interfered with detection more dramatically. All four of the signals presented at the 75 dBA level continued to be detected well even with the windows open, but for the 65 dBA signals there was a wide disparity in performance, with a few detected by more than 90 percent of participants and three others detected by only around 10 percent of participants.



**Figure 2. Percentage of participants who detected alerts under each ambient noise condition.**

Three-factor (signal, ambient noise background, vehicle type) ANOVAs were conducted for the measures of rated noticeability, rated urgency, and response time. The three parallel analyses yielded identical conclusions. Signal, noise condition, and their interaction were all statistically significant ( $p < 0.0001$ ). There was no main effect of vehicle type and no interaction of vehicle type with ambient noise background. There was a significant interaction of signal with vehicle type, though there was no clear pattern of effects. It may be expected that due to the complex and varied geometry of the vehicle cabin space and the nature of the reflective and absorbing materials, the differences in the acoustic space might differentially affect some signals. There was no three-way interaction. Figure 3 shows the group mean rating of urgency for each signal under each ambient noise condition. The figure shows that there were dramatic differences in perceived urgency, even among sounds that were equated for similar perceived loudness under relatively quiet conditions.



**Figure 3. Mean urgency rating for each combination of signal and ambient noise condition.**

The ambient noise condition influenced the category of meaning that a listener assigned to a particular alert. Participants had the option of classifying a given alert as “urgent crash warning,” “safety information,” “information not related to safety,” and “incoming personal communication.” As expected, the various alerts differed in terms of how they were interpreted, with some predominantly viewed as urgent crash warnings and others predominantly view as unrelated to safety at all. Multinomial logistic regression was used to analyze the perceived meaning classifications. 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. 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 analysis found significant effects of ambient noise ( $p = 0.0036$ ), acoustic signal ( $p < 0.0001$ ), and an interaction between these two factors ( $p < 0.0001$ ). The key outcome to note here is that

ambient noise condition had a significant effect on participants' categorization of signals, and the significant interaction term reveals that different signals were affected in different ways by the various ambient noise conditions, revealing a complex relationship between ambient noise and acoustic signals with regard to perceived category of meaning.

## **Study 2 Discussion**

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

This experiment was designed to provide an initial examination of the extent to which possible ambient noise conditions might interfere with signal detection and meaning. It was not intended to provide any systematic evaluation of signal features or parameters regarding their resistance to noise effects. However, based on the limited sample of sounds and conditions, it appeared that the predominant frequencies that characterize a signal may relate to perceived urgency under noise. Sounds with base frequencies below 1000 Hz generally performed worst and those with base or significant components above 1500 Hz performed best. However, this observation is based on a limited sample of sounds that also differed in a number of other respects, and so it should be considered tentative. A series of follow-up laboratory investigations conducted within NHTSA's CWIM research program replicated and extended the findings of this on-road study by investigating additional sounds, ambient noise conditions, and sound pressure levels.

## **CONCLUSIONS**

The National Highway Traffic Safety Administration's Crash Warning Interface Metrics program included a series of research studies intended to address research needs related to the driver-vehicle interface for advanced crash warning systems. The two complementary studies described in this paper investigated the signal characteristics that lead to people clearly distinguish urgent alert sounds from other message categories (Study 1), and how the effects of various in-vehicle ambient noise conditions affect the detection and perceived meaning of signals (Study 2).

In Study 1 (Categorical Perception of Alerts), a wide range of parameters was systematically tested, and four parameters were found to be most important in influencing categorization as a warning or safety alert: base frequency, number of harmonics, IBI, and peak to total time ratio. The cutoff values at which a sound becomes unambiguously perceived as an alarm were refined and validated in a series of tasks including a driving simulator task under cognitive load. The findings can potentially help designers create in-vehicle systems in which intended crash warnings are reliably perceived as urgent and critical and in which intended lower urgency message types do not lead to confusion with urgent warnings.

In Study 2 (Warning Perception in Ambient Noise Environments), a set of 15 sounds was presented to drivers under three ambient sound conditions. Results show that many sounds that were easily detected and perceived as urgent in a relatively quiet vehicle interior were much less likely to be detected or perceived as urgent in a louder vehicle interior (e.g., with windows open). Main effects of detection rate for sound stimulus suggest that sounds adjusted for equal loudness in a quiet environment are not necessarily equally loud in a vehicle environment. The presence of interactive effects between different sounds and ambient noise conditions suggests that some sounds are more resistant to the interfering effects of elevated ambient noise than others.

Taken together, the results of these two studies show that the acoustic characteristics that cause a sound to be unambiguously interpreted as an urgent alarm or other category are quantifiable, but that the ambient noise condition in which sounds are heard can have a significant effect on signal detection and interpretation. In-vehicle acoustic

signals, then, should be quickly and unambiguously interpreted with the meaning intended by the designer, and should maintain their detectability and meaning in a variety of ambient noise conditions. Study 2 found that sounds played at 75 dBA were substantially less impaired by elevated ambient noise than sounds played at 65 dBA, suggesting that alerts presented at approximately 65 dBA, as is common in production vehicles, might not always maintain their detectability and meaning under elevated ambient noise. The limited number of stimuli investigated in this study does not provide enough data to suggest an ideal loudness, nor suggest features that might make a given sound more detectable or more reliably categorized at a given loudness.

These studies have some notable limitations. As noted above, signal loudness has a substantial effect on signal detection and interpretation, especially in elevated ambient noise. Study 1 did not systematically manipulate signal loudness, and Study 2 only manipulated loudness at two levels for five signals. Study 2 also only investigated three ambient noise conditions. These studies also involved participants focused on an experimental task without the context of a real on-road information environment. As such, there were no conditions or hazards associated with signals, nor were participants expected to react or respond to signals (other than by providing a categorical response). Follow-up research conducted within NHTSA's CWIM program has continued the ambient noise line of research by replicating the on-road findings in a laboratory setting and extending the research to investigate additional sounds, ambient noise conditions, and sound pressure levels. A separate study conducted within NHTSA's Connected Vehicle program will extend the categorical perception research by investigating additional sounds, as well as haptic signals, in driving simulator and on-road experiments.

## REFERENCES

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

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

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

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

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